Hidden Markov Mixture Autoregressive Models: Stability and Moments

S.H.Alizadeh, S.Rezakhah*

Abstract

This paper introduces a new parsimonious structure for mixture of autoregressive models. The weighting coefficients are determined through latent random variables, following a hidden Markov model. We propose a dynamic programming algorithm for the application of forecasting. We also derive the limiting behavior of unconditional first moment of the process and an appropriate upper bound for the limiting value of the variance. This can be considered as long run behavior of the process. Finally we show convergence and stability of the second moment. Further, we illustrate the efficacy of the proposed model by simulation and forecasting.

MSC: primary 62M10, 60J10 secondary 60G25

Keywords and phrases. Hidden Markov Model, Mixture Autoregressive Model, Stability, Dynamic Programming, Forecasting.

1 Introduction

The most frequently used approaches to time series model building assume that the data under study are generated from a linear stochastic process. Linear models provide a number of appealing properties (such as physical interpretations, frequency domain analysis, asymptotic results, statistical inference and many others) [7]. Despite those advantages, it is well known that real-life systems are usually nonlinear, and certain features, such as limit-cycles, asymmetry [13],[17], conditional heteroscedasticity [9], flat stretches, bursts [14] and jump phenomena cannot be correctly captured by linear statistical models.

*Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran. Email: rezakhah@aut.ac.ir, sasan_alizadeh@aut.ac.ir
Since the Mixture Transition Distribution (MTD) was originally introduced by Raftery [19] for modeling high order Markov chains in the discrete state space, the broad family of this model have been extended and applied for modeling conditional distribution of observations in the context of nonlinear time series with arbitrary state spaces [3]. This model also has been extended to the mixture transition of Gaussian distributions, known as GMTD, which contains autoregressive model as a special case, for modeling flat stretches, bursts and outliers [14]. Mixture of Autoregressive (MAR) model (which has been proposed by Wong and Li [21]) is a flexible generalization of GMTD to model processes with multimodal conditional distributions and conditional heteroscedasticity. The important feature of MAR model is that it can be considered as the mixture of some stationary and non-stationary AR processes and remains stationary. For time series \( \{Y_t\}_{t=0}^{\infty}, Y_t \in \mathbb{R} \), the MAR\((K; p_1, p_2, \ldots, p_K)\) is defined as

\[
F(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_k \Phi\left(\frac{y_t - \phi_{k0} - \phi_{k1}y_{t-1} - \ldots - \phi_{kp_k}y_{t-p_k}}{\sigma_k}\right),
\]

in which \( y_t \) denotes a realization of \( Y_t \), and \( \mathcal{F}_t = \sigma\{Y_s : s \leq t\} \) and \( F(y_t|\mathcal{F}_{t-1}) \) is the conditional distribution of \( Y_t \) given information of \( \mathcal{F}_{t-1} \). Also \( \alpha_k, k = 1, \ldots, K \) are the weighting coefficients (i.e. \( \alpha_k > 0, k = 1, \ldots, K \) and \( \sum_{k=1}^{K} \alpha_k = 1 \)) and \( \Phi(.) \) is the cumulative distribution function of the standard normal distribution. This model is a mixture of \( K \) Gaussian AR\((p_k)\), \( k = 1, \ldots, K \) models [21].

The mixture of autoregressive conditional heteroscedasticity model was also proposed by Wong and Li [22] to capture the squared autocorrelation structure of observations. Berchtold [2] also introduced a new approach for modeling heteroscedastic time series with MTD model in which the variances of each Gaussian distributions depends on the past time series observations. For exhaustive review of MTD model see [3].

In the MTD models the contribution of distributions are always fixed and it is not sensitive to the past observations. However for real processes one might expect better forecast interval if additional information from the past were allowed to affect [9]. Another approach to study mixture models is to introduce some latent variables \( \{Z_t\}_{t=p+1}^{\infty} \), which are iid and \( Y_t \) given \( Z_t \) is independent of \( \{Z_s : s \neq t\} \). Each variable \( Z_t \) has a discrete distribution with support \( \{1, \cdots, K\} \) with probability masses \( P(Z_t = k) = \alpha_k, k = 1, \cdots, K \) as the weighting coefficients in the mixture model. Since these models do not consider the dependency structure of latent variables, the dynamics of weighting coefficients can not be modeled. For finite state
space time series, Bartolucci and Farcomeni [1] studied a generalization of mixture transition models with hidden Markov models.

In this paper, we propose a new approach to model conditional distribution of \( Y_t \) given past information for nonlinear time series in general state space (i.e. \( Y_t \in \mathbb{R} \)). We use latent Markov process as an appropriate tool to consider the effect of past information and build a parsimonious model; the idea of Markov switching models (see Hamilton [12], Mcculloch and Tsay [18]) for process \( \{Y_t\}_{t=0}^{\infty} \). Our new model includes the hidden Markov model (HMM) [6] as a special case and it also generalizes MAR model in a sensible way. This model makes use of the whole past information to maximize the posterior probability of \( Z_{t-1} \) (given observed \( Y_0, \cdots, Y_{t-1} \)) and predicts the probability of \( Z_t \) by the Markov assumption of the latent process. Although using all past observations could increase the complexity of the model, we propose a dynamic programming algorithm which reduces the volume of calculations for forecasting. We derive the limiting behavior of the first unconditional moment of the process, and obtain an upper bound for the limit of variance. We also investigate the existence and stability of the second moment.

This paper is organized as follows. Hidden Markov Mixture Autoregressive (HM-MAR) model is introduced in section 2. Section 3 is devoted to the statistical properties of the HM-MAR model. Section 4 analyzes the efficiency of the proposed model through simulation and comparison of the forecast errors with the MAR model. Section 5 concludes the paper.

2 Hidden Markov Mixture Autoregressive Model

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

\[ \rho = (\rho_1, \cdots, \rho_K)', \quad \rho_h = P(Z_p = h|y_0, \cdots, y_{p-1}) \quad h = 1, \ldots, K, \tag{2} \]

with transition probability matrix

\[ P = ||\pi_{i,j}||_{K \times K}, \tag{3} \]
in which
\[ \pi_{i,j} = P(Z_t = j | Z_{t-1} = i), \quad i, j \in \{1, ..., K\}, \]  
and invariant probability measure
\[ \mu = (\alpha_1, ..., \alpha_K)', \]  
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, \( p \) lagged observations in the AR processes, if the conditional distribution of \( Y_t \) given \( F_{t-1} \) is defined as follows:

i. For \( t = p \)
\[ F(y_{p}, Z_{p} = h|F_{p-1}) = \rho_h \Phi\left( \frac{y_{p} - a_{0,h} - a_{1,h} y_{p-1} - ... - a_{p,h} y_{0}}{\sigma_h} \right), \]  

ii. For \( t \geq p + 1 \)
\[ F(y_{t}|F_{t-1}) = \sum_{h=1}^{K} \alpha_{h}^{(t)} \Phi\left( \frac{y_{t} - a_{0,h} - a_{1,h} y_{t-1} - ... - a_{p,h} y_{t-p}}{\sigma_h} \right), \]  
where \( \alpha_{h}^{(t)} = P(Z_t = h|F_{t-1}) \) 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 and conditioning on \( Z_t \). We assume \( Y_t \) is \( p \)-tuple Markov, independent of \( \{Z_s, s \neq t\} \). In other words, by conditioning on \( \{Y_{t-1}, ..., Y_{t-p}\} \) 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. It makes use of the all past observations from \( Y_0 \) up to \( Y_{t-1} \). The hidden Markov assumption of the process \( \{Z_t\}_{t \geq p} \), enables us to build a parsimonious model.

The MAR model [21] can be considered as a special case of such a HM-MAR model [6,7], 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 \( i, j = 1, ..., K \). That is \( \{Z_t\}_{t=p+1}^{\infty} \) 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 [7] (i.e. \( Y_t \) given \( Z_t \), is independent of past observations).
3 Statistical Properties of the Model

In this section, we discuss the statistical properties of the HM-MAR model. We propose a dynamic programming approach to calculate conditional expectation and variance of the process. We also investigate the long run behavior of the first order HM-MAR($K, 1$) process, including limiting behavior of the unconditional first moment, and an appropriate upper bound for the limiting value of the variance. Finally convergence and stability of second moment is proved.

3.1 Forecasting

In HM-MAR model (6-7), the conditional expectation as the least square predictor (page 64 of [7]) of the process $Y_t$ for $t \geq p + 1$ is obtained by

$$E(Y_t | \mathcal{F}_{t-1}) = \sum_{h=1}^{K} \alpha_h^{(t)} (a_{0,h} + a_{1,h} y_{t-1} + \ldots + a_{p,h} y_{t-p}), \quad (8)$$

where $\alpha_h^{(t)}$ is measurable $\mathcal{F}_{t-1}$.

One of the main areas for modeling conditional heteroscedasticity (changes in the conditional variance) is the family of ARCH models [11], originally proposed by Engle [9] in the context of financial time series. In the class of MTD models, MAR [21] and MAR-ARCH [22] models also provide a mechanism to capture this effect. However in these models only changes in conditional mean of each distribution affect the conditional variance of process. The conditional variance of HM-MAR model is given by

$$\text{Var}(Y_t | \mathcal{F}_{t-1}) = \sum_{h=1}^{K} \alpha_h^{(t)} (\sigma_h^2 + (a_{0,h} + a_{1,h} y_{t-1} + \ldots + a_{p,h} y_{t-p})^2) -$$

$$\{ \sum_{h=1}^{K} \alpha_h^{(t)} (a_{0,h} + a_{1,h} y_{t-1} + \ldots + a_{p,h} y_{t-p}) \}^2$$

$$= \sum_{h=1}^{K} \alpha_h^{(t)} \sigma_h^2 + \sum_{h=1}^{K} \alpha_h^{(t)} \mu_{h,t}^2 - \{ \sum_{h=1}^{K} \alpha_h^{(t)} \mu_{h,t} \}^2 \quad (9)$$

in which $\mu_{h,t} = a_{0,h} + a_{1,h} y_{t-1} + \ldots + a_{p,h} y_{t-p}$ is the conditional mean of $h-$th distribution (i.e. $E[Y_t | Z_t = h, Y_{1}^{t-1}]$). Let $\mu_t$ be a random variable which takes values $\mu_{h,t}$ with probabilities $\alpha_h^{(t)}$ for $h = 1, \ldots, K$, then $\sum_{h=1}^{K} \alpha_h^{(t)} \mu_{h,t} - \{ \sum_{h=1}^{K} \alpha_h^{(t)} \mu_{h,t} \}^2$ can be interpreted as the conditional variance of $\mu_t$ given all past observations. This amount is
small (large) when all conditional means are equal (largely different). Relation (9) shows the impact of conditional mean $\mu_{h,t}$ and weighting coefficients $\alpha_{h}^{(t)}$ on the value of conditional variance of $Y_{t}$ given all past information. This is the merit of the HM-MAR model and its capability to model conditional heteroscedasticity as a function of simultaneous changes in the weighting coefficients as well as conditional mean of each distribution.

At each time step $t$, $\alpha_{h}^{(t)}$ (in equations (8) and (9)) can be determined via a dynamic programming method based on forward recursion algorithm, proposed in remark 3.1.

Remark 3.1. Let $y_{s}^r \equiv (y_{r}, \cdots, y_{s})$ for $s > r$, the weighting functions in the HM-MAR model (6-7) satisfy

$$\alpha_{h}^{(t)} = \sum_{m=1}^{K} \frac{F(y_{p}^{t-1}, Z_{t-1} = m|y_{0}^{p-1})\pi_{m,h}}{\sum_{m=1}^{K} F(y_{p}^{t-1}, Z_{t-1} = m|y_{0}^{p-1})},$$

where $F(y_{p}^{t}, z_{t}|y_{1}^{p-1})$ is calculated recursively as

$$F(y_{p}^{t}, Z_{t} = h|y_{0}^{p-1}) = \sum_{m} F(y_{p}^{t-1}, Z_{t-1} = m|y_{0}^{p-1})\pi_{m,h}\Phi(y_{t} - a_{0,h} - \sum_{i=1}^{p} a_{i,h}y_{t-i})\sigma_{h},$$

and recursion starts for $t = p$ by

$$F(y_{p}, Z_{p+1} = h|y_{0}^{p-1}) = \rho_{h}\Phi(y_{p} - a_{0,h} - \sum_{i=1}^{p} a_{i,h}y_{p-i})\sigma_{h},$$

Proof. As the hidden variables $\{Z_{t}\}_{t \geq p}$ have Markov structure in HM-MAR model, we have

$$\alpha_{h}^{(t)} = P(Z_{t} = h|y_{0}^{t-1}) = \sum_{m=1}^{K} P(Z_{t} = h, Z_{t-1} = m|y_{0}^{t-1})$$

$$= \sum_{m=1}^{K} P(Z_{t} = h|Z_{t-1} = m, y_{0}^{t-1})P(Z_{t-1} = m|y_{0}^{t-1})$$

$$= \sum_{m=1}^{K} P(Z_{t} = h|Z_{t-1} = m)P(Z_{t-1} = m|y_{0}^{t-1})$$

$$= \sum_{m=1}^{K} \frac{F(y_{0}^{t-1}, Z_{t-1} = m)\pi_{m,h}}{\sum_{m=1}^{K} F(y_{0}^{t-1}, Z_{t-1} = m)}.$$
where
\[
F(y_p^{t-1}, Z_{t-1} = m|y_0^{p-1}) = \sum_{j=1}^K F(y_p^{t-1}, Z_{t-1} = m, Z_{t-2} = j|y_0^{p-1}) =
\]
\[
\sum_{j=1}^K \sum_{m=1}^K \Phi(y_{t-1} - a_{0,m} - a_{1,m}y_{t-2} - \cdots - a_{p,m}y_{t-p-1}) \pi_{j,m} F(y_p^{t-2}, Z_{t-2} = j|y_0^{p-1}),
\]
in which the last equality implies by (7) and the recursion begins with (6).

Another characteristic of HM-MAR is modeling the all past observations and benefits from a dynamic programming approach. This will in turn minimize the volume of calculations for forecasting. The intermediate results and in fact the last state \(F(y_{p+1}, \ldots, y_t, Z_t = h|y_1, \ldots, y_p)\) is stored for different values of \(Z_t\) which could be used to update the process, see (10)(11).

### 3.2 Stability

In this section, we investigate the stability of moments for the nonlinear process \(\{Y_t\}_{t=0}^{\infty}\) that admits a HM-MAR(\(K, 1\)) model. This process is represented as a random coefficient autoregressive process of order one, in which the autoregressive coefficients are functions of the latent random variables \(\{Z_t\}_{t=1}^{\infty}\), (see Equations (2)-(5)). Let random variables and \(\sigma_Z\) respectively \(a_{i,Z_t}\) take values \(\{a_{i,1}, \ldots, a_{i,K}\}\) for \(i = 0, 1\), and \(\{\sigma_1, \ldots, \sigma_K\}\), where \(a_{i,j}\) and \(\sigma_j\), \(j = 1, \ldots, K\) are used in HM-MAR model (6)(7) with \(p = 1\). We consider
\[
Y_t = a_{0,Z_t} + a_{1,Z_t} Y_{t-1} + \sigma_Z \varepsilon_t, \quad (12)
\]
where \(\{\varepsilon_t\}_{t=1}^{\infty}\) is a Gaussian IID(0,1) process, independent of the hidden process \(\{Z_t\}_{t=1}^{\infty}\). The conditional distribution of the process \(Y_t\) in Equation (12) is determined as
\[
F(y_t|\mathcal{F}_{t-1}) = \sum_{h=1}^K P(Z_t = h|\mathcal{F}_{t-1}) F(y_t|Z_t = h, \mathcal{F}_{t-1}),
\]
in which \( P(z_t = h|\mathcal{F}_{t-1}) = \alpha_t^h \) is given by remark 3.1. By the Gaussian distribution of \( \varepsilon_t \) in (12), we have
\[
F(y_t|Z_t = h, \mathcal{F}_{t-1}) = \Phi\left(\frac{y_t - a_0,h - a_1,h y_{t-1}}{\sigma_h}\right).
\]
Thus (6-7) implies that \( \{Y_t\}_{t=0}^{\infty} \) admits HM-MAR\((K, 1)\) model.

Notice that the process \( \{Y_t\}_{t=0}^{\infty} \) is not necessarily a Markov process, however the extended process \( X = \{X_t\}_{t=1}^{\infty} \) with \( X_t = (Z_t, \bar{Y}_t = (Y_t, Y_{t-1}, \ldots, Y_{t-p})')' \) is Markov [23].

Timmermann [20] derived the moments of a class of stationary Markov switching models with state-dependent autoregressive dynamics and conditional mean, \( \mu_{Z_t} \). Our approach for deriving the limiting behavior of first and second moments of the process \( Y_t \) is not based on the stationary assumption of the model.

Let’s define the \( K \times K \) diagonal matrixes
\[
\phi_i = \text{diag}(a_{i,1}, \cdots, a_{i,K}), \quad i = 0, 1,
\]
\[
\sigma = \text{diag}(\sigma_1, \cdots, \sigma_K),
\]
for possible values of random variables \( a_i, Z_t \) and \( \sigma_{Z_t} \) in equation (12) where \( 1 = (1, \cdots, 1)' \) is a \( K \times 1 \) vector.

**Lemma 3.1.** Let \( \{Y_t\}_{t=0}^{\infty} \) be a HM-MAR\((K, 1)\) process defined by (12), then for \( n \geq 2 \)
\[
\begin{pmatrix}
E[\prod_{t=2}^{n} a_1, Z_t | Z_1 = 1] \\
\vdots \\
E[\prod_{t=2}^{n} a_1, Z_t | Z_1 = K]
\end{pmatrix} = (P \phi_1)^{n-1} 1.
\]

**Proof.** By the Markov property of \( \{Z_t\}_{t=1}^{\infty} \) we have that
\[
E[a_1, Z_t | \sigma\{Z_s, s \leq t - 1\}] = E[a_1, Z_t | Z_{t-1}].
\]
So
\[
E[\prod_{t=2}^{n} a_1, Z_t | Z_1 = k] = \sum_{Z_2,\cdots, Z_n} (\prod_{t=2}^{n} a_1, Z_t) P(Z_2, \cdots, Z_n | Z_1 = k)
\]
\[
= \sum_{Z_2,\cdots, Z_n} (\prod_{t=2}^{n} a_1, Z_t) P(Z_3, \cdots, Z_n | Z_1 = k) P(Z_2 | Z_1 = k)
\]

8
\[
= \sum_{Z_2} \{ \sum_{Z_3, \ldots, Z_n} (\prod_{t=3}^{n} a_{1,Z_t}) P(Z_3, \ldots, Z_n | Z_2) \} a_{1,Z_2} P(Z_2 | Z_1 = k) \\
= E[E[\prod_{t=3}^{n} a_{1,Z_t} | Z_2] a_{1,Z_2} | Z_1 = k].
\]

So for vector of conditional expectations of \( \prod_{t=2}^{n} a_{1,Z_t} \) given different values of \( Z_1 \), we have the following recursive equation

\[
\begin{bmatrix}
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_1 = 1] \\
\vdots \\
E[\prod_{t=2}^{k+1} a_{1,Z_t} | Z_1 = K] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_1 = 1] \\
\end{bmatrix} = 
\begin{bmatrix}
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = \pi_{11}] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = \pi_{1K}] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = 1] \\
\end{bmatrix} 
\begin{bmatrix}
a_{1,1} & 0 & \cdots \\
\vdots & \ddots & \vdots \\
a_{K1} & \cdots & \pi_{K1} \\
\end{bmatrix} 
\begin{bmatrix}
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = 1] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = K] \\
\end{bmatrix}
\]

\[
= P \phi_1 \begin{bmatrix}
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = 1] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = K] \\
\end{bmatrix},
\]

in which the recursion starts at \( t = n - 1 \) as

\[
\begin{bmatrix}
E[a_{1,Z_n} | Z_{n-1} = 1] \\
\vdots \\
E[a_{1,Z_n} | Z_{n-1} = K] \\
\end{bmatrix} = 
\begin{bmatrix}
\sum_{i=1}^{K} a_{1,i} \pi_{1k} \\
\vdots \\
\sum_{i=1}^{K} a_{1,i} \pi_{Kk} \\
\end{bmatrix} 
\begin{bmatrix}
\pi_{11} a_{1,1} & \cdots & \pi_{1K} a_{1,K} \\
\vdots & \ddots & \vdots \\
\pi_{K1} a_{1,1} & \cdots & \pi_{KK} a_{1,K} \\
\end{bmatrix} 
\begin{bmatrix}
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = 1] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,Z_t} | Z_2 = K] \\
\end{bmatrix}
\]

\[
= P \phi_1 1.
\]
Thus the solution of recursive equation (13) is given by

\[
\begin{pmatrix}
E[\prod_{t=2}^{n} a_{1,z_{t}} | Z_{1} = 1] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,z_{t}} | Z_{1} = K]
\end{pmatrix} = (P\phi_{1})^{n-1}1.
\]

\[
\text{Lemma 3.2.} \text{ Let } \{Z_{t}\}_{t=1}^{\infty} \text{ be a Markov chain starting with invariant probability measure } \mu \text{ defined by (5), then under conditions of the lemma 3.1}
\]

\[
E[\prod_{t=2}^{n} a_{1,Z_{t}a_{0,Z_{1}}}] = \mu'_{0}(P\phi_{1})^{n-1}1.
\]

Proof. By lemma 3.1, we have

\[
E[\prod_{t=2}^{n} a_{1,Z_{t}a_{0,Z_{1}}}] = E[E[\prod_{t=2}^{n} a_{1,Z_{t}} | Z_{1}]a_{0,Z_{1}}] = \\
= \sum_{k=1}^{K} \alpha_{k}a_{0,0}E[\prod_{t=2}^{n} a_{1,Z_{t}} | Z_{p} = k] = \\
= (\alpha_{1}, \ldots, \alpha_{K})'\begin{bmatrix}
a_{0,1} & 0 & \cdots & 0 \\
0 & a_{0,2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & a_{0,K}
\end{bmatrix} \begin{bmatrix}
E[\prod_{t=2}^{n} a_{1,Z_{t}} | Z_{p} = 1] \\
\vdots \\
E[\prod_{t=2}^{n} a_{1,Z_{t}} | Z_{p} = K]
\end{bmatrix} = \mu'_{0}(P\phi_{1})^{n-1}1.
\]

\[
\text{Lemma 3.3. If all eigenvalues of } P\phi_{1} \text{ lie inside the unite circle then under conditions of lemma 3.2}
\]

\[
i. \lim_{m \to \infty} E[\prod_{n=2}^{m+1} a_{1,Z_{n}a_{0,Z_{1}}}] = 0,
\]

\[
ii. \lim_{t \to \infty} \sum_{m=0}^{t} E[\prod_{n=2}^{m+1} a_{1,Z_{n}a_{0,Z_{1}}}] = \mu'_{0}(I - P\phi_{1})^{-1}1.
\]

Also if all eigenvalues of \( P\phi_{1}^{2} \) lie inside the unite circle then

\[
i. \lim_{m \to \infty} E[(\prod_{n=2}^{m+1} a_{1,Z_{n}a_{0,Z_{1}}})^{2}] = 0,
\]

\[
ii. \lim_{t \to \infty} \sum_{m=0}^{t} E[(\prod_{n=2}^{m+1} a_{1,Z_{n}a_{0,Z_{1}}})^{2}] = \mu'_{0}^{2}(I - P\phi_{1}^{2})^{-1}1.
\]
Proof. The first part is an immediate result of lemma 3.2 and Datta (page 508 of [8]) and for the second part:
\[
\lim_{t \to \infty} \sum_{m=0}^{t-1} E[m+1] \prod_{n=2}^{m+1} a_n Z_n a_0 Z_1 = \lim_{t \to \infty} \sum_{m=0}^{t} \mu' \phi_0 (P \phi_1)^m 1
\]
\[
= \mu' \phi_0 (I - P \phi_1)^{-1},
\]
in which the last equality holds by Datta (page 511 of [8]). The rest of proof can be done in a similar way by conducting a result similar to lemma 3.2 as
\[
E[(\prod_{n=2}^{m+1} a_n Z_n a_0 Z_1)^2] = \mu' \phi_0 (P \phi_1^2)^m 1.
\]

Lemma 3.4. If \( E[Y_0^2] < \infty \) then under conditions of lemma 3.3
\[
\lim_{t \to \infty} E[\prod_{i=1}^{t} a_1 Z_i Y_0] = 0.
\]

Proof. By Cauchy Schwarz inequality we have
\[
[Cov(\prod_{i=1}^{t} a_1 Z_i, Y_0)]^2 < Var(\prod_{i=1}^{t} a_1 Z_i) Var(Y_0),
\]
by lemma 3.3 we can deduce that
\[
\lim_{t \to \infty} Var(\prod_{i=1}^{t} a_1 Z_i) = 0,
\]
and since \( Var(Y_0) < \infty \), so
\[
\lim_{t \to \infty} Cov(\prod_{i=1}^{t} a_1 Z_i, Y_0) = 0,
\]
thus
\[
\lim_{t \to \infty} E[\prod_{i=1}^{t} a_1 Z_i Y_0] = \lim_{t \to \infty} E[\prod_{i=1}^{t} a_1 Z_i E[Y_0] = 0,
\]
in which the last equality can be verified by lemma 3.3 and the fact that \( E[Y_0] \) is finite by the assumption that \( E[Y_0^2] < \infty \) (page 274 of [5]).

Theorem 3.1. Let \( \{Y_t\}_{t=0}^{\infty} \) follows the HM-MAR(K,1) model, defined by (12), and the following assumptions hold
i. \( \{Z_t\}_{t>1} \) is an ergodic Markov chain starting from its invariant probability measure \( \mu \) specified in equation (5),

ii. \( E[Y_0^2] < \infty \),

iii. All eigenvalues of \( P_{\phi_1} \) and \( P_{\phi_2} \) lie inside the unit circle, then the process is asymptotically stable in mean and

\[
\lim_{t \to \infty} E[Y_t] = \mu \phi_0 (I - P_{\phi_1})^{-1}. \tag{16}
\]

**Proof.** Iterating equation (12), we get

\[
Y_t = a_{0, Z_t} + a_{1, Z_t} Y_{t-1} + \sigma Z_t \varepsilon_t
\]

\[
= a_{0, Z_t} + a_{1, Z_t} a_{0, Z_{t-1}} + a_{1, Z_t} \sigma Z_{t-1} \varepsilon_{t-1} + \sigma Z_t \varepsilon_t + a_{1, Z_t} a_{1, Z_{t-1}} Y_{t-2}
\]

\[
= \sum_{m=0}^{t-1} \prod_{i=0}^{m-1} a_{1, Z_{t-i}}(a_{0, Z_{t-m}} + \sigma Z_{t-m} \varepsilon_{t-m}) + \prod_{i=0}^{t-1} a_{1, Z_{t-i}} Y_0. \tag{17}
\]

Let \( u = t - i \) in (17) to get

\[
Y_t = \sum_{m=0}^{t-1} \prod_{u=t-m+1}^{t} a_{1, Z_u}(a_{0, Z_{t-m}} + \sigma Z_{t-m} \varepsilon_{t-m}) + \prod_{u=1}^{t} a_{1, Z_u} Y_0
\]

\[
= \sum_{m=0}^{t-1} \prod_{m+1}^{m+1} a_{1, Z_u}(a_{0, Z_{t-m}} + \sigma Z_{t-m} \varepsilon_{t-m}) + \prod_{u=1}^{t} a_{1, Z_u} Y_0, \tag{18}
\]

where the last equality follows from the strict stationarity property of \( \{Z_t\}_{t=1}^{\infty} \) (page 35 of [10]), which implies by assumption (i) of theorem. Also by the independence assumption of \( \{\varepsilon_t\} \) from \( \{Z_t\}_{t=1}^{\infty} \) in (12):

\[
\lim_{t \to \infty} E\left[\sum_{m=0}^{t-1} \prod_{u=t-m+2}^{t} a_{1, Z_u} \sigma Z_{t-m} \varepsilon_{t-m}\right] = \lim_{t \to \infty} E\left[\sum_{m=0}^{t-1} \prod_{u=2}^{t} a_{1, Z_u} \varepsilon_{t-m}\right] = 0. \tag{19}
\]

Thus by lemma 3.4 and (18) (19) we have that

\[
\lim_{t \to \infty} E[Y_t] = \lim_{t \to \infty} E\left[\sum_{m=0}^{t-1} \prod_{u=2}^{m+1} a_{1, Z_u} a_{0, Z_{t}}, \right]
\]

so by assumption (iii) and lemma 3.3 we get (16).

One interesting feature of Theorem 3.1 is that HM-MAR model could consist of some explosive (with \( a_1 \geq 1 \)) and non-explosive autoregressive processes and it remains asymptotically stable in mean.
Definition 3.1. Let $\lambda$ be the spectral radius of
\[ A \equiv 1(P\phi^1_0^1)'I = \text{diag}(E[a^2_1,Z_t | Z_{t-1} = 1], \ldots, E[a^2_1,Z_t | Z_{t-1} = K]). \]

Lemma 3.5. Let spectral radius $\lambda$ to be as in definition 3.1. If $\lambda$ lies inside the unit circle then under conditions of lemma 3.2
\[ \lim_{t \to \infty} E[(\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i} a_{0,Z_i})^2] < 2(1 + \mu'\phi_0^1)(1 - \lambda^{1/2})^2 < \infty. \] (20)

Furthermore if $E[Y_0^{2+\epsilon}] < \infty$, $\epsilon > 0$ then
\[ \lim_{t \to \infty} E[\prod_{i=1}^{t} a^2_{1,Z_i} Y_0^2] = 0. \]

Proof. By definition of spectral radius wee have that the absolute values of all eigenvalues of $A$ are less than or equal to $\lambda$, so by the lemma assumption about $\lambda$, we have that $E[a^2_1,Z_t | Z_{t-1} = k] \leq \lambda < 1$ for all values of $k = 1, \ldots, K$, thus by the method of iterative conditioning
\[ E[\prod_{u=2}^{m+1} a^2_{1,Z_u} a^2_{0,Z_1}] = E[E[\prod_{u=2}^{m+1} a^2_{1,Z_u} a^2_{0,Z_1} | \sigma\{Z_1^m\}]] \]
\[ = E[E[a^2_{1,Z_{m+1}} | \sigma\{Z_1^m\}]] \prod_{u=2}^{m} a^2_{1,Z_u} a^2_{0,Z_1} \]
\[ \leq \lambda E[\prod_{u=2}^{m} a^2_{1,Z_u} a^2_{0,Z_1}], \] (21)
in which $\sigma\{Z_1^m\} \equiv \sigma\{Z_1, \ldots, Z_m\}$. Iterating (21) we get
\[ E[\prod_{u=2}^{m+1} a^2_{1,Z_u} a^2_{0,Z_1}] \leq \lambda^m E[a^2_{1,Z_1}] = \lambda^m \mu'\phi_0^2. \] (22)
thus
\[ \lim_{t \to \infty} \frac{1}{m+1} \prod_{u=2}^{m+1} a^2_{1,Z_u} a^2_{0,Z_1} \leq \mu'\phi_0^2. \] (23)
Now by Cauchy Schwarz inequality,
\[ E^2((\prod_{i=2}^{m+1} a_{1,Z_i} a_{0,Z_1}) (\prod_{j=0}^{n+1} a_{1,Z_j} a_{0,Z_1})) \]
\[ \leq E[\prod_{i=2}^{m+1} a^2_{1,Z_i} a^2_{0,Z_1}] E[\prod_{j=0}^{n+1} a^2_{1,Z_j} a^2_{0,Z_1}], \]
thus
\[
E[(\prod_{i=2}^{m+1} a_{1,z_i} a_{0,z_1}) (\prod_{j=2}^{n+1} a_{1,z_j} a_{0,z_1})] \leq \mu' \phi_0^2 1 \lambda^{(m+n)/2},
\]
and summing up for different values of \( m \neq n = 0 \) to \( \infty \),
\[
\sum_{m=0}^{\infty} E[(\prod_{i=2}^{m+1} a_{1,z_i} a_{0,z_1}) (\prod_{j=2}^{n+1} a_{1,z_j} a_{0,z_1})] < \sum_{m=0}^{\infty} \mu' \phi_0 1 \lambda^{(m+n)/2}
\]
\[
< (\sum_{m=0}^{\infty} \mu' \phi_0^2 1 \lambda^{(m)/2})^2 = \left( \frac{\mu' \phi_0 1}{1 - \lambda^{1/2}} \right)^2.
\]
(24)

Now by (23) and (24) we have
\[
\lim_{t \to \infty} E[(\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} a_{0,z_1})^2] =
\sum_{m=0}^{\infty} E[(\prod_{i=2}^{m+1} a_{1,z_i} a_{0,z_1})^2] + 2 \sum_{m=0}^{\infty} E[(\prod_{i=2}^{m+1} a_{1,z_i} a_{0,z_1}) (\prod_{j=2}^{n+1} a_{1,z_j} a_{0,z_1})]
\]< \frac{\mu' \phi_0^2 1}{1 - \lambda} + 2 \left( \frac{\mu' \phi_0^2 1}{1 - \lambda^{1/2}} \right)^2 < 2 \left( 2 + \frac{1 + \mu' \phi_0^2 1}{1 - \lambda^{1/2}} \right)^2.
\]

Now by Holder inequality (page 80 of \([5]\)),
\[
E[a_{1,z_1} Y_0^2] < E^{1/u}[a_{1,z_1}^{2u}] E^{1/v}[Y_0^{2v}] = (\mu' \phi_1^{2u} 1)^{1/u} E^{1/v}[Y_0^{2v}] < \infty,
\]
in which \( u, v > 1 \) and \( 1/u + 1/v = 1 \), so for \( v = 1 + \epsilon / 2 \) we set \( u = v/(v - 1) \) and \( (\mu' \phi_1^{2u} 1)^{1/u} < \infty \). Thus by inequality (22) and the fact that \( \lambda < 1 \), we have
\[
\lim_{t \to \infty} E[\prod_{i=1}^{t} a_{1,z_i} Y_0^{2}] = \lim_{t \to \infty} \lambda^{t-1} E[a_{1,z_1} Y_0^{2}] = 0.
\]
\[
\text{Thus by lemmas 3.3 and 3.5, we got the following inequality}
\]
\[
\lim_{t \to \infty} \text{Var} \left( \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} a_{0,z_1} \right) < 2 \left( 1 + \frac{\mu' \phi_0^2 1}{1 - \lambda^{1/2}} \right)^2 (\mu' \phi_0 (I - P \phi_1)^{-1} 1)^2.
\]
(25)
Theorem 3.2. Let \( \{Y_i\}_{i=0}^{\infty} \) follow the HM-MAR(\( K, 1 \)) model defined by (12) with \( \lambda \) as in definition 3.1. If the conditions of theorem 3.1 hold and

i. \( E[Y_0^2 + \epsilon] < \infty, \quad \epsilon > 0, \)

ii. \( \lambda < 1, \)

then the process has finite second moment and

\[
\lim_{t \to \infty} E(Y_t^2) \leq 2 \frac{1 + \mu \phi_0^2}{1 - \lambda^{1/2}}^2 + \mu \sigma^2 (I - P \phi_1^2)^{-1} \mathbf{1}.
\]

Proof. Using (18) we have

\[
E[Y_t^2] = E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1} \}^2] + E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i \sigma Z_i \varepsilon_{t-m}} \}^2] +
E[\prod_{i=1}^{t} a_1^2 Z_i Y_0^2] + 2E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1} \} \prod_{i=1}^{t} a_{1,Z_i Y_0}] +
2E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1} + \prod_{i=1}^{t} a_{1,Z_i Y_0} \} (\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i \sigma Z_i \varepsilon_{t-m}})],
\]

by independence of Gaussian IID(0,1) process, \( \{\varepsilon_t\} \) from \( \{Z_t\} \), (as indicated in (12)), we have

\[
E[(\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1})(\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i \sigma Z_i \varepsilon_{t-m}})] = 0.
\]

Also by Cauchy Schwarz inequality we have that

\[
[Cov(\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1} \}, \prod_{i=1}^{t} a_{1,Z_i Y_0})]^2 \leq Var(\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1})Var(\prod_{i=1}^{t} a_{1,Z_i Y_0}),
\]

lemmas 3.4 and 3.5 imply that \( \lim_{t \to \infty} Var(\prod_{i=1}^{t} a_{1,Z_i Y_0}) = 0 \), so by (25) we have

\[
\lim_{t \to \infty} [Cov(\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1} \}, \prod_{i=1}^{t} a_{1,Z_i Y_0})]^2 = 0,
\]

\[
\text{lemmas 3.4 and 3.5 imply that } \lim_{t \to \infty} Var(\prod_{i=1}^{t} a_{1,Z_i Y_0}) = 0, \text{ so by (25) we have}
\]

\[
\lim_{t \to \infty} [Cov(\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,Z_i a_0 Z_1} \}, \prod_{i=1}^{t} a_{1,Z_i Y_0})]^2 = 0,
\]

15
so we get

$$
\lim_{t \to \infty} E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} (a_0, z_1) \} \prod_{i=1}^{t} a_{1,z_i} Y_0] \\
= \lim_{t \to \infty} E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} (a_0, z_1) \}] E[\prod_{i=1}^{t} a_{1,z_i} Y_0] = 0, \quad (29)
$$

in which the last equality follows by lemma 3.3 and lemma 3.4. By a similar method as for (29) we get

$$
\lim_{t \to \infty} E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} (a_0, z_1, \varepsilon_{t-m}) \} \prod_{i=1}^{t} a_{1,z_i} Y_0] = 0. \quad (30)
$$

Thus collecting results, by lemma 3.4 (27–30) we have

$$
\lim_{t \to \infty} E[Y_t] = \lim_{t \to \infty} \{ E[\{ \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} (a_0, z_1) \}^2] + E[\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} (a_0, z_1, \varepsilon_{t-m})] \}.
$$

Now by lemma 3.3, \( \lim_{t \to \infty} E[\sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} (a_0, z_1) \}^2] = \mu \sigma^2 (I - P \phi_1^2)^{-1} 1, \)

so using lemma 3.5 we get (26).

**Remark 3.2.** An immediate consequence of theorems 3.1 and 3.2 is that

$$
\lim_{t \to \infty} Var(Y_t) \leq 2(1 + \mu' \phi_0^2 1) (I - \lambda^{1/2})^{-1} 1 - (\mu \phi_0^2 (I - P \phi_1^2)^{-1} 1)^2.
$$

This result can be considered as an appropriate upper bound for the variance as we utilize inequality (20) for the first term of (27) by Cauchy Schwarz inequality.

**Theorem 3.3.** Let \( \{Y_t\}_{t=0}^{\infty} \) follows the HM-MAR(K,1) model defined by (12) and \( \phi_i^+ = \text{diag}(|a_{i,1}|, \ldots, |a_{i,K}|) \) for \( i = 0, 1 \). Also, let conditions of theorem 3.2 hold and all eigenvalues of \( P \phi_1^+ \) lie inside the unit circle, then \( E[\lim t \to \infty Y_t^2] \) exists and is finite.

**Proof.** Let random variable \( X \) be defined as

$$
X = \lim_{t \to \infty} X_t = \lim_{t \to \infty} \sum_{m=0}^{t-1} \prod_{i=2}^{m+1} a_{1,z_i} |(a_0, z_1) + |\sigma_z, \varepsilon_{t-m}| | \prod_{i=0}^{t} a_{1,z_i} Y_0 |.
$$

By monotone convergence theorem (theorem 16.2 of [5]) \( E[X^2] = \lim_{t \to \infty} E[X_t^2] \). By the assumption of theorem 3.2 we deduce that spectral radius of \( I(1 + P(\phi_1^+)^2) \) lies inside the unit circle, so by a similar method as used to obtain (20) in lemma 3.5 we have

\[
\lim_{t \to \infty} E[(\sum_{m=0}^{t-1} | \prod_{i=2}^{m+1} a_1,z_i,\sigma_{Z_i,\varepsilon_{t-m}}|)] < 2(\frac{1 + \mu'\sigma^2 1}{1 - \lambda^{1/2}})^2. \tag{31}
\]

So by (31), lemma 3.5 and Cauchy Schwarz inequality we have that

\[
\lim_{t \to \infty} E[|Y_t|] < \infty.
\]

By the assumption, all eigenvalues of \( P\phi_1^+ \) lie inside the unit circle, so by a similar method as used to obtain (15) we have that

\[
\lim_{t \to \infty} E[(\sum_{m=0}^{t-1} | \prod_{i=2}^{m+1} a_1,z_i,\sigma_{Z_i,\varepsilon_{t-m}}|)] \leq 2(\frac{1 + \mu'\sigma^2 1}{1 - \lambda^{1/2}})^2. \tag{32}
\]

Therefor using inequality (32) instead of (28) in the proof of theorem 3.2, relation (26) changes to, we get

\[
\lim_{t \to \infty} E[X_t^2] < 2\left(\frac{1 + \mu'\sigma^2 1}{1 - \lambda^{1/2}}\right)^2 + \mu\sigma^2(I - P\phi_1^+)^{-1}1 + 2(\frac{1 + \mu'\sigma^2 1}{1 - \lambda^{1/2}})^2.
\]

Thus \( X^2 \) is integrable, so \( X \) is integrable. Also by triangular inequality we have \( |Y_t| < X \) for all \( t \) and thus for all \( \omega \in \mathbb{R} \), \( \lim_{t \to \infty} Y_t = Y \), where

\[
Y = \{ \sum_{m=0}^{\infty} \prod_{i=2}^{m+1} a_1,z_i,\sigma_{Z_i,\varepsilon_{t-m}} + \prod_{i=0}^{\infty} a_1,z_i,\varepsilon_{t-m} \}
\]

So by continuous mapping theorem [4] we have that \( Y_t^2 \to Y^2 \) almost surely. Finally \( |Y_t| < X \) implies that \( |Y_t^2| < 1 + X^2 \), so by the integrability of \( X^2 \), and dominated convergence theorem (theorem 16.4 of [5]) we conclude that \( E[\lim_{t \to \infty} Y_t^2] \) exists and

\[
E[Y^2] = E[\lim_{t \to \infty} Y_t^2] = \lim_{t \to \infty} E[Y_t^2] < \infty.
\]
4 Simulation

The hidden process \( \{Z_t\} \) in (6-7) is assumed to follow a first order Markov structure, so HM-MAR can be considered as a generalization of MAR model. Clarifying, MAR model can be considered as HM-MAR model with independent hidden process \( \{Z_t\} \). However, HM-MAR model is more complex, using the past observations to determine the next coefficients, and demanding a longer calculation to estimate the parameters and dynamically updating the weighting coefficients.

In this section, we investigate the efficiency of these models for time series where \( \{Z_t\} \) follow a first order Markov process. To this end, 100 observations are generated from the following HM-MAR model:

\[
F(y_t|\mathcal{F}_{t-1}) = \alpha_1^{(t)} \Phi\left(\frac{y_t - 0.7y_{t-1} - 0.2y_{t-2}}{1}\right) + (1 - \alpha_1^{(t)}) \Phi\left(\frac{y_t - 0.5y_{t-1} - 0.2y_{t-2}}{1}\right)
\]

with \( \rho = (1, 0)' \), (that is starting from the first model, \( \Phi\left(\frac{y_t - 0.7y_{t-1} - 0.2y_{t-2}}{1}\right) \)), and transition probability matrix \( P = [0.8077, 0.1923; 0.7619, 0.2381] \). We used EM [16] algorithm to estimate the conditional probability of hidden variable \( Z_t \) given \( Y_1, \ldots, Y_T \) (i.e. \( P(Z_t|Y_1, \ldots, Y_T) \)), and Baum-Welch [15] algorithm to estimate the joint conditional probability of \( Z_t, Z_{t-1} \) given \( Y_1, \ldots, Y_T \) (i.e. \( P(Z_t, Z_{t-1}|Y_1, \ldots, Y_T) \)). Using these estimations we get the following HM-MAR model:

\[
\hat{F}_{HM-MAR}(y_t|\mathcal{F}_{t-1}) = \alpha_1^{(t)} \Phi\left(\frac{y_t - 0.6514y_{t-1} - 0.2973y_{t-2}}{0.9887}\right) +
\]
Table 1: Sum of absolute forecasting errors by MAR and HM-MAR models in 10 iterations

| Iterations | 1     | 2     | 3     | 4     | 5     |
|------------|-------|-------|-------|-------|-------|
| HM-MAR     | 27.7559 | 27.7590 | 27.7560 | 27.7559 | 27.7576 |
| MAR        | 75.1827 | 75.0973 | 75.2079 | 75.0839 | 75.2378 |

| Iterations | 6     | 7     | 8     | 9     | 10    |
|------------|-------|-------|-------|-------|-------|
| HM-MAR     | 27.7580 | 27.7567 | 27.7567 | 27.7566 | 27.7569 |
| MAR        | 75.1054 | 75.1464 | 75.1338 | 75.09641 | 75.2176 |

\[
(1 - \alpha_1^{(t)}) \Phi \left( \frac{y_t - 0.6468y_{t-1} - 0.3050y_{t-2}}{0.9875} \right).
\]

with \( \hat{\rho} = (0.7261, 0.2739)' \) and \( \hat{P} = [0.5905, 0.4095; 0.3331, 0.6669] \), and

\[
\hat{F}_{MAR}(y_t|\mathcal{F}_{t-1}) = 0.3732\Phi \left( \frac{y_t - 0.4042y_{t-1} - 0.7121y_{t-2}}{0.9773} \right) + 0.6278\Phi \left( \frac{y_t - 0.8176y_{t-1} - 0.0485y_{t-2}}{0.8640} \right),
\]

is the estimated MAR model. In figure 4, the left figure shows the sample path of forecasting errors by MAR model (blue) and forecasting errors by HM-MAR model (green). The right one presents the sample path of simulated HM-MAR model (red), forecasted observations by MAR (blue) and forecasted observations by HM-MAR model (green). We observe that HM-MAR model produces significantly smaller forecasting errors than MAR model and a better approximation for the time series. In table 4 sum of the absolute forecast errors for MAR and HM-MAR models for ten iterations are presented.

5 Summary and discussions

We proposed HM-MAR model as a flexible structure for modeling conditional distribution of \( Y_t \) given past observations \( (Y_1, \ldots, Y_{t-1}) \) in a nonlinear time series. We considered HM-MAR model as the mixture of some Gaussian distributions where the mean of each distribution follows an AR(p) model. Unlike the ordinary mixture models, the weighting coefficients determining the contribution of distributions are not of predefined fixed form (constant values). These values are conditional probabilities of a latent variable \( Z_t \) given past observations \( (Y_1, \ldots, Y_{t-1}) \). At each time step \( t \), the
coefficients are determined through maximizing the posterior probability of latent variable \( Z_{t-1} \) given past information. Latent variables are assumed to follow a Markov process to build a parsimonious model. A suitable application for HM-MAR model is when the process \( Y_t \) is a result of some processes, and the contribution of each process changes over time. If such effects are not present in time series then our model automatically will reduce to the ordinary mixture models. HM-MAR model will also lead to hidden Markov model for continuous process \( \{Y_t\} \) where \( p \) is zero (i.e. \( Y_t \) given \( Z_t \) is independent of past observations).

Although modeling the effect of all past information makes the model complicated, a dynamic programming method is proposed for forecasting. It is worth mentioning that it is still possible to study some properties of \( \{Y_t\} \), such as asymptotic behavior of first moment, existence and finiteness of second moment and deriving the upper bound of asymptotic variance of process. Although the variances of each distribution in the mixture model are constant, the conditional variance of the process in HM-MAR model is not fixed. This feature can be used to model conditional volatility effects frequently presented in financial time series. Another interesting feature is that the first order HM-MAR(\( K,1 \)) model can be considered as a mixture of some explosive autoregressive processes (i.e. \( a_{-1} > 1 \)) and the non-explosive ones (i.e. \( a_{-1} < 1 \)). However, it is still asymptotically stable in first and second order.

This work has the potential to be applied in the context of nonlinear time series by imposing hidden Markov property for the weighting coefficients of mixture model. Also it can elaborate further researches for extending the stability results to the case of HM-MAR(\( K,p \)), where the lag of autoregressive processes is of order \( p \). Stationarity and ergodicity are two major aspects. Finally this area of research can be expanded by considering other distributions besides the Gaussian as the underlying distribution of mixture model.

References

[1] Francesco Bartolucci and Alessio Farcomeni, *A note on the mixture transition distribution and hidden markov models*, Journal of Time Series Analysis 31 (2010), no. 2, 132–138.

[2] Andr Berchtold, *Mixture transition distribution (mtd) modeling of heteroscedastic time series*, Computational Statistics & Data Analysis 41 (2003), no. 3-4, 399 – 411.
[3] Andr Berchtold and Adrian E. Raftery, *The mixture transition distribution model for high-order markov chains and non-gaussian time series*, Statistical Science 17 (2002), no. 3, pp. 328–356 (English).

[4] P. Billingsley, *Convergence of probability measures*, Wiley series in probability and statistics: Probability and statistics, Wiley, 1999.

[5] ———, *Probability and Measure, 3rd ed.*, Wiley India Pvt. Ltd., 2008.

[6] C.M. Bishop, *Pattern recognition and machine learning*, Information science and statistics, Springer, 2006.

[7] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, 2009.

[8] K.B. Datta, *Matrix and Linear Algebra*, Prentice-Hall Of India Pvt. Ltd., 2004.

[9] Robert F Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation*, Econometrica 50 (1982), no. 4, 987–1007.

[10] J. Fan and Q. Yao, *Nonlinear time series: nonparametric and parametric methods*, Springer series in statistics, Springer, 2005.

[11] C. Francq and J.M. Zakoian, *GARCH Models: Structure, Statistical Inference and Financial Applications*, John Wiley & Sons, 2010.

[12] James D. Hamilton, *Analysis of time series subject to changes in regime*, Journal of Econometrics 45 (1990), no. 1-2, 39–70.

[13] Tze Leung Lai and Samuel Po-Shing Wong, *Stochastic neural networks with applications to nonlinear time series*, Journal of the American Statistical Association 96 (2001), no. 455, pp. 968–981 (English).

[14] Nhu D. Le, R. Douglas Martin, and Adrian E. Raftery, *Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models*, Journal of the American Statistical Association 91 (1996), no. 436, pp. 1504–1515 (English).

[15] 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.
[16] G.J. McLachlan and T. Krishnan, *The EM algorithm and extensions*, Wiley series in probability and statistics, Wiley-Interscience, 2008.

[17] M.C. Medeiros and A. Veiga, *A hybrid linear-neural model for time series forecasting*, Neural Networks, IEEE Transactions on 11 (2000), no. 6, 1402 – 1412.

[18] McCulloch R. E. and Tsay R. S., *Statistical analysis of economic time series via markov switching models*, Journal of Time Series Analysis 15 (1994), no. 6, 523 – 539.

[19] Adrian E. Raftery, *A model for high-order markov chains*, Journal of the Royal Statistical Society. Series B (Methodological) 47 (1985), no. 3, pp. 528–539 (English).

[20] Allan Timmermann, *Moments of markov switching models*, Journal of Econometrics 96 (2000), no. 1, 75–111.

[21] 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).

[22] ______, *On a mixture autoregressive conditional heteroscedastic model*, Journal of the American Statistical Association 96 (2001), no. 455, pp. 982–995 (English).

[23] J.-F. Yao and J.-G. Attali, *On stability of nonlinear ar processes with markov switching*, Advances in Applied Probability 32 (2000), no. 2, pp. 394–407 (English).