Quantitative Bio-Science 33(1), 21~27(2014)

Empirical Bayesian Markov Chain Model for Random Model Change

Zhi-Ming Luo[1], Tae Yoon Kim[1,*], Beom Joon Kim[2]

[1]Department of Statistics, Keimyung University, Daegu 704-701, Korea
[2]Department of Electronic Engineering, Keimyung University, Daegu 704-701, Korea
(Received April 20, 2014; Revised May 20, 2014; Accepted May 24, 2014)

ABSTRACT

In this short paper we consider long term binary prediction of stock market by using empirical Bayesian Markov chain model. Our empirical Bayesian MC model is designed to accommodate random change of MC over time. Surprisingly enough, it is shown that empirical Bayesian MC model is homogeneous and has limiting distribution though it accommodates random model change over time. Empirical works are given to illustrate usefulness of our results.

Key words : Empirical Bayesian Markov chain, Random model change, Stock market

Introduction

Long-run stock market directivity or binary prediction of ups and downs is an interesting problem. It is particularly useful when one wants to decide to buy or sell a specific stock commodity for long-term investment. It is often observed that meaningful short-term market directivity prediction is hard. For instance, [1] employs various time series models for binary prediction of stock market and notices that none of them produces reasonably meaningful prediction. See also [2] and the references therein. Recently [3], [4] and others showed that limiting stationary probabilities of MC are straightforwardly estimated by the frequentist view. As with [1] above, however, these long term estimators does not provide meaningful forecasting results either. Recall that the frequentist estimators assume a fixed MC model over time. These observations strongly suggest that some other approach is necessary for meaningful short-term or long-term binary prediction of stock market. One possible alternative is Bayesian approach that allows model to change randomly with time. This article proposes new empirical Bayesian Markov chain model to accommodate such random model change. As a basic model we employ Markov chain (MC) with two discretized states (up and down).

It is well known that limiting transition probabilities of MC, if exist, provide useful information of long-term behavior of a given stochastic system. Thus theoretically the long-run directivity prediction might be well addressed by obtaining limiting transition probabilities of the correctly assumed MC chain. In order to take account of possible random change of MC model over time, we introduce random mechanism that randomly selects one MC from various MCs available at each transition. This system is formally defined as random compound MC which will provide theoretical underpinning to our empirical Bayesian approach. As one of our main results it is shown that random compound MC is homogeneous. This is surprising since it indicates that random compound MC which randomly selects one MC from various MCs at each transition is still homogeneous. As a matter of fact, the random compound MC enables us to employ Bayesian approach for MC. Our study consists of two steps. First we will show that the random compound MC proposed is homogeneous and has stationary limiting probabilities. Second we will establish empirical Bayesian MC model. We also discuss various Bayesian MC examples for which the random compound MC is useful. As an empirical application of our results, Section 3 forecasts long-run direc-

* Correspondence should be addressed to Dr. Tae Yoon Kim, Department of Statistics, Keimyung University, Daegu 704-701, Korea. Tel: +82-53-580-5533, Fax: +82-53-580-5164, E-mail: tykim@kmu.ac.kr

tivity of the Korean stock price index (KOSPI). In Section 4, the concluding remarks are given to discuss possible future works from our empirical Bayesian MC model.

Bayesian Markov Chain Approach

1. Simple Markov chain

Let Z_t be the closing SPI on t-th day. If $Z_t > Z_{t-1}$, we let $Y_t=1$; if $Z_t < Z_{t-1}$, we let $Y_t=0$; if there is no change or $Z_t=Z_{t-1}$, we will discard it. That is, we have a chain $\{Y_t \mid t=0, 1, 2, \cdots\}$ with two states $S=\{1, 0\}$. We need the following assumptions to model this chain as simple MC for SPI.

• The operation of SPI is affected by random factors such as the global or local economics, politics, society status, and so on.

• Up or down of the closing SPI in a given day just depends on the state of the previous closing SPI.

• The probability with which the closing SPI moves from one state to another state is constant and independent of time. That is, the chain is homogeneous with stationary transition probabilities.

• Let $\mathbf{u}=(1-p, p)$ denote initial state probability vector with $P\{Y_0=0\}=1-p$ and $P\{Y_0=1\}=p$. Let

$$\mathbf{P}=\begin{pmatrix} 1-p_{01} & p_{01} \\ 1-p_{11} & p_{11} \end{pmatrix}$$

denote the transition probability matrix with $p_{00}=1-p_{01}$ and $p_{10}=1-p_{11}$. Then it is assumed that $0<p_{01}<1$ and $0<p_{11}<1$ so that the MC is ergodic (that is, it is irreducible, positive recurrent and aperiodic).

The ergodicity condition (A4) implies that the MC achieves an equilibrium;

$$\lim_n P_{ij}^n=\pi_j=\sum_{i=0}^{1}\pi_i P_{ij},\ \sum_{i=0}^{1}\pi_i=1 \tag{1}$$

where $P_{ij}^n=P(Y_{n+m}=j \mid Y_m=i)$ and the limiting probabilities π's are uniquely determined by

$$\pi_0=\frac{1-p_{11}}{1-p_{11}+p_{01}} \quad \text{and} \quad \pi_1=1-\frac{p_{01}}{1-p_{11}+p_{01}}.$$

Refer to Theorem 1.3 of [5] for instance.

2. Random compound Markov chain

For specification of the random compound MC, suppose that we have a set of independent MCs indexed by $i \in \mathrm{I}$ where the index set I may be a finite, countable or uncountable set, i.e.,

$$\mathbf{Y}_i=\{Y_{is} : s=0, 1, 2, \cdots\} \quad \text{for} \quad i \in \mathrm{I}$$

where each Markov chain has the same state space and is homogeneous, aperiodic, positive recurrent and irreducible. Note that the each MC here is distinguished by its unique transition probability. Define

$$\mathbf{X}=\{X_s=Y_{r_s s} : s=0, 1, 2, \cdots\} \tag{2}$$

where $\{r_s : s=0, 1, 2, \cdots\}$ is a stochastic process with state space I and is employed as a sampling scheme for the MCs. If r_s with $s=0, 1, 2, \cdots$, are iid random variables on I, we call $\{X_s\}$ random compound Markov chain.

Theorem 1 *The random compound Markov chain is a homogeneous, aperiodic, positive recurrent and irreducible chain.*

Proof. Given two states j, k in the state space of the chain $\{X_s\}$. Suppose r_s has a stationary distribution P_r on I.

$$\begin{aligned} &P[X_s=k \mid X_{s-1}=j] \\ &\quad =\int_{\mathrm{I}} P[X_s=k, r_s=i \mid X_{s-1}=j]\, dP_r=\int_{\mathrm{I}} p_{ijk}\, dP_r \end{aligned}$$

where p_{ijk} is the transition probability from state j to k in the i-th chain $\{Y_i\}$.

An immediate consequence of Theorem 1 is that we have limiting probabilities for random compound Markov chain though Markov chain is randomly selected at each transition and hence transition probabilities are random. If one employs Markov chain for $\{r_s\}$, then we have the well-known Markov switching model. Refer to [6]. Note that nonhomogeneous Markov chain results in such case.

3. Empirical Bayesian Markov chain model

For modeling empirical Bayesian MC, suppose that we have random compound Markov chain $\{X_i\}$ which is homogeneous. This chain can be described by a probability model

$$P\{X_0=j_0, X_1=j_1, \cdots, X_s=j_s\}=P\{X_0=j_0\}\, p_{j_0 j_1} \cdots p_{j_{s-1} j_s}$$

Since each transition probability p_{jk} is random variable, we have

$$p_{jk} : p(p_{jk} \mid \theta_{jk})$$

where function $p(\cdot)$ is the pdf of p_{jk} and θ_{jk} is the hyperparameter. The hyperparameter θ_{jk} can be empirically estimated from data and then the parameter p_{jk} can be estimated from the posterior distribution.

4. Examples

Example 1. Suppose that the random compound Markov chain $\{X_i\}$ has the simplest state space $S=\{0, 1\}$. Then its probability model is

$$\begin{aligned}&P\{X_0=j_0, X_1=j_1, \cdots, X_s=j_s\}\\&\quad=P\{X_0=j_0\}\,p_{01}^{y_{01}}(1-p_{01})^{n_0-y_{01}}p_{11}^{y_{11}}(1-p_{11})^{n_1-y_{11}}\end{aligned}$$

where y_{01} is the number of jumps from 0 to 1, y_{11} is the number of jumps from 1 to 1, n_0 is the number of jumps from 0 and n_1 is the number of jumps from 1. Then y_{01} and y_{11} can be modeled as binomial random variables

$$y_{01}: binomial(n_0, p_{01}) \quad \text{and} \quad y_{11}: binomial(n_1, p_{11}).$$

For p_{01}(similar formula for p_{11}), the empirical Bayes model specifies the prior

$$p_{01}: beta(\alpha_{01}, \beta_{01}).$$

The posterior estimator of p_{01} is the posterior mean

$$\begin{aligned}\hat{p}_{01}&=\frac{\alpha_{01}+y_{01}}{\alpha_{01}+\beta_{01}+n_0}\\&=\left(\frac{\alpha_{01}+\beta_{01}}{\alpha_{01}+\beta_{01}+n_0}\right)\left(\frac{\alpha_{01}}{\alpha_{01}+\beta_{01}}\right)+\left(1-\frac{\alpha_{01}+\beta_{01}}{\alpha_{01}+\beta_{01}+n_0}\right)\frac{y_{01}}{n_0}, \quad (3)\end{aligned}$$

or the weighted average of sample mean and prior mean. We estimate the hyperparameters from data by assuming that the marginal distribution of y_{01} (unconditional on p_{01}) is the beta-binomial

$$p(y_{01}\,|\,\alpha_{01}, \beta_{01})=\binom{n_0}{y_{01}}\frac{B(y_{01}+\alpha_{01}, n_0-y_{01}+\beta_{01})}{B(\alpha_{01}, \beta_{01})}.$$

Indeed the first two population moments are

$$\mu_1=\frac{n_0\alpha_{01}}{\alpha_{01}+\beta_{01}}$$

$$\mu_2=\frac{n_0\alpha_{01}[n_0(1+\alpha_{01})+\beta_{01}]}{(\alpha_{01}+\beta_{01})(1+\alpha_{01}+\beta_{01})}.$$

Letting the first two sample moments be $\hat{\mu}_1=m_1$ and $\hat{\mu}_2=m_2$, by the method of moments we have estimates

$$\hat{\alpha}_{01}=\frac{n_0m_1-m_2}{n_0(m_2/m_1-m_1-1)+m_1}$$

$$\hat{\beta}_{01}=\frac{(n_0-m_1)(n_0-m_2/m_1)}{n_0(m_2/m_1-m_1-1)+m_1}. \quad (4)$$

Example 2. Suppose that we have T independent random compound Markov chain $\mathbf{X}_t=\{X_{it}: i=1, \cdots, n_t\}$ for $t=1, 2, \cdots, T$. We can model these chains as

$$\begin{aligned}&y_{01t}: binomial(n_{0t}, p_{01t}),\ y_{11t}: binomial(n_{1t}, p_{11t})\\&i.i.d\ \ for \quad t=1, \cdots, T.\end{aligned}$$

and

$$\begin{aligned}&p_{01t}: beta(\alpha_{01}, \beta_{01}),\ p_{11t}: beta(\alpha_{11}, \beta_{11})\\&i.i.d\ \ for \quad t=1, \cdots, T.\end{aligned}$$

This example is established and stipulated on a set of independent samples or segments from the random component MC model where each t-th independent segment is assumed to have random transition probabilities whose distributions are identical for a given prior. Fitting the data (y_{01t}, n_{0t}), (y_{11t}, n_{1t}) for $t=1, \cdots, T$ to beta-binomial model and utilizing the numerical maximum likelihood estimate scheme, we can estimate the values of $\hat{\alpha}_{01}$, $\hat{\beta}_{01}$, $\hat{\alpha}_{11}$ and $\hat{\beta}_{11}$.

Example 3. When modelling regular Bayesian MC without empirical effort, $\mathbf{X}_t=\{X_{it}: i=1, \cdots, n_t\}$ has three parameters: p_t, the initial probability at state "1", p_{01t} and p_{11t}. Then letting $n_{0t}+n_{1t}=n_t$

$$\begin{aligned}&P\{X_{0t}=i_{0t}, X_{1t}=i_{1t}, \cdots, X_{n_t t}=i_{n_t t}\}\\&\quad=P\{X_{0t}=i_{0t}\}\,p_{i_{0t}i_{1t}}p_{i_{1t}i_{2t}}\cdots p_{i_{n_t-1,t}i_{n_t t}}\\&\quad=(p_t)^{i_{0t}}(1-p_t)^{1-i_{0t}}(p_{01t})^{y_{01t}}(1-p_{01t})^{n_{0t}-y_{01t}}\\&\quad\ \ (p_{11t})^{y_{11t}}(1-p_{11t})^{n_{1t}-y_{11t}}, \qquad (5)\end{aligned}$$

where i_{0t}, i_{1t}, $\cdots$, $i_{n_t t}$ are either "0" or "1". For mathematical convenience, we treat all initial probabilities to be same, i.e., $p_t=p$. Here notice that $X_{0t}: Bernoulli(p)$, $Y_{01t}: binomial(n_{0t}, p_{01t})$ and $Y_{11t}: binomial(n_{1t}, p_{11t})$ where $Bernoulli(p)$ denotes Bernoulli distribution with probability of success p and *binomial*(n, p) binomial distribution with n trials and probability of success p. Based on (??) we choose the conjugate prior distributions for the parameters p, p_{01t}, p_{11t}, i.e., the parameters p, p_{01t}, p_{11t} are assumed to have independent Beta distributions

$$p: Beta(\alpha, \beta),\ p_{01t}: Beta(\alpha_{01}, \beta_{01}),\ p_{11t}: Beta(\alpha_{11}, \beta_{11}).$$

Furthermore we assign a noninformative hyperprior distribution due to our ignorance of the unknown hyperparameters by assuming independent hyperparameter pairs (α, β), $(\alpha_{01}, \beta_{01})$, $(\alpha_{11}, \beta_{11})$. Since, as mentioned in [7] [pp. 128], a uniform prior density on (α, β) yields an improper posterior density, one reasonable alternative for hyperprior density is uniform on the transformed parameters $\left(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}\right)$, i.e.,

$$(\alpha, \beta): prior(\alpha, \beta) \propto (\alpha+\beta)^{-5/2},$$

$$(\alpha_{01}, \beta_{01}): prior(\alpha_{01}, \beta_{01}) \propto (\alpha_{01}+\beta_{01})^{-5/2},$$

$$(\alpha_{11}, \beta_{11}): prior(\alpha_{11}, \beta_{11}) \propto (\alpha_{11}+\beta_{11})^{-5/2},$$

Let $\mathbf{X}=(\mathbf{X}_1, \cdots, \mathbf{X}_T)$ be T segments of observations from random compound MC where $\mathbf{X}_t=(i_{0t}, \cdots, i_{n_t t})$ and $i_{0t}, \cdots, i_{n_t t}$ are either 0 or 1. Let $\mathbf{p_{01}}=(p_{011}, \cdots, p_{01T})$ and $\mathbf{p_{11}}=(p_{111}, \cdots, p_{11T})$ be T dimensional random vectors. Then the joint posterior distribution has an analytic form

$$\begin{aligned}
&p(p, \mathbf{p_{01}}, \mathbf{p_{11}}, \alpha, \beta, \alpha_{01}, \beta_{01}, \alpha_{11}, \beta_{11} | \mathbf{X}) \\
&\propto prior(\alpha, \beta) prior(\alpha_{01}, \beta_{01}) prior(\alpha_{11}, \beta_{11}) \\
&\quad p(p | \alpha, \beta)\ p(\mathbf{p_{01}} | \alpha_{01}, \beta_{01})\ p(\mathbf{p_{11}} | \alpha_{11}, \beta_{11}) \\
&\quad \times p(\mathbf{X} | p, \mathbf{p_{01}}, \mathbf{p_{11}}, \alpha, \beta, \alpha_{01}, \beta_{01}, \alpha_{11}, \beta_{11})
\end{aligned}$$

$$\begin{aligned}
&\propto prior(\alpha, \beta) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (p)^{\alpha-1}(1-p)^{\beta-1}(p)^{n_0}(1-p)^{T-n_0} \\
&\quad \times prior(\alpha_{01}, \beta_{01}) \prod_{t=1}^{T} \frac{\Gamma(\alpha_{01}+\beta_{01})}{\Gamma(\alpha_{01})\Gamma(\beta_{01})} (p_{01t})^{\alpha_{01}-1}(1-p_{01t})^{\beta_{01}-1} \\
&\quad \prod_{t=1}^{T} (p_{01t})^{y_{01t}}(1-p_{01t})^{n_{0t}-y_{01t}} \\
&\quad \times prior(\alpha_{11}, \beta_{11}) \prod_{t=1}^{T} \frac{\Gamma(\alpha_{11}+\beta_{11})}{\Gamma(\alpha_{11})\Gamma(\beta_{11})} (p_{11t})^{\alpha_{11}-1}(1-p_{11t})^{\beta_{11}-1} \\
&\quad \prod_{t=1}^{T} (p_{11t})^{y_{11t}}(1-p_{11t})^{n_{1t}-y_{11t}},
\end{aligned}$$

where $n_0=\sum_{t=1}^{T} i_{0t}$, i.e., the number of initial state="1"s in the T chains. Given (α, β), $(\alpha_{01}, \beta_{01})$ and $(\alpha_{11}, \beta_{11})$, the components of p, $\mathbf{p_{01}}$, $\mathbf{p_{11}}$ have independent conditional posterior densities as Beta densities

$$\begin{aligned}
&p(p | \alpha, \beta, \mathbf{X}) \\
&\quad = \frac{\Gamma(\alpha+\beta+T)}{\Gamma(\alpha+n_0)\Gamma(\beta+T-n_0)} (p)^{\alpha+n_0-1}(1-p)^{\beta+T-n_0-1},
\end{aligned}$$

$$\begin{aligned}
&p(\mathbf{p_{01}} | \alpha_{01}, \beta_{01}, \mathbf{X}) \\
&\quad = \prod_{t=1}^{T} \frac{\Gamma(\alpha_{01}+\beta_{01}+n_{0t})}{\Gamma(\alpha_{01}+y_{01t})\Gamma(\beta_{01}+n_{0t}-y_{01t})} \\
&\quad\quad (p_{01t})^{\alpha_{01}+y_{01t}-1}(1-p_{01t})^{\beta_{01}+n_{0t}-y_{01t}-1},
\end{aligned}$$

$$\begin{aligned}
&p(\mathbf{p_{11}} | \alpha_{11}, \beta_{11}, \mathbf{X}) \\
&\quad = \prod_{t=1}^{T} \frac{\Gamma(\alpha_{11}+\beta_{11}+n_{1t})}{\Gamma(\alpha_{11}+y_{11t})\Gamma(\beta_{11}+n_{1t}-y_{11t})} \\
&\quad\quad (p_{11t})^{\alpha_{11}+y_{11t}-1}(1-p_{11t})^{\beta_{11}+n_{1t}-y_{11t}-1},
\end{aligned}$$

respectively. Dividing the joint posterior density by the above conditional posterior densities, we have the marginal posterior densities of (α, β), $(\alpha_{01}, \beta_{01})$, and $(\alpha_{11}, \beta_{11})$, respectively:

$$\begin{aligned}
&p(\alpha, \beta | \mathbf{X}) \\
&\quad \propto prior(\alpha, \beta) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+n_0)\Gamma(\beta+T-n_0)}{\Gamma(\alpha+\beta+T)},
\end{aligned}$$

$$\begin{aligned}
&p(\alpha_{01}, \beta_{01} | \mathbf{X}) \\
&\quad \propto prior(\alpha_{01}, \beta_{01}) \\
&\quad\quad \prod_{t=1}^{T} \frac{\Gamma(\alpha_{01}+\beta_{01})}{\Gamma(\alpha_{01})\Gamma(\beta_{01})} \frac{\Gamma(\alpha_{01}+y_{01t})\Gamma(\beta_{01}+n_{0t}-y_{01t})}{\Gamma(\alpha_{01}+\beta_{01}+n_{0t})},
\end{aligned}$$

$$\begin{aligned}
&p(\alpha_{11}, \beta_{11} | \mathbf{X}) \\
&\quad \propto prior(\alpha_{11}, \beta_{11}) \\
&\quad\quad \prod_{t=1}^{T} \frac{\Gamma(\alpha_{11}+\beta_{11})}{\Gamma(\alpha_{11})\Gamma(\beta_{11})} \frac{\Gamma(\alpha_{11}+y_{11t})\Gamma(\beta_{11}+n_{1t}-y_{11t})}{\Gamma(\alpha_{11}+\beta_{11}+n_{1t})}. \qquad (6)
\end{aligned}$$

These functions cannot be simplified further analytically but are easy to compute by numerical method. Denote the means of these marginal distributions by $(\hat{\alpha}, \hat{\beta})$, $(\hat{\alpha}_{01}, \hat{\beta}_{01})$, and $(\hat{\alpha}_{11}, \hat{\beta}_{11})$, respectively. Then we may deduce

$$p: Beta(\hat{\alpha}, \hat{\beta}),\ p_{01}: Beta(\hat{\alpha}_{01}, \hat{\beta}_{01}),\ p_{11}: Beta(\hat{\alpha}_{11}, \hat{\beta}_{11}).$$

Using the modes of these densities as the estimator of parameters p, p_{01}, and p_{11}, we have

$$\hat{p}=\frac{\hat{\alpha}-1}{\hat{\alpha}+\hat{\beta}-2},\ \hat{p}_{01}=\frac{\hat{\alpha}_{01}-1}{\hat{\alpha}_{01}+\hat{\beta}_{01}-2},\ \hat{p}_{11}=\frac{\hat{\alpha}_{11}-1}{\hat{\alpha}_{11}+\hat{\beta}_{11}-2}.$$

5. Remarks

Remark 1. Regarding making inferences about transitional probabilities of MC, early efforts involve [8] where the maximum likelihood methods are used for testing homogeneous MC and [9] where a Bayesian analysis of homogeneous MC is presented using data in the form of aggregated proportions. Then an empirical Bayesian approach is introduced by [10] where extensions to non-homogeneous MC is made by viewing the problem as a parametric empirical Bayes problem in terms of [11]. Later works are proposed based on hierarchial Bayesian approach that considered the effects of covariates on transition probabilities. See, for instance, [12] and [13]. These approaches, however, have not considered estimation of limiting probabilities in their inferences.

Remark 2. In the Bayesian literature, the term Markov model used to refer to two different classes, i.e., parameter-driven and observation-driven Markov models in the sense of [14]. Both classes of models are used for analysis of categorical time series data. The observation-driven Markov models are the MCs where the Markov structure is on the observables such as the state occupancies of the individuals while the parameter-driven Markov models are the MCs where the parameters evolve over time according to MC (see, e.g., [15]). The latter is often termed as dynamic Bayesian models. Our Bayesian MC model is observation-driven Markov model as well as dynamic model since it employs the random selection mechanism, not parameters, via Bayesian approach. Note that the random selection of MC at each transition certainly renders dynamic character to the our Bayesian MC. This remark shows that our Bayesian model is more evolved than other Bayesian MC models.

Remark 3. Finally it is worth mentioning that our empirical Bayesian MC model used in Example 2 is different from the usual empirical Bayesian MC (see, e.g., [10]). In fact, the usual empirical Bayesian MC employs T sets of independent samples as T realizations of the Bayesian Morkov chain from a given prior and hence accommodates the T different underlying MCs (see, e.g., [10]). Our empirical Bayesian MC model employs T independent samples from the unique one random compound MC model (Examples 2 and 3).

Empirical Analysis of Stock Market Forecast

In this section we employ our empirical Bayesian MC to predict long run behavior of Korean stock market where the referenced data are 2009 KOSPI (Korean Stock Price Index). This approach is sensible particularly when one plans to long-term investment in KOSPI at the early stage of 2010. Using the sampling scheme of example 2, we choose KOSPI data from January, March, May, July, September and November in 2009 to have segmented data from random component MC. There is one month separation made between any two consecutive months, which makes the T=6 Markov chain samples to be independent. Recall independence among MC samples or segments is needed for sampling scheme of Example 2. All data obtained from these 6 segments are summarized in Table 1.

Table 1. Summary of 6 segments

	January	March	May	July	September	November
t	1	2	3	4	5	6
n_t	18	22	18	22	21	20
n_{0t}	7	8	9	6	10	10
n_{1t}	11	13	9	16	11	10
y_{01t}	4	5	5	3	6	6
y_{11t}	6	9	4	13	4	5
i_{0t}	1	0	1	1	1	0

Using data from the 6 months, we have (refer to examples 1 and 2 above)

$$\hat{\alpha}_{01}=2.46\times 10^8,\ \hat{\beta}_{01}=1.78\times 10^8,\ \hat{\alpha}_{11}=19.71,\ \hat{\beta}_{11}=14.35.$$

Now the transition probabilities of the month November (refer to Examples 1 and 2 above)

$$\hat{p}_{016}=0.58\quad \hat{p}_{016}=0.56.$$

That is, the estimated transition matrix of November is

$$\hat{\mathbf{P}}=\begin{pmatrix} 1-\hat{p}_{016} & \hat{p}_{016} \\ 1-\hat{p}_{116} & \hat{p}_{116} \end{pmatrix}=\begin{pmatrix} 0.42 & 0.58 \\ 0.44 & 0.56 \end{pmatrix}$$

with stationary probability vector as

$$\begin{pmatrix} \hat{p}_{0,\infty} \\ \hat{p}_{1,\infty} \end{pmatrix}=\begin{pmatrix} 0.431 \\ 0.569 \end{pmatrix}. \tag{7}$$

When one uses regular Baysian MC model (Example 3), we proceed to as follows. For the initial states, from Table 1 we know that there are $n_0=\sum_{t=1}^{6} i_{0t}=4$ ones (or 1s). Plugging $n_0=4$ and $T=6$ values in the expression of (??), we obtain an unnormalized posterior marginal density for (α, β). To estimate the mean of $p(\alpha, \beta\mid\mathbf{X})$, we discretize the continuous unnormalized density obtained via (??) over a grid. By using the contour and perspective plots, the density is found to be concentrated mostly in the range $(\alpha, \beta)\in[0.01, 1]\times[0.01, 1]$ and hence 100×100 grid is employed with its increment 0.01 on each coordinate. Computing the relative posterior marginal density on the grid, we normalize it by approximating the density as a step function over the grid and setting the total sum of probabilities over the grid to 1. Hence we can compute the posterior means based on the grid of (α, β) as

$$\hat{\alpha}=E(\alpha\mid\mathbf{y})\approx\sum_{(\alpha,\beta)}\alpha p(\alpha, \beta\mid\mathbf{y})\approx 0.243,$$

$$\hat{\beta}=E(\beta\mid\mathbf{y})\approx\sum_{(\alpha,\beta)}\beta p(\alpha, \beta\mid\mathbf{y})\approx 0.224.$$

Therefore, by Example 3, the estimate of parameter p is

$$\hat{p}=\frac{\hat{\alpha}-1}{\hat{\alpha}+\hat{\beta}-2}=\frac{0.243-1}{0.243+0.224-2}\approx 0.494.$$

Similarly, after plugging the values of Table 1 in the expressions of (??) and investigating their contour and perspective plots, we apply 400×400 grid on $[0.01, 4]\times[0.01, 4]$ with increment 0.01 on each coordinate for calculating densities $p(\alpha_{01}, \beta_{01}\mid\mathbf{X})$ and $p(\alpha_{11}, \beta_{11}\mid\mathbf{x})$. Furthermore, we calculate the posterior means based on the grids of $(\alpha_{01}, \beta_{01})$ and $(\alpha_{11}, \beta_{11})$, respectively, and as a result we have

$$\hat{\alpha}_{01}=E(\alpha_{01}|\mathbf{y})\approx\sum_{(\alpha_{01},\beta_{01})}\alpha_{01}p(\alpha_{01},\beta_{01}|\mathbf{y})\approx 2.455,$$

$$\hat{\beta}_{01}=E(\beta_{01}|\mathbf{y})\approx\sum_{(\alpha_{01},\beta_{01})}\beta_{01}p(\alpha_{01},\beta_{01}|\mathbf{y})\approx 2.066,$$

$$\hat{\alpha}_{11}=E(\alpha_{11}|\mathbf{y})\approx\sum_{(\alpha_{11},\beta_{11})}\alpha_{11}p(\alpha_{11},\beta_{11}|\mathbf{y})\approx 2.257,$$

$$\hat{\beta}_{11}=E(\beta_{11}|\mathbf{y})\approx\sum_{(\alpha_{11},\beta_{11})}\beta_{11}p(\alpha_{11},\beta_{11}|\mathbf{y})\approx 1.901.$$

Therefore, by Example 3 the estimates of parameters p_{01}^* and p_{11}^* are

$$\hat{p}_{01}=\frac{\hat{\alpha}_{01}-1}{\hat{\alpha}_{01}+\hat{\beta}_{01}-2}=\frac{2.455-1}{2.455+2.066-2}\approx 0.577,$$

$$\hat{p}_{11}=\frac{\hat{\alpha}_{11}-1}{\hat{\alpha}_{11}+\hat{\beta}_{11}-2}=\frac{2.257-1}{2.257+1.901-2}\approx 0.582,$$

respectively. That is, the estimated transition matrix is

$$\hat{\mathbf{P}}=\begin{pmatrix}1-\hat{p}_{01} & \hat{p}_{01}\\ 1-\hat{p}_{11} & \hat{p}_{11}\end{pmatrix}=\begin{pmatrix}0.423 & 0.577\\ 0.418 & 0.582\end{pmatrix}$$

with stationary probability vector as

$$\begin{pmatrix}\hat{p}_{0,\infty}\\ \hat{p}_{1,\infty}\end{pmatrix}=\begin{pmatrix}0.420\\ 0.580\end{pmatrix}. \tag{8}$$

With this long-run forecasting probabilities at hand, one might decide to make long term investment into KOSPI at the early stage of 2010 since $\hat{p}_{0,\infty}<\hat{p}_{1,\infty}$. For empirical evaluation of our methodology against 2010, we calculate the state probabilities from 2010 data i.e., there are total 251 observations with 139 “up”s and 112 “down”s. Using these, the state probabilities can be simply estimated as

$$\begin{pmatrix}\hat{p}_{0,2010}\\ \hat{p}_{1,2010}\end{pmatrix}=\begin{pmatrix}\frac{112}{251}\\ \frac{139}{251}\end{pmatrix}=\begin{pmatrix}0.446\\ 0.554\end{pmatrix}. \tag{9}$$

It is interesting to see that the estimated stationary probabilities from 2009 via our methodology are close to the real state probabilities for 2010.

Another interesting episode noticeable from (7)-(9) is related to random walk hypothesis of stock market. It is easy to see that pure random walk model produces

$$\begin{pmatrix}p_{0,\infty}\\ p_{1,\infty}\end{pmatrix}=\begin{pmatrix}0.5\\ 0.5\end{pmatrix} \tag{10}$$

and that the estimated limiting probabilities in (7) are closer to (10) than (8). This indicates that empirical Bayesian approach utilizing more data from stock market produces limiting probabilities closer to random walk hypothesis.

Conclusion

In this paper we propose Bayesian Markov chain model to accommodate possible random model change. As an application we provide long-term prediction of stock market. Empirical study reports a promising result. Of course our results may be extended to MC model application in other areas. Also it would be another interesting issue if one extends our model to hierarchical regression model for estimating limiting probabilities.

Acknowledgements

TY Kim’s work is supported by the Basic Science Research Program through the National Research Foundation of Korea (KRF-2011-0015936).

References

1. Choi MS, Baek JS, Hwang SY. An analysis of categorical time series driven by clipping GARCH processes. Korean J of Appl Stat 2010;23:683-692.
2. Zhen X, Basawa IV. Estimation for binary models generated by Gaussian autoregressive processes. J Stat Comput Sim 2010; 80:1041-1051.
3. Hao F. The applications of Markov prediction method in stock market. Friends Sci 2006;6:78-81.
4. Zhang D, Zhang X. Study on forecasting the stock market trend based on stochastic analysis method. Int J Bus and Manage 2009;4:163-170.
5. Karlin S, Taylor HM. A First course in stochastic processes. 2nd ed. New York: Academic Press; 1975.
6. Chib S. Calculating posterior distributions and modal estimates in Markov mixture models. J Econometrics 1996;75:79-97.
7. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed. London: Chapman Hall; 2004.
8. Anderson TW, Goodman LA. Statsitical inference about Markov Chain. Ann Math Stat 1957;28:89-110.
9. Lee TC, Judge GG, Zeller N. Maximum likelihood and Bayesian estimation of transition probabilities. J Am Stat Assoc

1968;63:1162-1179.
10. Meshkani MR, Billard L. Empirical Bayes estimators for a finite Markov chain. Biometrika 1992;79:185-193.
11. Morris CN. Parametric empirical Bayes inference: theory and applications. J Am Stat Assoc 1983;78:47-55.
12. Muenz I, Rubinstein I. Markov models for covariate dependence of binary sequence. Biometrics 1985;78:47-55.
13. Sung M, Soyer R, Nahn M. Bayesian analysis of non-homogeneous Markov chain: applications to mental health data. Stat Med 2007;26:3000-3017.
14. Cox DR. Statistical analysis of time series: some recent developments. Scand J Stat 1981;8:93-115.
15. Cargoni C, Mueller P, West M. Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. J Am Stat Assoc 1997;92:640-647.