

Constrained Estimation of Mixture Vector Autoregressive Model

C.S. Wong^a

^a*Department of Finance, The Chinese University of Hong Kong, Shatin, Hong Kong*
Email: albertw@baf.msmail.cuhk.edu.hk

Abstract: A mixture vector autoregressive model has recently been introduced to the literature. This model consists of a mixture of finite number of vector autoregressive models. The first- and second-order stationarity conditions have been derived and an EM algorithm for estimation has been proposed. A mixture of stationary and nonstationary vector autoregressive components may still result in a stationary model. With the shape-changing predictive distribution, flexible conditional moment and autocorrelation structures, the mixture vector autoregressive model represents an attractive candidate for nonlinear multiple time series modeling. However, previous methodology has two limitations. First, method for computing or approximating the standard errors of the estimates has not been investigated. The approximation of the standard errors based on numerical procedures could be seriously inaccurate. Second, as a multivariate model, the MVAR model contains a large number of parameters in its specification. Estimation efficiency for short time series may be poor. Thus, it is important to consider procedures that eliminate unnecessary parameters and estimate the models with parameter constraints.

The contribution of this paper is threefold. First, a form of parameter constraints is introduced with an efficient EM algorithm for estimation. Second, an accurate method for computing standard errors is presented for the model with and without parameter constraints. Lastly, a hypothesis-testing approach based on likelihood ratio tests is proposed, which aids in the selection of unnecessary parameters and leads to the greater efficiency at the estimation. An example employing U.S. Treasury constant maturity rates illustrates the applicability of the mixture vector autoregressive model with parameter constraints, and the importance of using a reliable method to compute standard errors.

Keywords: EM algorithm, interest rate, likelihood ratio test, non-linear time-series model

1 INTRODUCTION

Over the past decade, several new time series models that generalize the idea of finite mixture distribution (Titterton et al., 1985) to the context of practical nonlinear time series model building have been proposed. Le et al. (1996) introduced Gaussian mixture transition distribution (GMTD) models to capture the flat stretches, bursts and outliers in univariate time series. The mixture autoregressive (MAR) models introduced by Wong and Li (2000) can be considered as a generalization of the GMTD models. MAR models have several advantages over other nonlinear time series models. First, the mixture of a nonstationary autoregressive (AR) component with a stationary AR component can result in a stationary process. Second, the conditional distributions of time series given the past history are changing over time. Lastly, MAR models can capture conditional heteroscedasticity (Engle, 1982), which is common in some financial time series. Several extensions of these models have been proposed: i.e., including conditional heteroscedasticity in the components, allowing the mixing proportion to be changing over time, and replacing the Gaussian assumption with the Student t -assumption. See Wong (2011), Wong and Li (2001a,b), and Wong et al. (2009) for more details.

Fong et al. (2007) recently generalized MAR models to mixture vector autoregressive (MVAR) models for multiple time series modeling. These models consist of a mixture of K vector autoregressive (VAR) models. These researchers gave the first- and second-order stationarity conditions and proposed an EM algorithm for estimation. Similarly to the univariate case, a stationary MVAR model may consist of a mixture of stationary and nonstationary VAR components. Although the MVAR model is a promising candidate for nonlinear multiple time series modeling, their methodology has two limitations. First, they presented no method for computing or approximating the standard errors of the estimates. The approximation of the standard errors based on numerical procedures could be seriously inaccurate which may be due to the irregularity of the log-likelihood surface for the EM algorithm, such as the existence of local maxima and flat likelihood surface. Reliance on numerical procedures to obtain standard errors may thus result in false conclusions. Second, as a multivariate model, the MVAR model contains a large number of parameters in its specification. Estimation efficiency for short time series may be poor. Thus, it is important to consider procedures that eliminate unnecessary parameters and estimate the models with parameter constraints.

In this paper, we introduce a form of parameter constraints for MVAR models that is similar to that presented for VAR models in Lütkepohl (2006). We call MVAR models with and without parameter constraints constrained and full MVAR models, respectively. We derive an EM algorithm for estimation in the constrained case. The standard errors of the estimates for both the full and constrained MVAR models are computed using Louis' (1982) method. In addition to the criterion-based model selection approach, we also propose a hypothesis-testing approach based on likelihood ratio tests for identifying the unnecessary parameters in MVAR models. Extensive simulation studies are conducted to verify the applicability of our approaches, especially in the case of short time series. Our proposed methodology is applied to the same dataset considered by Fong et al. (2007) and Tsay (2005), which demonstrates the usefulness of constrained estimation and the importance of approximating the standard errors accurately in MVAR modeling.

2 THE MIXTURE VECTOR AUTOREGRESSIVE MODEL

The K -component MVAR model for n -dimensional vector $\mathbf{Y}_t = (Y_{t,1}, \dots, Y_{t,n})'$, denoted by $\text{MVAR}(n, K; p_1, \dots, p_K)$, is defined as

$$F(\mathbf{Y}_t | \mathcal{F}_{t-1}) = \sum_{k=1}^K \alpha_k N \left[\boldsymbol{\Sigma}_k^{-1/2} (\mathbf{Y}_t - \boldsymbol{\Phi}_{k0} - \boldsymbol{\Phi}_{k1} \mathbf{Y}_{t-1} - \dots - \boldsymbol{\Phi}_{kp_k} \mathbf{Y}_{t-p_k}) \right]. \quad (1)$$

Here, $F(\mathbf{Y}_t | \mathcal{F}_{t-1})$ is the conditional cumulative distribution function of \mathbf{Y}_t , given past information; \mathcal{F}_t is the information set up to time t ; $N(\cdot)$ is the conditional cumulative distribution function of the standard multivariate Gaussian distribution; $\boldsymbol{\Phi}_{k0}$ is an n -dimensional vector, and $\boldsymbol{\Phi}_{ki}$ ($i = 1, \dots, p_k$) are the $n \times n$ autoregressive coefficient matrices; $\boldsymbol{\Sigma}_k$ is the $n \times n$ positive definite variance-covariance matrix for the k th component; $\alpha_1 + \dots + \alpha_K = 1$; and $\alpha_k > 0$ ($k = 1, \dots, K$). Let $p = \max(p_1, \dots, p_K)$.

We write $S_{i_1 i_2}$ for the (i_1, i_2) element of matrix \mathbf{S} and e_i for the i th element of the vector \mathbf{e} . We also write \mathbf{I}_n for the $n \times n$ identity matrix unless otherwise stated. The determinant of matrix \mathbf{S} is written as $|\mathbf{S}|$. The vectorization operator, vec , is defined as a column vector created from a matrix, say \mathbf{A} , by stacking the column vectors of \mathbf{A} one after another, and the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is defined as $(A_{ij} \mathbf{B})$. The half-vectorization operator, vech , stacks the column vectors in the lower triangle of a matrix. As an example, if $\mathbf{S} = (S_{ij})$ is a 2×2 matrix, then $\text{vech } \mathbf{S} = (S_{11} \ S_{21} \ S_{22})'$ and $\text{vec } \mathbf{S} = (S_{11} \ S_{21} \ S_{12} \ S_{22})'$. Here, the transpose of a vector or matrix is denoted by the symbol $'$. See Magnus and Neudecker (1988) for the relationships between the various vector and matrix operators.

2.1 Estimation of Full Model

Suppose that the observations $\mathbf{Y}_1, \dots, \mathbf{Y}_T$, are generated from the MVAR model (1), which is referred to as the full MVAR model hereafter. Define $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{K-1})'$, $\boldsymbol{\theta}_k = \text{vec } \boldsymbol{\Theta}_k$ where $\boldsymbol{\Theta}_k = (\boldsymbol{\Phi}_{k0}, \boldsymbol{\Phi}_{k1}, \dots, \boldsymbol{\Phi}_{kp_k})$ ($k = 1, \dots, K$), $\boldsymbol{\omega}_k = \text{vech } \boldsymbol{\Sigma}_k$ ($k = 1, \dots, K$), and the vector of the parameters is defined as $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\theta}'_1, \boldsymbol{\omega}'_1, \dots, \boldsymbol{\theta}'_K, \boldsymbol{\omega}'_K)'$. Let $\mathbf{Z}_t = (Z_{t,1}, \dots, Z_{t,K})'$ ($t = 1, \dots, T$) be the unobservable random vectors that indicate from which component \mathbf{Y}_t evolves, where $Z_{t,k} = 1$ if \mathbf{Y}_t evolves from the k th component and $Z_{t,k} = 0$ otherwise.

Given \mathbf{Z}_t , the log-likelihood of the complete data $(\mathbf{Y}_t, \mathbf{Z}_t)$ conditional on the first p observations is

$$\ell = \sum_{t=p+1}^T \left(\sum_{k=1}^K Z_{t,k} \log \alpha_k - \frac{1}{2} \sum_{k=1}^K Z_{t,k} \log |\boldsymbol{\Sigma}_k| - \frac{1}{2} \sum_{k=1}^K Z_{t,k} \mathbf{e}'_{kt} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_{kt} \right), \tag{2}$$

where $\mathbf{e}_{kt} = \mathbf{Y}_t - \boldsymbol{\Theta}_k \mathbf{X}_{kt}$ and $\mathbf{X}_{kt} = (1, \mathbf{Y}'_{t-1}, \dots, \mathbf{Y}'_{t-p_k})'$. The iterative EM algorithm for estimation can be found in Fong et al. (2007).

2.2 Estimation of Constrained Model

A form of linear constraints can be imposed on the parameters of the MVAR model (1). We use the form similar to the constraints imposed on the VAR model by Lütkepohl (2006, sec. 5.2.1). For $k = 1, \dots, K$, let m_k be the number of free parameters in the autoregressive coefficient matrices of component k , and $\boldsymbol{\theta}_k^{(r)}$ be the m_k -vector of free parameters. The constraints under consideration are

$$\boldsymbol{\theta}_k = \text{vec } \boldsymbol{\Theta}_k = \mathbf{R}_k \boldsymbol{\theta}_k^{(r)} + \text{vec } \mathbf{r}_k \quad (k = 1, \dots, K), \tag{3}$$

where \mathbf{R}_k s are known $n(np_k + 1) \times m_k$ matrices and \mathbf{r}_k s are known constant matrices with the same dimensions as $\boldsymbol{\Theta}_k$. We refer to the MVAR model (1) together with the constraints (3) as the constrained MVAR model. For each set of constraints imposed by m_k , \mathbf{R}_k , and \mathbf{r}_k ($k = 1, \dots, K$), there exists a $\{(K-1) + Kn(np_k + 1) + Kn(n+1)/2\} \times \{(K-1) + \sum_{k=1}^K m_k + Kn(n+1)/2\}$ matrix \mathbf{R} and a $\{(K-1) + Kn(np_k + 1) + Kn(n+1)/2\}$ -vector \mathbf{r} such that

$$\boldsymbol{\theta} = \mathbf{R} \boldsymbol{\theta}^{(r)} + \mathbf{r}, \tag{4}$$

where $\boldsymbol{\theta}^{(r)} = (\boldsymbol{\alpha}', \boldsymbol{\theta}_1^{(r)'}, \boldsymbol{\omega}'_1, \dots, \boldsymbol{\theta}_K^{(r)'}, \boldsymbol{\omega}'_K)'$.

As an example, suppose that the first component of an MVAR(2,2;1,1) model is constrained to have an upper triangular autoregressive matrix (i.e., upper triangular $\boldsymbol{\Phi}_{11}$). We then have $m_1 = 5$; $m_2 = 6$; \mathbf{R}_1 is a 6×5 matrix with $R_{1,11} = R_{1,22} = R_{1,33} = R_{1,54} = R_{1,65} = 1$ and $R_{1,ij} = 0$ otherwise; $\mathbf{R}_2 = \mathbf{I}_6$; $\boldsymbol{\theta}_1^{(r)} = (\Phi_{10,1}, \Phi_{10,2}, \Phi_{11,11}, \Phi_{11,12}, \Phi_{11,22})'$; $\boldsymbol{\theta}_2^{(r)} = \boldsymbol{\theta}_2$; $\mathbf{r}_1 = \mathbf{0}$; $\mathbf{r}_2 = \mathbf{0}$; \mathbf{R} is a block diagonal matrix with the elements or matrices $1, \mathbf{R}_1, \mathbf{I}_3, \mathbf{R}_2, \mathbf{I}_3$ on the diagonal; and $\mathbf{r} = \mathbf{0}$.

The conditional log-likelihood of the complete data $(\mathbf{Y}_t, \mathbf{Z}_t)$ for the constrained model is the same as in (2), except that it is being defined for the new vector of parameters $\boldsymbol{\theta}^{(r)}$. The log-likelihood function is maximized through an iterative EM procedure (Dempster et al., 1977) in which the E step and the M step are iterated until the likelihood converges.

E step The missing data \mathbf{Z}_t are replaced by their expectations, conditional on parameter $\boldsymbol{\theta}$ and on observed data $\mathbf{Y}_1, \dots, \mathbf{Y}_T$. The conditional expectation of the k th component of \mathbf{Z}_t is the conditional probability that observation \mathbf{Y}_t comes from the k th component of the mixture distribution, conditional on $\boldsymbol{\theta}$ and $\mathbf{Y}_1, \dots, \mathbf{Y}_T$. Let $\tau_{t,k}$ be the conditional expectation of the k th component of \mathbf{Z}_t . Then,

$$\tau_{t,k} = \frac{\alpha_k |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{e}'_{kt} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_{kt}\right)}{\sum_{k=1}^K \alpha_k |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{e}'_{kt} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_{kt}\right)} \quad (k = 1, \dots, K). \quad (5)$$

M step Suppose that the missing data are known. The estimates of parameters $\boldsymbol{\theta}$ can be obtained by setting the first derivatives of ℓ to zero. Then,

$$\hat{\alpha}_k = \frac{1}{T-p} \sum_{t=p+1}^T \tau_{t,k} \quad (k = 1, \dots, K), \quad (6)$$

$$\hat{\boldsymbol{\theta}}_k^{(r)} = \left\{ \mathbf{R}'_k \left[\left(\sum_{t=p+1}^T \tau_{t,k} \mathbf{X}_{kt} \mathbf{X}'_{kt} \right) \otimes \boldsymbol{\Sigma}_k^{-1} \right] \mathbf{R}_k \right\}^{-1} \times \left\{ \mathbf{R}'_k \sum_{t=p+1}^T \tau_{t,k} \text{vec} \left[\boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_t - \mathbf{r}_k \mathbf{X}_{kt}) \mathbf{X}'_{kt} \right] \right\} \quad (k = 1, \dots, K), \quad (7)$$

$$\hat{\boldsymbol{\Sigma}}_k = \left(\sum_{t=p+1}^T \tau_{t,k} \right)^{-1} \sum_{t=p+1}^T \tau_{t,k} \hat{\mathbf{e}}_{kt} \hat{\mathbf{e}}'_{kt} \quad (k = 1, \dots, K), \quad (8)$$

where $\hat{\mathbf{e}}_{kt} = \mathbf{Y}_t - \hat{\boldsymbol{\Theta}}_k \mathbf{X}_{kt}$. Due to the interdependence between $\hat{\boldsymbol{\theta}}_k^{(r)}$ and $\hat{\boldsymbol{\Sigma}}_k$, equations (7) and (8) are iterated until convergence.

2.3 Observed Information Matrix

The standard errors of the parameter estimates can be obtained by the missing information principle introduced by Louis (1982). The observed information matrix of the full model, \mathbf{I} , can be computed from the complete information matrix, \mathbf{I}_c , and the missing information matrix, \mathbf{I}_m , with the relation

$$\mathbf{I} = \mathbf{I}_c - \mathbf{I}_m = \mathbf{E} \left(-\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \middle| \boldsymbol{\theta}, \mathbf{Y} \right)_{\hat{\boldsymbol{\theta}}} - \text{var} \left(\frac{\partial \ell}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{\theta}, \mathbf{Y} \right)_{\hat{\boldsymbol{\theta}}}. \quad (9)$$

For the constrained model, the observed information matrix, $\mathbf{I}^{(r)}$, can be computed from \mathbf{I} with

$$\mathbf{I}^{(r)} = \mathbf{R}' \mathbf{I} \mathbf{R}. \quad (10)$$

The variance matrix of estimates $\hat{\boldsymbol{\theta}}$ ($\hat{\boldsymbol{\theta}}^{(r)}$) is given by the inverse of observed information matrix \mathbf{I} ($\mathbf{I}^{(r)}$). The variance of estimate $\hat{\alpha}_K = 1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_{K-1}$ is given by $\sum_{k=1}^{K-1} \text{var}(\hat{\alpha}_k) + \sum_{k=1}^{K-1} \sum_{l=1, l \neq k}^{K-1} \text{cov}(\hat{\alpha}_k, \hat{\alpha}_l)$.

3 MODEL SELECTION AND HYPOTHESIS TESTING

Model selection for MVAR models comprises several aspects. First, the number of components, K , is an important parameter to be determined. Second, we have to select the orders of the vector autoregressive components, p_k s. Finally, we may need to omit some of the insignificant parameters in the autoregressive coefficient matrices. Note that for univariate mixture time series models, the last aspect of model selection is of less interest, as the efficiency gain in parameter estimation by dropping insignificant parameters is minimal.

The testing and model selection problems for K are difficult to handle, as they correspond to testing problems with nuisance parameters, which do not exist under the null hypothesis (Davies, 1977, 1987).

In particular, the likelihood ratio statistic for a hypothesis such as $\Phi_{1i} = \Phi_{2i}$ for all $i = 0, \dots, p$, does not have the standard χ^2 null distribution. Fong et al. (2007) suggested the use of the Bayesian information criterion (BIC) (Schwarz, 1978) to solve this problem. Through simulation experiments, they illustrated that the BIC exhibits satisfactory performance in the selection of K . In the context of the constrained estimation of MVAR models, the BIC is defined as $-2\ell^* + \log(T-p)\{\sum_{k=1}^K m_k + Kn(n+2)/2 + K - 1\}$, where $\ell^* = \sum_{t=p+1}^T \log f(\mathbf{Y}_t | \mathcal{F}_{t-1})$ is the log-likelihood based on $f(\mathbf{Y}_t | \mathcal{F}_{t-1})$, the first derivatives of (1).

After the number of components, K , in the MVAR model has been selected, the BIC can be used to select parameters p_k s, as well as the insignificant parameters to be dropped in the autoregressive coefficient matrices. Alternatively, classical likelihood ratio tests can be performed when the hypotheses are specified with embedded models, such as that a single parameter in an autoregressive coefficient matrix is zero and the autoregressive coefficient matrix at a particular lag is zero. The likelihood ratio statistics are defined as $-2(\ell_0^* - \ell_1^*)$ where ℓ_0^* and ℓ_1^* are the maximized log-likelihood ℓ^* under the null and alternative hypotheses, respectively, and they should have the standard χ^2 null distributions. The availability of the likelihood ratio test enhances the procedure for parameter selection and hence improves estimation efficiency by reducing the number of free parameters. Care must be taken, however, because two MVAR models with the same orders may not be embedded, as the model structures may be completely different. For example, a two-component MVAR model with $\alpha_1 = \alpha_2 = 0.5$ is considered to have a different structure to that of a two-component model with $\alpha_1 = 0.95$ and $\alpha_2 = 0.05$.

4 SIMULATION STUDIES

We have performed a number of simulation experiments for assessing the performance of the constrained estimation, the model selection with the BIC and the likelihood ratio tests. The results reveal that the EM estimation procedure for constrained MVAR models has small biases and reasonable standard errors even with small sample, the performance of the BIC in the selection of K and p_k s is satisfactory, and the likelihood ratio tests have reasonable empirical sizes and power.

5 EXAMPLE: U.S. TREASURY CONSTANT MATURITY RATES

The constrained MVAR modeling and the importance of approximating standard errors via Louis' method are illustrated using monthly data on 1-year and 3-year U.S. Treasury constant maturity rates from April 1953 to January 2001. These data can be obtained from the Federal Reserve Bank of St. Louis at *stlouisfed.org*. Fong et al. (2007) and Tsay (2005, sec. 8.4) also investigated these data and drawn slightly different conclusions.

The original series are clearly nonstationary. The log-differenced series seem to be stationary with no large difference in variability over time despite the two large extreme values. Let $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2})'$ be the vector of the log-differenced, or, equivalently, the percentage change in, 1-year and 3-year Treasury constant maturity rates. The log-differenced series have a length of 573. The log-differenced data exhibit heavy-tailedness in comparison with a bivariate normal distribution. The excess kurtoses are 5.34 and 1.77 for $Y_{t,1}$ and $Y_{t,2}$, respectively. Models, which imply a marginal bivariate normal distribution, such as the VAR model, may not capture the features of this data well.

We consider two- and three-component MVAR models with the maximum allowable order of four for the autoregressive components. The number of parameters in the most complicated models are 43 and 65 for the MVAR(2,2;4,4) and MVAR(2,3;4,4,4) models, respectively. We start our selection process with the value of p_k s. The initial values for the EM estimation are $\alpha_k = 1/K$ and $\Sigma_k = \mathbf{I}_2$ ($k = 1, \dots, K$), and random numbers are generated from the uniform distribution on the interval $[-0.8, 0.8]$ for the parameters in θ_k ($k = 1, \dots, K$). For each model, 50 sets of estimation results are obtained with different initial values. The estimation results based on different initial values can differ substantially due to the presence of local maxima in the log-likelihood. The best two-component model is the MVAR(2,2;2,1) model with a BIC of -6509.4 . The likelihood ratio tests against other candidate two-component models confirm that this model is the most parsimonious one that is still statistically acceptable. After examining all possible three-component models, the MVAR(2,3;2,2,1) model with a BIC of -6502.9 is selected. As the MVAR(2,2;2,1) model has a smaller BIC, we examine it in greater detail. The full MVAR(2,2;2,1)

Table 1. Estimated full MVAR(2,2;2,1) model

Parameter	α_1	$\Sigma_{1,11}$	$\Sigma_{1,12}$	$\Phi_{10,1}$	$\Phi_{11,11}$	$\Phi_{11,12}$	$\Phi_{12,11}$	$\Phi_{12,12}$
Estimate	0.375	0.00708	0.00451	-0.003	0.359	0.476	-0.056	-0.203
s.e.	0.051	0.00089	0.00057	0.006	0.181	0.242	0.174	0.239
Parameter		$\Sigma_{1,21}$	$\Sigma_{1,22}$	$\Phi_{10,2}$	$\Phi_{11,21}$	$\Phi_{11,22}$	$\Phi_{12,21}$	$\Phi_{12,22}$
Estimate		0.00451	0.00367	-0.001	0.290	0.295	0.092	-0.469
s.e.		0.00057	0.00043	0.005	0.132	0.177	0.126	0.176
Parameter	α_2	$\Sigma_{2,11}$	$\Sigma_{2,12}$	$\Phi_{20,1}$	$\Phi_{21,11}$	$\Phi_{21,12}$		
Estimate	0.625	0.00128	0.00114	0.003	0.031	0.256		
s.e.	0.051	0.00018	0.00014	0.002	0.108	0.144		
Parameter		$\Sigma_{2,21}$	$\Sigma_{2,22}$	$\Phi_{20,2}$	$\Phi_{21,21}$	$\Phi_{21,22}$		
Estimate		0.00114	0.00117	0.001	-0.065	0.347		
s.e.		0.00014	0.00013	0.002	0.103	0.131		

s.e. = standard error.

Table 2. Estimated constrained MVAR(2,2;2,1) model

Parameter	α_1	$\Sigma_{1,11}$	$\Sigma_{1,12}$	$\Phi_{10,1}$	$\Phi_{11,11}$	$\Phi_{11,12}$	$\Phi_{12,11}$	$\Phi_{12,12}$
Estimate	0.358	0.00749	0.00477	-0.003	0.668	0.000	-0.189	0.000
s.e.	0.051	0.00096	0.00061	0.007	0.084	–	0.066	–
Parameter		$\Sigma_{1,21}$	$\Sigma_{1,22}$	$\Phi_{10,2}$	$\Phi_{11,21}$	$\Phi_{11,22}$	$\Phi_{12,21}$	$\Phi_{12,22}$
Estimate		0.00477	0.00385	-0.001	0.479	0.000	0.000	-0.338
s.e.		0.00061	0.00046	0.005	0.062	–	–	0.070
Parameter	α_2	$\Sigma_{2,11}$	$\Sigma_{2,12}$	$\Phi_{20,1}$	$\Phi_{21,11}$	$\Phi_{21,12}$		
Estimate	0.642	0.00132	0.00117	0.003	0.082	0.229		
s.e.	0.051	0.00018	0.00014	0.002	0.036	0.083		
Parameter		$\Sigma_{2,21}$	$\Sigma_{2,22}$	$\Phi_{20,2}$	$\Phi_{21,21}$	$\Phi_{21,22}$		
Estimate		0.00117	0.00119	0.001	0.000	0.296		
s.e.		0.00014	0.00013	0.002	–	0.058		

s.e. = standard error.

model is shown in Table 1.

The parameters in the autoregressive matrices with small absolute estimate to standard error ratios are dropped sequentially starting with the off-diagonal elements. The new models are estimated with the methodology employed in Sections 2.2 and 2.3. Likelihood ratio tests are used to check whether the dropped parameters should be retained in the models. Based on this strategy, the sequence of parameters dropped is $\Phi_{21,21}$, $\Phi_{12,21}$, $\Phi_{12,12}$, $\Phi_{11,22}$, and $\Phi_{11,12}$. The estimated constrained MVAR(2,2;2,1) model with 18 parameters is shown in Table 2. The values of the log-likelihood and BIC are 3325.04 and -6535.9 , respectively. A likelihood ratio test against the full MVAR(2,2;2,1) model gives a statistic of 5.22 with a p -value of 0.390, which indicates that the constrained model is preferable.

In the constrained MVAR(2,2;2,1) model, the log-differenced 1-year and 3-year interest rates are modeled with two VAR components with different degrees of variability. The first component consists of larger variances ($\hat{\Sigma}_{1,11} = 0.00749$ and $\hat{\Sigma}_{1,22} = 0.00385$) and a mixing proportion of 0.358. In this component, the 3-year interest rate depends on the past values of the 1-year interest rate but not vice versa. Changes in the short-term interest rate have some influence on the long-term interest rate in a volatile interest rate market. For the second component, with a mixing proportion of 0.642, the variances are smaller, with $\hat{\Sigma}_{2,11} = 0.00132$ and $\hat{\Sigma}_{2,22} = 0.00119$. In this component, the 1-year interest rate depends on the past values of the 3-year interest rate but not vice versa. Changes in the long-term interest rate have some influence on the short-term interest rate in a peaceful interest rate market.

Our conclusion is somewhat different from that of Tsay (2005, p. 374), who suggested that the 3-year

interest rate does not depend on the past values of the 1-year interest rate. His conclusion is based on VAR modeling, which may be inappropriate as the log-differenced interest rate exhibits heavy-tailedness. Fong et al. (2007) suggested that the 3-year interest rate depends on the past values of the 1-year interest rate but not vice versa in the first component, and that the two interest rate series are mutually correlated in the second component, based on a full MVAR(2,2;3,4) model. However, they computed the standard errors of the estimates based on a numerical procedure that may be inappropriate because of the irregularity in the log-likelihood surface. For example, the estimates and standard errors for $\text{vec } \Phi_{24}$ are $(-0.169, -0.090, 0.196, 0.096)'$ and $(0.012, 0.015, 0.030, 0.022)'$, respectively in their estimated model (Fong et al., 2007). Using Louis' (1982) method which is presented in Section 2.3, the standard errors are $(0.102, 0.093, 0.118, 0.113)'$, which implies that none of the parameters in $\text{vec } \Phi_{24}$ has a large estimate to standard error ratio. This finding illustrates the importance of using Louis' method to compute the standard errors after EM estimation.

REFERENCES

- Davies, R. B (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.
- Davies, R. B (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.
- Dempster, A. P., N. M. Laird and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *Journal of the Royal Statistical Society B* 39, 1–38.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Fong, P., W. K. Li, C. W. Yau and C. S. Wong, (2007). On a mixture vector autoregressive model. *The Canadian Journal of Statistics* 35, 135–150.
- Le, N. D., R. D. Martin, and A. E. Raftery (1996). Modelling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association* 91, 1504–1514.
- Louis, T. A (1996). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society B* 44, 226–233.
- Lütkepohl, H. (2006) *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- Magnus, J. R. and H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: Wiley.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.
- Titterton, D. M., A. F. M. Smith, and U. E. Makov (1985) *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.
- Tsay, R. S. (2005) *Analysis of Financial Time Series*. New York: Wiley.
- Wong, C. S. (2011). Modeling Hong Kong's stock index with the Student t -mixture autoregressive model. *Mathematics and Computers in Simulation* 81, 1334–1343.
- Wong, C. S., W. S. Chan and P. L. Kam (2009). A Student t -mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika* 96, 751–760.
- Wong, C. S. and W. K. Li (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society B* 62, 95–115.
- Wong, C. S. and W. K. Li (2001a). On a logistic mixture autoregressive model. *Biometrika* 88, 833–846.
- Wong, C. S. and W. K. Li (2001b). On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association* 96, 982–95.