

Modeling Panel Time Series with Mixture Autoregressive Model

Shusong Jin and Wai Keung Li
The University of Hong Kong

Abstract: This paper considers the mixture autoregressive panel (MARP) model. This model can capture the burst and multi-modal phenomenon in some panel data sets. It also enlarges the stationarity region of the traditional AR model. An estimation method based on the EM algorithm is proposed and the assumption required of the model is quite low. To illustrate the method, we fitted the MARP model to the gray-sided voles data. Another MARP model with less restriction is also proposed.

Key words: EM algorithm, finite mixture model, gray-sided voles data, nonlinear time series, panel data.

1. Introduction

Panel time series data are collections of similar time series variables. A general linear dynamic model for a panel time series $\{X_{jt}\}$, $j = 1, \dots, M$, $t = 1, \dots, T_j$ is given by

$$X_{jt} = \phi_{0j} + \sum_{k=1}^p \phi_{kj} X_{jt-k} + \eta_t + \lambda_j + \beta^T W_{jt} + \varepsilon_{jt}, \quad (1.1)$$

where t denotes time and j indicate the individuals. Moreover, η_t represents effects for all of the series and λ_j denotes the individual impact of each of the series. The other variables consist of the vector of explanatory variables W_{jt} and the error term ε_{jt} . The latter are assumed to be independent identically distributed. For details, see Hsiao (1986, p. 71) and Hjellvik and Tjøstheim (1999). Panel data have now become more and more important both in the natural science and in macroeconomic study. For instance, gray-sided voles data are yearly observations from ninety-one different positions. In macroeconomics, two very famous data sets are the National Longitudinal Survey of Labor Market Experience (NLS) and the Michigan Panel Study of Income Dynamics (PSID). Traditionally, in panel data, there are many different individuals observed through time, but the number of observations of each individual is relatively small. In many situations, time series data display non-linear features such as flat stretches, bursts and change

points. At the same time, there could also be many missing datum points and outliers inside a time series data set. These are no exceptions for panel time series data. All these non-linearity and missing information make the modeling of panel time series more difficult than ordinary time series data. Many papers tried to overcome the difficulties in panel data modeling. Hjellvik and Tjøstheim (1999) treated the common factor η_t in model (1.1) as nuisance parameters and estimated them by averaging the observations across all the series at fixed time point t . Fu *et al.* (2002b) assumed that the series were contemporaneously correlated and proposed a contemporaneous correlated threshold model for panel data analysis. In the nonparametric and semiparametric literature, Hjellvik *et al.* (2004) considered a nonparametric estimation and testing procedure of the nonlinearity for panel data. Lin and Ying (2001) and Fan and Li (2004) suggested to model the panel data with semiparametric method. Recently, Jin and Li (2005) considered the modeling of a contemporaneously correlated panel data with partial linear regression model. These papers partly solved the problem of change points, nonlinearity and intercorrelation. However, other features such as flat stretches, bursts and multi-modality are still untouched.

Le *et al.* (1996) introduced the Gaussian mixture transition distribution (GMTD) model. With this model, one can capture the flat stretches, bursts and outliers in regression analysis. Wong and Li (2000) extended GMTD model to time series analysis and suggested the mixture autoregressive model (MAR) to catch the multi-modal phenomena. An N -component univariate MAR model is defined almost surely by equation (1.2).

$$F(x_t|\mathcal{F}_{t-1}) = \sum_{i=1}^N \pi_i \Phi \left(\frac{x_t - \phi_{i0} - \phi_{i1}x_{t-1} - \cdots - \phi_{ip_i}x_{t-p_i}}{\sigma_i} \right), \quad (1.2)$$

where $F(x_t|\mathcal{F}_{t-1})$ is the conditional cumulative distribution function of X_t given the past information, evaluated at x_t . The σ -field \mathcal{F}_{t-1} denotes the past information set up to time $t-1$ and the function $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and the summation of the probability $\sum_{i=1}^N \pi_i = 1$. For details, see Wong and Li (2000). From our experience, many panel time series appeared to be multi-modal. For example, Figure 1 shows four typical series in the gray-sided voles data. The four series are stationary but it is quite obvious that they are multi-modal. Hence, developing a mixture model to capture the multi-modal characteristics in panel time series would be of importance.

This paper is arranged as follows. In section 2, we propose the panel mixture autoregressive model which assumes that all the panels have the same mixture probability π_i , $i = 1, \dots, N$ and one of the components has a common autoregressive structure. We develop an estimation method based on the EM algorithm. In

section 3, a simulation experiment with different series length was performed to check our model and the estimation method. The minimum sample size requirement needed for the application of our model is also suggested. The gray-sided voles data are studied in section 4 as a real example of the application of our model. In section 5, we will relax the restriction that the series has the same mixture probabilities and extend the MARP model to a more flexible model. Lastly, we will draw some conclusions in section 6.

2. Model and Estimation

Our N -component finite mixture autoregressive model for a panel time series (MARP) X_{jt} , $j = 1, \dots, M$, $t = 1, \dots, T_j$ in conditional cumulative distribution function (CDF) form, is given by

$$F(X_{jt}|\mathcal{F}_{t-1}) = \sum_{i=1}^{N-1} \pi_i \Phi \left(\frac{X_{jt} - \phi_{0ij} - \sum_{k=1}^p \phi_{kij} X_{j,t-k}}{\sigma_{ij}} \right) + \pi_N \Phi \left(\frac{X_{jt} - \phi_{0N} - \sum_{k=1}^p \phi_{kN} X_{j,t-k}}{\sigma_{Nj}} \right), \quad (2.1)$$

where $i = 1, \dots, N$ is the index of the components, $j = 1, \dots, M$ is the index of the series, $t = 1, \dots, T_j$ is the time index, k is the index of the lags and p is the autoregressive order. Here, we assume that the order of each component in each series is the same. If the orders p'_{ij} are not the same, we can set $p = \max_{ij} \{p'_{ij}\}$ and those $\phi_{kij} = 0$ when $k > p'_{ij}$. In the conditional CDF, the random variable X_{jt} is evaluated at x_{jt} given the past information up to time $t - 1$ which is denoted by \mathcal{F}_{t-1} . The random variable X_{jt} is taken from the i th component with probability π_i where $\sum_{i=1}^N \pi_i = 1$ and $\pi_N = 1 - \pi_1 - \dots - \pi_{N-1}$. For identifiability we assume $\pi_1 \leq \pi_2 \leq \dots \leq \pi_N$ (see McLachlan and Peel (2000), 2000). This assumption is only for the theoretical purpose and is not used as the restriction in our model estimation. As usual, $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Although we assume that the order p are the same, however, the series length T_j are not assumed to be same. In panel data analysis, the data often have some common features. For example, in section 4 of this paper, we studied the gray-sided voles data which were collected from different positions in Hokkaido, Japan. It is reasonable to believe that the different series would follow some common pattern. To capture the common structure among the series, we assume that the last component in each of the series has the same coefficients ϕ_{kN} , $k = 0, \dots, p$. The error terms ε_{ijt} are assumed to be mutually independent and from the normal distribution $N(0, \sigma_{ij}^2)$. Although the noises of each component are assumed to be Gaussian, the composition noise of each series by finite mixture are non-Gaussian and can

display more flexibility in modeling. This gives us much freedom to fit some complicated data set.

Model (2.1) assumes that the mixture probabilities are the same across the panel. It seems that this condition is a bit strict. In time series, to compensate the internal dependence of the data set, we often require a bit longer series than that in the independent case so that we can obtain good parameter estimates. However, in traditional panel time series data, each of the series contains only a few observations. For example, in the gray-sided voles data, each series contains only 31 observations. This sample size is also a bit small for statistical analysis even in the independent cases. To obtain accurate estimation, it is reasonable to assume that all these series have the same mixing probability so that we can use all the information we can obtain. When series length is a bit longer, we can remove this assumption. We will relax this restriction in Section 5.

Another advantage of the MARP model is that this model will widen the regions of the first order and second order stationarity. Wong and Li (2000) proved that the necessary and sufficient condition for the first order is that the roots z_1, \dots, z_p of the equation $1 - \sum_{k=1}^p (\sum_{i=1}^{N-1} \pi_i \phi_{ki}) z^{-i} - \sum_{k=1}^p \pi_N \phi_{kN} z^{-i} = 0$ all lie inside the unit circle. The necessary and sufficient condition of second order stationarity is $|\sum_{i=1}^{N-1} \pi_i \phi_{1ij}^2 + \pi_N \phi_{1N}^2| < 1$. For details, see Wong and Li (2000). These results show that the mixture of non-stationary process and stationary process can be stationary.

Let $\boldsymbol{\theta}_{ij} = (\phi_{0ij}, \dots, \phi_{pij}, \sigma_{ij}^2)^T$, $\boldsymbol{\theta}_j = (\boldsymbol{\theta}_{1j}^T, \dots, \boldsymbol{\theta}_{N-1,j}^T)^T$, $\boldsymbol{\theta}_{NL} = (\phi_{0N}, \dots, \phi_{pN}, \sigma_{N1}^2, \dots, \sigma_{NM}^2)^T$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_M^T, \boldsymbol{\theta}_{NL}^T)^T$. Here the subscript L in $\boldsymbol{\theta}_{NL}$ stands for "last" component so that we can avoid the confusion of the parameters in the component. Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)^T$. Let $\mathbf{Z}_{jt} = (Z_{1jt}, Z_{2jt}, \dots, Z_{Njt})^T$, $\mathbf{Z}_j = (\mathbf{Z}_{j1}^T, \mathbf{Z}_{j2}^T, \dots, \mathbf{Z}_{jT_j}^T)^T$ and $\mathbf{Z} = (\mathbf{Z}_1^T, \dots, \mathbf{Z}_M^T)^T$. The vector \mathbf{Z}_{jt} contains the unobservable random variable Z_{ijt} where $Z_{ijt} = 1$ when at time t , in the j th series, the X_{jt} comes from the i th component of the conditional distribution function and $Z_{ijt} = 0$ otherwise. Then the parameters are naturally divided into two groups, $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ with \mathbf{Z} together. With the help of Z_{ijt} , we can rewrite model (2.1) as (2.2).

$$X_{jt} = \sum_{i=1}^{N-1} Z_{ijt}(\phi_{0ij} + \sum_{k=1}^p \phi_{kij} X_{j,t-k} + \varepsilon_{ijt}) + Z_{Njt}(\phi_{0N} + \sum_{k=1}^p \phi_{kN} X_{j,t-k} + \varepsilon_{Njt}), \quad (2.2)$$

where ε_{ijt} is the white noise process corresponding to the i th component of the j th series. The $\{\varepsilon_{ijt}\}$ are also independent for $i = 1, \dots, N$, $j = 1, \dots, M$ and all t . If we study equation (2.2) closely, we will find that the MARP model is actually a mixture of N Gaussian AR models for each of the series. The conditional mean

of series j is

$$E(X_{jt}|\mathcal{F}_{t-1}) = \sum_{i=1}^{N-1} \pi_i(\phi_{0ij} + \sum_{k=1}^p \phi_{kij}x_{j,t-k}) + \pi_N(\phi_{0N} + \sum_{k=1}^p \phi_{kN}x_{j,t-k}).$$

Since the conditional mean depends on the past information of the time series, the shape of the conditional distributions will change from time to time and can be uni-modal or multi-modal. It can be proved that for a stationary mixture autoregressive time series X_t , the marginal distribution of X_t is also a mixture distribution. Due to this nature, the marginal distribution could be multi-modal if the conditional distribution is multi-modal. McLachlan and Peel (2000) showed some examples of uni-modal and multi-modal distributions in static situations and Wong and Li (2000) demonstrated the presence of multi-modality in dynamic situations. Note that for mixtures, the conditional mean could take the value at a local minimum point of the conditional mixture density and therefore it could be a misleading predictor of the future values.

Since \mathbf{Z} is the vector of unobservable random variables, maximizing the log-likelihood function directly from the samples would be difficult. To overcome this difficulty, Dempster *et al.* (1977) suggested the EM algorithm which substitutes the unobservable variables by their expectations and then maximize the log-likelihood function. Using the expectation and maximization iteratively, one can obtain the estimates of the parameters when they converge. To use the EM algorithm, we should first find the log-likelihood function. The conditional log-likelihood of the j th series involving Z_{ijt} is given by

$$\begin{aligned} l_j &= \sum_{t=p+1}^{T_j} l_{jt} \\ &= \sum_{t=p+1}^{T_j} \left\{ \sum_{i=1}^N Z_{ijt} \log(\pi_i) - \frac{1}{2} \sum_{i=1}^N Z_{ijt} \log(\sigma_{ij}^2) - \sum_{i=1}^N \frac{Z_{ijt} \varepsilon_{ijt}^2}{2\sigma_{ij}^2} \right\}, \end{aligned}$$

The global conditional log-likelihood function given \mathbf{Z} for the entire panel is

$$\begin{aligned} l &= \sum_{j=1}^M l_j = \sum_{j=1}^M \sum_{t=p+1}^{T_j} l_{jt} \\ &= \sum_{j=1}^M \sum_{t=p+1}^{T_j} \left\{ \sum_{i=1}^N Z_{ijt} \log(\pi_i) - \frac{1}{2} \sum_{i=1}^N Z_{ijt} \log(\sigma_{ij}^2) - \sum_{i=1}^N \frac{Z_{ijt} \varepsilon_{ijt}^2}{2\sigma_{ij}^2} \right\}. \quad (2.3) \end{aligned}$$

Since the first derivatives of the log-likelihood with respect to the interested parameters can be obtained straightforwardly, we skip them here. With the first

derivatives, we can apply the EM algorithm. The iterative procedure of the EM algorithm can be divided into two steps – E-step and M-step respectively.

E-step In this step, the parameter vector θ are assumed to be known. Then the missing data \mathbf{Z} will be estimated by their conditional expectation respecting to θ and the observation vector $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_M^T)^T$ where $\mathbf{X}_j = (X_{j1}, \dots, X_{jT_j})^T$. Let τ_{ijt} denote the conditional expectation of Z_{ijt} . Similar to Le *et al.* (1996) and Wong and Li (2000), the τ_{ijt} can be expressed as

$$\tau_{ijt} = \frac{\pi_i / \sqrt{\sigma_{ij}^2} \cdot \exp\{-\varepsilon_{ijt}^2 / (2\sigma_{ij}^2)\}}{\sum_{i'=1}^N \pi_{i'} / \sqrt{\sigma_{i'j}^2} \cdot \exp\{-\varepsilon_{i'jt}^2 / (2\sigma_{i'j}^2)\}}, \quad (2.4)$$

where $i = 1, \dots, N$, $j = 1, \dots, M$, $t = 1, \dots, T_j$ and ε_{ijt} has the same meaning as before.

M-step In this step, we replace Z_{ijt} by its conditional expectation τ_{ijt} . Therefore, π_i can be estimated by the mean of τ_{ijt} , $j = 1, \dots, M$.

$$\hat{\pi}_i = \frac{\sum_{j=1}^M \sum_{t=p+1}^{T_j} \tau_{ijt}}{\sum_{j=1}^M (T_j - p)}.$$

We estimate the parameter vector θ by maximizing the global log-likelihood function l . This can be done by setting the first derivatives to zero and then solve the system of equations. The system can be simplified as follows. Let

$$\begin{aligned} \hat{\sigma}_{ij}^2 &= \frac{\sum_{t=p+1}^{T_j} \tau_{ijt} \varepsilon_{ijt}^2}{\sum_{t=p+1}^{T_j} \tau_{ijt}} \\ &= \frac{\sum_{t=p+1}^{T_j} \tau_{ijt} (X_{jt} - \hat{\phi}_{0ij} - \sum_{k=1}^p \hat{\phi}_{kij} X_{j,t-k})^2}{\sum_{t=p+1}^{T_j} \tau_{ijt}}, \end{aligned} \quad (2.5)$$

where $(\hat{\phi}_{011}, \hat{\phi}_{111}, \dots, \hat{\phi}_{p,N-1,M})$ is determined by another system of equations:

$$\left(\sum_{t=p+1}^{T_j} \tau_{ijt} \right) \hat{\phi}_{0ij} + \sum_{k=1}^p \left(\sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-k} \right) \hat{\phi}_{kij} = \sum_{t=p+1}^{T_j} \tau_{ijt} X_{jt}, \quad (2.6)$$

$$\begin{aligned} &\left(\sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-k'} \right) \hat{\phi}_{0ij} + \sum_{k=1}^p \left(\sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-k} X_{j,t-k'} \right) \hat{\phi}_{kij} \\ &= \sum_{t=p+1}^{T_j} \tau_{ijt} X_{jt} X_{j,t-k'}, \end{aligned} \quad (2.7)$$

where $i = 1, \dots, N-1, j = 1, \dots, M, k' = 1, \dots, p$. For the N th component, since the ϕ_{ki} are common to all the series, the estimation equations are a bit different from (2.5) — (2.7). By pooling all available information, we can write the equation system of $\hat{\phi}_{0N}, \dots, \hat{\phi}_{pN}, \hat{\sigma}_{N1}^2, \dots, \hat{\sigma}_{NM}^2$ as follows:

$$\begin{aligned} \hat{\sigma}_{Nj}^2 &= \frac{\sum_{t=p+1}^{T_j} \tau_{Njt} \varepsilon_{Njt}^2}{\sum_{t=p+1}^{T_j} \tau_{Njt}} \\ &= \frac{\sum_{t=p+1}^{T_j} \tau_{Njt} (X_{jt} - \hat{\phi}_{0N} - \sum_{k=1}^p \hat{\phi}_{kN} X_{j,t-k})^2}{\sum_{t=p+1}^{T_j} \tau_{Njt}}, \end{aligned} \tag{2.8}$$

$$\begin{aligned} &\sum_{j=1}^M \left[\frac{\sum_{t=p+1}^{T_j} \tau_{Njt}}{\hat{\sigma}_{Nj}^2} \right] \hat{\phi}_{0N} + \sum_{k=1}^p \left\{ \left[\sum_{j=1}^M \frac{\sum_{t=p+1}^{T_j} \tau_{Njt} X_{j,t-k}}{\hat{\sigma}_{Nj}^2} \right] \hat{\phi}_{kN} \right\} \\ &= \sum_{j=1}^M \left[\frac{\sum_{t=p+1}^{T_j} \tau_{Njt} X_{jt}}{\hat{\sigma}_{Nj}^2} \right], \end{aligned} \tag{2.9}$$

$$\begin{aligned} &\sum_{j=1}^M \left[\frac{\sum_{t=p+1}^{T_j} \tau_{N,j,t}}{\hat{\sigma}_{Nj}^2} \right] \hat{\phi}_{0,N} + \sum_{k=1}^p \left\{ \left[\sum_{j=1}^M \frac{\sum_{t=p+1}^{T_j} \tau_{N,j,t} X_{j,t-k}}{\hat{\sigma}_{N,j}^2} \right] \hat{\phi}_{k,N} \right\} \\ &= \sum_{j=1}^M \left[\frac{\sum_{t=p+1}^{T_j} \tau_{N,j,t} X_{j,t}}{\hat{\sigma}_{N,j}^2} \right], \end{aligned} \tag{2.10}$$

Studying the system of equations carefully, we find that, by fixing j , to the i th component, equations (2.6) and (2.7) contain $p + 1$ linear equations with $p + 1$ unknown variables $\phi_{0ij}, \dots, \phi_{pij}$. Hence, the parameters of the i th component in the j th series, $i = 1, \dots, N - 1$ can be solved separately from the other components. Let \mathbf{A}_{ij} be the matrix below:

$$\begin{bmatrix} \sum_{t=p+1}^{T_j} \tau_{ijt} & \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-1} & \cdots & \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-p} \\ \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-1} & \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-1} X_{j,t-1} & \cdots & \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-p} X_{j,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-p} & \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-1} X_{j,t-p} & \cdots & \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t-p} X_{j,t-p} \end{bmatrix}$$

and let $\hat{\phi}_{ij}$ and \mathbf{b}_{ij} be

$$\begin{aligned} \hat{\phi}_{ij}^T &= \left[\hat{\phi}_{0ij}, \hat{\phi}_{1ij}, \dots, \hat{\phi}_{pij} \right] \\ \mathbf{b}_{ij}^T &= \left[\sum_{t=p+1}^{T_j} \tau_{ijt} X_{jt}, \sum_{t=p+1}^{T_j} \tau_{ijt} X_{jt} X_{j,t-1}, \dots, \sum_{t=p+1}^{T_j} \tau_{ijt} X_{j,t} X_{j,t-p} \right]. \end{aligned}$$

Then equations (2.6) and (2.7) and the solution of the system can be written in matrix form as in (2.11) and (2.12) respectively.

$$\mathbf{A}_{ij} \hat{\phi}_{ij} = \mathbf{b}_{ij}, \quad (2.11)$$

$$\hat{\phi}_{ij} = \mathbf{A}_{ij}^{-1} \mathbf{b}_{ij}, \quad (2.12)$$

and therefore σ_{ij}^2 can be easily evaluated with equation (2.5).

The analysis of panel time series often encounter the difficulty that the number of observations in each time series is limited. Many researchers pooled the data from different series which shared a common structure, see for instance, Hjellvik and Tjøstheim (1999), Stenseth *et al.* (1999), Yao *et al.* (2000) and Fu *et al.* (2002a). They obtained more accurate results with the increase of the sample size. For the parameters of the N th component, they also have common parameters ϕ_{kN} , $k = 1, \dots, p$. Hence, we pooled the observations X_{jt} and τ_{Njt} , $j = 1, \dots, M$ to estimate ϕ_{kN} , $k = 0, \dots, p$. Equations (2.8) — (2.10) form a $(1 + p + M)$ -equation system with $1 + p + M$ unknown variables. Since this equation system is nonlinear, a numerical method is needed to find the roots of the system. However, by substituting (2.8) to (2.9) and (2.10), the latter two equations contain only $1 + p$ unknown variables which will be easier to solve than directly solving the $1 + p + M$ equations together since p is much smaller than M in most of the cases.

To obtain estimates of the parameters, we iterate the E-step and M-step until convergence. The performance of the EM algorithm will be shown in Section 3. Louis (1982) proposed the missing information principle by which the observed information matrix I can be computed by subtracting I_m , the missing information matrix from I_c , the complete information matrix.

$$\begin{aligned} I &= I_c - I_m \\ &= E \left(-\frac{\partial^2 l}{\partial \boldsymbol{\theta}^2} \middle| \boldsymbol{\theta}, \mathbf{X} \right)_{\hat{\boldsymbol{\theta}}} - \text{var} \left(\frac{\partial l}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{\theta}, \mathbf{X} \right)_{\hat{\boldsymbol{\theta}}}. \end{aligned} \quad (2.13)$$

Since the variance matrix of the estimates $\hat{\boldsymbol{\theta}}$ equals to the inverse of the observed information matrix I , equation (2.13) can be used to find the theoretical large sample variance of the estimates. All the entries in I_c and I_m are available from the authors. Model selection is an important problem in time series modeling. However, since the length of each series in the panel data is often very short, a high order model is not preferred. Setting the order $p = 1$ and the number of components $N = 2$ will be a wise choice. Hence, we will not pursue this problem.

3. Simulation Study

To investigate the performance of the proposed estimation procedure, we do a Monte Carlo simulation ¹. As mentioned before, for a panel data set which the length of each series is short, a low order model is preferred. Hence, we simulated the data set from a twenty-series panel model. Each series is a two-component AR(1) mixture time series. The model is

$$X_{jt} = Z_{1jt}(\phi_{01j} + \phi_{11j}X_{j,t-1} + \varepsilon_{1jt}) + Z_{2jt}(\phi_{02} + \phi_{12}X_{j,t-1} + \varepsilon_{2jt}). \quad (3.1)$$

where $Z_{1jt} = 1$ with probability π_1 if X_{jt} is from the first component of j th series. Otherwise, $Z_{1jt} = 0$. The noise terms ε_{1jt} and ε_{2jt} followed two zero-mean normal distributions with variances σ_{1j}^2 and σ_{2j}^2 . The parameter values of ϕ_{01j} , ϕ_{11j} , σ_{1j}^2 and σ_{2j}^2 for $j = 1, \dots, 20$ are listed in Table 1 respectively. The mixture probability $\pi_1 = 0.4$ and $\pi_2 = 1 - \pi_1 = 0.6$. The parameters of the common structure ϕ_{02} and ϕ_{12} were 1.5 and -0.5 respectively. We did four experiments with different series length to study the finite sample properties of the proposed estimation method. The series lengths T_j are around 1100, 100, 50, 30 in these four experiments respectively. Each experiment was replicated 1000 times. The results are listed in Table 2.

Table 1: True values of the parameters of model (3.1)

$\phi_{0,1,1}$	$\phi_{0,1,2}$	$\phi_{0,1,3}$	$\phi_{0,1,4}$	$\phi_{0,1,5}$	$\phi_{0,1,6}$	$\phi_{0,1,7}$	$\phi_{0,1,8}$	$\phi_{0,1,9}$	$\phi_{0,1,10}$
-1.8	-1.6	-1.4	-1.2	-1.0	-0.8	-0.6	-0.5	-0.4	-0.2
$\phi_{0,1,11}$	$\phi_{0,1,12}$	$\phi_{0,1,13}$	$\phi_{0,1,14}$	$\phi_{0,1,15}$	$\phi_{0,1,16}$	$\phi_{0,1,17}$	$\phi_{0,1,18}$	$\phi_{0,1,19}$	$\phi_{0,1,20}$
0.2	0.4	0.5	0.6	0.8	1.0	1.2	1.4	1.6	1.8
$\phi_{1,1,1}$	$\phi_{1,1,2}$	$\phi_{1,1,3}$	$\phi_{1,1,4}$	$\phi_{1,1,5}$	$\phi_{1,1,6}$	$\phi_{1,1,7}$	$\phi_{1,1,8}$	$\phi_{1,1,9}$	$\phi_{1,1,10}$
-0.2	0.4	0.6	-0.8	-0.2	0.4	0.6	-0.8	-0.2	0.4
$\phi_{1,1,11}$	$\phi_{1,1,12}$	$\phi_{1,1,13}$	$\phi_{1,1,14}$	$\phi_{1,1,15}$	$\phi_{1,1,16}$	$\phi_{1,1,17}$	$\phi_{1,1,18}$	$\phi_{1,1,19}$	$\phi_{1,1,20}$
0.6	-0.8	-0.2	0.4	0.6	-0.8	-0.2	0.4	0.6	-0.8
$\sigma_{1,1}^2$	$\sigma_{1,2}^2$	$\sigma_{1,3}^2$	$\sigma_{1,4}^2$	$\sigma_{1,5}^2$	$\sigma_{1,6}^2$	$\sigma_{1,7}^2$	$\sigma_{1,8}^2$	$\sigma_{1,9}^2$	$\sigma_{1,10}^2$
1.5	1.0	1.5	2.0	2.5	1.5	1.0	1.5	2.0	2.5
$\sigma_{1,11}^2$	$\sigma_{1,12}^2$	$\sigma_{1,13}^2$	$\sigma_{1,14}^2$	$\sigma_{1,15}^2$	$\sigma_{1,16}^2$	$\sigma_{1,17}^2$	$\sigma_{1,18}^2$	$\sigma_{1,19}^2$	$\sigma_{1,20}^2$
1.5	1.0	1.5	2.0	2.5	1.5	1.0	1.5	2.0	2.5
$\sigma_{2,1}^2$	$\sigma_{2,2}^2$	$\sigma_{2,3}^2$	$\sigma_{2,4}^2$	$\sigma_{2,5}^2$	$\sigma_{2,6}^2$	$\sigma_{2,7}^2$	$\sigma_{2,8}^2$	$\sigma_{2,9}^2$	$\sigma_{2,10}^2$
1.5	1.5	1.5	1.5	1.0	2.5	2.0	1.0	2.0	1.5
$\sigma_{2,11}^2$	$\sigma_{2,12}^2$	$\sigma_{2,13}^2$	$\sigma_{2,14}^2$	$\sigma_{2,15}^2$	$\sigma_{2,16}^2$	$\sigma_{2,17}^2$	$\sigma_{2,18}^2$	$\sigma_{2,19}^2$	$\sigma_{2,20}^2$
1.5	1.5	1.5	1.5	1.5	1.0	1.5	1.5	1.5	2.0

¹The simulation programs of model (3.1) and model (5.4) can be obtained from the authors upon request

Table 2: Estimation results of the simulation of model (3.1) with different series length (1000 replications)

	1000 < T_j < 1100		80 < T_j < 100		40 < T_j < 50		20 < T_j < 30		
	True	est.	std.	est.	std.	est.	std.	est.	std.
π_1	0.4	0.4130	0.1219	0.4161	0.1727	0.4128	0.2152	0.4176	0.2190
$\phi_{0,1,1}$	-1.8	-1.7796	0.1814	-1.7395	0.3689	-1.7504	0.5213	-1.6795	0.7798
$\phi_{0,1,2}$	-1.6	-1.5796	0.2097	-1.5674	0.3617	-1.5927	0.4470	-1.5651	0.7040
$\phi_{0,1,3}$	-1.4	-1.3786	0.2150	-1.3993	0.3890	-1.4126	0.5583	-1.3520	0.7759
$\phi_{0,1,4}$	-1.2	-1.1942	0.1097	-1.1681	0.4026	-1.2137	0.5345	-1.1683	0.6963
$\phi_{1,1,1}$	-0.2	-0.1863	0.1671	-0.1803	0.2229	-0.1802	0.2791	-0.2069	0.3215
$\phi_{1,1,2}$	0.4	0.4106	0.1149	0.3901	0.1602	0.3801	0.2515	0.3482	0.3320
$\phi_{1,1,3}$	0.6	0.6020	0.0486	0.5655	0.1577	0.5601	0.2492	0.5013	0.3783
$\phi_{1,1,4}$	-0.8	-0.7760	0.2339	-0.7569	0.2768	-0.7471	0.3135	-0.7698	0.3614
$\phi_{0,2}$	1.5	1.4795	0.1996	1.4788	0.2124	1.4711	0.2121	1.4786	0.2565
$\phi_{1,2}$	-0.5	-0.4949	0.2096	-0.4906	0.2123	-0.4817	0.2352	-0.4915	0.2613
$\sigma_{1,1}^2$	1.5	1.4765	0.2167	1.4393	0.6091	1.3311	0.8478	1.0632	0.8469
$\sigma_{2,1}^2$	1.5	1.4914	0.1844	1.5168	0.4058	1.5376	0.5951	1.5585	0.8670
$\sigma_{1,2}^2$	1.0	0.9825	0.1366	0.9396	0.4003	0.8358	0.5085	0.6733	0.6045
$\sigma_{2,2}^2$	1.5	1.4896	0.1853	1.4506	0.4065	1.4649	0.5962	1.5440	0.9301
$\sigma_{1,3}^2$	1.5	1.4779	0.1997	1.3741	0.5592	1.2166	0.7496	1.0009	0.9090
$\sigma_{2,3}^2$	1.5	1.4838	0.1829	1.4750	0.4177	1.4585	0.5711	1.4538	0.7959
$\sigma_{1,4}^2$	2.0	1.9709	0.2824	1.8740	0.7762	1.6414	0.9245	1.4004	1.1049
$\sigma_{2,4}^2$	1.5	1.4875	0.1843	1.4592	0.4405	1.4789	0.5385	1.4912	0.8455

Only the results of series 1–4 and the common structure are shown in this table.

Other estimates of the parameters can be obtained from the authors.

est. means the average value of the estimates

std. means standard deviation of the estimates.

Table 3: Theoretical large sample standard errors of model (3.1) with length about 1100

π_1	0.0884			$\phi_{0,2}$	0.1863	ϕ_{12}	0.1915
$\phi_{0,1,1}$	0.1743	$\phi_{1,1,1}$	0.1503	$\phi_{0,1,11}$	0.0892	$\phi_{1,1,11}$	0.0915
$\phi_{0,1,2}$	0.1615	$\phi_{1,1,2}$	0.1005	$\phi_{0,1,12}$	0.1322	$\phi_{1,1,12}$	0.1589
$\phi_{0,1,3}$	0.1597	$\phi_{1,1,3}$	0.0517	$\phi_{0,1,13}$	0.0971	$\phi_{1,1,13}$	0.1204
$\phi_{0,1,4}$	0.1027	$\phi_{1,1,4}$	0.1732	$\phi_{0,1,14}$	0.1027	$\phi_{1,1,14}$	0.1123
$\phi_{0,1,5}$	0.1238	$\phi_{1,1,5}$	0.1621	$\phi_{0,1,15}$	0.1039	$\phi_{1,1,15}$	0.1023
$\phi_{0,1,6}$	0.1317	$\phi_{1,1,6}$	0.1017	$\phi_{0,1,16}$	0.1857	$\phi_{1,1,16}$	0.2075
$\phi_{0,1,7}$	0.1293	$\phi_{1,1,7}$	0.1124	$\phi_{0,1,17}$	0.1524	$\phi_{1,1,17}$	0.2291
$\phi_{0,1,8}$	0.0915	$\phi_{1,1,8}$	0.2027	$\phi_{0,1,18}$	0.1417	$\phi_{1,1,18}$	0.1108
$\phi_{0,1,9}$	0.0979	$\phi_{1,1,9}$	0.2301	$\phi_{0,1,19}$	0.1398	$\phi_{1,1,19}$	0.1813
$\phi_{0,1,10}$	0.1203	$\phi_{1,1,10}$	0.1029	$\phi_{0,1,20}$	0.2145	$\phi_{1,1,20}$	0.1210

From the results in Table 2, we can find that the estimation results are very good when series length was longer than 100. In the case that series length was shorter than 30, the estimate method was still applicable, though the standard errors were a bit large. In this extreme case, one can obtain acceptable results by applying our model. Therefore, the lower bound of the series length to apply our model and estimation method seems to be about 30. When better result is required, a series length of about 50 for each of the series is recommended.

Using the information matrix (2.13), we also calculated the theoretical standard error for the case that the length of each series is about 1100. The results for the $\hat{\phi}$ s and $\hat{\pi}$ are listed in Table 3. The results for the $\hat{\sigma}^2$ s are similar and we choose not to list them. The results are in general close to the empirical standard errors in Table 2 suggesting the accuracy of (2.13) in large samples. Hence, our estimation procedure seems useful and can provide good results.

4. A Real Example

This data set consists of the number of yearly trapped gray-sided voles (*Clethrionomys rufocanus*) in ninety-one different stations in forested regions of northern Hokkaido, Japan. Hokkaido, the northern-most island of Japan displays similar bio-geographically feature as the neighboring Asian mainland more than the other Japanese islands (Tatewaki 1958; Kondo 1993). The average temperature in August and February is above $20^{\circ}C$ and below $-7^{\circ}C$. The amount of snow in the winter is largely correlated to the distance to the sea shore. The forests in which the voles were trapped were under the management of the Forestry Agency of the Japanese Government. Among them, 76% is indigenous forests and 24% is planted forest. The gray-sided vole is a forested habitat mammal which can be found in both types of forests. The observation period spanning from 1962 to 1992, over 31 years. This data set has been widely studied in mathematical ecology and panel data literature such as Bjornstard *et al.* (1996) and Stenseth *et al.* (1996). The ninety-one series can be divided into three groups with 16, 41 and 34 series in group one, two and three respectively. For details, see Hjellvik and Tjøstheim (1999) and Hjellvik *et al.* (2004). Figure 2 shows the series in group two after logarithm transformation. Hjellvik and Tjøstheim (1999) developed an autoregressive model with an intersection term to interpret the correlation among these panels. Hjellvik *et al.* (2004) developed a nonparametric estimation method to model the intercorrelation and a test to check the linearity assumption. Jin and Li (2005) assumed that the different series were correlated contemporaneously and proposed a semiparametric model to fit the data set. All these papers did not take the multi-modal phenomenon into account. Stenseth *et al.* (1996) tested the linearity of the data set and found that it was rejected at 5% level. But they still employed linear models to fit the data. However,

after the linear fitting, they found a clear indication of bimodality which in fact violate their underlying assumption. The histogram (Figure 1) of four typical series partially shows the bimodal characteristics. In many biological papers (for example, Bjornstard *et al.* (1996); Stenseth *et al.* (1996); Ota (1984) which studied the gray-sided voles, it has been pointed out that the population of the voles not only correlated with the previous population size but also correlated to the distance from the Sea of Japan. The latter factor affects the climate directly and the food of the gray-sided voles indirectly. The impact of these unobservable factors will randomly vary from year to year and they often followed different patterns. Hence, it seems appropriate to fit the data with MARP so that we can catch these features.

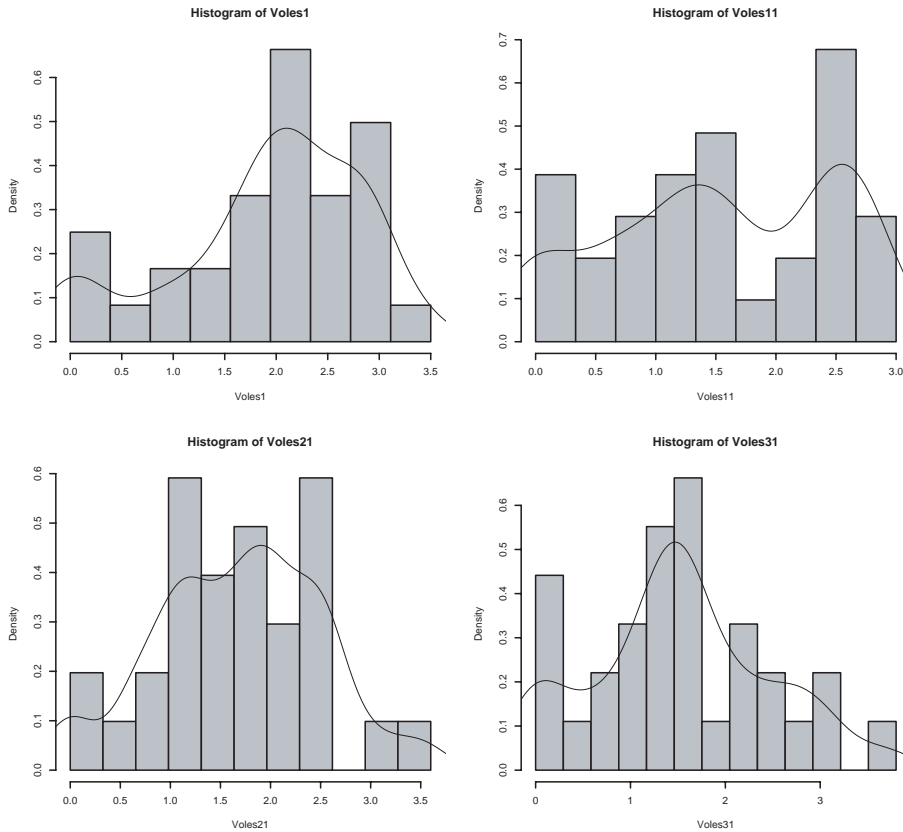


Figure 1: Histograms of four typical series in the gray-sided voles data. The solid lines stand for the kernel smooth of the data with a Gaussian kernel.

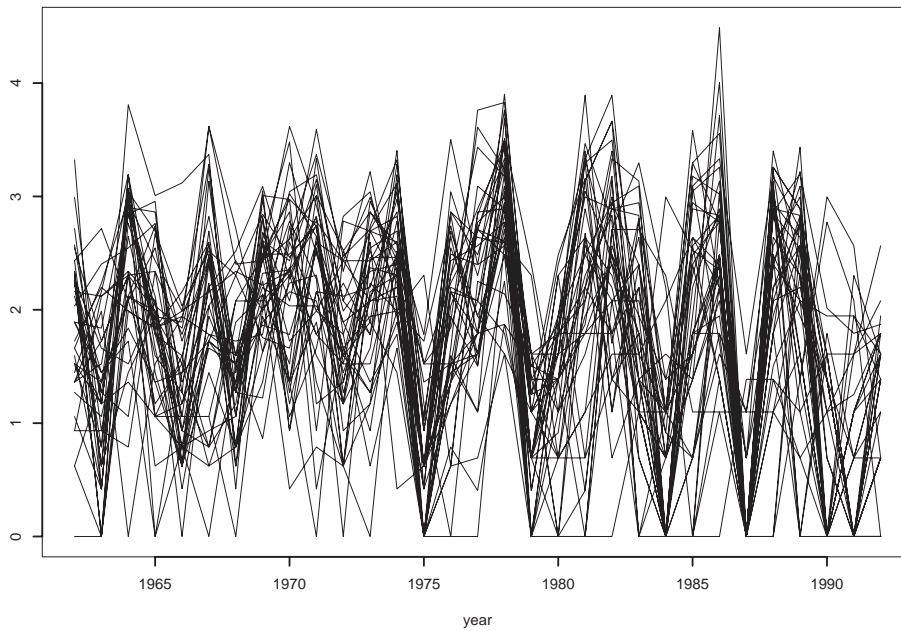


Figure 2: Logarithms of yearly caught gray-sided voles in group two (41 series)

Similar to Hjellvik and Tjøstheim (1999), we also used the series in group two which contained 41 series. The length of each series was 31. Among the 1271 observations, there were 174 (13.7%) zeroes. Hjellvik and Tjøstheim (1999) proposed the model (4.1)

$$X_{jt} = \sum_{k=1}^p \phi_k X_{j,t-k} + \eta_t + \varepsilon_{jt} \quad (4.1)$$

to model the inter-correlation effect of this data set. Here ϕ_k denotes the common autoregressive structure and η_t is the possibly nonlinear effect influencing all the series at time t . In their analysis, they first treated η_t as nuisance parameters and estimate them by the averages of X_{it} over i and kept t fixed. They discarded the nonlinearity property since they believed that it was very weak. Clearly, the assumption that all the series follows the same autoregressive model is too restrictive and ignoring the nonlinearity may be inappropriate. Slightly loosening the restriction of the model would help us to find new implicit pattern inside the data. In fact, by ignoring η_t , model (4.1) is a special case of model (2.1) or (2.2) if we set the probability $\pi_N = 1 - \sum_{i=1}^{N-1} \pi_i = 1$. Although we have increased the number of parameters, we have gained much flexibility by relaxing the restrictive assumption of model (4.1).

Similar to Hjellvik and Tjøstheim (1999), for convenience, we added a value of one to each datum point and then took logarithm. After that, we fitted the

Table 4: Estimation results of the gray-sided voles

j	$\hat{\phi}_{0,1,j}$	$\hat{\phi}_{1,1,j}$	$\hat{\sigma}_{1,j}^2$	$\hat{\sigma}_{2,j}^2$	j	$\hat{\phi}_{0,1,j}$	$\hat{\phi}_{1,1,j}$	$\hat{\sigma}_{1,j}^2$	$\hat{\sigma}_{2,j}^2$
1	0.2316	0.2851	0.7044	0.3337	22	0.1631	0.4461	0.6014	0.5635
2	0.1843	0.0906	0.3005	0.6532	23	0.2495	0.4005	0.6388	0.7968
3	0.0376	0.3504	0.0409	0.6122	24	-0.4734	0.5391	0.4458	1.0540
4	0.6686	0.0746	0.9945	0.1613	25	0.3493	-0.0013	0.3050	0.5406
5	0.2995	-0.0047	0.6914	0.3446	26	-0.4661	0.3774	0.3937	0.8385
6	0.6382	-0.0046	0.6914	1.2020	27	0.2417	0.1789	0.4399	0.5613
7	0.4448	0.4903	0.4188	0.3593	28	0.3353	0.3591	0.6999	0.4249
8	0.7982	0.2304	0.5479	0.7458	29	0.7151	0.0979	0.5903	1.8287
9	1.6010	0.1459	1.7405	0.1468	30	0.6724	-0.2239	0.3860	1.5016
10	-0.0824	0.4068	1.0775	0.6044	31	0.4859	0.0354	0.2827	0.7723
11	0.0224	0.2350	0.2872	0.3846	32	0.0909	0.1715	0.3269	0.8516
12	-0.1211	0.3056	0.3751	0.4536	33	0.0033	-0.0196	0.0009	0.7933
13	0.3247	0.1342	0.2154	0.7005	34	0.0758	0.2526	0.3380	0.4760
14	0.2391	0.1559	0.5886	0.4502	35	0.5854	-0.2575	0.5418	0.6168
15	0.0203	0.2343	0.5318	0.4412	36	0.0871	0.0597	0.7440	0.6244
16	0.1733	0.3664	0.4733	0.3983	37	0.4881	0.2851	0.4114	0.6064
17	0.2834	0.1905	0.6145	1.5468	38	0.5064	0.4054	0.5948	0.4797
18	0.0248	0.1922	0.6376	0.7068	39	0.5129	-0.0687	0.4598	0.6776
19	0.4857	0.3052	0.5526	0.3914	40	0.0137	0.2387	0.3313	0.5982
20	0.0457	0.2364	0.2944	0.6779	41	0.1176	0.4271	0.5291	1.3643
21	0.5559	-0.2377	0.7167	0.6099					
	$\hat{\phi}_{0,2}$	1.1513	$\hat{\phi}_{1,2}$	0.6402		$\hat{\pi}$	0.3015		

data with model (2.2). In fact, we only used the first 30 observations in each series to fit our model. The last observation in all the series are used to check the coverage of our model. Table 4 shows the result. Since all these voles were found on the same island, it was reasonable to believe that they should follow some common pattern which were modeled by the common mixture component whereas the regional variations may be reflected by the individual component. From the results in Table 4, we found that the autoregressive coefficients of the common structure are significantly different from those of the first component in each of these 41 series which suggested that equation (2.1) could be an appropriate model. Together with the result of $\hat{\pi} = 0.3015$, we believed that the mixture phenomenon was quite significant which was consistent with the histogram plots (Figure 1) of these series. An interpretation of the result is that the data follow the common component about 70 % of the time whereas in the other 30 % time, the dependent structure follows a pattern which is dictated by the environment, possibly the food supplies. From our analysis, we found that among these 41 series, the individual components showed three different patterns. The twenty-first, thirtieth and thirty-fifth series displayed negative autocorrelation in the

individual components. These suggested that the regional factors of these three districts may have negative effect on the number of the voles. It is possible that the strict survival environments could not support too many voles. The other twenty-nine series showed positive autocorrelation in the individual components suggesting that the environment could be relatively good. The food might be sufficient and competition within species might be mild. The remaining nine series had insignificant autocorrelation with the previous observation. The ϕ_{11j} were around zero suggesting that the number of voles is a constant.

To compare our model with the one of Hjellvik and Tjøstheim (1999), we checked the coverage of the one-step forecasting. Since our model was a bit complicated, we had to use simulation to find the empirical quantiles. For example, let $X_{1,30}$ be the thirtieth observation of the first series. From the results in Table 4, we knew that $X_{1,31}$ would follow a mixture distribution. It would come from the $N(0.2316 + 0.2851X_{1,30}, 0.7040)$ distribution with probability 0.3015 and come from the $N(1.1513 + 0.6402X_{1,30}, 0.3337)$ distribution with probability 0.6985. Note that all X_{jt} have been log-transformed. Therefore, we can generate a random number U following a uniform (0,1) distribution. If U was smaller than 0.3015, we picked $\tilde{X}_{1,31}$ from the first normal distribution as a simulated $X_{1,31}$. Otherwise, we picked $\tilde{X}_{1,31}$ from the second normal distribution. Repeating this 10,000 times, we obtained a sample following the mixture distribution. The empirical 95% predictive interval was obtained by (a, b) where a was the empirical 2.5% quantile and b was the empirical 97.5% quantile. With the same method, we obtained the empirical predictive intervals for all these 41 series. Comparing to the corresponding interval, we found that there were 3 last observations among these 41 series outside 95% one-step ahead predictive intervals, a coverage of about 7.3%. Considering the small sample size, this coverage was still very near the nominal 95%. There were 3 different model orders in Hjellvik and Tjøstheim (1999), one, two and three respectively. In general, higher order model would be more accurate than lower order model. Hence, comparing our model with their model whose order was three could be more cogent than comparing to the other two. By checking the coverage of their model, we found that there was no observation outside the 95% intervals. This indicated that their model leads to conservative predictive intervals. Another commonly used method for model comparison is BIC (Schwarz, 1978). This criterion adds heavier penalty than AIC (Akaike, 1974). Since the number of parameters are large, using BIC can be a challenge to our model. Although panel data are multivariate time series, since the residual of our model were hard to define due to the characteristics of the finite mixture, we used a simpler version of BIC which was the same as that in the univariate case,

$$\begin{aligned} \text{BIC} &= -2 \log(\text{maximum likelihood}) + \log(\text{number of data points}) \\ &\quad \times (\text{number of estimated parameters}). \end{aligned}$$

The values of the log-likelihood were -1082.10 and -1310.27 for our model and that of Hjellvik and Tjøstheim (1999) respectively. The BICs were 2680.22 and 2715.40 respectively for these two different models. Although the difference between the two BICs are not very large, it is an evidence that our model better fits the data.

5. Extension

In this section, we relax the assumption that the mixture parameters π_i are common to every series. For simplicity, we only show the two component and order one case. The more general cases can be extended straightforwardly. For notation consistency, we still use three indices as the subscripts of the parameters. The model is modified as

$$\begin{aligned} F(X_{jt}|\mathcal{F}_{t-1}) &= \pi_{1j} \Phi \left(\frac{X_{jt} - \phi_{01j} - \phi_{11j} X_{j,t-1}}{\sigma_{1j}} \right) \\ &\quad + \pi_{2j} \Phi \left(\frac{X_{jt} - \phi_{02} - \phi_{12} X_{j,t-1}}{\sigma_{2j}} \right), \end{aligned} \quad (5.1)$$

where $\pi_{1j} + \pi_{2j} = 1$ for each j , $j = 1, \dots, M$. The conditional log-likelihood functions, for each j given \mathbf{Z}_j which has the same meaning as section 2, are

$$l_j = \sum_{t=p+1}^{T_j} \left\{ \sum_{i=1}^2 Z_{ijt} \log(\pi_{ij}) - \frac{1}{2} \sum_{i=1}^2 Z_{ijt} \log(\sigma_{ij}^2) - \sum_{i=1}^2 \frac{Z_{ijt} \varepsilon_{ijt}^2}{2\sigma_{ij}^2} \right\},$$

where $\varepsilon_{1jt} = X_{jt} - \phi_{01j} - \phi_{11j} X_{j,t-1}$ and $\varepsilon_{2jt} = X_{jt} - \phi_{02} - \phi_{12} X_{j,t-1}$. The global conditional log-likelihood function given \mathbf{Z} is

$$l = \sum_{j=1}^M \sum_{t=p+1}^{T_j} \left\{ \sum_{i=1}^2 Z_{ijt} \log(\pi_{ij}) - \frac{1}{2} \sum_{i=1}^2 Z_{ijt} \log(\sigma_{ij}^2) - \sum_{i=1}^2 \frac{Z_{ijt} \varepsilon_{ijt}^2}{2\sigma_{ij}^2} \right\}.$$

Since the first derivatives of the log-likelihood function l can be obtained straightforwardly, we skip them here.

We show the estimation method with the EM algorithm briefly since the method is almost the same as the one in section 2.

Table 5: True values of the parameters of model (5.4)

$\phi_{0,1,1}$	$\phi_{0,1,2}$	$\phi_{0,1,3}$	$\phi_{0,1,4}$	$\phi_{0,1,5}$	$\phi_{0,1,6}$	$\phi_{0,1,7}$	$\phi_{0,1,8}$	$\phi_{0,1,9}$	$\phi_{0,1,10}$
-1.8	-1.6	-1.4	-1.2	-1.0	-0.8	-0.6	-0.5	-0.4	-0.2
$\phi_{0,1,11}$	$\phi_{0,1,12}$	$\phi_{0,1,13}$	$\phi_{0,1,14}$	$\phi_{0,1,15}$	$\phi_{0,1,16}$	$\phi_{0,1,17}$	$\phi_{0,1,18}$	$\phi_{0,1,19}$	$\phi_{0,1,20}$
0.2	0.4	0.5	0.6	0.8	1.0	1.2	1.4	1.6	1.8
$\phi_{1,1,1}$	$\phi_{1,1,2}$	$\phi_{1,1,3}$	$\phi_{1,1,4}$	$\phi_{1,1,5}$	$\phi_{1,1,6}$	$\phi_{1,1,7}$	$\phi_{1,1,8}$	$\phi_{1,1,9}$	$\phi_{1,1,10}$
-0.2	0.4	0.6	-0.8	-0.2	0.4	0.6	-0.8	-0.2	0.4
$\phi_{1,1,11}$	$\phi_{1,1,12}$	$\phi_{1,1,13}$	$\phi_{1,1,14}$	$\phi_{1,1,15}$	$\phi_{1,1,16}$	$\phi_{1,1,17}$	$\phi_{1,1,18}$	$\phi_{1,1,19}$	$\phi_{1,1,20}$
0.6	-0.8	-0.2	0.4	0.6	-0.8	-0.2	0.4	0.6	-0.8
$\sigma_{1,1}^2$	$\sigma_{1,2}^2$	$\sigma_{1,3}^2$	$\sigma_{1,4}^2$	$\sigma_{1,5}^2$	$\sigma_{1,6}^2$	$\sigma_{1,7}^2$	$\sigma_{1,8}^2$	$\sigma_{1,9}^2$	$\sigma_{1,10}^2$
1.5	1.0	1.5	2.0	2.5	1.5	1.0	1.5	2.0	2.5
$\sigma_{1,11}^2$	$\sigma_{1,12}^2$	$\sigma_{1,13}^2$	$\sigma_{1,14}^2$	$\sigma_{1,15}^2$	$\sigma_{1,16}^2$	$\sigma_{1,17}^2$	$\sigma_{1,18}^2$	$\sigma_{1,19}^2$	$\sigma_{1,20}^2$
1.5	1.0	1.5	2.0	2.5	1.5	1.0	1.5	2.0	2.5
$\sigma_{2,1}^2$	$\sigma_{2,2}^2$	$\sigma_{2,3}^2$	$\sigma_{2,4}^2$	$\sigma_{2,5}^2$	$\sigma_{2,6}^2$	$\sigma_{2,7}^2$	$\sigma_{2,8}^2$	$\sigma_{2,9}^2$	$\sigma_{2,10}^2$
1.5	1.5	1.5	1.5	1.0	2.5	2.0	1.0	2.0	1.5
$\sigma_{2,11}^2$	$\sigma_{2,12}^2$	$\sigma_{2,13}^2$	$\sigma_{2,14}^2$	$\sigma_{2,15}^2$	$\sigma_{2,16}^2$	$\sigma_{2,17}^2$	$\sigma_{2,18}^2$	$\sigma_{2,19}^2$	$\sigma_{2,20}^2$
1.5	1.5	1.5	1.5	1.5	1.0	1.5	1.5	1.5	2.0
$\pi_{1,1}$	$\pi_{1,2}$	$\pi_{1,3}$	$\pi_{1,4}$	$\pi_{1,5}$	$\pi_{1,6}$	$\pi_{1,7}$	$\pi_{1,8}$	$\pi_{1,9}$	$\pi_{1,10}$
0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.20	0.25	0.30
$\pi_{1,11}$	$\pi_{1,12}$	$\pi_{1,13}$	$\pi_{1,14}$	$\pi_{1,15}$	$\pi_{1,16}$	$\pi_{1,17}$	$\pi_{1,18}$	$\pi_{1,19}$	$\pi_{1,20}$
0.35	0.40	0.45	0.50	0.20	0.25	0.30	0.35	0.40	0.45

E-step Suppose the parameter vector θ is known, then

$$\tau_{ijt} = \frac{\pi_{ij} / \sqrt{\sigma_{ij}^2} \cdot \exp\{-\varepsilon_{ijt}^2 / (2\sigma_{ij}^2)\}}{\sum_{i'=1}^2 \pi_{i'j} / \sqrt{\sigma_{i'j}^2} \cdot \exp\{-\varepsilon_{i'jt}^2 / (2\sigma_{i'j}^2)\}}, \quad (5.2)$$

where $i = 1, 2$, $j = 1, \dots, M$ and $t = 1, \dots, T_j$.

M-step Suppose that the missing data are known. The estimates of the parameters θ can then be obtained by maximizing the log-likelihood function l . The M-step equations become

$$\hat{\pi}_{ij} = \frac{\sum_{t=p+1}^{T_j} \tau_{ijt}}{T_j - p}. \quad (5.3)$$

The estimation equations of the other parameters are the same as those in section 2.

Following the missing information principle Louis (1982), the information matrix equation (2.13) still holds. We skip the derivation of the entries in I_c and I_m since the method is just the same except for those entries involving π_{ij} .

Table 6: Estimation results of the simulation of model (5.4) with different series length (1000 replications)

	True	1000 < T_j < 1100		80 < T_j < 100		40 < T_j < 50		20 < T_j < 30	
		est.	std.	est.	std.	est.	std.	est.	std.
$\pi_{1,1}$	0.20	0.1947	0.1747	0.2023	0.2926	0.2453	0.3589	0.2639	0.4085
$\pi_{1,2}$	0.25	0.2465	0.1776	0.2522	0.2851	0.2567	0.3487	0.2835	0.3653
$\pi_{1,3}$	0.30	0.2985	0.1931	0.3130	0.3081	0.3105	0.3654	0.3023	0.3707
$\pi_{1,4}$	0.35	0.3518	0.2173	0.3624	0.3538	0.3751	0.3975	0.4128	0.4438
$\phi_{0,1,1}$	-1.8	-1.7915	0.5097	-1.7824	0.8011	-1.6305	0.9471	-1.5808	1.0337
$\phi_{0,1,2}$	-1.6	-1.5915	0.4249	-1.6617	0.6294	-1.5183	0.8539	-1.5572	0.9045
$\phi_{0,1,3}$	-1.4	-1.3750	0.4146	-1.3692	0.6391	-1.4656	0.7931	-1.3687	1.0438
$\phi_{0,1,4}$	-1.2	-1.1673	0.4624	-1.2178	0.7686	-1.2021	0.8436	-1.0528	0.8822
$\phi_{1,1,1}$	-0.2	-0.1997	0.2346	-0.1729	0.5051	-0.2969	0.6099	-0.2499	0.7735
$\phi_{1,1,2}$	0.4	0.3960	0.2521	0.3994	0.4163	0.2542	0.6094	0.4026	0.7004
$\phi_{1,1,3}$	0.6	0.5868	0.2719	0.5778	0.4597	0.5193	0.6400	0.5471	0.7325
$\phi_{1,1,4}$	-0.8	-0.7944	0.2979	-0.7832	0.4414	-0.8012	0.5098	-0.8022	0.5641
$\phi_{0,2}$	1.5	1.4771	0.4427	1.4762	0.4571	1.4942	0.4717	1.5289	0.4325
$\phi_{1,2}$	-0.5	-0.4952	0.4236	-0.4946	0.3267	-0.4914	0.3891	-0.5012	0.4985
$\sigma_{1,1}^2$	1.5	1.4583	0.5528	1.2002	0.9541	1.1046	0.9831	0.6277	0.9021
$\sigma_{2,1}^2$	1.5	1.4907	0.4151	1.5357	0.5856	1.4229	0.7051	1.3662	0.8337
$\sigma_{1,2}^2$	1.0	0.9812	0.4153	0.7376	0.7063	0.6664	0.8027	0.3581	0.7272
$\sigma_{2,2}^2$	1.5	1.4790	0.4157	1.4826	0.6248	1.4025	0.6634	1.3075	0.7762
$\sigma_{1,3}^2$	1.5	1.4924	0.4802	1.2603	0.8378	0.9745	0.8422	0.6980	0.9617
$\sigma_{2,3}^2$	1.5	1.4818	0.4166	1.4297	0.5902	1.4036	0.7067	1.3772	0.8223
$\sigma_{1,4}^2$	2.0	1.9812	0.5896	1.7268	0.9220	1.5535	0.9891	1.3507	1.0365
$\sigma_{2,4}^2$	1.5	1.4713	0.4264	1.4709	0.6978	1.3687	0.7621	1.3320	0.8472

Only the results of series 1–4 and the common structure are shown in this table.

Other estimates of the parameters can be obtained from the authors.

est. means the average value of the estimates,

std. means standard deviation of the estimates.

To investigate the performance of model (5.1), we did a simulation study. The underlying model considered was

$$X_{jt} = Z_{1jt}(\phi_{01j} + \phi_{11j}X_{j,t-1} + \varepsilon_{1jt}) + Z_{2jt}(\phi_{02} + \phi_{12}X_{j,t-1} + \varepsilon_{2jt}). \quad (5.4)$$

where $Z_{1jt} = 1$ with probability π_{1j} if X_{jt} was from the first component of j th series. Otherwise, $Z_{1jt} = 0$. The parameter values of ϕ_{0ij} , ϕ_{1ij} , σ_{1j} , σ_{2j} and π_{1j} are listed in Table 5. The parameters of the common structure ϕ_{02} and ϕ_{12} were 1.5 and -0.5 respectively. It can be seen that the parameters are almost the same as model (3.1) except the parameters π_{1j} . We also did four experiments with different series length to study the finite sample properties of the proposed estimation method. The series length is also the same as before, say around

1100, 100, 50 and 30 respectively. Each experiment was replicated 1000 times. The results are listed in Table 6. From the results, we find almost the same conclusion as Table 2. The difference is that the standard deviations of the ϕ_j and σ_{ij}^2 estimates become larger in Table 6 than those in Table 2. However, the bias of the estimates seems acceptable.

Table 7: Estimation results of the gray-sided voles with model (5.1)

j	$\hat{\phi}_{0,1,j}$	$\hat{\phi}_{1,1,j}$	$\hat{\sigma}_{1,j}^2$	$\hat{\sigma}_{2,j}^2$	$\hat{\pi}_j$	j	$\hat{\phi}_{0,1,j}$	$\hat{\phi}_{1,1,j}$	$\hat{\sigma}_{1,j}^2$	$\hat{\sigma}_{2,j}^2$	$\hat{\pi}_j$
1	0.2401	0.2829	0.7194	0.3546	0.3179	22	0.1660	0.4471	0.6249	0.5714	0.3822
2	0.3329	0.0979	0.3106	0.6711	0.3742	23	0.2537	0.2759	0.6624	0.8143	0.3179
3	0.0424	0.3529	0.1270	0.6306	0.3231	24	-0.4686	0.5411	0.4561	1.0670	0.3061
4	0.6922	0.0865	1.0117	0.2473	0.3875	25	0.3542	0.0079	0.3095	0.5494	0.4500
5	0.3177	0.0074	0.7025	0.3572	0.3850	26	-0.4506	0.3902	0.4051	0.8546	0.3153
6	0.6506	-0.0033	0.7049	1.2178	0.3825	27	0.2474	0.1846	0.4477	0.5789	0.5125
7	0.4570	0.4927	1.4423	0.3664	0.3173	28	0.3408	0.3589	0.7159	0.4408	0.7014
8	1.3056	0.2304	0.7540	14.6978	0.9503	29	0.7302	0.1020	0.6046	1.8470	0.3193
9	1.6166	0.1460	1.7641	0.1638	0.9491	30	0.6959	-0.2143	0.3960	1.5187	0.2998
10	-0.0680	0.4106	1.0944	0.6156	0.3026	31	0.5034	0.1029	0.2946	0.7909	0.3726
11	0.0362	0.2435	0.2930	0.4026	0.3802	32	0.1047	0.1738	0.3360	0.8649	0.3125
12	-0.1077	0.3186	0.3846	0.4696	0.3093	33	0.0156	-0.0067	0.0003	0.8117	0.1929
13	0.3351	0.1479	0.2217	0.7145	0.3164	34	0.0820	0.2511	0.3474	0.4901	0.3187
14	0.2564	0.1611	0.6017	0.4578	0.3129	35	0.6089	-0.1784	0.5539	0.6284	0.3145
15	0.0439	0.3075	0.5416	0.4647	0.3183	36	0.1107	0.1372	0.7569	0.6480	0.3165
16	0.1798	0.3741	0.4838	0.4073	0.3144	37	0.4897	0.2865	0.4292	0.6191	0.3783
17	0.2968	0.1922	0.6291	1.5648	0.3187	38	0.9362	0.4045	1.1183	0.4975	0.9403
18	0.0418	0.2629	0.6424	0.7248	0.4433	39	0.5260	-0.0535	0.4713	0.6924	0.3081
19	0.4900	0.3064	0.5692	0.4051	0.3107	40	0.0916	0.2418	0.3385	0.6074	0.4444
20	0.0510	0.2387	0.3004	0.6932	0.3124	41	0.1245	0.4280	0.5458	1.3827	0.5125
21	0.8919	-0.2142	0.7332	0.6247	0.3019		$\hat{\phi}_{02}$	1.5025	$\hat{\phi}_{12}$	0.4822	

We also applied model (5.1) to the gray-sided voles data. As before, we only used the first 30 observations in each series and left the last one to check the coverage of the forecasting value. The results are listed in Table 7. The results of series 8, 9 and 38 showed a very large $\hat{\pi}$ which were close to 1. Considering the short series length, we can accept some bad estimates. For series 8, the second component has a very large variance about 14. Note that the mixture probability of this series is 0.95, there were only two observations coming from the second component. We checked the two observations and found that they were exactly the two missing data points in this series. Hence, the variance of the second component was just an artifact from the Fortran program and would be meaningless. Other series showed a significant mixture phenomenon. By checking the coverage of the forecasting, we found that there were also 3 observations outside the 95% one-step ahead predictive interval, that is, a coverage about 7.3%. However, although the log-likelihood was -963.38 which were smaller than

that of model (3.1) and the one in Hjellvik and Tøstheim (1999), the number of parameters was 207 which increased the BIC to 3399.52. Because of the large number of parameters, we have some reservations on fitting gray-sided voles data with model (5.1).

6. Conclusion

In this paper, we considered a mixture autoregressive panel (MARP) model which can capture the multi-modal phenomenon in some data sets. We suggested to estimate the MARP model by the EM algorithm. The simulation studies showed that we can obtain quite satisfying estimates if the series length of each series is 50 or longer. Even if the series length is around 30, the estimates still appeared to be acceptable. We illustrated the model by applying it to the gray-sided voles data. All these results show that the sample size requirement of our model is quite low and the performance of the model is satisfying so that it may be of some potential in applications.

Acknowledgement

The authors would like to thank the Hong Kong Research Grants Council (HKU7028/00P) for partial support. The authors would like to thank Professors V. Hjellvik and N. Stenseth for providing us the data set of the gray-sided voles. The authors would also like to thank a referee and the editor for comments that led to the improvement of the paper.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* **AC-19**, 716-723.
- Bjørnstad, O. N., Champley, S., Stenseth, N. C. and Saitoh, T. (1996). Cyclicity and stability of grey-sided voles, *clethrionomys rufocanus* of hokkaido: Spectral and principal components analysis. *Philosophical Transactions Royal Society London B* **351**, 867-875.
- 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.
- Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association* **99**, 710-723.

- Fu, B., Li, W. K. and Fung, W. K. On contemporaneously correlated panel threshold models. Working papers, Department of Statistics and Actuarial Science, The University of Hong Kong.
- Fu, B., Li, W. K. and Fung, W. K. (2002). Testing model adequacy for dynamic panel data with intercorrelation. *Biometrika* **89**, 591-601.
- Hjellvik, V. and Chen, R. and Tjøstheim, D. (2004). Nonparametric estimation and testing in panels of intercorrelated time series. *Journal of Time Series Analysis* **25**, 831-872.
- Hjellvik, V. and Tjøstheim, D. (1999). Modelling panels of intercorrelated autoregressive time series. *Biometrika* **86**, 573-590.
- Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge University Press.
- Jin, S. and Li, W. K. (2005). Modeling contemporaneously correlated panel data with partial linear regression. Technical report 404, Dept. of Stat. and Actu. Sci., Univ. of Hong Kong.
- Le, N. D., Martin, R. D. and Raftery, A. E. (1996). Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association* **91**, 1504-1514.
- Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **96**, 103-113.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B* **44**, 226-233.
- McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models* John Wiley and Sons.
- Ota, K. (ed.) (1984). *Study on Wild Murid Rodents in Hokkaido*. Hokkaido University Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.
- Stenseth, N. C., Bjørnstad, O. N. and Saitoh, T. (1996). A gradient from stable to cyclic populations of *clethrionomys rufocamus* in Hokkaido, Japan. *Proceedings Royal Society London B* **263**, 1117-1126.
- Stenseth, N. C., Chan, K. S., Tong, H., Boonstra, R., Boutin, S., Krebs, C. J., Post, E., O'Donoghue, M., Yoccoz, N. G., Forchhammer, M. C. and Hurrell, J. W. (1999). Common dynamic structure of Canada lynx populations within three climatic regions. *Science* **285**, 1071-1073.
- Wong, C. S. and Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society Series B* **62**, 95-115.
- Yao, Q., Tong, H., Finkenstadt, B. and Stenseth, N. C. (2000). Common structure in panels of short ecological time-series. *Proceedings of the Royal Society of London Series B* **267**, 2459-2467.

Received May 27, 2005; accepted December 6, 2005.

Shusong Jin
Department of Statistics
School of Management
Fudan University
Guosun Road 670
Shanghai 200433
China
jinss@fudan.edu.cn

Wai Keung Li
Department of Statistics and Actuarial Science
The University of Hong Kong
Hong Kong
hrntlwk@hku.hk