Article:

Triantafyllopoulos, K. (2008) Missing observation analysis for matrix-variate time series data. Statistics and Probability Letters, 78 (16). pp. 2647-2653. ISSN 0167-7152

https://doi.org/10.1016/j.spl.2008.03.033

Missing observation analysis for matrix-variate time series data

K. Triantafyllopoulos*

May 25, 2008

Abstract

Bayesian inference is developed for matrix-variate dynamic linear models (MV-DLMs), in order to allow missing observation analysis, of any sub-vector or sub-matrix of the observation time series matrix. We propose modifications of the inverted Wishart and matrix t distributions, replacing the scalar degrees of freedom by a diagonal matrix of degrees of freedom. The MV-DLM is then re-defined and modifications of the updating algorithm for missing observations are suggested.

Some key words: Bayesian forecasting, dynamic models, inverted Wishart distribution, state space models.

1 Introduction

Suppose that, in the notation of West and Harrison (1997, Chapter 16), the $p \times r$ matrix-variate time series $\{y_t\}$ follows a matrix-variate dynamic linear model (MV-DLM) so that

$$y_t' = F_t'\Theta_t + \epsilon_t' \quad \text{and} \quad \Theta_t = G_t\Theta_{t-1} + \omega_t, \tag{1}$$

where F_t is a $d \times r$ design matrix, G_t is a $d \times d$ evolution matrix and Θ_t is a $d \times p$ state matrix. Conditional on a $p \times p$ covariance matrix Σ, the innovations ϵ_t and ω_t follow, respectively, matrix-variate normal distributions (Dawid, 1981), i.e.

$$\epsilon_t|\Sigma \sim N_{r\times p}(0, V_t, \Sigma) \quad \text{and} \quad \omega_t|\Sigma \sim N_{d\times p}(0, W_t, \Sigma).$$

This is equivalent to writing $\text{vec}(\epsilon_t)|\Sigma \sim N_{rp}(0, \Sigma \otimes V_t)$ and $\text{vec}(\omega_t)|\Sigma \sim N_{dp}(0, \Sigma \otimes W_t)$, where vec(.) denotes the column stacking operator of a matrix, $\otimes$ denotes the Kronecker product of two matrices and $N_{rp}(.,.)$ denotes the multivariate normal distribution.

*Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, UK, email: `k.triantafyllopoulos@sheffield.ac.uk`

We assume that the innovation series $\{\epsilon_t\}$ and $\{\omega_t\}$ are internally and mutually uncorrelated and also they are uncorrelated with the assumed initial priors

$$\Theta_0|\Sigma \sim N_{d\times p}(m_0, P_0, \Sigma) \quad \text{and} \quad \Sigma \sim IW_p(n_0, n_0S_0), \tag{2}$$

for some known m_0, P_0, n_0 and S_0. Here $\Sigma \sim IW_p(n_0, n_0S_0)$ denotes the inverted Wishart distribution with n_0 degrees of freedom and parameter matrix n_0S_0. The covariance matrices V_t and W_t are assumed known; usually $V_t = I_r$ (the $r \times r$ identity matrix) and W_t can be specified using discount factors as in West and Harrison (1997, Chapter 6). Alternatively, $W_t = W$ may be considered time-invariant and it can be estimated from the data using the EM algorithm (Dempster *et al.*, 1977; Shumway and Stoffer, 1982). With the above initial priors (2) the posterior distribution of $\Theta_t|\Sigma, y_1, \ldots, y_t$ is a matrix-variate normal distribution and the posterior distribution of $\Sigma|y_1, \ldots, y_t$ is an inverted Wishart distribution with degrees of freedom $n_t = n_{t-1}+1$ and a parameter matrix n_tS_t, which are calculated recurrently (West and Harrison, 1997, Chapter 16).

Missing data in time series are typically handled by evaluating the likelihood function (Jones, 1980; Ljung, 1982; Shumway and Stoffer, 1982; Harvey and Pierse, 1984; Wincek and Reinsel, 1984; Kohn and Ansley, 1986; Ljung, 1993; Gómez and Maravall, 1994; Luceño, 1994; Luceño, 1997). In the context of model (1) a major obstacle in inference is when a sub-vector or sub-matrix $\widetilde{y}_t$ of y_t is missing at time t. Then the scalar degrees of freedom of the inverted Wishart distribution of $\Sigma|y_1, \ldots, y_t$, are incapable to include the information of the observed part of y_t, but to exclude the influence of the missing part $\widetilde{y}_t$. For example consider $p = 2$ and $r = 1$ or $y_t = [y_{1t} \ y_{2t}]'$ and suppose that at time t, y_{1t} is missing ($\widetilde{y}_t = y_{1t}$), while y_{2t} is observed. Let n_{t-1} denote the degrees of freedom of the inverted Wishart distribution of $\Sigma|y_1, \ldots, y_{t-1}$. One question is how one should update n_t, since the information at time t is partial (one component observed and one missing). Likewise, given this partial information at time t, another question is how to estimate the off-diagonal elements of Σ, which leads to the estimation of the covariance of y_{1t} and y_{2t}.

In this paper, introducing several degrees of freedom that form a diagonal matrix, we propose modifications to the inverted Wishart and matrix t distributions. We prove the conjugacy between these distributions and we discuss modifications in the recursions of the posterior moments in the presence of missing data. This approach does not require to order all missing observations in one matrix (Shumway and Stoffer, 1982; Luceño, 1997) and therefore it can be applied for sequential purposes as new data are observed.

2 Matrix-variate dynamic linear models

2.1 Modified inverted Wishart distribution

Suppose that Σ is a $p\times p$ random covariance matrix, S, R are $p\times p$ covariance matrices and N is a $p\times p$ diagonal matrix with positive diagonal elements. Let $\mathrm{tr}(.)$, $\mathrm{etr}(.)$ and $|.|$ denote the trace, the exponent of the trace and the determinant of a square matrix, respectively. The density of the inverted Wishart distribution is given by

$$p(\Sigma) = c|R|^{(k-p-1)/2}|\Sigma|^{-k/2}\mathrm{etr}\left(-\frac{1}{2}R\Sigma^{-1}\right), \tag{3}$$

from which it is deduced that

$$\int_\Omega |\Sigma|^{-k/2}\mathrm{etr}\left(-\frac{1}{2}R\Sigma^{-1}\right)\, d\Sigma = c^{-1}|R|^{-(k-p-1)/2}, \tag{4}$$

with $\Omega = \{\Sigma \in \mathbb{R}^{p\times p} : \Sigma > 0\}$, $c^{-1} = 2^{(k-p-1)p/2}\Gamma_p\{(k-p-1)/2\}$, and $k > 2p$, where $\Gamma_p(.)$ is the multivariate gamma function.

Lemma 1. *The function*

$$p(\Sigma) = c|\Sigma|^{-\{v+tr(N)/(2p)\}}etr\left(-\frac{1}{2}N^{1/2}SN^{1/2}\Sigma^{-1}\right), \tag{5}$$

where c does not depend on Σ, is a density function.

Proof. If the following bijective transformation is applied

$$R = N^{1/2}SN^{1/2} \quad \text{and} \quad k = 2v + \frac{\mathrm{tr}(N)}{p}, \tag{6}$$

then (5) is directly obtained from (3). □

From the above bijection and the Wishart integral, we can see that the normalizing constant c is

$$c = c_0|S|^{\{2v+\mathrm{tr}(N)/p-p-1\}/2}\left(\prod_{j=1}^{p} n_j\right)^{\{2v+\mathrm{tr}(N)/p-p-1\}/2},$$

where

$$c_0^{-1} = 2^{\{2v+\mathrm{tr}(N)/p-p-1\}p/2}\Gamma_p\left\{\frac{2v+\mathrm{tr}\,(N)\,/p-p-1}{2}\right\},$$

for $N = \mathrm{diag}(n_1, \dots, n_p)$ and $n_i > 0$ $(i = 1, \dots, p)$.

Density (5) proposes a modification of the inverted Wishart distribution in order to incorporate a diagonal matrix of degrees of freedom. The modification consists of a bijective transform of the two distributions. We will then say that Σ follows the *modified inverted*

Wishart distribution and we will write $\Sigma \sim MIW_p(S, N, v)$, where v is a scalar hyperparameter. Note that when $n_1 = \cdots = n_p = n$ and $v = p$, the above distribution reduces to an inverted Wishart distribution with n degrees of freedom.

With k and R as defined in equation (6), the mean of Σ is

$$E(\Sigma) = \frac{R}{k - 2p - 2} = \left\{ \frac{\operatorname{tr}(N)}{p} + 2v - 2p - 2 \right\}^{-1} N^{1/2} S N^{1/2},$$

for $p^{-1}\operatorname{tr}(N) > 2p - 2v + 2$. The next result gives the distribution of a MIW matrix conditional on a normal matrix.

Proposition 1. *Let Y be an $r \times p$ random matrix that follows a matrix normal distribution, conditional on Σ, and Σ a $p \times p$ covariance random matrix that follows a modified inverted Wishart distribution, written $Y|\Sigma \sim N_{r\times p}(m, P, \Sigma)$ and $\Sigma \sim MIW_p(S, N, v)$ respectively, for some known quantities m, P, S, N, and v. Then, the conditional distribution of Σ given Y, is*

$$\Sigma|Y \sim MIW_p(S^*, N^*, v),$$

where $N^{*1/2} S^* N^{*1/2} = (Y - m)' P^{-1} (Y - m) + N^{1/2} S N^{1/2}$ and $N^* = N + rI_p$.

Proof. Form the joint distribution of Y and Σ and write

$$\begin{aligned} p(\Sigma|Y) &\propto p(Y, \Sigma) = p(Y|\Sigma)p(\Sigma) \\ &\propto |\Sigma|^{-\{v+r/2+\operatorname{tr}(N)/(2p)\}} \operatorname{etr}\bigg[-\frac{1}{2}\{(Y - f)' Q^{-1} (Y - f) \\ &\quad + N^{1/2} S N^{1/2}\}\Sigma^{-1} \bigg], \end{aligned} \tag{7}$$

which is sufficient for the proof with the definition of S^* and N^*. □

In the context of Proposition 1 the joint distribution of Y and Σ is referred to as joint normal modified inverted Wishart distribution with notation $Y, \Sigma \sim NMIW_{r\times p,p}(m, P, S, N, v)$, for m, P, S, N, and v as defined in Proposition 1. The next result gives the marginal distribution of Y. First we give some background material on the matrix t distribution.

Let X be an $r \times p$ random matrix. Then, the matrix t distribution is defined by

$$p(X) = c|Q + (X - M)' P^{-1} (X - M)|^{-(k+r+p-1)/2}, \tag{8}$$

with

$$c = \frac{\Gamma_p\{(k + r + p - 1)/2\} |Q|^{(k+p-1)/2} |P|^{-p/2}}{\pi^{rp/2} \Gamma_p\{(k + p - 1)/2\}},$$

where M is an $r \times p$ matrix, P a $r \times r$ covariance matrix, Q a $p \times p$ covariance matrix, and k any positive real number.

Proposition 2. *Let Y be an $r \times p$ random matrix that follows a matrix normal distribution conditional on Σ, and Σ be a $p \times p$ covariance random matrix that follows a modified inverted Wishart distribution, written $Y|\Sigma \sim N_{r\times p}(f, Q, \Sigma)$, and $\Sigma \sim MIW_p(S, N, v)$ respectively, for known quantities f, Q, S, N, and v. Then, the marginal distribution of Y is*

$$p(Y) = c|N^{1/2}SN^{1/2} + (Y - f)'Q^{-1}(Y - f)|^{-\{2v+tr(N)/p+d-p-1\}/2}, \tag{9}$$

which by analogy of the MIW distribution, is a modification of the matrix t distribution and it is written as $MT(f, Q, S, N, v)$.

Proof. The joint distribution of Y and Σ is given by equation (7). Hence, the marginal distribution of Y is

$$p(Y) = \int_{\Omega} p(Y, \Sigma)\, d\Sigma,$$

where $\Omega = \{\Sigma \in \mathbb{R}^{p\times p} : \Sigma > 0\}$. Set $R = (Y - f)'Q^{-1}(Y - f) + N^{1/2}SN^{1/2}$ and $k = 2v + r + \mathrm{tr}(N)/p$ and from equation (4) we have equation (9). □

The distribution of Proposition (2) can be derived from the matrix t distribution (see equation (8)). The normalizing constant c of (9) is obtainable from (8) as

$$c = \frac{\pi^{pr/2}\Gamma_p\{(k + p - 1)/2\}}{\Gamma_p\{(k + r + p - 1)/2\}}|S|^{(k+p-1)/2}\left(\prod_{j=1}^{p} n_j\right)^{(k+p-1)/2}|Q|^{-p/2},$$

where $N = \mathrm{diag}(n_1, \dots, n_p)$ and $k = 2v - 2p + \mathrm{tr}(N)/p$. Note that if all the diagonal elements of N are the same (i.e. $n_1 = \cdots = n_p = n$) and $v = p$, then the above distribution reduces to a matrix t distribution with n degrees of freedom.

Finally we give the marginal distribution of Σ. Consider the following partition of Σ, S, and N

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}, \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix}, \quad N = \begin{bmatrix} N_1 & 0 \\ 0' & N_2 \end{bmatrix},$$

where Σ_{11}, S_{11} and N_{11} have dimension $q \times q$, for some $1 \le q < p$. The next result gives the marginal distribution of Σ_{11}.

Proposition 3. *If $\Sigma \sim MIW_p(S, N, v)$, under the above partition of Σ the distribution of Σ_{11} is $\Sigma_{11} \sim MIW_q(S_{11}, N_{11}, v_1)$, where $v_1 = v - p + q + 2^{-1}p^{-1}tr(N) - 2^{-1}q^{-1}tr(N_1)$.*

Proof. The proof suggests the adoption of transformation (6) together with the partition of R in (3) as

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{12}' & R_{22} \end{bmatrix}.$$

Using marginalization properties of the inverted Wishart distribution, upon noticing

$$N^{1/2}SN^{1/2} = \left[\begin{array}{cc} N_1^{1/2}S_{11}N_1^{1/2} & N_1^{1/2}S_{12}N_2^{1/2} \\ N_2^{1/2}S_{12}'N_1^{1/2} & N_2^{1/2}S_{22}N_2^{1/2} \end{array} \right],$$

we get $\Sigma_{11} \sim MIW_q(S_{11}, N_1, v_1)$, with v_1 as required. $\square$

A similar result can be obtained for Σ_{22}. Consequently, if we write $\Sigma = \{\sigma_{ij}\}$ $(1 \leq i, j \leq p)$ and $N = \text{diag}(n_1, \ldots, n_p)$, then the diagonal variances σ_{ii} follow modified inverted Wishart distributions, $\sigma_{ii} \sim MIW_1(s_{ii}, n_i, v_i)$, where $v_i = v - p + 1 + 2^{-1}p^{-1}\text{tr}(N) - 2^{-1}n_i$. These in fact are inverted gamma distributions $\sigma_{ii} \sim IG(v_i + n_i/2 - 1, n_i s_{ii}/2)$. Note that if $n_1 = \cdots = n_p = n$ and $v = p$, then we have that $\sigma_{ii} \sim IG(n/2, ns_{ii}/2)$ (the inverted gamma distribution used in West and Harrison (1997) when $p = 1$).

We close this section with a brief discussion on an earlier study proposing the incorporation of several degrees of freedom for inverted Wishart matrices (Brown *et al.*, 1994). This approach is based on breaking the degrees of freedom on blocks and requiring for each block the marginal density of the covariance matrix to follow an inverted Wishart distribution. However, in that framework the conjugacy between the normal and that distribution is lost and as a result the proposed estimation procedure may be slow and probably not suitable for time series application. Relevant inferential issues of that approach are discussed in Garthwaite and Al-Awadhi (2001). Our proposal of the MIW distribution retains the desired conjugacy and it leads to relevant modifications of the matrix t distribution, which provides the forecast distribution. Furthermore, the MIW density leads to fast computationally efficient algorithms, which are suitable for sequential model monitoring and expert intervention (Salvador and Gargallo, 2004). Finally, according to Proposition 3, the marginal distributions of MIW matrices are also MIW, which means that several degrees of freedom are included in the marginal models too, something that is not the case in the approach of Brown *et al.* (1994).

2.2 Matrix-variate dynamic linear models revisited

We consider model (1), but now we replace the initial priors (2) by the priors

$$\Theta_0|\Sigma \sim N_{d\times p}(m_0, P_0, \Sigma_0) \quad \text{and} \quad \Sigma_0 \sim MIW_p(S_0, N_0, p), \tag{10}$$

for some known m_0, P_0, S_0 and N_0. Practically we have replaced the inverted Wishart prior by the MIW and so, for each $t = 1, \ldots, T$, we use p degrees of freedom $n_{1t}, \ldots, n_{pt}$ in order to estimate $\Sigma|y^t$, where y^t denotes the information set, comprising of observed data $y_1, \ldots, y_t$. The next result provides the posterior and forecast distributions of the new MV-DLM.

Proposition 4. *One-step forecast and posterior distributions in the model (1) with the initial priors (10), are given, for each t, as follows.*

(a) *Posterior at* $t-1$: $\quad \Theta_{t-1}, \Sigma | y^{t-1} \sim NMIW_{d\times p,p}(m_{t-1}, P_{t-1}, S_{t-1}, N_{t-1}, p)$, *for some* m_{t-1}, P_{t-1}, S_{t-1} *and* N_{t-1}.

(b) *Prior at* t : $\quad \Theta_t, \Sigma | y^{t-1} \sim NMIW_{d\times p,p}(a_t, R_t, S_{t-1}, N_{t-1}, p)$, *where* $a_t = G_t m_{t-1}$ *and* $R_t = G_t P_{t-1} G_t' + W_t$.

(c) *One-step forecast at t:* $\quad y_t' | \Sigma, y^{t-1} \sim N_{r\times p}(f_t', Q_t, \Sigma)$,
with marginal: $\quad y_t' | y^{t-1} \sim MT_{r\times p}(f_t', Q_t, S_{t-1}, N_{t-1}, p)$,
where $f_t' = F_t' a_t$ *and* $Q_t = F_t' R_t F_t + V_t$.

(d) *Posterior at* t : $\quad \Theta_t, \Sigma | y^t \sim NMIW_{d\times p,p}(m_t, P_t, S_t, N_t, p)$,
with

$$m_t = a_t + A_t e_t', \quad P_t = R_t - A_t Q_t A_t',$$
$$N_t = N_{t-1} + rI_p, \quad N_t^{1/2} S_t N_t^{1/2} = N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} + e_t Q_t^{-1} e_t',$$
$$A_t = R_t F_t Q_t^{-1}, \quad \textit{and} \quad e_t = y_t - f_t.$$

The proof of this result follows immediately from Propositions 1 and 2. For $t = 1$, (a) coincides with the priors (10). From Proposition 2, the marginal posterior of $\Theta_t | y^t$ is $\Theta_t | y^t \sim MT_{d\times p}(m_t, P_t, S_t, N_t, p)$. Thus the above proposition gives a recursive algorithm for the estimation and forecasting of the system for all $t = 1, \ldots, T$.

Proposition 4 gives a generalization of the updating recursions of matrix-variate dynamic models (West and Harrison, 1997, Chapter 16). The main difference of the two algorithms is that the scalar degrees of freedom n_t of the standard recursions are replaced by N_t in the above proposition and that the inverted Wishart distribution is replaced by the modified inverted Wishart distribution (in order to account for the matrix of degrees of freedom). As a result the classical Bayesian updating of West and Harrison (1997) is obtained as a special case of the distributional results of Proposition 4, by setting $N_t = n_t I_p = \text{diag}(n_t, \ldots, n_t)$ ($t = 0, 1, \ldots, T$), where n_t represent the scalar degrees of freedom of the inverted Wishart distribution of $\Sigma_t | y^t$ and n_0 is the initial degrees of freedom.

3 Missing observations

In this section we consider missing observations at random. Our approach is based on excluding any missing values of the calculation of the updating equations (state and forecast distributions) thus excluding the unknown influence of these unobserved variables. This approach is explained for univariate dynamic models in West and Harrison (1997, Chapters 4,10).

The univariate dynamic linear model with unknown observational variance is obtained from model (1) for $p = r = 1$. In this case the posterior recursions of m_t, P_t and S_t of West and Harrison (1997, Chapter 4) follow from Proposition 4 as a special case. Now suppose that at time t the scalar observation y_t is missing so that $y^t = y^{t-1}$. It is then obvious that the posterior distribution of Θ_t equals its prior distribution (since no information comes in to the system at time t). Then we have $m_t = a_t$, $P_t = R_t$, $S_t = S_{t-1}$ and $N_t = n_t = n_{t-1} = N_{t-1}$. To incorporate this into the updating equations of the posterior means and variances, we can write $m_t = a_t - A_t e_t u_t$, $P_t = R_t - A_t A_t' Q_t u_t$, $n_t S_t = n_{t-1} S_{t-1} + e_t^2 u_t / Q_t$ and $n_t = n_{t-1} + u_t$, where u_t is zero, if y_t is missing and $u_t = 1$, if y_t is observed. So when $p = 1$ the inclusion of u_t in the posterior recursions leads to identical analysis as in West and Harrison (1997) and in references therein. The introduction of u_t in the recursions automates the posterior/prior updating in the presence of missing values and it motivates the case for $p, r \geq 1$.

Moving to the multivariate case, first we consider model (1) as defined in the previous section with $r = 1$. Assume that we observe all the $p \times 1$ vectors y_i, $i = 1, \ldots, t-1$. At time t some observations are missing (sub-vectors of y_t, or the entire y_t). To distinguish the former from the latter case we have the following definition.

Definition 1. *A partial missing observation vector is said to be any strictly sub-vector of the observation vector that is missing. If the entire observation vector is missing it is referred to as full missing observation vector.*

Considering the MV-DLM (1), it is clear that in the case of a full missing vector we have

$$\Theta_t, \Sigma | y^t \sim NMIW_{d \times p, p}(m_t, P_t, S_t, N_t, p), \tag{11}$$

where $m_t = a_t$, $P_t = R_t$, $S_t = S_{t-1}$, $N_t = N_{t-1}$, since no information comes in at time t. This equation relates to the standard posterior distribution of West and Harrison (1997) by setting $N_t = \text{diag}(n_t, \ldots, n_t)$, for a scalar $n_t > 0$ and evidently reducing the MIW distribution by a IW distribution. If one starts with a prior $N_0 = \text{diag}(n_0, \ldots, n_0)$, and assuming that at some time t, there is a full missing vector y_t, then it is clear that the posterior (11) equals to the posterior of $\Theta_t, \Sigma | y^t$ using the standard recursions (West and Harrison, 1997). Any differences between the two algorithms is highlighted only by observing partial missing vectors and this has been the motivation of the new algorithm.

Define a $p \times p$ diagonal matrix $U_t = \text{diag}(i_{1t}, \ldots, i_{pt})$ with

$$i_{jt} = \begin{cases} 1 & \text{if } y_{jt} \text{ is observed,} \\ 0 & \text{if } y_{jt} \text{ is missing,} \end{cases}$$

for all $1 \leq j \leq p$, where $y_t = [y_{1t} \ \cdots \ y_{pt}]'$.

Then, the posterior distribution (11) still applies with recurrences

$$m_t = a_t + A_t e_t' U_t \tag{12}$$

$$P_t = R_t - A_t A_t' Q_t u_t \tag{13}$$

$$N_t = N_{t-1} + U_t \tag{14}$$

$$N_t^{1/2} S_t N_t^{1/2} = N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} + U_t e_t Q_t^{-1} e_t' U_t, \tag{15}$$

where $u_t = \text{tr}(U_t)/p$. Some explanation for the above formulae are in order.

First note that if no missing observation occurs $U_t = I_p$, $u_t = 1$ and we have the standard recurrences as in Proposition 4. On the other extreme (full missing vector), $U_t = 0$, $u_t = 0$ and we have equation (11). Consider now the case of partial missing observations. Equation (14) is the natural extension of the single degrees of freedom updating, see West and Harrison (1997, Chapter 16). For equation (12) note that the zero's of the main diagonal of U_t convey the idea that the corresponding to the missing values elements of m_t remain unchanged and equal to a_t. For example, consider the case of $p = 2$, $d = 2$ and assume that you observe y_{1t}, but y_{2t} is missing. Then

$$m_t = a_t + \begin{bmatrix} A_{1t}(y_{1t} - f_{1t}) & 0 \\ A_{2t}(y_{1t} - f_{1t}) & 0 \end{bmatrix},$$

where $A_t = [A_{1t} \ A_{2t}]'$. The zero's on the right hand side reveal that the second column of m_t is the same as the second column of a_t. Similar comments apply for equations (13) and (15).

Considering the case of $r \geq 2$, we define U_{kt} to be the diagonal matrix $U_{kt} = \text{diag}(i_{1k,t}, \ldots, i_{pk,t})$ with

$$i_{jk,t} = \begin{cases} 1 & \text{if } y_{jk,t} \text{ is observed,} \\ 0 & \text{if } y_{jk,t} \text{ is missing,} \end{cases}$$

where $y_t = \{y_{jk,t}\}$, $(j = 1, \ldots, p; k = 1, \ldots, r)$.

Then, the moments of equation (11) can be updated via

$$m_t = a_t + A_t e_t' \prod_{k=1}^{r} U_{kt}, \quad P_t = R_t - A_t Q_t A_t' u_t, \quad N_t = N_{t-1} + \sum_{k=1}^{r} U_{kt}$$

$$N_t^{1/2} S_t N_t^{1/2} = N_{t-1}^{1/2} S_{t-1} N_{t-1}^{1/2} + \left(\prod_{k=1}^{r} U_{kt} \right) e_t Q_t^{-1} e_t' \left(\prod_{k=1}^{r} U_{kt} \right),$$

where $u_t = \text{tr}(\prod_{k=1}^{r} U_{kt})/p$. Similar comments as in the case of $r = 1$ apply. Definition 1 is trivially extended in the case when observations form a matrix ($r \geq 2$).

We illustrate the proposed methodology by considering simulated data, consisting of 100 bivariate time series $y_1, \ldots, y_{100}$, generated from a local level model $y_t = [y_{1t} \ y_{2t}]' = \psi_t + \epsilon_t$ and $\psi_t = \psi_{t-1} + \zeta_t$, where ψ_0, ϵ_t and ζ_t are all simulated from bivariate normal distributions. The

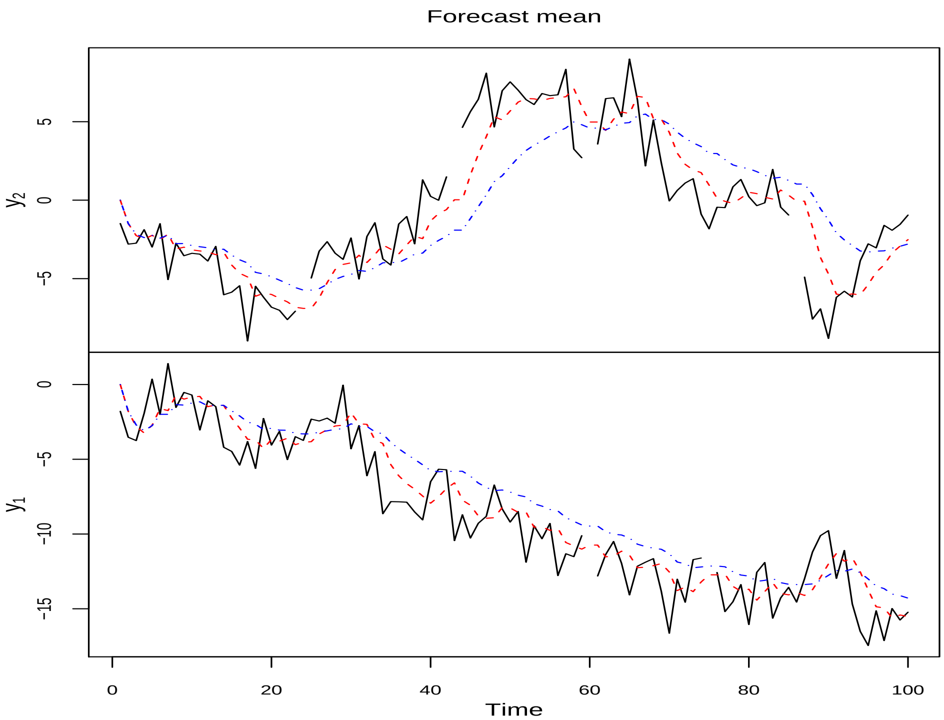

Figure 1: Simulated bivariate time series (solid line) with one-step forecasts from (a) the standard DLM recursions (dotted/dashed line) and (b) the new DLM recursions (dashed line). The gaps indicate missing values.

correlation of ϵ_{1t} and ϵ_{2t} is set to 0.8, while the elements of ζ_t are uncorrelated. This model is a special case of model (1) with $\Theta_t' F_t = \psi_t$ and $G_t = I_2$. Figure 1 (solid line) shows the simulated data; the gaps in this figure indicate missing values at times $t = 24, 43, 60, 75, 86$. At times $t = 24, 43, 86$, y_{t2} is only missing (partial missing vectors), at time $t = 75$, y_{t1} is only missing (partial missing vector) and at time $t = 60$, both y_{t1}, y_{t2} are missing (full missing vector). For this data set, we compare the performance of recursions (12)-(15) with that of the classic or old recursions of West and Harrison (1997), which assume that when there is at least one missing value we set $U_t = 0$ and $u_t = 0$. For example using the old recursions, for $t = 24$ one would set $U_{24} = 0$ and $u_{24} = 0$, losing the "partial" information of $y_{24,1} = -3.739$, which is observed. On the other hand, the new recursions would suggest for $t = 24$ to set

$$U_{24} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad u_{24} = 1/2.$$

Figure 1 shows the one-step forecast mean of $\{y_t\}$ using the new recursions (dashed line) and the old recursions (dotted/dashed line). We observe that the new method produces a clear improvement in the forecasts as the old recursions provide poor forecasts, especially in the low panel of Figure 1 (for $\{y_{1t}\}$). What is really happening in this case is that, under the old recursions, the missing values of y_{2t} affect the recursions for y_{1t}, since the observed information at y_{1t} is wrongly "masked" or "ignored" for the points of time when y_{2t} is missing. On the other hand, the new recursions use the explicit information from each sub-vector of y_t and thus the new recursions result in a notably more accurate forecast performance. This is backed by the mean square standardized forecast error vector, which for the new recursions is $[1.300\ 1.825]'$, while for the old recursions is $[1.545\ 2.182]'$. Under the old recursions we can not obtain an estimate of the covariance between an observed y_{1t} and a missing y_{2t}. However, this is indeed obtained under the proposed new recursions and so the respective correlations at points of time where there are gaps are 0.633 (at $t = 24$), 0.779 (at $t = 43$), 0.812 (at $t = 75$) and 0.809 (at $t = 86$); the mean of these correlations is 0.792, which is close to the real 0.8 under the simulation experiment.

Acknowledgements

I am grateful to Jeff Harrison for useful discussions on the topic of missing data in time series. I would like to thank a referee for helpful comments.

References

[1] Brown, P.J., Le, N.D. and Zidek, J.V. (1994) Inference for a covariance matrix. In *Aspects of Uncertainty* (eds P.R. Freeman, A.F.M. Smith). Chichester: Wiley.

[2] Dawid, A.P. (1981), Some matrix-variate distribution theory: notational considerations and a Bayesian application, *Biometrika*, **68**, 265-274.

[3] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B*, **39**, 1-38.

[4] Garthwaite, P.H. and Al-Awadhi, S.A. (2001) Non-conjugate prior distribution assessment for multivariate normal sampling. *Journal of the Royal Statistical Society Series B*, **63**, 95-110.

[5] Gómez, V. and Marvall, D. (1994) Estimation, prediction, and interpolation for nonstationary series with the Kalman filter. *Journal of the American Statistical Association*, **89**, 611-624.

[6] Harvey, A.C. and Pierse, R.G. (1984) Estimating missing observations in economic time series. *Journal of the American Statistical Association*, **79**, 125-131.

[7] Jones, R.H. (1980) Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, **22**, 125-131.

[8] Kohn, R. and Ansley, C.F. (1986) Estimation, prediction, and interpolation for ARIMA models with missing observations. *Journal of the American Statistical Association*, **81**, 751-761.

[9] Ljung, G.M. (1982) The likelihood function of a stationary Gaussian autoregressive-moving-average process with missing observations. *Biometrika*, **69**, 265-268.

[10] Ljung, G.M. (1993) On outlier detection in time series. *Journal of the Royal Statistical Society B*, **55**, 559-567.

[11] Luceño, A. (1994) A fast algorithm for the exact likelihood of stationary and partially nonstationary vector autoregressive moving average processes. *Biometrika*, **81**, 555-565.

[12] Luceño, A. (1997) Estimation of missing values in possibly partially nonstationary vector time series. *Biometrika*, **84**, 495-499.

[13] Salvador, M. and Gargallo, P. (2004). Automatic monitoring and intervention in multivariate dynamic linear models. *Computational Statistics and Data Analysis*, **47**, 401-431.

[14] Shumway, R.H. and Stoffer, D.S. (1982) An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, **3**, 253-264.

[15] West, M. and Harrison, P.J. (1997) *Bayesian Forecasting and Dynamic Models*, Springer-Verlag (second edition), New York.

[16] Wincek, M.A. and Reinsel, G.C. (1986) An exact maximum likelihood estimation procedure for regression-ARMA time series models with possibly nonconsecutive data. *Journal of the Royal Statistical Society B*, **48**, 303-313.