ISyE 6416 – Basic Statistical Methods - Spring 2016
Bonus Project: "Big" Data Analytics
Final Report

Team Member Names: Ziqi Yang, Feng Gao

Project Title: Hidden Markov Model for Stock Market Index

Problem Statement

In financial world, the market behaves differently when it is in "good" state or "bad" state. One key difference is the average market return can be higher in a good state and lower in the bad state. Another difference is that the market volatility, which can be represented by the volatility of SP500 index, varies by states. Usually, when the market is doing badly, the volatility will be higher than it is doing better. Acknowledging which states we are in is important to all market participants such as financial institutes and investors, since it was not only a signal to what the market will behave, but the mean return predicted by state will have an impact on market confidence, interest rate and other asset's returns while the volatility predicted by state also can be used for VIX (CBOE volatility index) pricing, option pricing and so on. In this project, we are going to use Hidden Markov Model to analyze which state the market is in and trying to obtain the mean return and volatility of each state.

Data Source

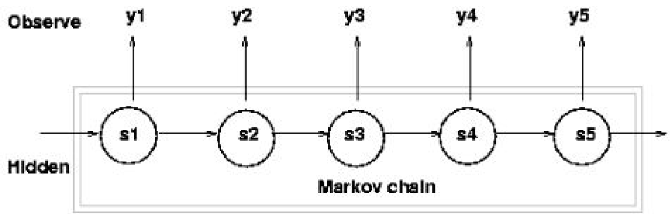
The data we will use is the SP500 index from Jan 1st, 1990 to Dec 31st, 2015, which can be obtained from Yahoo Finance. When pricing option, another input is the risk free rate. We can use the 6 months Treasury bill and it can be obtained from US. Department of Treasury.

figure1: S&P500 Index

Methodology

1. Hidden Markov Model

In hidden Markov model, the system is assumed to be a Markov process with hidden states. The underlying states are unobserved and follow Markov chain process with certain transition probability. Observations are dependent on the hidden states and visible.



Graph 1: Hidden Markov Model

Graph 1 shows the process of hidden Markov model with S being the hidden states and y being the observations.

2. Black-Scholes Formula

In Black-Scholes model, stock prices follow geometric Brownian Motion. The price process of the underlying stock is

$$S_t = S_0 \exp[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t]$$

where S_t is the stock price at time t, S_0 is the initial stock price, μ is the drift, σ is the volatility and B_t is the Brownian Motion. Taking logarithm on both sides of the equation above we get

$$\log(S_t) = \log(S_0) + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t$$

The logarithm of stock price at time t follows the normal distribution with expectation $\mathbb{E}\log(S_t) = \log(S_0) + (\mu - \sigma^2/2)t$. The volatility of stock price σ and the drift μ are assumed to remain constant across time in Black-Scholes model. However, in reality, they will change with the overall conditions of the stock market. Combined with the Markov chain of economic situation, Black-Scholes model is modified to be Markov Black-Scholes model, which can accommodate the change in states of the underlying financial assets.

3. Forward-Backward Algorithm

The forward-backward algorithm is developed to compute the posterior marginal of the hidden state variables based on observations. Since the parameters of the Markov Black-Scholes model including the drift μ, the volatility σ and the transition probability P are unknown, we apply forward-backward algorithm, which can help to estimate these parameters. The algorithm has two passes, in which the forward probability and backward probability are calculated respectively. The forward probability is represented as

$$\alpha_k(t) = \mathbb{P}(O_1, \dots, O_t, I_t = k)$$

and backward probability is represented as

$$\beta_k(t) = \mathbb{P}(O_{t+1}, \dots, O_T, I_t = k)$$

where O_j is the observation at time j and I_j stands for the state of the underlying stock at time j. The forward probability $\alpha_k(t)$ and backward probability $\beta_k(t)$ can be calculated in a recursive way. Imposing the probabilities into likelihood functions and transition equations,

$$L_k(t) = \mathbb{P}(S_t = k|O) = \frac{\alpha_k(t)\beta_k(t)}{\mathbb{P}(O)} \propto \alpha_k(t)\beta_k(t)$$

$$H_{k,l}(t) = \mathbb{P}(S_t = k, S_{t+1} = l|O) = \frac{\alpha_k(t)a_{k,l}b_l(O_{t+1})\beta_l(t+1)}{\mathbb{P}(O)}$$

where $a_{k,l}$ is the probability of transiting from state k to state l and $b_l(O_{t+1})$ is the probability of observation O_{t+1} given state l. Then we can get the estimates of parameters in the normal distribution and the transition matrix P.

4. Viterbi Algorithm in Hidden Markov Model

The Viterbi algorithm is used to find the most likely sequence of the hidden states based on a series of observations in hidden Markov model. In this algorithm, the maximum likelihood estimation procedure is implemented in a recursive way, which makes it efficient to calculate the corresponding probabilities. In each step, the algorithm incorporates one more observation in the data series and the complexity is $\mathrm{O(kt)}$ if the total number of states is k. The recursive process is shown as below.

$$\mathbb{P}(S_0, S_1 \dots S_T, O_1 \dots O_T) = \mathbb{P}(S_0)\mathbb{P}(S_1 \dots S_T, O_1 \dots O_T|S_0)$$
$$= \pi_{S_0}\mathbb{P}(S_1, O_1|S_0)\mathbb{P}(S_2 \dots S_T, O_2 \dots O_T|S_1, O_1, S_0) = \cdots = \pi_{S_0}\{\Pi_{i=1}^{T} a_{S_{i-1},S_i} b(O_i|S_i)\}$$

At each iteration, the formula has the similar form and the likelihood function finally becomes a product of T terms by induction.

Taking logarithm on both sides of the equation above, we can simplify it into

$$\log\mathbb{P}(S_0, S_1 \dots S_T, O_1 \dots O_T) = log\pi_{S_0} + \Sigma_{i=1}^{T}(loga_{S_{i-1},S_i} + logb(O_i|S_i)$$

By this transformation, the maximum likelihood problem is converted into the shortest path problem. We can adopt Dijkstra's algorithm or Bellman-Ford algorithm to find the shortest path, which is also the most likely sequence of the hidden states.

Preprocess and Assumptions

1. Data Preprocess:

Returns on S&P500 market index are calculated from the historical data from 1990 to 2015 by simple holding period return method

$$r_{t-1} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

where r is the return and P represents the price of the market index. Returns calculated are monthly returns and needed to be transformed into annual returns in final results generally.

2. States of Market

Two states are considered in the equity market. In the good state, companies are usually able to make profit and have plenty of chances to expand their business. Therefore, the average return is expected to be high with relatively low volatility. In the opposite, average return should be low and volatility will increase in the bad time.

Evaluation and Final Results

1. Average Return and Volatility of S&P500

Applying the forward-backward method to maximize the likelihood function, we get the estimates of the average returns and corresponding volatilities in both good and bad economic states.

S&P500 Index	Good State	Bad State
Average Return	0.0202	-0.0135
Volatility of Return	0.0268	0.0257

Table1: Average Return and Volatility in Different States

The returns and volatilities can be used to determine the prices of various derivatives and to identify arbitrage opportunities.

2. Transition between States

The transition matrix represents the probabilities by which the market will transit into another state or stay the same. It is given by the algorithm as below

States	Good	Bad
Good	0.6894	0.3106
Bad	0.4156	0.5844

Table2: Transition Matrix

The transition probabilities shown above is roughly accordant with the market performance. It is more likely that the good state will persist than the bad one.

3. Path of Economic States

According to the Viterbi Algorithm, we uncover the hidden states of the market every month. From the path we can find that there are clusters of good states and bad states in different time period. For example, in dot-com bubble crisis and 2008 financial crisis, most of the states are bad.

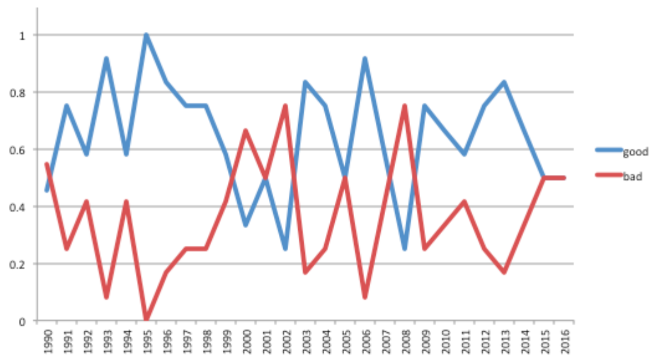


Figure2: Good vs Bad States

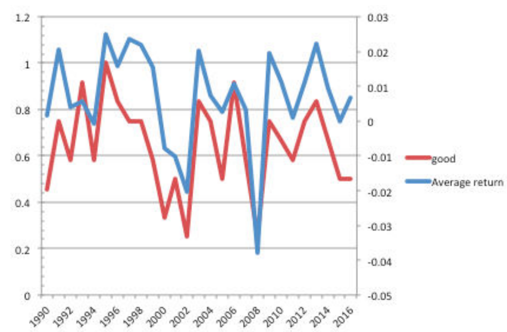


Figure3: Good States with S&P500 Return

Figure 2 demonstrates the percentage of good states and bad states every year. Figure 3 shows the percentage of good states along with the average return of S&P500. They move in the same trend, indicating that the algorithm works correctly.

Conclusion

From our analysis, this HMM model is useful in replicating the real trend of the market to some extent. Firstly, when looking at the percentage of good months in a year predicted by this model, we can find it follows the trend of averaged return of that specific year. From the perspective of transition probability, when we are in the good state, there is an approximately 70% chance to stay in the good state; while in bad state, half of the chance the market will be better in the following year. It is similar to the business cycle in many research papers. More importantly, we figure out there is a significant difference between the mean returns of the market in good states and bad states, which is the same as what we experience in the real situation. This model, however, has a limited ability to distinguish the volatilities. According to our result, both the good and bad states have a similar volatility in Gaussian model, but the true standard deviation is around 0.0368 in good state and 0.0535 in bad state. Additionally, there are many assumptions in HMM model, which might not be true in the real world. For example, the HMM model assumes the state of next period does not depend on how long the current state we were in. It is not true under general business cycle. Another example might be the state might not be just good or bad but can have many possibilities. In other words, each time interval might in a unique state. Thus, further study, for example, using the state space model, should be followed to remove those assumptions.

Further Study with Hidden State Space Model and Kalman Filter

Hidden Markov Model and State Space Model are similar in the sense that they all have the underlying hidden process which determines the series of observations. The difference is that in Hidden Markov Model, the underlying process have certain number of states and the transition probabilities are determined. However, in State Space Model, the underlying state variable changes continuously over time and follows the stochastic process. To estimate the parameters under state space model, we need to adopt the Kalman Filtering technique.

In the further study, we decide to apply the state space model to the stochastic commodity pricing models and estimate the parameters using Kalman Filtering method.