International Review of Economics and Finance 18 (2009) 306–317

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

# Application of hidden Markov switching moving average model in the stock markets: Theory and empirical evidence[☆]

Shih-Kuei Lin [a], Shin-Yun Wang [b,*], Pei-Ling Tsai [c]

[a] Department of Finance, National University of Kaohsiung, Kaohsiung 81148, Taiwan, ROC
[b] Department of Finance, National Dong Hwa University, Hualien 97401, Taiwan, ROC
[c] Department of Business Administration , National Dong Hwa University, Hualien 97401, Taiwan, ROC

ARTICLE INFO

*Article history:*
Received 16 December 2007
Received in revised form 3 June 2008
Accepted 9 June 2008
Available online 7 August 2008

*JEL classification:*
C22

*Keywords:*
Stock return mean reversion
Hidden Markov chains
Moving average
EM algorithm

ABSTRACT

In this paper, we propose a hidden Markov switching moving average model (MS-MA model) to extend the moving average model when the dynamic process of stock returns is predictable. That is, hidden Markov chain can be utilized to better describe the stock return dynamics when moving averages are correlated. Based on the MS-MA model, a recursive method of EM algorithm for parameter estimation is proposed and a numerical analysis is demonstrated. Furthermore, we empirically test the hidden Markov chain model using Dow Jones thirty stocks' data. The empirical results show that the dynamic process of stock returns exhibits MS-MA property, meaning the moving averages of stock returns are correlated. Therefore, the MS-MA model allows us to better understand and to predict stock return stochastic process. This model also helps in pricing equity derivatives.



## 1. Introduction

Stock returns are not predictable when they follow random walk. However, stock returns may be predictable if they exhibit mean-reversion property. In the Black–Scholes options pricing model (1973), stock price changes are assumed to follow a geometric Brownian motion, thus stock price changes include a stochastic and a non-stochastic components, where the stochastic component assumes that stock price changes are independent, i.e., Cov($S_{t-1}$, $S_t$), where $S_t$ is the stock price at time $t$. However, in recent years many empirical studies have pointed out that stock returns are autocorrelated. For example, Fama and French (1988) find stock returns of 3–5 months horizon are negatively autocorrelated. Potreba and Summers (1988) find that short-term stock returns are positively autocorrelated due to the trading of noise traders. Jegadeesh (1991) also find stock return mean-reversion in the UK and US markets. Similar findings are reported in Beveridge and Oickle (1997), Cassano (1999), Chelley-Steeley (2001), and Chen, Su, and Huang (2008). Since models of moving average process (MA) have been used to explain the stock return mean-reversion, this study employs MA(1) to deduce the Markov chain model, leaving MA ($q$) to future studies.

If stock returns are hidden Markov chain correlated, i.e., moving averages vary over time, then Markov switching moving average (MS-MA) model can be used to improve the moving average model. The Markov chain was initially proposed by Markov (1906), which is composed of probable state and state transition probability. Markov discusses only a finite state Markov chain, and it was applied in the studies of physics. Wiener (1923) establishes a continuous time Markov chain, and Martin (1975) adds Bayes theory to it. The hidden Markov chain is initially proposed by Baum and Petrie (1966), in which the hidden variable could be obtained from an observable random variable set. In more recent years, the hidden Markov chain has been widely applied in speech

☆ This project is supported by NSC grant number 95-2416-H-259-021.
* Corresponding author. Department of Finance, National Dong Hwa University, No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan, ROC. Tel.: +886 3 8633136; fax: +886 3 8633130.
*E-mail address:* gracew@mail.ndhu.edu.tw (S.-Y. Wang).


doi:10.1016/j.iref.2008.06.010

recognition (Korgh, Brown, Mian, Sjolander and Haussler, 1994), signal processing (Elliott, Aggoun and Moore, 1995), DNA recognition (Churchill, 1989) and image processing (Asa, Eikvil and Huseby, 1999), though less frequently used in the pricing of financial derivatives. Nevertheless, Liao and Chen (2006) finds that if stock returns are autocorrelated, then an MA(1)-based options pricing model can be used to estimate the options value. Finally, Kim, Nelson and Startz (1998) suggest the use of Markov switching model to describe time-series dynamics when the data present heteroskedasticity. They find that test procedures which ignore the pattern of heteroskedasticity tend to reject the null of no mean reversion too often. Their model is a three state Markov-switching process estimated by the Bayesian Gibbs sampling approach.

When estimating parameters using the MS-MA model, a state is determined based on observed values. Since random state variables cannot be observed in advance, as this will cause difficulties when estimating parameters, we employ the Expectation-Maximization algorithm (EM algorithm) to estimate the maximal likelihood parameters. The EM algorithm is a very efficient numeric method. Bilmes (1998) mentions that the EM algorithm can be primarily used in the following two scenarios: (1) constraints during the observation process results in incomplete data set or missing value; and (2) difficulty in finding the parameter estimates that maximize the likelihood function. To solve these problems, one can assume missing parameter values for the empirical purpose. To simplify the solutions, the algorithm allows the incomplete-data likelihood function to become a complete-data likelihood function after incorporating preset parameters. As such, this method facilitates parameter estimation in the later stage; hence, a closed form solution can be obtained using regressive parameter estimation method.

The EM algorithm is initially proposed by Newcomb (1886) to study the parameter estimation of equal-variance mixture normal distribution models. McKendrick (1926) applies the EM algorithm in the medical industry, while Baum and Petrie (1966), Baum and Eagon (1967), and Baum, Petrie, Soules and Weiss (1970) study the application of EM algorithm in the Markov model. These studies put forward some convergence results, which are the predecessors of the applications of the EM algorithm in the hidden Markov model. Later, Orchard and Woodbury (1972) propose the missing information principle, which is very similar to the basic concept of the EM algorithm. Orchard and Woodbury (1972) also introduce general applications of missing information principle, creating the complete-data likelihood functions and the incomplete-data likelihood functions, and conclude that maximal likelihood estimate is a constant. Other EM algorithm literature include Hartley (1958); Buck (1960); Efron (1967); Blight (1970); and Chen and Finberg, (1974). In short, most of the related literature before 1977 discusses only the EM algorithm application. The importance of the EM regressive theory, especially the converged local maximum likelihood estimates derived from the EM regression is not established until Dempster, Laird and Rubin (1977).

In this paper, we use the MS-MA model to derive the parameters of stock return generating process, and to test whether the empirical data is consistent with the MS-MA phenomenon. The rest of this study is organized as follows. Section 2 introduces the MS-MA model; Section 3 shows the estimation method and test of the EM algorithm. Section 4 presents numerical analyses and use the EM algorithm to estimate the MS-MA model parameters. Thirty Dow Jones Industrial Average (DJIA) stocks are utilized for parameter estimation and to find if stock returns exhibit the hidden Markov chain properties. Section 5 concludes.

## 2. A hidden Markov switching moving average model

The MS-MA model combines geometric Brownian motion and moving average components. When average stock returns are hidden Markov chain correlated with that of the previous $n$ periods, i.e., when correlations of average stock returns are time-varying, then the log stock returns follow the MS-MA process. The MS-MA model can be expressed as

$$R_t = \mu + \sigma Z_t + \theta_i \sigma z_{t-1}, \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, I, \tag{1}$$

where $R_t$ denotes stock return at time $t$, $\mu$ is the constant expected stock return, $\sigma$ is the standard deviation of stock return, $\theta_i$ denotes the correlation factor of the stock return between time $t-1$ and time $t$, and $|\theta_i|<1$, $Z \sim N(0,1)$, $R_t \sim N(\mu+\theta_i \sigma z_{t-1}, \sigma^2)$, $t=1,2,\cdots,T$, $i \in O$, $O$ is a state space, $O=\{1, 2,\cdots, I\}$, and $i$ is the correlation state. Assuming that $i$ has two states, 1 and 2 ($\theta_1>\theta_2$), then this model is called the MS-MA model. If the correlation with previous period is high, then $\theta_i=\theta_1$; if low then $\theta_i=\theta_2$.

The state variable in Eq. (1) is a hidden random variable, representing a joint probability distribution of state and log stock return. Let $R_t$ be an observable discrete or continuous random variable, and $S_t$ a hidden state discrete random variable with $I$ possible values, $O=\{1, 2,\cdots, I\}$. In a hidden Markov model, there are two assumptions regarding the relation between these two variables. First, state at time $t$ is only correlated with state at time $t-1$, unrelated with any other time periods. That is,

$$P(S_t|S_{t-1}, R_{t-1}, \dots, S_1, R_1) = P(S_t|S_{t-1}). \tag{2}$$

Second, log stock return at time $t$ is only correlated with state at time $t$, i.e.,

$$P(R_t|S_T, R_T, S_{T-1}, R_{T-1}, \dots, S_{t+1}, R_{t+1}, S_t, S_{t-1}, R_{t-1}, \dots, S_1, R_1) = P(R_t|S_t). \tag{3}$$

The MS-MA model contains observable and unobservable continuous or discrete random variables. In this study, $S_t$ is an unobserved (hidden) discrete random variable at time $t$. This paper further assumes that the underlying hidden Markov chain defined by $P(S_t|S_{t-1})$ is consistent with the notion that state $i$ is correlated only with state $j$. Therefore, the time-dependent stochastic transition probability is $P=(p_{ij})=P(s_{t+1}=j|s_t=i)$. To facilitate model inference, let the initial value of state at $t=1$ be $\pi_i$, and $\pi_i=P(s_1=i)$, $i \in O$. A particular sequence of states is described by $\tilde{S}=(s_1, s_2,\cdots, s_T)$ where $s_t \in O$ is the state at time $t$. An observed log stock return sequence $R$ is described as $\tilde{R}=(R_1=r_1, R_2=r_2,\cdots, R_T=r_T)$. The probability of log stock return vector at time $t$ for state $i$ can be described by $b_i(r_t)=p(R_t=r_t|S_t=i)$. The complete collection of parameters for all observation distributions is represented by $\tilde{B}=\{b_i(\cdot)\}$.

## 3. Estimation and test

EM algorithm can be primarily used in the following two cases: (1) constraints during the observation process results in incomplete data set or missing value; and (2) difficulty in finding the parameter estimates that maximize the likelihood function. To solve these problems, one can assume missing parameter values for the empirical purpose. To simplify the solutions, the algorithm allows the incomplete-data likelihood function to become a complete-data likelihood function after incorporating preset parameters. As such, this method facilitates parameter estimation in the latter stage; hence, a closed form solution can be obtained using regressive parameter estimation method. Since the state setting in the MS-MA model depends on observed values, yet the state random variable is unobservable, therefore, the EM algorithm is employed to estimate the maximal likelihood parameters.

Let observed random variable set be $\tilde{R}=(r_1, r_2,\cdots, r_T)$, and unobserved hidden random variable set be $\tilde{S}=(s_1, s_2,\cdots, s_T)$, then the complete-data likelihood function $L(\Theta|R, S)$can be expressed as

$$L(\Theta|R,S) = P(R,S|\Theta) = \pi_{s_0}\prod_{t=1}^{T} p_{s_{t-1}s_t} b_i(r_t), \tag{4}$$

where $\Theta$ denotes sample space, $\Theta=(P, B, \pi)=(p_{ij}, \mu, \sigma, \theta_i, \pi_i;\ 0<p_{ij}<1, \mu\in R, \sigma^2>0, |\theta_i|<1, 0<\pi_i<1, i\in O, j\in O$. From the complete-data likelihood function, Eq. (4), we can obtain the incomplete data likelihood function $L_c(\Theta|R)$as

$$L_c(\Theta|R) = P(R|\Theta) = \sum_{s_0,s_1,\cdots,s_t=1}^{I} P(R,S|\Theta) = \sum_{s_0,s_1,\ldots,s_t=1}^{I} \pi_{s_0}\prod_{t=1}^{T} p_{s_{t-1}s_t} b_i(r_t). \tag{5}$$

### 3.1. Expectation-maximization algorithm

The EM algorithm is a general method of finding the maximum likelihood parameter estimates from a given data set when the data are incomplete or have missing values (Bilmes, 1998). The EM algorithm has two main applications: (1) when the data have missing values due to limitations in the observation process; and (2) when optimizing the likelihood function is analytically intractable but one can simplify the likelihood function by assuming the existence of additional but hidden parameters and their values. In this paper, the EM algorithm is used to derive the sample space $\Theta$ parameters, $p_{ij}$, $\mu$, $\sigma$, $\theta_i$, $\pi_i$. The EM algorithm includes two steps: E-step (expectation step) and M-step (maximization step).

#### 3.1.1. E-step

In the E-step, we take logarithm of the complete-data likelihood function, $\log P(R, S|\Theta)$, given observable stock return $R$ and last period parameter $\Theta^{(k)}$. We then take the conditional expectation of the unobserved hidden random variable $S$ under known stock return $R$ and last period parameter $\Theta^{(k)}$. That is, we define

$$Q\left(\Theta,\Theta^{(k)}\right) = E\left[\log P(R,S|\Theta)|R,\Theta^{(k)}\right]. \tag{6}$$

Substituting Eq. (5) into Eq. (6), function $Q$ can be modified as

$$\begin{aligned} Q\left(\Theta,\Theta^{(k)}\right) &= \sum_{s\in S}\log\pi_{s_0}P\left(S|R,\Theta^{(k)}\right) + \sum_{s\in S}\left(\sum_{t=1}^{T}\log p_{s_{t-1}s_t}\right)P\left(S|R,\Theta^{(k)}\right) \\ &\quad + \sum_{s\in S}\left(\sum_{t=1}^{T}\log b_{s_t}(r_t)\right)P\left(S|R,\Theta^{(k)}\right) \\ &= \sum_{i=1}^{I}\log\pi_i P\left(s_0 = i|R,\Theta^{(k)}\right) + \sum_{i=1}^{I}\sum_{j=1}^{I}\left(\sum_{t=1}^{T}\log p_{ij}\right)P\left(s_{t-1}=i, s_t=j|R,\Theta^{(k)}\right) \\ &\quad + \sum_{i=1}^{I}\left(\sum_{t=1}^{T}\log b_i(r_t)\right)P\left(s_t = i|R,\Theta^{(k)}\right). \end{aligned} \tag{7}$$

The $\sum_{i=1}^{I}\log\pi_i P(s_0=i|R, \Theta^{(k)})$ is used to derive the initial value $\pi_i$, the $\sum_{i=1}^{I}\sum_{j=1}^{I}(\sum_{t=1}^{T}\log p_{ij})P(s_{t-1}=i, s_t=j|R, \Theta^{(k)})$ is used to derive the transition probability $p_{ij}$, and the $\sum_{i=1}^{I}(\sum_{t=1}^{T}\log b_i(r_t))P(s_t=i|R, \Theta^{(k)})$ is used to derive the constant expected stock return, $\mu$; the volatility of stock return $\sigma$; and the correlation factor of the stock returns between time $t-1$ and time $t$, $\theta_i$.

#### 3.1.2. M-step

The M-step is used to maximize function $Q$,

$$\Theta^{(k+1)} = \mathrm{argmax}Q\left(\Theta,\Theta^{(k)}\right).$$

When we continue the iteration process using the above two steps, likelihood function will continue to increase until it converges to local maximum. Lagrange multiplier $\gamma$ is used to derive the initial value $\pi_i$, the transition probability $p_{ij}$ and the probability of log stock return vector at a particular time $t$ for state $i$, described as $b_i(r_t)$. To find the initial value $\pi_i$, we introduce the Lagrange multiplier λ with the constraint $\sum_{i=1}^{I} \pi_i=1$, and solves the following equation

$$\frac{\partial}{\partial \pi_i}\left[\sum_{i=1}^{I} \log \pi_i P\left(s_0=i|R,\Theta^{(k)}\right)+\gamma\left(\sum_{i=1}^{I} \pi_i-1\right)\right]=0,$$

We obtain

$$\pi_i=\frac{P\left(s_0=i|R,\Theta^{(k)}\right)}{\sum_{i=1}^{I} P\left(s_0=i|R,\Theta^{(k)}\right)}. \tag{8}$$

By parity of reasoning, the transition probability $p_{ij}$ is

$$p_{ij}=\frac{\sum_{t=1}^{T} P\left(s_{t-1}=i,s_t=j|R,\Theta^{(k)}\right)}{\sum_{t=1}^{T} P\left(s_t=j|R,\Theta^{(k)}\right)}. \tag{9}$$

Finally, based on Eq. (7), $\sum_{i=1}^{I}(\sum_{t=1}^{T} \log b_i(r_t))P(s_t=i|R,\Theta^{(k)})$ can be rewritten as

$$\begin{aligned}&\sum_{i=1}^{I}\left(\sum_{t=1}^{T} \log b_i(r_t)\right)P\left(s_t=i|R,\Theta^{(k)}\right)\\&\quad=\sum_{i=1}^{I}\left\{\sum_{t=1}^{T}\log\left[\frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{1}{2}\sigma^{-2}(r_t-\mu-\theta_i\sigma z_{t-1})^2\right]\right]\right\}P\left(s_t=i|R,\Theta^{(k)}\right).\end{aligned} \tag{10}$$

Take partial derivative of Eq. (10) with respect to the constant expected stock return $\mu$ and set it equal to 0, we can derive the estimation equation of $\mu$ as

$$\hat{\mu}=\frac{\sum_{i=1}^{I}\sum_{t=1}^{T}\left(r_t-\hat{\theta}_i\hat{\sigma}z_{t-1}\right)P\left(s_t=i|R,\Theta^{(k)}\right)}{\sum_{i=1}^{I}\sum_{t=1}^{T} P\left(s_t=i|R,\Theta^{(k)}\right)}. \tag{11}$$

Similarly, if we take partial derivative of Eq. (10) with respect to the correlation factor of the stock returns between time $t-1$ and time $t$, $\theta_i$; and the volatility of stock return, $\sigma$; and set it to 0, we can derive the estimation equations of $\theta_i$ and $\sigma$ as

$$\hat{\theta}_i=\frac{\sum_{t=1}^{T}(r_t-\hat{\mu})z_{t-1}P\left(s_t=i|R,\Theta^{(k)}\right)}{\hat{\sigma}\sum_{t=1}^{T} z_{t-1}^2 P\left(s_t=i|R,\Theta^{(k)}\right)}, \tag{12}$$

$$\hat{\sigma}=\frac{\sum_{i=1}^{I}\sum_{t=1}^{T}\left[\hat{\sigma}^{-1}\left(r_t-\hat{\mu}-\hat{\theta}_i\hat{\sigma}z_{t-1}\right)^2+\hat{\theta}_i z_{t-1}\left(r_t-\hat{\mu}-\hat{\theta}_i\hat{\sigma}z_{t-1}\right)\right]P\left(s_t=i|R,\Theta^{(k)}\right)}{\sum_{i=1}^{I}\sum_{t=1}^{T} P\left(s_t=i|R,\Theta^{(k)}\right)}. \tag{13}$$

Appendix A details the derivation process for the initial value $\pi_i$ and the transition probability $p_{ij}$; and Appendix B for the expected stock return $\mu$; the volatility of stock returns $\sigma$; and the correlation factor of the stock returns between time $t-1$ and time $t$, $\theta_i$.

*3.1.3. Efficient calculation of desired quantities*

To better estimate parameters in function $Q$, we adopt the Baum–Welch algorithm, which is a forward–backward algorithm. An initial value is given to facilitate the parameter estimation. We then iterate using the forward–backward algorithm for the estimation of $P(\tilde{r}_T|\Theta)$, and followed by $\gamma_i(t)$ and $\xi_{ij}(t)$. The $\gamma_i(t)$ and $\xi_{ij}(t)$ can be substituted into Eqs. (8), (9), (11)–(13) to estimate parameters $\pi_i$, $p_{ij}$, $\mu$, $\sigma$, and $\theta_i$. The , $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\theta}$ estimated by the MA(1) can be used as the initial values to estimate the expected stock return $\mu$; the volatility of stock returns $\sigma$; and the correlation factor of the stock returns between time $t-1$ and $t$, $\theta_i$. For $\pi_i$ and the transition probability $p_{ij}$, $1/I$ is the initial value. $\Theta^{(k)}$ can be written as

$$\Theta^{(k)}=\left(\pi_i^{(k)},p_{ij}^{(k)},\mu^{(k)},\sigma^{(k)},\theta_i^{(k)}\right),\quad i,j\in O.$$

The following algorithms use $k$-period parameters to estimate all k+1 period parameters. Forward step can be used to estimate $P(\tilde{r}_T|\Theta)$. Assume $\alpha_i(t)=P(r_1, r_2,\ldots, r_t, s_t=i|\Theta)$, $P(\tilde{r}_T|\Theta)$, can be estimated in three steps.

Step 1: Calculate $\alpha_i(1)$, $i \in O$,

$$\begin{aligned}\alpha_i(1) &= P(r_1, s_1 = i|\Theta) \\ &= P(r_1|s_1 = i, \Theta)P(s_1 = i|\Theta) \\ &= \pi_i \phi(r_1; \mu, \sigma^2).\end{aligned} \tag{14}$$

Step 2: Calculate $\alpha_j(t)$, $t$=2,3,⋯,$T$, $j \in O$,

$$\begin{aligned}\alpha_j(t) &= \left[\sum_{i=1}^{I} \alpha_i(t-1)p_{ij}\right] b_j(r_t) \\ &= \left[\sum_{i=1}^{I} \alpha_i(t-1)p_{ij}\right] \phi\left(r_t; \mu + \theta_{q_t}\sigma z_{t-1}, \sigma^2\right).\end{aligned} \tag{15}$$

Step 3: Derive $P(\tilde{r}_T|\Theta)$

$$P(\tilde{r}_T|\Theta) = \sum_{i=1}^{I} \alpha_i(T).$$

Backward step can be used to estimate $P(\tilde{r}_T|\Theta)$. Assume $\beta_i(t)$=$P(r_{t+1}, r_{t+2},\ldots, r_T|s_t$=$i,\Theta)$, $P(\tilde{r}|\Theta)$ can be estimated in three steps.

Step 1: At time $T$, as all states are known, $\beta_i(T)$=1, $i \in O$.
Step 2: Calculate $\beta_i(t)$, $t$=1, 2,…, $T-1$, $i \in O$,

$$\beta_i(t) = P(r_{t+1}, r_{t+2}, \ldots, r_T|s_t = i, \Theta) = \sum_{j=1}^{I} p_{ij}\phi\left(r_{t+1}; \mu + \theta_{q_{t+1}}\sigma z_t, \sigma^2\right)\beta_j(t+1). \tag{16}$$

Step 3: Derive $P(\tilde{r}_T|\Theta)$

$$P(\tilde{r}_T|\Theta) = \sum_{i=1}^{I} \beta_i(1)\pi_i b_i(r_1) = \sum_{i=1}^{I} \beta_i(1)\pi_i\phi\left(r_1 : \mu, \sigma^2\right).$$

This paper uses $\alpha_i(t)$ and $\beta_i(t)$ estimated in the forward and backward steps to calculate $\gamma_i(t)$. First, $\gamma_i(t)$ is defined as

$$\gamma_i(t) = P\left(s_t = i|\tilde{r}_T, \Theta^{(k)}\right).$$

Using the Bayes theorem, $\gamma_i(t)$ can be written as

$$\gamma_i(t) = P\left(s_t = i|\tilde{r}_T, \Theta^{(k)}\right) = \frac{P\left(\tilde{r}_T, s_t = i|\Theta^{(k)}\right)}{P\left(\tilde{r}_T|\Theta^{(k)}\right)} = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{I} \alpha_j(t)\beta_j(t).} \tag{17}$$

We also use $\alpha_i(t)$, $\beta_i(t)$ and $\beta_j(t+1)$ calculated in the forward and backward steps to estimate $\xi_{ij}(t)$, where $\xi_{ij}(t)$ is defined as

$$\xi_{ij}(t) = P\left(s_t = i, s_{t+1} = j|\tilde{r}_T, \Theta^{(k)}\right).$$

Using the Bayes theorem, $\xi_i(t)$ can be written as

$$\begin{aligned}\xi_{ij}(t) &= \frac{p\left(s_t = i|\tilde{r}_T, \Theta^{(k)}\right)p\left(r_{t+1}, \ldots, r_T, s_{t+1} = j|s_t = i, \Theta^{(k)}\right)}{p\left(r_{t+1}, \ldots, r_T|s_t = i, \Theta^{(k)}\right)} \\ &= \frac{\gamma_i(t)p_{ij}b_j(r_{t+1})\beta_j(t+1)}{\beta_i(t)} = \frac{\gamma_i(t)p_{ij}\beta_j(t+1)\phi\left(r_{t+1}; \mu + \theta_{q_{t+1}}\sigma z_t, \sigma^2\right)}{\beta_i(t)}.\end{aligned} \tag{18}$$

To estimate the following parameters, we can simply use relative frequencies. This paper defines updated rules as follows:

$$\pi_i^{(k+1)} = P(s_1 = i) = \gamma_i(1),$$

$$p_{ij}^{(k+1)} = \frac{\sum_{t=1}^{T} P\left(s_{t-1} = i, s_t = j|\hat{r}_T, \Theta^{(k+1)}\right)}{\sum_{t=1}^{T} P\left(s_t = j|\hat{r}_T, \Theta^{(k+1)}\right)} = \frac{\sum_{t=1}^{T-1} \xi_{ij}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)}, \quad t = 1, 2, \ldots, T, \quad i,j \in O.$$

Therefore,

$$\widehat{\mu}^{(k+1)} = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T}\left(r_t-\hat{\theta}_i^{(k)}\hat{\sigma}^{(k)}z_{t-1}\right)P\left(s_t=i|R,\Theta^{(k+1)}\right)}{\sum_{i=1}^{I}\sum_{t=1}^{T}P\left(s_t=i|R,\Theta^{(k+1)}\right)} = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T}\left(r_t-\hat{\theta}_i^{(k)}\hat{\sigma}^{(k)}z_{t-1}\right)\gamma_i(t)}{\sum_{i=1}^{I}\sum_{t=1}^{T}\gamma_i(t)},$$

$$\hat{\theta}_i^{(k+1)} = \frac{\sum_{t=1}^{T}\left(r_t-\hat{\mu}^{(k+1)}\right)z_{t-1}P\left(s_t=i|R,\Theta^{(k+1)}\right)}{\hat{\sigma}^{(k)}\sum_{t=1}^{T}z_{t-1}^2P\left(s_t=i|R,\Theta^{(k+!)}\right)} = \frac{\sum_{t=1}^{T}\left(r_t-\hat{\mu}^{(k+1)}\right)z_{t-1}\gamma_i(t)}{\hat{\sigma}^{(k)}\sum_{t=1}^{T}z_{t-1}^2\gamma_i(t)},$$

$$\hat{\sigma}^{(k+1)} = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T}\begin{bmatrix}\hat{\sigma}^{-1,(k)}\left(r_t-\hat{\mu}^{(k+1)}-\hat{\theta}_i^{(k+1)}\hat{\sigma}^{(k)}z_{t-1}\right)^2\\ +\hat{\theta}_i^{(k+1)}z_{t-1}\left(r_t-\hat{\mu}-\hat{\theta}_i^{(k+1)}\hat{\sigma}^{(k)}z_{t-1}\right)\end{bmatrix}\gamma_i(t)}{\sum_{i=1}^{I}\sum_{t=1}^{T}\gamma_i(t)}.$$

### 3.2. Likelihood ratio test (LRT)

Because the main objective of this paper is to derive the parameters of stochastic stock return generating process using MS-MA model when stock returns exhibit hidden Markov property, likelihood ratio test is used to examine whether stock returns follow hidden Markov phenomenon. The null hypothesis of the likelihood ratio test, $H_0$, is that likelihood function $L_c(\Theta|R)$ follows MA(1) model, and the alternative hypothesis, $H_a$, is that incomplete-data likelihood function $L_c(\Theta|R)$ follows MS-MA model. The likelihood ratio test can be expressed as

$$-2\ln\left[\frac{L(\Omega_0)}{L(\Omega_a)}\right] = -2\ln\left[\frac{L(\Theta|R)}{L_c(\Theta|R)}\right] \overset{\text{asy}}{\rightarrow} \chi^2_{(n-r)}, \tag{19}$$

where $L(\Omega_0)$ is the estimated likelihood value with constraint, $H_0$; and $L(\Omega_a)$ is the estimated likelihood value without constraint, $H_a$. Moreover, $n$ is the degree of freedom for $L(\Omega_a)$; and $r$ is the degree of freedom for $L(\Omega_0)$. In this paper, the estimation function for the MS-MA model has five parameters, $\pi_i$, $p_{ij}$, $\mu$, $\sigma$, $\theta_i$ where the degree of freedom for $\pi_i$ is $I-1$, the degree of freedom for $p_{ij}$ is $I(I-1)$, $\mu$ and $\sigma$ each has one degree of freedom, and the degree of freedom for $\theta_i$ is $I-1$. Hence, the degree of freedom for $L(\Omega_a)$, $n$, is $(2+I)(I-1)+2$, and the degree of freedom for $L(\Omega_0)$, $r$, is 3. Thus, there are a total of $(2+I)(I-1)-1$ degrees of freedom for the likelihood ratio test.

## 4. Numerical analysis and empirical evidence

### 4.1. Numerical analysis

In this section, we use a numerical analysis to check the convergence of model parameters. Assuming there are two states for stock returns; preset parameter values (true values) are: $\pi_1$=0.5, the transition probability $p_{11}=p_{22}$=0.99, the expected stock return $\mu$=0.0003, the volatility of stock returns $\sigma$=0.01, the correlation factors of the stock returns between time $t-1$ and $t$, $\theta_1$=0.5 and $\theta_2$=0.1, and time length is 1,000 days. We first substitute these numbers into Eq. (5) to simulate stock returns. EM algorithm is then employed to estimate model parameters. After 200 loop runs, we check if the estimated parameter values are close to the preset true values. We report some results in Table 1, Figs. 1 and 2. The dotted lines denote true values, the solid lines denote the estimated parameter values after convergence. As shown, the estimated values are very close to the true values, indicating that EM algorithm is able to calculate the MS-MA model parameter values effectively. Furthermore, after 1,000 runs of 500 converging loops, the estimated parameter values are: $\hat{\pi}_1$=0.51188, $\hat{p}_{11}$=0.97728, $\hat{p}_{22}$=0.98255, $\hat{\mu}$=0.00032, $\hat{\sigma}$=0.00998, $\hat{\theta}_1$=0.50211 and $\hat{\theta}_2$=0.08831, which are also very close to the true values.

**Table 1**
MS-MA model parameter estimation

| | $\pi_1$ | $p_{11}$ | $p_{22}$ | $\mu$ | $\sigma$ | $\theta_1$ | $\theta_2$ |
|---|---|---|---|---|---|---|---|
| True value | 0.5000 | 0.9900 | 0.9900 | 0.0003 | 0.0100 | 0.5000 | 0.1000 |
| Simulation value (1000 runs) | 0.5119 | 0.9773 | 0.9826 | 0.0003 | 0.0100 | 0.5021 | 0.0883 |
| Standard deviation | 0.1871 | 0.0706 | 0.0497 | 0.0004 | 0.0002 | 0.0236 | 0.0991 |

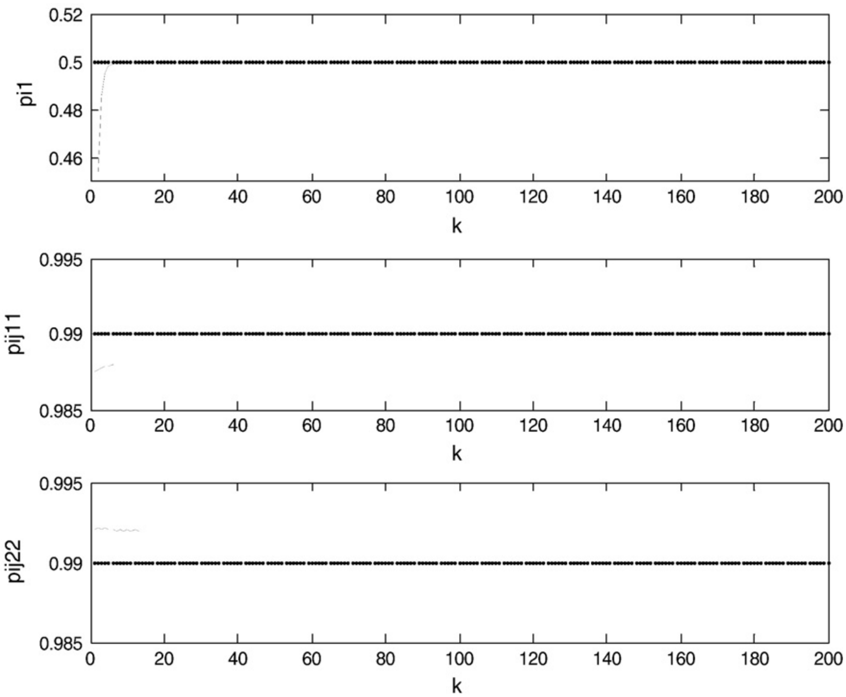


**Fig. 1.** EM algorithm is used to estimate $\pi_1$, $p_{11}$ and $p_{22}$. Assuming the initial value $\pi_1=0.5$ and the transition probability $p_{11}=p_{22}=0.99$, results converged after 200 loops.

## *4.2. Empirical evidence*

We use daily data of 30 DJIA component stocks from September of 2004 to December of 2006 (608 trading days) for the empirical analysis. In this empirical analysis we show evidence that stock returns exhibit MS-MA property. Data of stock returns are obtained from the Datastream database.

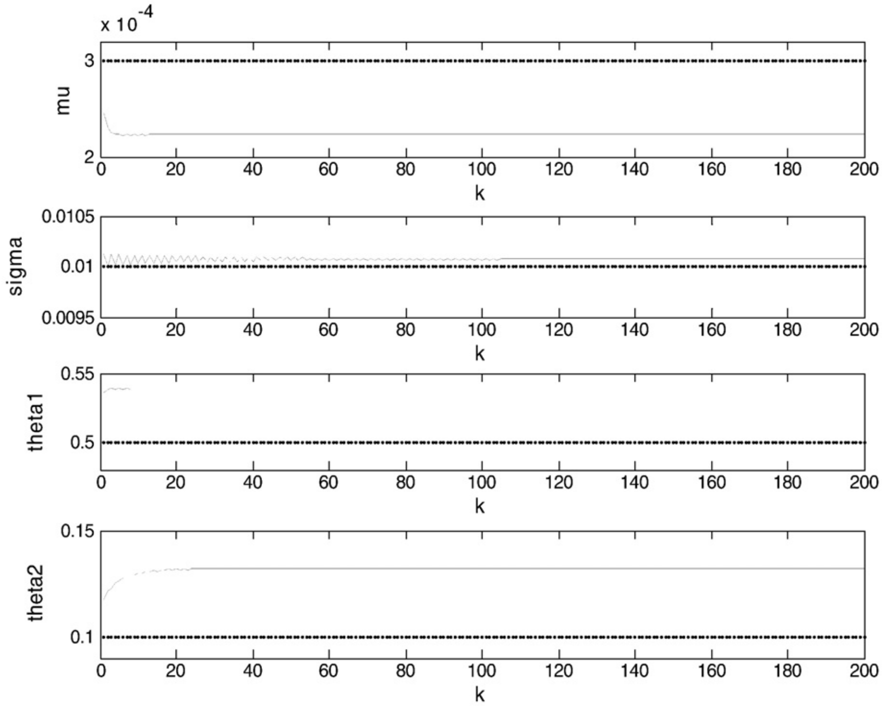


**Fig. 2.** EM algorithm is used to estimate $\mu$, $\sigma$, $\theta_1$ and $\theta_2$. Assuming the expected stock return $\mu=0.0003$, the volatility of stock returns $\sigma=0.01$, the correlation factors of the stock returns between time $t-1$ and $t$, $\theta_1=0.5$ and $\theta_2=0.1$, respectively, results converged after 200 loops.

As shown in Table 2 using likelihood test, All DJIA component stocks exhibit hidden Markov property — only Wal-Mart is significant at the 5% level; all other 29 stocks (96.67% of the total sample) are significant at the 1% level or better. Based upon our sample, therefore, hidden Markov property is evident in the stock return stochastic process. The effect of the MS-MA model, therefore, must be considered in predicting stock returns. This MS-MA effect should also be considered when pricing equity derivatives.

**Table 2**
Parameter estimates and test results for the MS-MA model using DJIA stocks

| Company | $\pi_1$ | $p_{11}$ | $p_{22}$ | $\mu$ | $\sigma$ | $\theta_1$ | $\theta_2$ | LRT |
|---|---|---|---|---|---|---|---|---|
| Alcoa | 0.4891 | 0.9911 | 0.9886 | −0.0001 | 0.0157 | 0.1055 | 0.0539 | 13.1580** (0.0043) |
| American Express | 0.4952 | 0.9898 | 0.9902 | 0.0003 | 0.0114 | 0.0152 | −0.0035 | 13.3480** (0.0039) |
| AT&T | 0.6138 | 0.9922 | 0.9872 | 0.0005 | 0.0099 | 0.1897 | 0.0814 | 12.9760** (0.0047) |
| Boeing | 0.9995 | 0.9924 | 0.9842 | 0.0010 | 0.0131 | 0.0331 | −0.4395 | 19.7624** (0.0002) |
| Caterpillar | 0.5348 | 0.9905 | 0.9893 | 0.0010 | 0.0164 | 0.0773 | 0.0483 | 12.9154** (0.0048) |
| Citigroup Inc. | 0.4617 | 0.9899 | 0.9900 | 0.0003 | 0.0087 | 0.0416 | 0.0070 | 13.3803** (0.0039) |
| Coca Cola | 0.5169 | 0.9899 | 0.9901 | 0.0011 | 0.0143 | −0.0286 | 0.0063 | 13.3561** (0.0039) |
| Dupont | 0.4455 | 0.9907 | 0.9892 | 0.0003 | 0.0109 | 0.0739 | 0.0423 | 12.9558** (0.0047) |
| Eastman Kodak | 0.4950 | 0.9901 | 0.9899 | −0.0001 | 0.0175 | −0.0308 | −0.0225 | 12.9966** (0.0046) |
| Exxon Mobil | 0.5000 | 0.9900 | 0.9900 | 0.0009 | 0.0127 | −0.0040 | −0.0041 | 12.9982** (0.0046) |
| General Electric | 0.5211 | 0.9898 | 0.9902 | 0.0002 | 0.0085 | 0.0520 | −0.0118 | 13.6363** (0.0034) |
| General Motors | 0.4974 | 0.9907 | 0.9893 | −0.0005 | 0.0254 | 0.0859 | 0.0727 | 14.4708** (0.0023) |
| Hewlett Packard | 0.4965 | 0.9904 | 0.9895 | 0.0015 | 0.0156 | −0.0683 | −0.0448 | 13.6211** (0.0035) |
| Home Depot | 0.4948 | 0.9900 | 0.9899 | 0.0002 | 0.0123 | 0.0249 | 0.0107 | 12.9045** (0.0048) |
| Honeywell International | 0.4993 | 0.9900 | 0.9900 | 0.0004 | 0.0123 | 0.0026 | −0.0041 | 12.7949** (0.0051) |
| IBM | 0.5337 | 0.9901 | 0.9898 | 0.0002 | 0.0101 | 0.0488 | 0.0329 | 12.7130** (0.0053) |
| Intel | 0.4951 | 0.9901 | 0.9899 | 0.0000 | 0.0154 | 0.0363 | 0.0305 | 13.2370** (0.0042) |
| International Paper | 0.5097 | 0.9899 | 0.9901 | −0.0003 | 0.0129 | 0.0140 | 0.0029 | 12.9765** (0.0047) |
| Johnson & Johnson | 0.4658 | 0.9898 | 0.9903 | 0.0002 | 0.0085 | −0.0108 | −0.0300 | 14.0103** (0.0029) |
| J. P. Morgan | 0.4460 | 0.9900 | 0.9899 | 0.0003 | 0.0096 | 0.0062 | −0.0050 | 13.0453** (0.0045) |
| McDonalds | 0.4469 | 0.9916 | 0.9881 | 0.0008 | 0.0123 | −0.1032 | −0.0778 | 14.0868** (0.0028) |
| Merck | 0.0001 | 0.9999 | 0.9872 | 0.0000 | 0.0203 | −0.0580 | 0.0069 | 13.5139** (0.0036) |
| Microsoft | 0.4994 | 0.9901 | 0.9899 | 0.0002 | 0.0111 | −0.0258 | −0.0166 | 12.8137** (0.0051) |
| 3M | 0.5156 | 0.9900 | 0.9899 | −0.0001 | 0.0111 | 0.0275 | 0.0273 | 12.9859** (0.0047) |
| Philip Morris | 0.0000 | 0.9990 | 0.9883 | 0.0009 | 0.0115 | −0.1115 | 0.2816 | 25.8419** (0.0000) |
| Procter Gamble | 0.5399 | 0.9917 | 0.9872 | 0.0003 | 0.0097 | −0.1589 | −0.0843 | 14.5350** (0.0023) |
| SBC Communication | 0.9328 | 0.9433 | 0.9923 | 0.0008 | 0.0161 | −0.5132 | 0.0308 | 44.3268** (0.0000) |
| United Technologies | 0.0000 | 0.9867 | 0.9976 | 0.0005 | 0.0101 | 0.5478 | −0.0858 | 40.6179** (0.0000) |
| Wal-Mart Stores | 0.9999 | 0.9891 | 0.9853 | 0.0002 | 0.0130 | −0.0631 | −0.0446 | 11.1137* (0.0149) |
| Walt Disney | 0.4975 | 0.9901 | 0.9899 | 0.0007 | 0.0115 | −0.0191 | −0.0100 | 12.9181** (0.0048) |

Note. 1. ** and * denote significant at the 1% and 5% levels respectively.
2. (.) is the $p$ value.

## 5. Conclusions

In recent years, many empirical studies find autocorrelation in stock returns, hence moving average method have been used to explain the stock return mean-reversion (e.g., Fama and French, 1988; Potreba and Summers, 1988; Jegadeesh, 1991; Beveridge and Oickle, 1997; Cassano, 1999; Chelley-Steeley, 2001). However, when stock return autocorrelation of two adjacent periods varies over time, stock returns are hidden Markov correlated. Therefore, this paper studies the case of simultaneous existence of mean-reversion and hidden Markov phenomenon in stock returns. That is, when stock returns exhibit both the MA and MS properties, then stock return stochastic process can be better structured in a MS-MA model. We use the EM algorithm to estimate the maximal likelihood parameter values, and use numerical analysis to check whether the EM algorithm effectively calculates the MS-MA model parameters. Finally empirical analysis is conducted on the 30 DJIA components stocks to investigate whether there is a hidden Markov phenomenon in stock return stochastic process.

Numerical analysis results indicate that parameter values calculated from the EM algorithm are very close to the preset actual values, meaning that EM algorithm is capable of calculating parameter values of the MS-MA model effectively. Moreover, empirical results reveal that all 30 DJIA component stocks exhibit MS-MA property, hence the MS-MA model needs to be considered to better understand stock return stochastic process and to predict stock returns. This model also helps in pricing equity derivatives.

## Appendix A. Inference of the initial value $\pi_i$ and the transition probability $p_{ij}$

For the $\sum_{i=1}^{I} \log \pi_i P(s_0=i|R, \Theta^{(k)})$ in Eq. (7), we can use Lagrange multiplier $\gamma$, and let $\sum_{i=1}^{I} \pi_i=1$ to derive the initial value $\pi_i$,

$$\frac{\partial}{\partial \pi_i}\left[\sum_{i=1}^{I} \log \pi_i P\left(s_0=i|R,\Theta^{(k)}\right)+\gamma\left(\sum_{i=1}^{I} \pi_i-1\right)\right]=0. \tag{A.1}$$

Taking partial derivative of Eq. (A.1) with respect to $\pi_i$; setting it to 0, then

$$\frac{1}{\pi_i}P\left(s_0=i|R,\Theta^{(k)}\right)+\gamma=0.$$

Taking transpose, it becomes

$$\pi_i=\frac{P\left(s_0=i|R,\Theta^{(k)}\right)}{-\gamma}. \tag{A.2}$$

Because $\sum_{i=1}^{I} \pi_i=1$, then

$$\sum_{i=1}^{I} \pi_i=\sum_{i=1}^{I} \frac{P\left(s_0=i|R,\Theta^{(k)}\right)}{-\gamma}=1,$$

and

$$-\gamma=\sum_{i=1}^{I} P\left(s_0=i|R,\Theta^{(k)}\right).$$

Thus, the estimator of the initial value $\pi_i$ can be obtained as

$$\hat{\pi}_i=\frac{P\left(s_0=i|R,\Theta^{(k)}\right)}{\sum_{i=1}^{I} P\left(s_0=i|R,\Theta^{(k)}\right)} \tag{A.3}$$

For the $\sum_{i=1}^{I}\sum_{j=1}^{I}(\sum_{t=1}^{T} \log p_{ij})P(s_{t-1}=i, s_t=j|R, \Theta^{(k)})$ in Eq. (7), we can use Lagrange multiplier $\gamma$ and let $\sum_{j=1}^{I} p_{ij}=1$ to derive the transition probability $p_{ij}$,

$$\frac{\partial}{\partial p_{ij}}\left[\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{t=1}^{T} \log p_{ij} P\left(s_{t-1}=i,s_t=j|R,\Theta^{(k)}\right)+\gamma\left(\sum_{j=1}^{I} p_{ij}-1\right)\right]=0. \tag{A.4}$$

Taking partial derivative of Eq. (A.4) with respect to the transition probability $p_{ij}$; setting it to 0, we obtain

$$\frac{1}{p_{ij}}\sum_{t=1}^{T} P\left(s_{t-1}=i,s_t=j|R,\Theta^{(k)}\right)+\gamma=0.$$

Taking transpose, it can be modified as

$$p_{ij} = \frac{\sum_{t=1}^{T} P\left(s_{t-1} = i, s_t = j | R, \Theta^{(k)}\right)}{-\gamma}. \tag{A.5}$$

Because $\sum_{j=1}^{I} p_{ij}=1$, then

$$\sum_{j=1}^{I} p_{ij} = \sum_{j=1}^{I} \frac{\sum_{t=1}^{T} P\left(s_{t-1} = i, s_t = j | R, \Theta^{(k)}\right)}{-\gamma} = 1,$$

one obtains

$$-\gamma = \sum_{i=1}^{I} \sum_{t=1}^{T} P\left(s_{t-1} = i, s_t = j | R, \Theta^{(k)}\right) = \sum_{t=1}^{T} P\left(s_t = j | R, \Theta^{(k)}\right).$$

Thus, the estimator of the transition probability $p_{ij}$ can be obtained as

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} P\left(s_{t-1} = i, s_t = j | R, \Theta^{(k)}\right)}{\sum_{t=1}^{T} P\left(s_t = j | R, \Theta^{(k)}\right)}. \tag{A.6}$$

## Appendix B. Derivation of the expected stock return, $\mu$, the correlation factor of the stock returns between time $t-1$ and $t$, $\theta_i$, and the volatility of stock returns, $\sigma$

We can use $\sum_{i=1}^{I}(\sum_{t=1}^{T} \log b_i(r_t))P(s_t=i| R, \Theta^{(k)})$ in Eq. (7) to derive the expected stock return, $\mu$; the volatility of stock returns, $\sigma$; and the correlation factor of the stock returns between time $t-1$ and, $t$ $\theta_i$. Because

$$\begin{aligned} b_i(r_t) = \phi\left(r_t : \mu + \delta_i z_{t-1}, \sigma^2\right) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{[r_t-(\mu+\theta_i\sigma z_{t-1})]^2}{2\sigma^2}\right] \\ &= \frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}} \exp\left[-\frac{1}{2}(\sigma^2)^{-1}(r_t-(\mu+\theta_i\sigma z_{t-1}))^2\right], \end{aligned} \tag{B.1}$$

then $\sum_{i=1}^{I}(\sum_{t=1}^{T} \log b_i(r_t))P(s_t=i| R, \Theta^{(k)})$ can be modified as

$$\begin{aligned} &\sum_{i=1}^{I}\left(\sum_{t=1}^{T} \log b_i(r_t)\right) P\left(s_t = i | R, \Theta^{(k)}\right) \\ &= \sum_{i=1}^{I}\left\{\sum_{t=1}^{T} \log\left[\frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}} \exp\left[-\frac{1}{2}(\sigma^2)^{-1}(r_t-(\mu+\theta_i\sigma z_{t-1}))^2\right]\right]\right\} P\left(s_t = i | R, \Theta^{(k)}\right). \end{aligned} \tag{B.2}$$

Taking partial derivative with respect to the expected stock return $\mu$ in Eq. (B.2); setting it to 0, the equation for deriving the expected stock return $\mu$ becomes

$$\frac{\partial}{\partial \mu}\sum_{i=1}^{I}\left\{\sum_{t=1}^{T} \log\left[\frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}} \exp\left[-\frac{1}{2}(\sigma^2)^{-1}(r_t-(\mu+\theta_i\sigma z_{t-1}))^2\right]\right]\right\} P\left(s_t = i | R, \Theta^{(k)}\right) = 0,$$

after differentiation, we have

$$\mu(\sigma^2)^{-1}\sum_{i=1}^{I}\sum_{t=1}^{T} P\left(s_t = i | R, \Theta^{(k)}\right) = (\sigma^2)^{-1}\sum_{i=1}^{I}\sum_{t=1}^{T} (r_t-\theta_i\sigma z_{t-1}) P\left(s_t = i | R, \Theta^{(k)}\right).$$

The estimator of the expected stock return $\mu$ can be obtained by transposing the above equation

$$\hat{\mu} = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T}\left(r_t-\hat{\theta}_i\hat{\sigma} z_{t-1}\right) P\left(s_t = i | R, \Theta^{(k)}\right)}{\sum_{i=1}^{I}\sum_{t=1}^{T} P\left(s_t = i | R, \Theta^{(k)}\right)}. \tag{B.3}$$

Taking partial derivative with respect to the correlation factor of the stock returns between time $t-1$ and $t$, $\theta_i$ in Eq. (B.2); setting it to 0, equation to derive the correlation factor of the stock returns becomes

$$\frac{\partial}{\partial\theta_i}\sum_{i=1}^{I}\left\{\sum_{t=1}^{T}\log\left[\frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}}\exp\left[-\frac{1}{2}\left(\sigma^2\right)^{-1}(r_t-(\mu+\theta_i\sigma z_{t-1}))^2\right]\right]\right\}P\left(s_t=i|R,\Theta^{(k)}\right)=0,$$

after differentiation, it becomes

$$\theta_i\sum_{t=1}^{T}\sigma^2 z_{t-1}^2 P\left(s_t=i|R,\Theta^{(k)}\right)=\sum_{t=1}^{T}(r_t-\mu)\sigma z_{t-1}P\left(s_t=i|R,\Theta^{(k)}\right).$$

Transposing the above equation, estimator of the correlation factor of the stock returns between time $t-1$ and time, $t$ $\theta_i$, can be obtained as

$$\hat{\theta}_i=\frac{\sum_{t=1}^{T}(r_t-\hat{\mu})z_{t-1}P\left(s_t=i|R,\Theta^{(k)}\right)}{\hat{\sigma}\sum_{t=1}^{T}z_{t-1}^2 P\left(s_t=i|R,\Theta^{(k)}\right)}. \tag{B.4}$$

Taking partial derivative with respect to the volatility of stock return, $\sigma$, in Eq. (B.2); setting it to 0, equation to estimate the volatility of stock return, $\sigma$, can be derived as

$$\frac{\partial}{\partial\sigma}\sum_{i=1}^{I}\left\{\sum_{t=1}^{T}\log\left[\frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}}\exp\left[-\frac{1}{2}\left(\sigma^2\right)^{-1}(r_t-(\mu+\theta_i\sigma z_{t-1}))^2\right]\right]\right\}P\left(s_t=i|R,\Theta^{(k)}\right)=0,$$

after differentiation, we have

$$\frac{1}{\sigma}\sum_{i=1}^{I}\sum_{t=1}^{T}\left[\sigma^{-1}(r_t-\mu-\theta_i\sigma z_{t-1})^2+\theta_i z_{t-1}(r_t-\mu-\theta_i\sigma z_{t-1})\right]P\left(s_t=i|R,\Theta^{(k)}\right)=\sum_{i=1}^{I}\sum_{t=1}^{T}P\left(s_t=i|R,\Theta^{(k)}\right).$$

Transposing the above equation, the estimator of the volatility of stock return $\sigma$ can be obtained as

$$\hat{\sigma}=\frac{\sum_{i=1}^{I}\sum_{t=1}^{T}\left[\hat{\sigma}^{-1}\left(r_t-\hat{\mu}-\hat{\theta}_i\hat{\sigma}z_{t-1}\right)^2+\hat{\theta}_i z_{t-1}\left(r_t-\hat{\mu}-\hat{\theta}_i\hat{\sigma}z_{t-1}\right)\right]P\left(s_t=i|R,\Theta^{(k)}\right)}{\sum_{i=1}^{I}\sum_{t=1}^{T}P\left(s_t=i|R,\Theta^{(k)}\right)}. \tag{B.5}$$

## References


Asa, K., Eikvil, L., & Huseby, R. B. (1999). Applications of hidden Markov chains in image analysis. *The Journal of Pattern Recognition Society*, *32*, 703–713.

Baum, L. E., & Eagon, J. A. (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, *73*, 360–363.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains? *Annals of the Institute of Statistical Mathematics*, *37*, 1554–1563.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of hidden Markov chains. *Annals of Mathematical Statistics*, *41*, 164–171.

Beveridge, S., & Oickle, C. (1997). Long memory in the Canadian stock market. *Applied Financial Economics*, *7*, 667–672.

Bilmes, J. A. (1998, April). *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.* International Computer Science Institute.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*, 637–654.

Blight, B. J. N. (1970). Estimation from a censored sample for the exponential family. *Biometrika*, *57*, 389–395.

Buck, S. F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society B*, *22*, 302–306.

Cassano, M. A. (1999). Learning and mean reversion in asset returns. *Quarterly Review of Economics and Finance*, *39*, 529–545.

Chelly-Steeley, P. (2001). Mean reversion in the horizon returns of U.K. portfolios. *Journal of Business Finance & Accounting*, *28*, 107–126.

Chen, T. T., & Finberg, S. E. (1974). Two-dimensional contingency tables with both completely cross-classified data. *Biometrics*, *30*, 629–642.

Chen, C. R., Su, Y., & Huang, Y. (2008). Hourly Index Return Autocorrelation and Conditional Volatility in an EAR-GJR-GARCH Model with Generalized Error Distribution. *Journal of Empirical Finance*, *15*, 789–798.

Churchill, G. A. (1989). Stochastic models for heterogeneous DNA sequences? *Bulletin of Mathematical Biology*, *51*, 79–94.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistic Society*, *39*(1), 1–38.

Efron, B. (1967). The two sample problem with censored data. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 4* (pp. 831–853). Berkely, California: University of California Press.

Elliott, R. J., Aggoun, L., & Moore, J. B. (1995). *Hidden Markov Models: Estimation and Control.* New York: Springer.

Fama, E. F., & French, K. R. (1988). Common Factors in the Serial Correlation of Stock Returns, Center for Research in Security Prices. working paper, 200 (University of Chicago, Chicago, IL).

Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, *14*, 174–194.

Jegadeesh, N. (1991). Seasonality in stock price mean reversion: evidence from the U.S. and the U.K.. *Journal of Finance*, *45*, 1427–1444.

Kim, C. J., Nelson, C. R., & Startz, R. (1998). Testing for mean reversion in heteroskedastic data based on Gibbs-sampling-augmented randomization. *Journal of Empirical Finance*, *5*, 131–154.

Korgh, A., Brown, M., Mian, I. S., Sjolander, K., & Haussler, D. (1994). Hidden Markov models in computational biology: applications to protein modeling? *Journal of Molecular Biology*, *235*, 1501−1513.
Liao, S. L., & Chen, C. C. (2006). The valuation of European options when asset returns are autocorrelated. *The Journal of Futures Markets*, *26*(1), 85−102.
Markov, A. A. (1906). Extension of the law of large numbers to dependent events (Russian). *Bulletin de la Société Physico-Mathémathique de Kasan*, *15*(2), 135−156.
Martin, D. L. (1975). Interest rates and occupational choice. *Journal of Human Resources*, *10*(4), 537−543.
McKendrick, A. G. (1926). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, *44*, 98−130.
Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, *8*, 343−366.
Orchard, T., & Woodbury, M. A. (1972). A missing information principle: theory and applications. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, 1* (pp. 697−715). Berkeley, California: University of California Press.
Potreba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics*, *22*, 27−59.
Wiener, N. (1923). Differential space. *Journal of Mathematical Physics*, *2*, 131−174.