Hidden Markov Mixture Autoregressive Models: Parameter Estimation

S.H.Alizadeh, S.Rezakhah*

Abstract

This report introduces a parsimonious structure for mixture of autoregressive models, where the weighting coefficients are determined through latent random variables as functions of all past observations. These variables follow a hidden Markov model. We modify EM and Baum-Welch algorithms to estimate the parameters of the model.

MSC: primary 62M10, 60J10 secondary 60G25

Keywords and phrases. Hidden Markov Model, Mixture Autoregressive Model, Parameter Estimation.

1 Hidden Markov Mixture Autoregressive Model

Let $Y = \{Y_t\}_{t=0}^{\infty}$ be a sequence of continuous random variables, where y_t is a realization of Y_t. Also let $\mathcal{F}_t = \sigma\{Y_s : s \leq t\}$ represents the sigma-field of all information up to time t, $F(y_t|\mathcal{F}_{t-1})$ the conditional distribution function of Y_t given past information and $\alpha_h^{(t)} \equiv \alpha_h^{(t)}(y_1, ..., y_{t-1})$. In addition $\{Z_t\}_{t\geq p}$ denotes a hidden or latent process which construct a positive recurrent Markov chain on a finite set $E = \{1, 2, ..., K\}$, with the initial conditional probabilities

$$\boldsymbol{\rho} = (\rho_1, \cdots, \rho_K)', \;\; \rho_h = P(Z_p = h|y_0, \cdots, y_{p-1}) \qquad h = 1, ..., K, \tag{1}$$

and transition probability matrix

$$P = \|\pi_{i,j}\|_{K\times K}, \tag{2}$$

*Faculty of Mathematics and Computer Science ,Amirkabir University of Technology, Tehran, Iran. Email:rezakhah@aut.ac.ir, sasan_alizadeh@aut.ac.ir

in which

$$\pi_{i,j} = P(Z_t = j|Z_{t-1} = i), \qquad i, j \in \{1, ..., K\}. \tag{3}$$

Also invariant probability measure is denoted by

$$\boldsymbol{\mu} = (\alpha_1, ..., \alpha_K)', \tag{4}$$

where $\alpha_j = \lim_{t\to\infty} P(Z_t = j)$.

We consider $\{Y_t\}_{t=0}^{\infty}$ to have a Hidden Markov-Mixture Autoregressive, HM-MAR(K,p), model with K normal distributions, and p lagged observations in the AR processes, if the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ follows

i. For $t = p$

$$F(y_p, Z_p = h|\mathcal{F}_{p-1}) = \rho_h \Phi(\frac{y_p - a_{0,h} - a_{1,h}y_{p-1} - ... - a_{p,h}y_0}{\sigma_h}), \tag{5}$$

ii. For $t \geq p+1$

$$F(y_t|\mathcal{F}_{t-1}) = \sum_{h=1}^{K} \alpha_h^{(t)} \Phi(\frac{y_t - a_{0,h} - a_{1,h}y_{t-1} - ... - a_{p,h}y_{t-p}}{\sigma_h}), \tag{6}$$

where $\alpha_h^{(t)} = P(Z_t = h|\mathcal{F}_{t-1})$ and $\Phi(.)$ is the standard normal distribution function.

In fact latent random variables $\{Z_t\}_{t=p+1}^{\infty}$ determine the contribution of distributions in the mixture model. Also conditioning on Z_t, $\{Y_t, t \in \mathbb{N}\}$ is p-tuple Markov, independent of $\{Z_s,\ s \neq t\}$. So by conditioning on $\{Y_{t-1}, \cdots, Y_{t-p}\}$ and Z_t, Y_t is independent of $\{Y_s,\ s < t-p\}$ and $\{Z_s,\ s \neq t\}$.

The novelty of HM-MAR model is that the contribution of each distribution in the mixture structure is not of predefined fixed form. Although HM-MAR model uses all past observations from Y_0 to Y_{t-1} but the hidden Markov assumption of the process $\{Z_t\}_{t\geq p}$, enables us to build a parsimonious model.

The MAR model [3] can be considered as a special case of such a HM-MAR model (5-6), in which the transition matrix P of the process $\{Z_t\}_{t\geq p}$ has K identical rows (i.e. $p(Z_t = i|Z_{t-1} = j) = \alpha_i$ for all i, $j = 1, ..., K$. Thus $\{Z_t\}_{t=p+1}^{\infty}$ are independent and identically distributed) with $p(Z_t = i|Z_{t-1} = j) = \alpha_i$.

HM-MAR model will also lead to hidden Markov model in general state space where p is considered to be zero in (6) (i.e. Y_t given Z_t, is independent of past observations).

2 Estimation

In this section, we discuss estimation of parameters of a HM-MAR(K, p) model. A new algorithm is proposed based on modification of Baum-Welch [1] and EM [2] algorithms. Baum welch algorithm was originally proposed in the context of Hidden Markov Models for parameter estimation (For a comprehensive review see MacDonal and Zucchini [1]). In HMM each observation just depends on a state of a hidden variable, however in HM-MAR, past observations have also effect on next time series observation. First we justify that the modification of Baum-Welch algorithm is correct and then modify the EM algorithm for the case where the latent variable follows a Hidden Markov process.

Let denote $\boldsymbol{A}_j = (a_{0,j}, \cdots, a_{p,j})'$ then $\theta = \{\boldsymbol{A}_j, \sigma_j, \rho_j, \pi_{mn}, m, n, j = 1, \cdots, K\}$ constitutes the parameter set of HM-MAR model, which includes $\{K^2 + (p+2)K\}$ parameters. As Y_t given Z_t forms a p-tuple Markov in HM-MAR model, its conditional distribution can be written as

$$F(y_t|y_0...y_{t-1}, z_t) = \prod_{k=1}^{K} \Phi(\frac{y_t - \mathbf{Y}'_{t-1}A_k}{\sigma_k})^{I(z_t=k)}, \tag{7}$$

where $\mathbf{Y}_t = (1, y_t, \cdots, y_{t-p+1})'$, also the conditional distribution $P(z_t|z_{t-1})$ is given by

$$P(Z_t = z_t|Z_{t-1} = z_{t-1}) = \prod_j \prod_k \pi_{j,k}^{I(z_t=k)I(z_{t-1}=j)}. \tag{8}$$

2.1 Extension of Baum-Welch Algorithm

Lemma 2.1. *Let $\{y_t\}_{t=0}^T$ be a set of time series observations and $\{Z_t\}$ be a set of correct predictor indexes, in ARSNN next time series observations just depends on the last correct predictor. That is for $t \leq k \leq T$*

$$\begin{aligned} F(y_{t+1}, \cdots, y_K|y_1, \cdots, y_t,\ \{Z_s\}_{s\in\mathbb{N},\, s\leq t}) \\ = F(y_{t+1}, \cdots, y_K|y_1, \cdots, y_t, Z_t) \end{aligned} \tag{9}$$

Proof. Considering the homogeneous hidden Markov structure assumption of $\{Z_t\}$ in HM-MAR model (5-6) and the assumption that y_t given we have information about the Z_t, just depends on p lagged time series observations through 7, we use the method of induction to prove (9). So for $k = t+1$ we have that

$$F(y_{t+1}|y_1, \cdots, y_t,\ \{Z_s\}_{s\in\mathbb{N},\, s\leq t})$$

$$
\begin{aligned}
&= \sum_{j=1}^{K} F(y_{t+1}, Z_{t+1} = j | y_1, \cdots, y_t, \{Z_s\}_{s\in\mathbb{N},\, s\le t}) \\
&= \sum_{j=1}^{K} F(y_{t+1} | y_1, \cdots, y_t, Z_{t+1} = j) P(Z_{t+1} = j | Z_t),
\end{aligned}
$$

which is independent of $\{Z_{t-i},\ i \in \mathbb{N}, i > 1\}$. Now assume that equation (9) holds for $t+1 < \ell < T$, that is

$$
\begin{aligned}
F(y_{t+1}, \cdots, y_\ell | y_1, \cdots, y_t, &\{Z_s\}_{s\in\mathbb{N},\, s\le t}) \\
&= F(y_{t+1}, \cdots, y_\ell | y_1, \cdots, y_t, Z_t) \qquad (10)
\end{aligned}
$$

We show that (9) is valid for $k = \ell + 1$

$$
\begin{aligned}
&F(y_{t+1}, \cdots, y_\ell, y_{\ell+1} | y_1, \cdots, y_t, \{Z_s\}_{s\in\mathbb{N},\, s\le t}) = \\
&\sum_{j=1}^{K} F(y_{\ell+1} | y_1, \cdots, y_\ell, Z_{\ell+1} = j) P(Z_{\ell+1} | y_1, \cdots, y_\ell, \{Z_s\}_{s\in\mathbb{N},\, s\le t}) \times \\
&F(y_\ell | y_1, \cdots, y_t, \{Z_s\}_{s\in\mathbb{N},\, s\le t}) \\
= &\sum_{j=1}^{K} F(y_{\ell+1} | y_1, \cdots, y_\ell, Z_{\ell+1} = j) P(Z_{\ell+1} | Z_t) \times \\
&F(y_\ell | y_1, \cdots, y_t, \{Z_s\}_{s\in\mathbb{N},\, s\le t}),
\end{aligned}
$$

which is independent of $\{Z_{t-i}\}_{i\ge 1}$ by the induction's assumption (10). □

Theorem 2.1. *Let for $t > p$*

$$
\alpha_t(h) = F(y_{p+1} ... y_t, Z_t = h | y_1 ... y_p), \qquad (11)
$$

$$
\beta_t(h) = F(y_{t+1} ... y_T | y_1 ... y_t, Z_t = h), \qquad (12)
$$

then $\alpha_t(h)$ and $\beta_t(h)$ can be calculated by Baum-welch forward backward recursions as

$$
\begin{aligned}
\alpha_{t+1}(h) &= \sum_{m} \pi_{m,h} \alpha_t(m) \Phi(\frac{y_{t+1} - \mathbf{Y}_t' A_h}{\sigma_h}) \\
\beta_t(h) &= \sum_{j=1}^{K} \pi_{h,j} \beta_{t+1}(j) \Phi(\frac{y_{t+1} - \mathbf{Y}_{t-1}' A_k}{\sigma_j}). \qquad (13)
\end{aligned}
$$

And the forward recursion starts with $\alpha_{p+1}(h) = \rho_h \Phi\{(y_{p+1} - \mathbf{Y}_p' A_h)/\sigma_h\}$ and backward recursion starts at $\beta_T(h) = 1$, in which $\Phi(.)$ is the standard normal distribution function.

Proof. $\alpha_t(h)$ in equation (11) can be written as

$$\begin{aligned}
\alpha_{t+1}(h) &= \sum_m F(y_{p+1}...y_{t+1}, Z_t = m, Z_{t+1} = h|y_1...y_p) \\
&= \sum_m p(Z_{t+1} = h|Z_t = m, y_1...y_t) \times F(y_{t+1}|y_1...y_t, Z_t = m, Z_{t+1} = h) \\
&\times \; F(y_{p+1}...y_t, Z_t = m|y_1...y_p) \\
&= \sum_m \pi_{m,h}\alpha_t(m)\Phi(\frac{y_t - \mathbf{Y}'_{t-1}A_h}{\sigma_h}) \qquad (14)
\end{aligned}$$

Also by lemma 2.1, for $\beta_t(h)$ in equation (12) we have

$$\begin{aligned}
\beta_t(h) &= \sum_{j=1}^{K} F(Z_{t+1} = j, y_{t+1}...y_T|y_1...y_t, Z_t = h) \\
&= \sum_{j=1}^{K} F(y_{t+2}...y_T|y_1...y_t, y_{t+1}, Z_t = h, Z_{t+1} = j) \times \\
&\quad F(y_{t+1}|y_1...y_t, Z_t = h, Z_{t+1} = j)p(Z_{t+1} = j|y_1...y_t, Z_t = h) \\
&= \sum_{j=1}^{K} \pi_{h,j}\beta_{t+1}(j)\Phi(\frac{y_{t+1} - \mathbf{Y}'_{t-1}A_j}{\sigma_j}) \qquad (15)
\end{aligned}$$

□

2.2 Modification of EM Algorithm

The EM algorithm is used for maximization of completed data log-likelihood. By completed data we mean that the set of time series observations $\{y_t\}_{t=1}^T$ augmented with the latent set of correct predictor indicators $\{z_t\}_{t=p+1}^T$ (i.e. $\{\{y_t\}_{t=1}^T, \{z_t\}_{t=p+1}^T\}$). So this log-likelihood, by the method of iterative conditioning, can be represented as

$$\begin{aligned}
\ell^*(\theta) &= \log F(y_{p+1}...y_T, z_{p+1}...z_T|y_1...y_p) \\
&= \sum_{t=p+1}^{T} \log(F(y_t|y_{t-1}, \cdots, y_0, z_t)) + \sum_{t=p+2}^{T} \log(P(z_t|z_{t-1}, \cdots, Zp, y_{t-1}, \cdots, y_0)) + \\
&\quad \log P(z_{p+1}|y_1, \cdots, y_p) \\
&= \sum_{t=p+1}^{T} \sum_k I(z_t = k) \log \Phi(\frac{y_t - \mathbf{Y}'_{t-1}A_k}{\sigma_k}) +
\end{aligned}$$

$$\sum_{t=p+2}^{T}\sum_{k}\sum_{j} I(z_t = k)I(z_{t-1} = j)\log \pi_{j,k} +$$
$$\sum_{k} I(z_{p+1} = k)\log \rho_k,$$

where the last equality holds by (7) and the Markov property of $\{Z_t\}$ with transition probabilities in (8). It is clear that $\sum_{t=p+2}^{T} I(z_t = k)I(z_{t-1} = j)$ is equal to the number of transitions from state j to state k. At the E-step, the algorithm computes the conditional expected value of each $I(z_t = k)$ and $I(z_t = k)I(z_{t-1} = j)$ given the observed data.

$$E[\ell^*(\theta)|y_1, \cdots, y_T] = \sum_{t=p+2}^{T}\sum_{k}\sum_{j} P(z_t = k, z_{t-1} = j|y_1...y_T)\log \pi_{j,k} +$$
$$\sum_{t=p+1}^{T}\sum_{k} P(z_t = k|y_1...y_T)\{-\log(\sqrt{2\pi}) - \log(\sigma_k) - \frac{(y_t - \mathbf{Y}'_{t-1}A_k)^2}{2\sigma_k^2})\}$$
$$+\sum_{k} P(z_{p+1} = k|y_1...y_T)\log \rho_k. \tag{16}$$

Last equation holds by linear property of expectation and since $\Phi\{(y_t - \mathbf{Y}'_{t-1}A_k)\sigma_k\}$ is measurable with respect to $\sigma\{Y_1, ..., Y_T\}$. Also $E(I(z_t = k)|y_1...y_T) = P(z_t = k|y_1...y_T)$ and $E(I(z_t = k)I(z_{t-1} = j|y_1...y_T) = P(z_t = k, z_{t-1} = j|y_1...y_T)$. These posterior probabilities can be obtained by the following lemma

Lemma 2.2. *$P(Z_t = h|Y_1, \cdots, Y_T)$ and $P(Z_t = j, Z_{t-1} = i|Y_1, \cdots, Y_T)$ in equation (16) can be calculated as*

$$P(Z_t = h|Y_1, \cdots, Y_T) = \frac{\alpha_t(h)\beta_t(h)}{F(Y_{p+1}, \cdots, Y_T|Y_1, \cdots, Y_p)},$$
$$P(Z_t = j, Z_{t-1} = i|Y_1, \cdots, Y_T) = \frac{\beta_t(j)\pi_{ij}\alpha_{t-1}(i)}{F(Y_{p+1}, \cdots, Y_T|Y_1, \cdots, Y_p)} \times$$
$$\Phi(\frac{y_t - \mathbf{Y}'_{t-1}A_k}{\sigma_j})$$

in which $F(Y_{p+1}, \cdots, Y_T|Y_1, \cdots, Y_p) = \sum_{j=1}^{K} \alpha_T(j)$ and $\{\alpha_t(.), \beta_t(.)\}_{t=p+1}^{T}$ are calculated by theorem 2.1.

Proof. Using equations (11) and (12) we have

$$P(Z_t = h|Y_1, \cdots, Y_T) = \frac{F(Z_t = h, Y_1, \cdots, Y_T)}{F(Y_1, \cdots, Y_T)}$$

$$
\begin{aligned}
&= F(Z_t = h, Y_{p+1}, \cdots, Y_t | Y_1, \cdots, Y_p) \times \\
&\quad F(Y_{t+1}, \cdots, Y_T | Z_t = h, Y_1, \cdots, Y_t) \times \frac{F(Y_1, \cdots, Y_p)}{F(Y_1, \cdots, Y_T)} \\
&= \frac{\alpha_t(h)\beta_t(h)}{F(Y_{p+1}, \cdots, Y_T | Y_1, \cdots, Y_p)}, \qquad (17)
\end{aligned}
$$

in which

$$
F(Y_{p+1}, \cdots, Y_T | Y_1, \cdots, Y_p) = \sum_{j=1}^{K} F(Y_{p+1}, \cdots, Y_T, Z_t = j | Y_1, \cdots, Y_p) = \sum_{j=1}^{K} \alpha_T(j)
$$

and by lemma 2.1, (7) and Markov property of $\{Z_t\}$ we have that

$$
\begin{aligned}
& P(Z_t = j, Z_{t-1} = i | Y_1, \cdots, Y_T) = \\
= \; & F(y_{t+1}, \cdots, y_T | y_1, \cdots, y_t, z_t, z_{t-1}) F(y_t | y_1, \cdots, y_{t-1}, z_t, z_{t-1}) \times \\
& \frac{P(z_t | y_1, \cdots, y_{t-1}, z_{t-1}) F(y_{p+1}, \cdots, y_{t-1}, z_{t-1} | y_1, \cdots, y_p)}{F(y_1, \cdots, y_T)} \\
= \; & \frac{\beta_t(j)\pi_{ij}\alpha_{t-1}(i)}{F(Y_{p+1}, \cdots, Y_T | Y_1, \cdots, Y_p)} \Phi(\frac{y_t - \mathbf{Y}'_{t-1} A_k}{\sigma_j})
\end{aligned}
$$

□

In the M-step, roots of equation $\partial E[\ell^*(\theta) | y_1, \cdots, y_T] / \partial \theta_i = 0$, $\theta_i \in \theta$, are calculated

Theorem 2.2. *Let* $\widetilde{\mathbf{Y}} = (\mathbf{Y}_p, \cdots, \mathbf{Y}_{T-1})$, $\bar{\mathbf{Y}} = (y_{p+1}, \cdots, y_T)'$ *and* $\mathbf{P}_k = diag(P(Z_{p+1} = k | y_1 ... y_T), \cdots, P(Z_T = k | y_1 ... y_T))$, *then maximum likelihood estimate of the parameters HM-MAR are given by*

$$
\hat{A}_k = (\widetilde{\mathbf{Y}} \mathbf{P}_k \widetilde{\mathbf{Y}}')^{-1} \widetilde{\mathbf{Y}} \mathbf{P}_k \bar{\mathbf{Y}} \qquad (18)
$$

$$
\hat{\sigma}_k^2 = \frac{\{\bar{\mathbf{Y}}' \mathbf{P}_k (\mathbf{I} - \widetilde{\mathbf{Y}}' (\widetilde{\mathbf{Y}} \mathbf{P}_k \widetilde{\mathbf{Y}}')^{-1}) \bar{\mathbf{Y}} \widetilde{\mathbf{Y}} \mathbf{P}_k - 2 \bar{\mathbf{Y}}' \mathbf{P}_k \widetilde{\mathbf{Y}}' (\widetilde{\mathbf{Y}} \mathbf{P}_k \widetilde{\mathbf{Y}}')^{-1} \widetilde{\mathbf{Y}} \mathbf{P}_k \bar{\mathbf{Y}}\}}{tr(\mathbf{P}_k)} \qquad (19)
$$

$$
\hat{\pi}_{j,i} = \frac{\sum_{t=p+2}^{T} P(Z_t = i, Z_{t-1} = j | y_1, \cdots, y_T)}{\sum_{t=p+2}^{T} P(Z_{t-1} = j | y_1, \cdots, y_T)} \qquad (20)
$$

$$
\hat{\rho}_j = \frac{\sum_{t=P+1}^{T} P(Z_t = j | Y_1, \cdots, Y_T)}{T - P}. \qquad (21)
$$

Proof. calculating $\partial E[\ell^*(\theta) | y_1, \cdots, y_T] / \partial \phi_k = 0$, we obtain

$$
\sum_{t=p+1}^{T} P(z_t = k | y_1 ... y_T) \mathbf{Y}_{t-1} (y_t - \mathbf{Y}'_{t-1} A_k) = 0
$$

$$\Rightarrow \widetilde{\mathbf{Y}}\mathbf{P}_k\bar{\mathbf{Y}} - \widetilde{\mathbf{Y}}\mathbf{P}_k\widetilde{\mathbf{Y}}'A_k = 0$$
$$\Rightarrow \hat{A}_k = (\widetilde{\mathbf{Y}}\mathbf{P}_k\widetilde{\mathbf{Y}}')^{-1}\widetilde{\mathbf{Y}}\mathbf{P}_k\bar{\mathbf{Y}} \tag{22}$$

calculating $\partial E[\ell^*(\theta)|y_1,\cdots,y_T]/\partial\sigma_k = 0$, we obtain

$$\sum_{t=p+1}^{T} P(z_t = k|y_1...y_T)(-\frac{1}{\sigma_k} + \frac{(y_t - \mathbf{Y}'_{t-1}A_k)^2}{\sigma_k^3}) = 0$$
$$\Rightarrow tr(\mathbf{P}_k)\sigma_k^2 = (\bar{\mathbf{Y}} - \widetilde{\mathbf{Y}}'A_k)'\mathbf{P}_k(\bar{\mathbf{Y}} - \widetilde{\mathbf{Y}}'A_k)$$
$$= \bar{\mathbf{Y}}'\mathbf{P}_k\bar{\mathbf{Y}} - 2\bar{\mathbf{Y}}'\mathbf{P}_k\widetilde{\mathbf{Y}}'A_k + A_k'\widetilde{\mathbf{Y}}\mathbf{P}_k\widetilde{\mathbf{Y}}'A_k \tag{23}$$

Since $(\bar{\mathbf{Y}}'\mathbf{P}_k\widetilde{\mathbf{Y}}'A_k)' = A_k'\widetilde{\mathbf{Y}}\mathbf{P}_k\bar{\mathbf{Y}}$. Replacing $\hat{A}_k$ from equation (22), we obtain equation (19) for $\hat{\sigma}_k^2$.

Since for each $j = 1,\cdots,K$ in the transition matrix P of Markov process Z_t, $\sum_{i=1}^{K}\pi_{j,i} = 1$ thus

$$\pi_{j,K} = 1 - \sum_{i=1}^{K-1}\pi_{j,i}. \tag{24}$$

Calculating the roots of equation $\partial E[\ell^*(\theta)|y_1,\cdots,y_T]/\partial\pi_{j,i} = 0$, by equation (24), we have

$$\pi_{j,i} = \pi_{j,K}\frac{\sum_{t=p+2}^{T} P(Z_t = i, Z_{t-1} = j|y_1,\cdots,y_T)}{\sum_{t=p+2}^{T} P(Z_t = K, Z_{t-1} = j|y_1,\cdots,y_T)}$$
$$\Rightarrow \hat{\pi}_{j,i} = \frac{\sum_{t=p+2}^{T} P(Z_t = i, Z_{t-1} = j|y_1,\cdots,y_T)}{\sum_{t=p+2}^{T} P(Z_{t-1} = j|y_1,\cdots,y_T)} \tag{25}$$

In a similar way we obtain equation (21) for $\hat{\rho}_j$ for $j = 1,\cdots,K$. □

2.3 Learning

A brief summary of HM-MAR(K,P) parameter estimation algorithm is as follows:

1. For t=1 to T do

$$\mathbf{Y}_t = (y_t, \cdots, y_{t-p+1})'$$

2. Let

$$\widetilde{\mathbf{Y}} = (\mathbf{Y}_p, \cdots, \mathbf{Y}_{T-1})$$
$$\bar{\mathbf{Y}} = (y_{p+1}, \cdots, y_T)'$$

3. For h=1 to K do

$$A_h = (w_1^{1,h}, \cdots, w_p^{1,h})'$$

4. Let

$$\rho_h = P(Z_{p+1} = h | y_1, \cdots, y_p)$$
$$\theta = \{A_j,\ \sigma_j,\ \rho_j,\ \pi_{mn},\ m, n, j = 1, \cdots, K\}$$

5. Initialize θ randomly.

6. do while none of the parameters of θ changes

   (a) $\alpha_{p+1}(h) = \rho_h \Phi(\frac{y_{p+1} - \mathbf{Y}'_p A_h}{\sigma_h})$

   (b) $\beta_T(h) = 1$

   (c) For t=1 to T do

   - $\alpha_{t+1}(h) = \sum_m \pi_{m,h} \alpha_t(m) \Phi(\frac{y_{t+1} - \mathbf{Y}'_t A_h}{\sigma_h})$
   - $\beta_{T-t}(h) = \sum_{j=1}^{K} \pi_{h,j} \beta_{T-t+1}(j) \Phi(\frac{y_{t+1} - \mathbf{Y}'_{t-1} A_j}{\sigma_j})$

   (d) $F(Y_{p+1}^T | Y_1^p) = \sum_{j=1}^{K} \alpha_T(j)$

   (e) For t=1 to T

   - $P(Z_t = h | Y_1, \cdots, Y_T) = \frac{\alpha_t(h)\beta_t(h)}{F(Y_{p+1}^T | Y_1^p)}$
   - $P(Z_t = j, Z_{t-1} = i | Y_1, \cdots, Y_T) = \frac{\pi_{ij}\beta_t(j)\alpha_{t-1}(i)}{F(Y_{p+1}^T | Y_1^p)} \Phi(\frac{y_t - \mathbf{Y}'_{t-1} A_j}{\sigma_j})$

   (f) and

   $$\mathbf{P}_k = diag(P(Z_{p+1} = k | y_1 ... y_T), \cdots, P(Z_T = k | y_1 ... y_T)).$$

   (g) set the maximum likelihood estimate as

   - $\hat{A}_k = (\widetilde{\mathbf{Y}} \mathbf{P}_k \widetilde{\mathbf{Y}}')^{-1} \widetilde{\mathbf{Y}} \mathbf{P}_k \bar{\mathbf{Y}}$
   - $\hat{\sigma}_k^2 = \frac{\bar{\mathbf{Y}}' \mathbf{P}_k (\mathbf{I} - \widetilde{\mathbf{Y}}' (\widetilde{\mathbf{Y}} \mathbf{P}_k \widetilde{\mathbf{Y}}')^{-1}) \bar{\mathbf{Y}} \widetilde{\mathbf{Y}} \mathbf{P}_k - 2 \bar{\mathbf{Y}}' \mathbf{P}_k \widetilde{\mathbf{Y}}' (\widetilde{\mathbf{Y}} \mathbf{P}_k \widetilde{\mathbf{Y}}')^{-1} \widetilde{\mathbf{Y}} \mathbf{P}_k \bar{\mathbf{Y}}}{tr(\mathbf{P}_k)}$
   - $\hat{\pi}_{j,i} = \frac{\sum_{t=p+2}^{T} P(Z_t = i, Z_{t-1} = j | y_1, \cdots, y_T)}{tr(\mathbf{P}_j)}$
   - $\hat{\rho}_j = \frac{\sum_{t=P+1}^{T} P(Z_t = j | Y_1, \cdots, Y_T)}{T - P}$

the convergence of training algorithm is issued by the convergence of all expectation maximization algorithms [2].

Remark 2.1. *If all rows of the transition probability matrix, P (3), of hidden Markov chain* $\{Z_t\}$ *are estimated to be equal, then* $\{Z_t\}$ *are independent (i.e.* $P(z_{t+1}=j|z_t=i)=P(z_{t+1}=j)$*) and*

$$\begin{aligned}
\alpha_{t+1}(i) &= \sum_{j=1}^{K} P(Z_{t+1}=i|z_t=j)P(z_t=j|y_1,\cdots,y_t) \\
&= P(z_{t+1}=i)\sum_{j=1}^{K} P(z_t=j|y_1,\cdots,y_t) = P(z_{t+1}=i) \\
&= P(z_{t+1}=i)\sum_{j=1}^{K} \alpha_t(j) = P(z_{t+1}=i) \qquad (26)
\end{aligned}$$

which implies that the weighting coefficients of HM-MAR model can be considered to be fix after parameter estimation. Thus HM-MAR model will result in a MAR model automatically without any further parameter adjustment.

References

[1] I.L. MacDonald and W. Zucchini, *Hidden Markov and other models for discrete-valued time series*, Monographs on statistics and applied probability, Chapman & Hall, 1997.

[2] G.J. McLachlan and T. Krishnan, *The EM algorithm and extensions*, Wiley series in probability and statistics, Wiley-Interscience, 2008.

[3] Chun Shan Wong and Wai Keung Li, *On a mixture autoregressive model*, Journal of the Royal Statistical Society. Series B (Statistical Methodology) **62** (2000), no. 1, pp. 95–115 (English).