[cs229 Project]

Stock Forecasting using Hidden Markov Processes

Joohyung Lee, Minyong Shin

1. Introduction

In finance and economics, time series is usually modeled as a geometric Brownian motion with drift. Especially, in financial engineering field, the stock model, which is also modeled as geometric Brownian motion, is widely used for modeling derivatives. In this model, the stock return is modeled as a Brownian motion with draft.

$$\frac{S_{t+1}-S_t}{S_t} = R_t = \mu\Delta t + \sigma Z_{\Delta t} \sim \mathrm{N}(\mu\Delta t, \sigma^2\Delta t)$$

Here the coefficient of the drift, μ and volatility, σ are constant here. This model reflects the distribution of the actual data indeed. As we can see the Figure1 below, the overall data distribution follows.

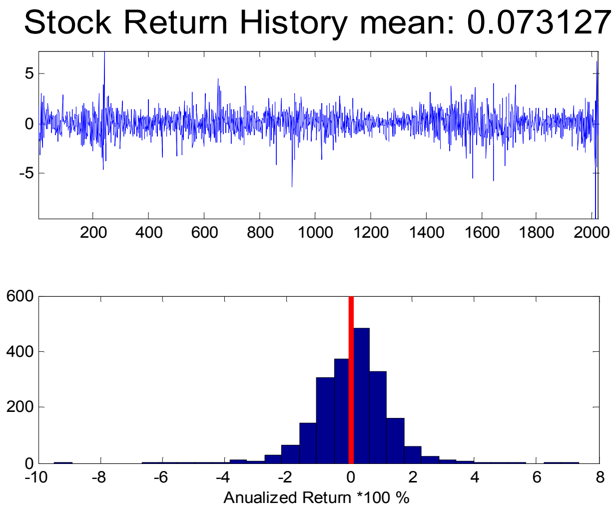


Figure1. Weekly stock return(S&P500) from 16 Mar 1970.

However, if we are in Bear market which is the times such as Internet Bubble or recent economic recession or Bull market in which we can observe obvious economic growth, it is difficult to discern we are in which situation. Also, in those times, the mean and variance of the stock will be totally different. The following Figure 2 shows this.

We define these region of time as a regime whose mean and variance are explicitly different from other region of time. This regime can represent economic situation. If we can figure out this regime, in other words, current economic situation, we can forecast better than using constant mean and variance. In this project, we would like to construct this regime and utilize it for the stock forecasting using one of the machine learning algorithms.

For this, we model the stock series has Gaussian distribution in a regime and each regime is modeled Hidden Markov Model (HMM) to obtain the discrete economic situations. In this model, each economics states is changing by a transition matrix which we need to estimates. Thus, we extract economic situations from the stock data itself.

As an extended model, we used another economic data for estimating economic change. This data can be the market information data such as Volatility Index or Macro-economic data such as GDP. In this project, we used default data for this.

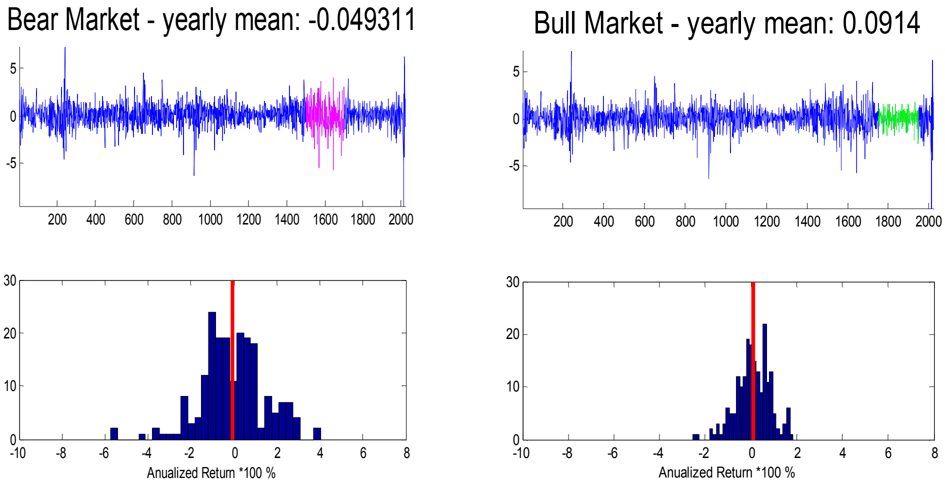


Figure 2. Stock return (S&P500) in Bear market and Bull market

In Figure 3, we can observe the obvious increase of number of defaults in the Bear market and the small and stable number of defaults in Bull market. Using this data, we discretize the economic situation and vary the transition for mean and variance states of stock.

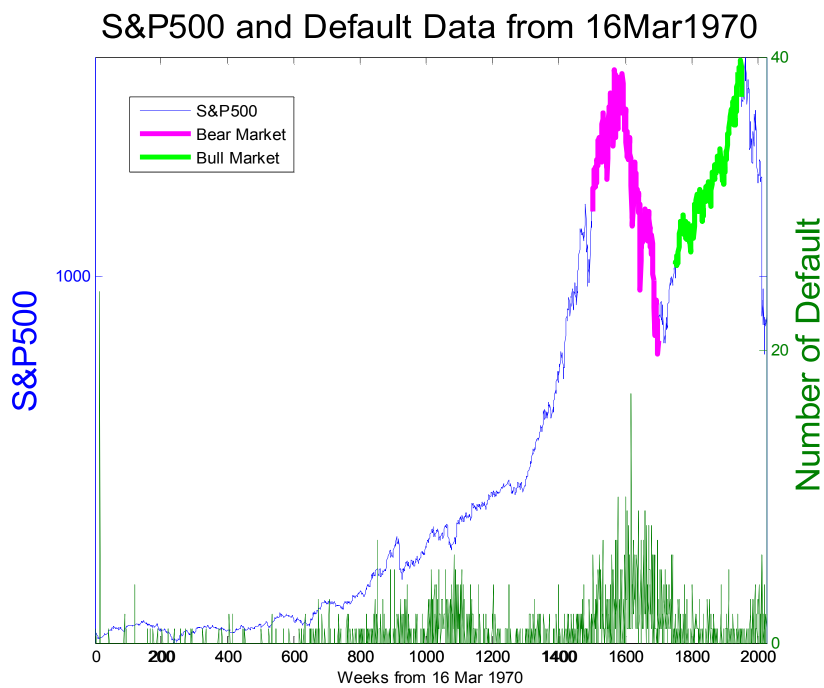


Figure3. Stock Index with Default data : red line represents Bear market, in which the increase of the default number is observed and light green line represents Bull market, in which low and stable number of default is observed.

2. Modeling and Method

2.1 Modeling

2.1.1 Single HMM

As we briefly introduced, we model the stock return series as a mixture of Gaussian distribution and discrete Markov chain. In a certain regime, stock series follow geometric Brownian motion with drift, which means stock returns have Gaussian distribution and the regimes are changing by the discrete Markov Process.

For example, while, in good economic situation, stock return has higher mean and smaller variance, it has lower mean and larger variance in bad economic situation. In this case, we have two economic states, i.e. good and bad, and each state has different mean and variance. The regime changing occurs randomly in this model. However, the changing probability is fixed and consistent thought the stock return series. We use a transition matrix for this.

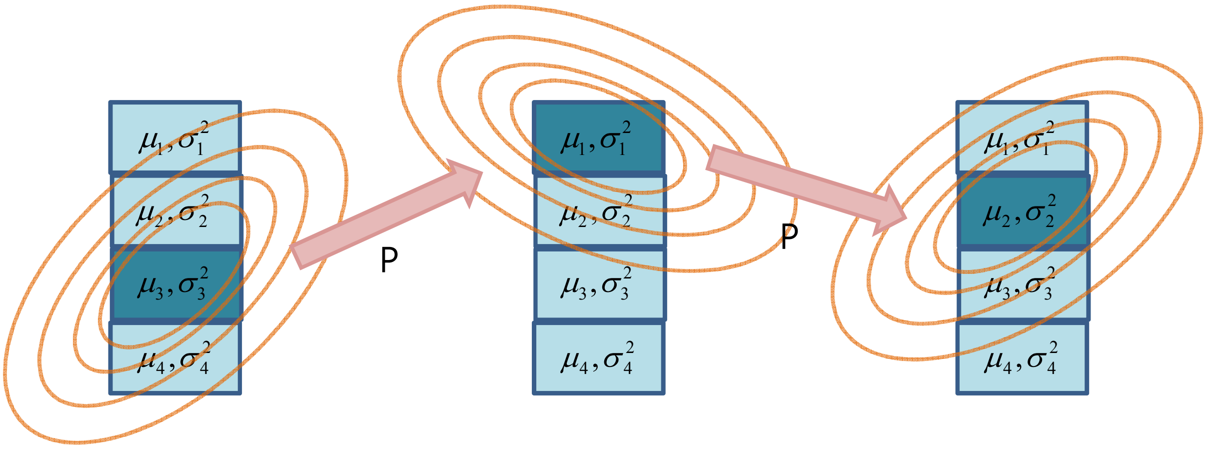


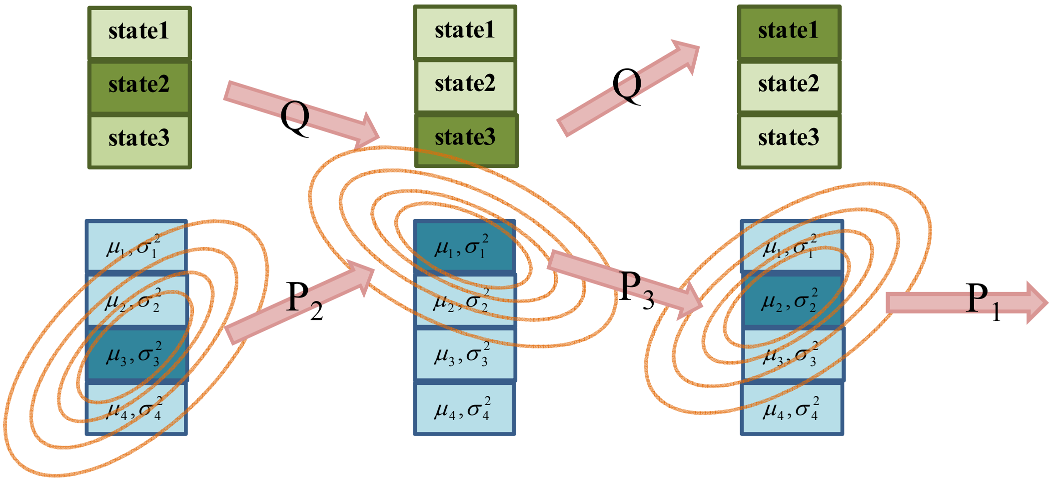
Figure 4. Single Hidden Markov Model: the discrete Markov chain represent economic situation and each economic situation had different mean and variance of stock return.

2.1.2 Double HMM

Single HMM identifies the economic situation and the mean and variance states of stock return. This implies any change of economic situation reflected on the stocks mean and variance simultaneously. This is reasonable considering fast and very responsive market but with this model there is no way to improve prediction. Thus, for the purpose of predictability improvement, we need to separate the economic situation from the mean and variance states so that the mean and variance can follow the signal from the economic situation. By modeling like this, we expect any economic situation changes prior to the mean and variance state change.

Thus, in double HMM, we modeled the default event as mixture of Poisson distribution with discrete Markov Chain and use this model to indicate the change of economics. We assume in every discrete economic situation, it has different Poisson intensity and the number of defaults is generated by the default intensity. We can obtain the economic situation without the stock return data.

In each economic situation, it has different transition matrix for the mean and variance states. This allows us to vary the probability from one state to another by the economic situations, which gives us more degree of freedom to fit the data and, once the parameters are learned, its predictability would be better.



2.2 Method : EM algorithm

Since we modeled stock return and the number of defaults as the mixture of Gaussian and discrete Markov Chain and Poisson and discrete Markov Chain respectively, the natural choice of method should Expectation and Maximization algorithm. In this chapter, we present the detail of derivation of the two models.

Definition : X_t – State of Mean and variance model at time t.

Y_t – State of Default intensity model at time t.

R_t – The stock return at time t.

Z_t – Number of Defaults at time t.

2.2.1 Single HMM

We first specify the parameters to determine.

$$\pi_i = P(X_0 = i)$$
$$p_{ij} = P(X_t = j \mid X_{t-1} = i)$$
$$N(\mu_{X_t}, \sigma^2_{X_t}) = P(R_t = r_t \mid X_t = x_t; \theta) = f_{X_t}(r_t)$$
$$\theta := \{\pi_i, p_{ij}, \mu_i, \sigma_i\}$$

We would like to find the parameter set, that maximizing posterior Likelihood as below.

$$\max_\theta P\{R_0 = r_0, \ldots, R_N = r_N\}$$

The Complete Likelihood function is defined as below.

$$L_c(\theta) = P\{R_0 = r_0, \ldots, R_N = r_N, X_0 = x_0, \ldots, X_N = x_N; \theta\}$$
$$= P(X_0 = x_0;\theta) P(R_0 = r_0 \mid X_0 = x_0;\theta) \ldots P(X_N \mid X_{N-1};\theta) P(R_N = r_N \mid X_N = x_N;\theta)$$
$$= P(X_0 = x_0;\theta) P(R_0 = r_0 \mid X_0 = x_0;\theta) \prod_{t=1}^{N} P(X_t \mid X_{t-1};\theta) P(R_t = r_t \mid X_t = x_t;\theta)$$

$$\log L_c(\theta) = \log \pi_{x_0} + \sum_{t=1}^{N} \log p_{x_t x_{t-1}} + \sum_{t=0}^{N} \left(-\frac{1}{2} \log(2\pi\sigma^2_{X_t}) - \frac{(r_t - \mu_{X_t})^2}{2\sigma^2_{X_t}} \right)$$

The Expectation of log-likelihood function is optimized at Maximization step.

$$E[\log L_c(\theta) \mid R_1 ... R_N ; \widetilde{\theta}] = \sum_i P(X_0 = i \mid R_0 ... R_N ; \widetilde{\theta}) \log \pi_i$$

$$+ \sum_{i,j} \sum_{t=1}^{N} P(X_t = j, X_{t-1} = i \mid R_0 ... R_N ; \widetilde{\theta}) \log p_{ij}$$

$$+ \sum_{i} \sum_{t=0}^{N} P(X_t = i \mid R_1 ... R_N ; \widetilde{\theta}) \left(-\frac{1}{2} \log(2\pi\sigma_i^2) - \frac{(r_t - \mu_i)^2}{2\sigma_i^2} \right)$$

- **Expectation Step**

We start with introducing the forward and backward probabilities, alpha and beta respectively.

$$\alpha_t(j) = P(X_t = j; R_0 = r_0, ... R_t = r_t)$$
$$\beta_t(j) = P(R_{t+1} = r_{t+1}, ... R_N = r_N \mid X_t = j)$$

Their values can be obtained recursively.

$$\alpha_{t+1}(j) = P(X_{t+1} = j; R_0 = r_0, ... R_{t+1} = r_{t+1}) = \sum_i P(X_{t+1} = j, X_t = i; R_0 = r_0, ... R_{t+1} = r_{t+1})$$

$$= \sum_i P(R_{t+1} = r_{t+1} \mid X_{t+1} = j, X_t = i, R_0 = r_0, ... R_t = r_t) P(X_{t+1} = j \mid X_t = i, R_0 = r_0, ... R_t = r_t)$$
$$P(X_t = i; R_0 = r_0, ... R_t = r_t)$$

$$= \sum_i p_{ij} f_j(r_{t+1}) \alpha_t(i)$$

$$\beta_t(j) = P(R_{t+1} = r_{t+1}, ... R_N = r_N \mid X_t = j)$$

$$= \sum_k P(X_{t+1} = k, R_{t+1} = r_{t+1}, ... R_N = r_N \mid X_t = j)$$

$$= \sum_k P(R_{t+2} = r_{t+2}, ... R_N = r_N \mid X_{t+1} = k) P(R_{t+1} = r_{t+1} \mid X_{t+1} = k) P(X_{t+1} = k \mid X_t = j)$$

$$= \sum_k \beta_{t+1}(k) p_{jk} f_k(r_{t+1})$$

Then, $\alpha_t(j)\beta_t(j) = P(X_t = j, R_0 = r_0, ... R_N = r_N)$

$$\therefore d_t(j) = P(X_t = j \mid R_0 = r_0, ... R_N = r_N) = \frac{\alpha_t(j)\beta_t(j)}{\sum_j \alpha_t(j)\beta_t(j)}$$

$$\alpha_t(i) p_{ij} f_j(r_{t+1}) \beta_{t+1}(j) = P(X_t = i; R_0 = r_0, ... R_t = r_t) P(X_{t+1} = j \mid X_t = i) P(R_{t+1} = r_{t+1} \mid X_{t+1} = j)$$
$$P(R_{t+2} = r_{t+2}, ... R_N = r_N \mid X_{t+1} = j)$$
$$= P(X_{t+1} = j, X_t = i, R_0 = r_0, ... R_N = r_N)$$

$$\therefore e_t(i,j) = P(X_{t+1} = j, X_t = i \mid R_0 = r_0, ... R_N = r_N) = \frac{\alpha_t(i) p_{ij} f_j(r_{t+1}) \beta_{t+1}(j)}{\sum_{i,j} \alpha_t(i) p_{ij} f_{j(t+1)}(r_{t+1}) \beta_{t+1}(j)}$$

- **Maximization Step**

$$\max_{\theta} E[\log L_c(\theta) \mid R_1 ... R_N ; \widetilde{\theta}]$$

$$= \max_{\theta} \sum_i d_0(i) \log \pi_i + \sum_{i,j} \sum_{t=1}^{N} e_{t-1}(i,j) \log p_{ij} + \sum_i \sum_{t=0}^{N} d_t(i) \left(-\frac{1}{2} \log(2\pi\sigma_i^2) - \frac{(r_t - \mu_i)^2}{2\sigma_i^2} \right)$$

$$s.t. \sum_i \pi_i = 1, \sum_j p_{ij} = 1$$

The Closed form solution for above maximization problem exists.

$$\pi_i^* = d_0(i) \quad , \quad p_{ij}^* = \frac{\sum_{t=0}^{N-1} e_t(i,j)}{\sum_j \sum_{t=0}^{N-1} e_t(i,j)} \quad , \quad \mu_i^* = \frac{\sum_{t=0}^{N} d_t(i) r_t}{\sum_{t=0}^{N} d_t(i)} \quad , \quad \sigma_i^* = \frac{\sum_{t=0}^{N} d_t(i)(r_t - \mu_i^*)^2}{\sum_{t=0}^{N} d_t(i)}$$

2.2.2 Double HMM

We run separate EM algorithm for Default data and using that information to run EM algorithm in Mean-variance data. To avoid redundancy, this part does not contain any derivation steps but include the key equation to implement the algorithm.

For Default data, everything is similar with previous EM algorithm except the probability of resulted default number is Poisson distribution, and the default intensity lambda is introduced as a parameter to determine.

$$\psi_i = P(Y_0 = i)$$
$$q_{ij} = P(Y_t = j \mid Y_{t-1} = i)$$
$$poisson(\lambda_i) = P(Z_t = z_t \mid Y_t = i; \theta_d) = f_{Y_t}^d(z_t)$$
$$\theta_d := \{\psi_i, q_{ij}, \lambda_i\}$$

- Expectation Step of Default State

$$\alpha_t^d(j) = P(Y_t = j; Z_0 = z_0, ... Z_t = z_t)$$
$$\beta_t^d(j) = P(Z_{t+1} = z_{t+1}, ... Z_N = z_N \mid Y_t = j)$$
$$\alpha_{t+1}^d(j) = \sum_i q_{ij} f_j^d(z_{t+1}) \alpha_t^d(i) \quad , \quad \beta_t^d(j) = \sum_k q_{jk} f_k^d(z_{t+1}) \beta_{t+1}^d(k)$$
$$\therefore d_t^d(j) = P(Y_t = j \mid Z_0 = z_0, ... Z_N = z_N) = \frac{\alpha_t^d(j) \beta_t^d(j)}{\sum_j \alpha_t^d(j) \beta_t^d(j)}$$
$$\therefore e_t^d(i,j) = P(Y_{t+1} = j, Y_t = i \mid Z_0 = z_0, ... Z_N = z_N) = \frac{\alpha_t^d(i) q_{ij} f_j^d(z_{t+1}) \beta_{t+1}^d(j)}{\sum_{i,j} \alpha_t^d(i) q_{ij} f_j^d(z_{t+1}) \beta_{t+1}^d(j)}$$

- Maximization Step of Default State

Similar with previous EM algorithm, the Expectation of log likelihood function defined as below.

$$\max_{\theta} E[\log L_c(\theta) \mid Z_1 ... Z_N ; \widetilde{\theta}_d]$$

$$= \max_{\theta} \sum_i d_0^d(i) \log \pi_i^d + \sum_{i,j} \sum_{t=1}^{N} e_{t-1}^d(i,j) \log q_{ij} + \sum_i \sum_{t=0}^{N} d_t^d(i) \left(z_t \log(\lambda_{d_t}) - \lambda_{d_t} - \log(z_t!) \right)$$

$$s.t. \sum_i \psi_i = 1, \sum_j q_{ij} = 1$$

Closed solution can be obtained with below equations.

$$\psi_i^* = d_0^d(i) \quad , \quad q_{ij}^* = \frac{\sum_{t=0}^{N-1} e_t^d(i,j)}{\sum_j \sum_{t=0}^{N-1} e_t^d(i,j)} \quad , \quad \lambda_i^* = \frac{\sum_{t=0}^{N} d_t^d(i) z_t}{\sum_{t=0}^{N} d_t^d(i)}$$

We bring above probability $d_t^d(j) = P(Y_t = j \mid Z_0 = z_0, ... Z_N = z_N)$ as our estimation of economic state, and assume the Probability transition matrix altered along the economic state.

Our goal is finding the parameter set, that maximizing posterior Likelihood given additional default data.

$$\max_{\theta} P\{R_0 = r_0, ..., R_N = r_N \mid Z_0 = z_0, ..., Z_N = z_N\}$$

- Expectation Step of Stock return state

We start with introducing the forward and backward probabilities, respectively, by

$$\alpha_t(j) = P(X_t = j; R_0 = r_0, ... R_t = r_t)$$

$$\beta_t(j) = P(R_{t+1} = r_{t+1}, ... R_N = r_N \mid X_t = j)$$

which can be calculated recursively,

$$\alpha_{t+1}(j) = \sum_i \sum_k p_{ij}^k d_t^d(k) f_j(r_{t+1}) \alpha_t(i)$$

$$\beta_t(j) = \sum_k \beta_{t+1}(k) \sum_i p_{jk}^i d_t^d(i) f_k(r_{t+1})$$

$$\therefore d_t(j) = P(X_t = j \mid R_0 = r_0, ... R_N = r_N) = \frac{\alpha_t(j)\beta_t(j)}{\sum_j \alpha_t(j)\beta_t(j)}$$

$$\therefore e_t(i,j,k) = P(X_{t+1} = j, X_t = i, Y_t = k \mid R_0 = r_0, ... R_N = r_N)$$

$$= \frac{\alpha_t(i) p_{ij}^k f_j(r_{t+1}) \beta_{t+1}(j)}{\sum_{i,j} \alpha_t(i) p_{ij}^k f_{j(t+1)}(r_{t+1}) \beta_{t+1}(j)} d_t^d(k)$$

- Maximization Step of Stock return state

$$\max_{\theta} E[\log L_c(\theta) \mid R_1 ... R_N ; \tilde{\theta}]$$

$$= \max_{\theta} \sum_i d_0(i) \log \pi_i + \sum_{i,j,k} \sum_{t=1}^{N} e_{t-1}(i,j,k) \log p_{ij}^k + \sum_i \sum_{t=0}^{N} d_t(i) \left(-\frac{1}{2} \log(2\pi\sigma_i^2) - \frac{(r_t - \mu_i)^2}{2\sigma_i^2} \right)$$

$$s.t. \sum_i \pi_i = 1, \sum_j p_{ij} = 1$$

$$\pi_i^* = d_0(i) \quad , \quad p_{ij}^* = \frac{\sum_{t=0}^{N-1} e_t(i,j,k)}{\sum_j \sum_{t=0}^{N-1} e_t(i,j,k)} \quad , \quad \mu_i^* = \frac{\sum_{t=0}^{N} d_t(i) r_t}{\sum_{t=0}^{N} d_t(i)} \quad , \quad \sigma_i^* = \frac{\sum_{t=0}^{N} d_t(i)(r_t - \mu_i^*)^2}{\sum_{t=0}^{N} d_t(i)}$$

3. Simulation

3.1 Data and Procedure

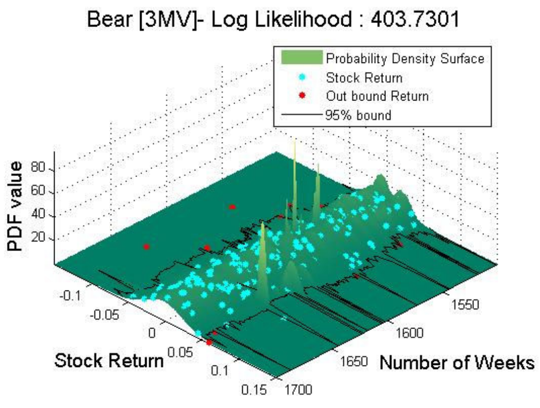
For stock data, we used weekly S&P 500 price from 16 Mar 1970 from http\\:finance.yahoo.com and for the default data, we used the weekly number of default data supplied from Risk Management class (MS&E 347).
We predict stock movements in Bear market regime and Bull market regime which are presented in Figure 4 with red line and light green line respective. Those regimes correspond to from week number 1501 to 1700 and week number 1751 to 1950. To forecast the stock return in the next week, we used 2 years horizon, which is 104 weeks prior to the prediction week as a training set. Once, the parameters are set through the learning algorithm, we forecast the stock return by giving the distribution of stock return of the next week.

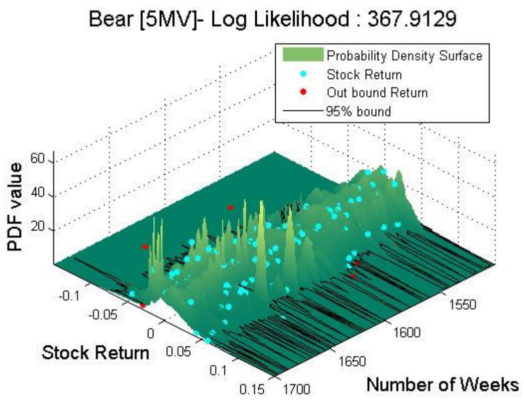
We perform the simulation with 3 states and 5 states for both modes. Thus, we simulated 3 states and 5 states of Single HMM, and 3 by 3 and 5 by 5 states of Double HMM.

As a performance measure to compare two models, we used sum of logarithm of the probability that the stock return is equal to the actual one given our predicted distribution.

3.2 Simulation Result

In each simulated graph, the prediction distribution is presented as a surface and the actual stock return is drawn as points.





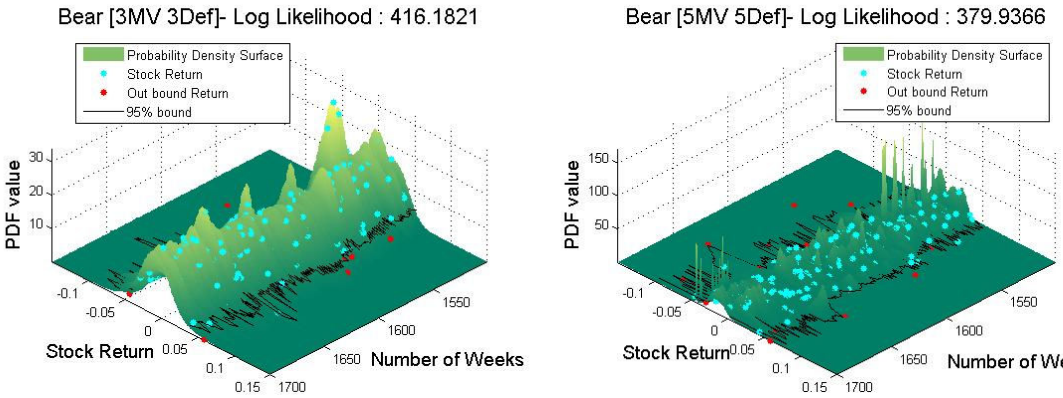


Figure 5. Bear Market prediction : upper row panels for Single HMM and lower row panels for Double HMM

Since the 3D surface stands for the predicted probability distribution in Figure 5, if the light blue dots (the actual returns) sit on the higher hills of the surface, it means our prediction is good. This performance measure calculated in each case.

As we can see in Figure 5, Double HMM performs better than the Single HMM when the number of states is the same. Also, 3 states case performs better than the case of 5 states. This is possible because larger number of states is generating fine many states which cause higher peaks and this leads poor predictability. In other words, fine states division can lead higher belief in small range of return and this belief is usually wrong.

In the Bull market (Figure 6), we can observe similar pattern. Obviously, the variance is much smaller than that of Bear market and also the measure is higher than the measure of Bear market.

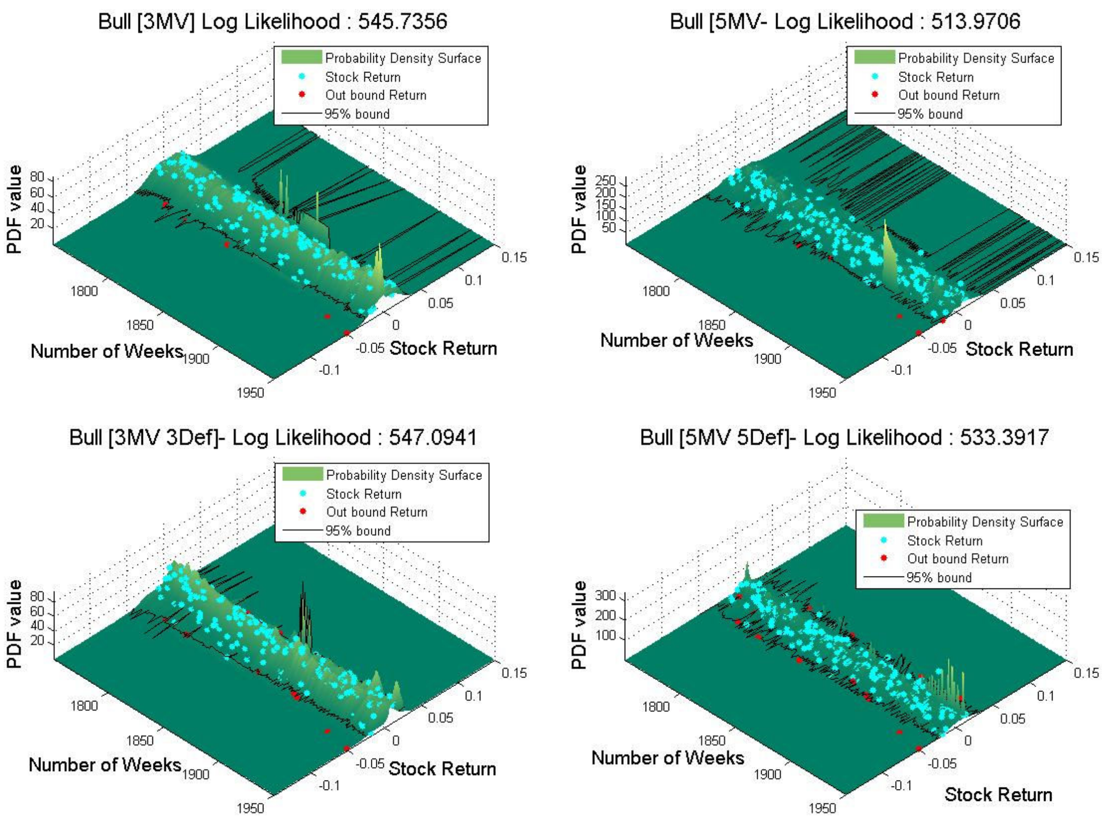


Figure 6. Bull Market prediction : upper row panels for Single HMM and lower row panels for Double HMM

The following figure shows the transition matrix for the 3 by 3 Double HMM.

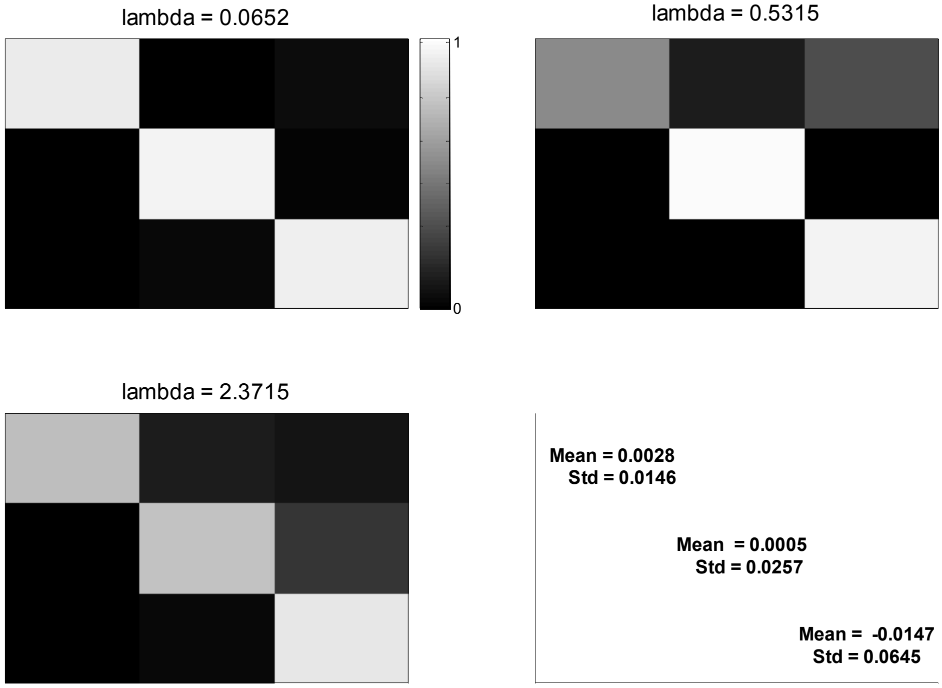


Figure 7. Transition matrix of mean variance states for the three economic states. The large value of lambda means bad situation economically.

In Double HMM setting, transition matrix has changed as the default state lambda varying. Above figure shows the EM-learning value of transition matrix into image form, brighter section is bigger value. As we see above, the transition matrix varies along the lambda. At the left upper plot, lambda value is the smallest, which means economic situation is good, and every state in mean-variance of stock return are stable and there is almost no probability of state changing. So if economic state is good, the state of mean-variance stays in where it was. Bigger lambda means the economic situation is severe, as in left lower plot, the transition matrix shows there is probability from good mean-variance to worse mean-variance, but no probability of getting better.
Therefore, transition matrix behavior is intuitively correct and it supports our previous assumption - transition probability of mean-variance is varying along the default state – as well.

4. Conclusion and Future works

In this project, we modeled the stock return as a mixture of Gaussian and discrete Markov Chain in order to improve the predictability of the stock model. Also, we introduced another economic data to present Double HMM which runs the Markov Chain of the economic states separately, which gives model more degree of freedom. As a result, we can verify that the Double HMM can predict better than the Single HMM and its transition matrix verifies the effectiveness of Double HMM.
As a future works, we can test the effectiveness of other economic data. In this project, we used default data for the economic states. However, the volatility index data can be a good candidate to extract the economic situation because it can gives us direct estimation of variance. As we mentioned in the introduction, the macroeconomic data is also directly connected to the economic situation too. By testing the effectiveness of those candidates, we can improve the predictability more. Moreover, if we can construct a choosing algorithm that chooses the most effective candidate by learning, we might build efficient trading machine.

5. Reference

[1] Azizpour, S. and Giesecke, K. 2008. Self-Exciting Corporate Defaults: Contagion vs.

Frailty. Working paper.
[2] Giesecke, K. 2004. Credit Risk Modeling and Valuation: An Introduction. Credit Risk: Models and Management, Vol. 2. D. Shimko (Ed.), Risk Books
[3] Daley, D. J. and Vere-Jones, D. 2007. An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. 2nd ed. Springer.
[4] CS 229 Lecture Note, Andrew Ng.