IME625A: Introduction to Stochastic Processes
PROJECT REPORT

HMM for Portfolio Management with Mortgage-Backed Securities ETF's

Submitted by:
Shripad Yadav(18807740)
Md. Kaif Faiyaz(190491)
Sanskar Shah(190763)

Indian Institute of Technology, Kanpur, India

August 18, 2022

1 THE PROBLEM

We want to look at the various growth and regression periods in the US using Mortgage Backed Securities prices as they depend on several macroeconomic indicators such as GDP, inflation rate, interest rate etc. Now to tackle this problem we employ **HMM aka Hidden Markov Model** to our benefit as it has been widely used in various financial predictions such as stock prices.

2 INTRODUCTION

2.1 Hidden Markov Models

A stochastic signal detection model in which observation sequences are supposed to be created from a hidden state sequence is known as the hidden Markov model.. This concept is derived from Markov Chains. Here, we assume that observations are generated by a hidden sequence and a Markov Process simulates it. It simply states that an observed event will be attributed to a series of probability distributions rather than its step-by-step status. HMM's major purpose is to observe a Markov chain's hidden states in order to learn about it. In the case of a Markov process X with hidden states Y, the HMM establishes that the probability distribution of Y for each time stamp must not be influenced by the history of X at that time.

The assumptions we make in a Hidden Markov model are as follows : Firstly, a hidden state generates an observation. These hidden states have a finite number of states and obey the Markov Properties. Between the states, the transition probability matrix is constant.

2.2 Toy Example for HMM

Here we are trying to understand HMM through an example:
Let us start our discussion with a guy 'X' who lives in a hypothetical town 'Y'. And let's assume there exists only two types of weather in the town, rainy and sunny. We also assume that on a given day, only one of them will occur. And the weather tomorrow only depends on today's weather. So we can easily model this as a Markov Chain with Transition Probabilities.
Now we add one more thing. We say on a given day, he has two moods, Happy and Sad. We also assume he has only one mood on any given day. Moreover, his mood depends on the weather of that day. Now the problem we have is we don't live in X's town. So we do not know weather on that day. However we can contact X and know his mood. So the states of the Markov Chain are unknown or Hidden to us. But we can observe some variables which are dependent on the states. And this is called a Hidden Markov Model. In other words, we can say a hidden Markov Model consists of ordinary Markov and a set of Observed Variables. Also, the observed variable depends only on the current state.

Given is the Transition Probability Matrix of the states, that is weather.

$$\begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$$

$P(sad|Sunny) = 0.75.P(happy|Sunny) = 0.25$
$P(sad|Rainy) = 0.40.P(happy|Rainy) = 0.60.$
Let S be Sunny and R be Rainy.

Hence, Now we have an Observation Probability Matrix as shown below.

$$\begin{bmatrix} 0.75 & 0.25 \\ 0.40 & 0.60 \end{bmatrix}$$

Now let us say we are guessing his mood for the next 3 days is Sad, Happy,Happy.Let us call this observation O. Also we know that the first day is a sunny day. So the probability of observing P(O).

Now we have four states which are possible, which will be S,S,S, S,S,R, S,R,S, S,S,S
$P(C1) = P(S, S, S) = 1 * 0.7 * 0.7 = 0.49$
$P(C2) = P(S, S, R) = 1 * 0.7 * 0.3 = 0.21$
$P(C3) = P(S, R, S) = 1 * 0.3 * 0.2 = 0.06$

$P(C4) = P(S, R, R) = 1 * 0.3 * 0.8 = 0.24$
And
$P(O, C1) = P(Sad|S) * P(Happy|S) * P(Happy|S) = 0.75 * 0.25 * 0.25 = 0.046875$
$P(O, C2) = P(Sad|S) * P(Happy|S) * P(Happy|R) = 0.75 * 0.25 * 0.6 = 0.1125$
$P(O, C3) = P(Sad|S) * P(Happy|R) * P(Happy|S) = 0.75 * 0.6 * 0.25 = 0.1125$
$P(O, C4) = P(Sad|S) * P(Happy|R) * P(Happy|R) = 0.75 * 0.6 * 0.6 = 0.27$

Therefore $P(O) = 0.046875 * (0.49) + 0.1125 * (0.21) + 0.1125 * (0.06) + 0.27 * (0.24) = 0.118144$.
Probability of observation coming sad, happy, happy.

Also, now we try to find out the best fit hidden sequence:
This would be $maxP(O, Ci) * P(Ci)$
From our earlier observation, we notice that the 4^{th} sequence i.e. C4 maximizes the probability of our observation. Hence it will be $P(O|C4) = P(O, C4) * P(C4) = 0.27 * 0.24 = 0.0648$

2.3 Brief Introduction to Mortgage Backed Securities

The word "mortgage" refers to a loan used to buy or keep a house, land, or other piece of real estate. Basically, the procedure of pledging anything as a guarantee or security against a loan is referred to as a mortgage. Mortgage Backed Security (MBS) is a kind of Asset Backed Security in which an investment is made up of a collection of House Loans which were purchased from banks which issued them.
MBS investors are paid on a regular basis, much like bond holders are paid on a regular basis.
Let us say 'A' is some kind of Investment Bank. Let's imagine A has purchased a large number of mortgages. Buying a mortgage basically means the home owners who borrow the money and instead of money which should've been gone to the mortgage broker whom the home owner thinks he got the loan from, or the servicer who they think they are paying to. But now, since the investment bank A has bought the mortgages, the bank 'A' has essentially become a lender to the home-owners and so, the mortgage payments of the home-owner will now come to the investment banks. But since the investment bank doesn't want to be the end holder as there is risk involved. As a result, it primarily seeks to serve as an intermediate medium. So the bank sets up a special entity/ corporation and sticks all the mortgages inside of that entity. This effectively makes the special entity/corporation the mortgage owner. To begin with, all of the mortgages are owned by investment banks. So now, the investment bank can sell the shares of the entity/ corporation to the investors. And now, these shares would be called Mortgage Backed Security.
These are part of a general class of Asset Backed Securities since all of the money flowing from the home owners would go to the mortgages. The asset is the right to payment that they hold. Now if by chance, any of the home/property owner is unable to repay the payment, then instead of payment, the entity will have the rights over their property and so the reason it is a Mortgage Backed Security.

Mortgage-backed securities (MBS) transforms a bank into a go-between(intermediary) for homebuyers and the investor community. Customers can get mortgages from a bank, which they can subsequently sell at a discount to be included in an MBS. The bank counts the transaction as a gain on its balance sheet, and also, it suffers no losses if the homebuyer fails later.

3 ALGORITHMS

The notations and symbols used in the algorithms are as follows :
Length of observation data, T
Number of states, N
Number of symbols per state, M
Observation sequence, $O = O_t, 1 <= t <= T$
Hidden state sequence, $Q = q_t, 1 <= t <= T$
Possible values of each state, $S = S_i, 1 <= t <= N$
Possible symbols per state, $V = v_k, 1 <= k <= M$
Transition matrix, $A = (a_{ij})$ where (a_{ij}) is the probability of being in state S_i at time t given that the observation is in

state S_i at time $t-1$,

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i), 1 <= i, j <= N.$$

Vector of initial probability, $p = (p_i), 1 <= i <= N,$, where p_i is the probability of being in state S_i at time $t = 1$,

$$p_i = P(q_1 = S_i), 1 <= i <= N.$$

Observation probability matrix, $B = (b_{ik})$, where b_{ik} is the probability of being in symbol v_k of an observation O_t given that the observation is in state S_i,

$$b_{ik} = b_i(k) = P(O_t = v_k | q_t = S_i), 1 <= i <= N, 1 <= k <= M.$$

Probability of observation $P(O|\lambda)$. This probability is called the likelihood of the HMM

Three Main Problems which HMM can solve :
In a generic scenario, we shall describe the three most important HMM issues :
1.Computing the probability of observations when we have the observation data and the HMM parameters.
2.Finding the best fit corresponding state when we have the observation data as well as the HMM parameters.
3.We try to calibrate the model parameters when we have the observation data in order to maximize the probability observation

Now we look into some algorithms to solve the above defined problems in HMM

3.1 Forward Algorithm

The forward algorithm is employed to find the conditional probability of observation on HMM's parameter's, P(O|λ). For initializing at time,t=1

$$\alpha_t(i) = p_i b_i(O_1) \tag{1}$$

Here we calculate the likelihood of the model by computing conditional probability of observation given last state is S_i :

$$\alpha_t(i) = P(O_1, O_2, ..., O_t, q_t = S_i | \lambda); T\epsilon(1, T) \tag{2}$$

Then employing Total probability over all the hidden states we get:

$$P(O_1, O_2, ..., O_T | \lambda) = \Sigma P(O_1, O_2, ..., O_T, q_T = S_i | \lambda) \tag{3}$$

Recursively we have:

$$\alpha_{t+1}(i) = (\Sigma \alpha_t(i) a_{ij}) b_j(O_{t+1}); T\epsilon[1, T) \tag{4}$$

The final probability :

$$P(O|\lambda) = \Sigma \alpha_T(i) \tag{5}$$

3.2 Backward Algorithm

It is along the lines of forward algorithm with the only exception that it starts iteration from the final time taken under consideration:

$$\beta_T(i) = 1 \tag{6}$$

for i over all the states.

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, ..., O_T | q_t = S_i \lambda) \tag{7}$$

The recursive formula to calculate this :

$$\beta_t(i) = (\Sigma \beta_{t+1}(i) a_{ij}) b_j(O_{t+1}); T\epsilon[1, T); i\epsilon[1, N] \tag{8}$$

$$P(O|\lambda) = \Sigma p_i b_i(O_1) \beta_1(i) \tag{9}$$

3.3 Viterbi Algorithm

For tackling the 2nd problem mentioned above this algorithm is used:the best choice of hidden states given the observation. The mission here is to find the best sequence of states when (O, λ) are given to maximize the probability of observation $P(O|\lambda)$.
We define function
$\delta_t(j)$ with $1 <= t <= T$ & $1 <= j <= N$ as

$$\delta_t(j) = max1 <= j <= N(P(q_1, q_2, ..., q_t = S_j, O_1, O_2, ..., O_t|\lambda)) \tag{10}$$

. The following steps are for the algorithmic steps:

$$\delta_t(i) = p_i b_i(O_1) \tag{11}$$

$$\delta_{t+1}(i) = (max_i \delta_t(i) a_{ij} b_j(O_{t+2}); T\epsilon[1, T) \tag{12}$$

$$q_T* = argmax_i(\delta_T(i)) \tag{13}$$

3.4 Baum-Welch Algorithm

Finally we look into fine-tuning the algorithmic parameters for the given observation sequence.The parameter set is defined to be :

$$\lambda = (A, B, p) \tag{14}$$

The algorithmic steps to follow are:

$$P(O|\lambda) = \Sigma\alpha_t(i)\beta_t(i); i\epsilon[1, N] \tag{15}$$

We define $\gamma_t(i)$ as the probability of being in state S_i at time t given the observation sequence:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} \tag{16}$$

Another jargon we throw in for computation is:

$$\eta_t(i, j) = P(q_t = S_i, q_{t+1} = S_j|O, \lambda) \tag{17}$$

and hence $\eta_t(i, j)$ becomes:

$$\eta_t(i, j) = \frac{\alpha_t(i)\beta_t(i)a_{ij}b_j(Ot + 1)}{P(O|\lambda)} \tag{18}$$

Finally the steps are: 1) Initializing the parameters , and the tolerance of the closeness of the data predicted wrt the original data.

$$p_i* = \gamma_1(i) \tag{19}$$

$$a_{ij}* = \frac{\Sigma\eta_t(i, j)}{\Sigma\gamma_t(i)}; j\epsilon[1, N] \tag{20}$$

$$b_{ik}* = \frac{\Sigma|(O_t = v_k)\gamma_t(i)}{\Sigma\gamma_t(i)}; k\epsilon[1, M] \tag{21}$$

$$\Delta = |P(O|\lambda*) - P(O|\lambda)| \tag{22}$$

4 IMPLEMENTATION

The HMM Was implemented using the R programming language. The close price dataset for **MBB** and **VMBS** was taken from **yahoofinance.com**, Two models were implemented, one had two hidden states and the other had four hidden states. We have tried to write down the algorithms from scratch but as it was taking a lot of time we resorted to the HMM package already present in R and have obtained the plots as follows. Here the hidden states considered are 2 namely growth and regression.
Now we take 2 observation states for 1 set and then 4 observation states as follow: For 2 observation state:
(i) if today's price is more than yesterday's price .

(ii) if today's price is less than yesterday's price.
For 4 observations:
(i) if change in price wrt yesterday's price is more than 2%
(ii) if change in price wrt yesterday's price is between 0% and 2%
(iii) if change in price wrt yesterday's price is between 0% and -2%
(iv) if change in price wrt yesterday's price is less than -2%
We used the Baum Welch algorithm to implement the estimation of parameters and then Viterbi algorithm to predict the best fit hidden states.The following graphs of Viterbi sequences were obtained for the datasets.

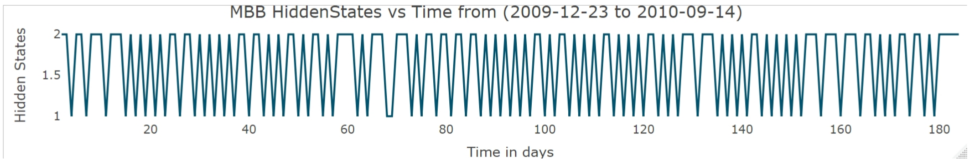

Figure 1: Two Hidden states, four symbols plot for MBB dataset from 23/12/2009 to 14/09/2010

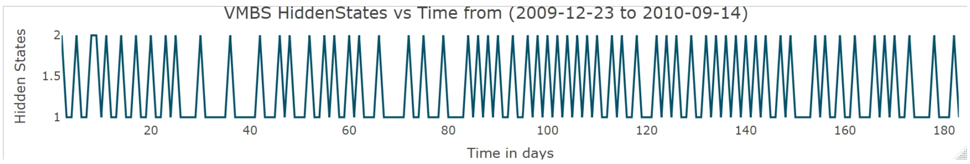

Figure 2: Two Hidden states, four symbols plot for VMBS dataset from 23/12/2009 to 14/09/2010

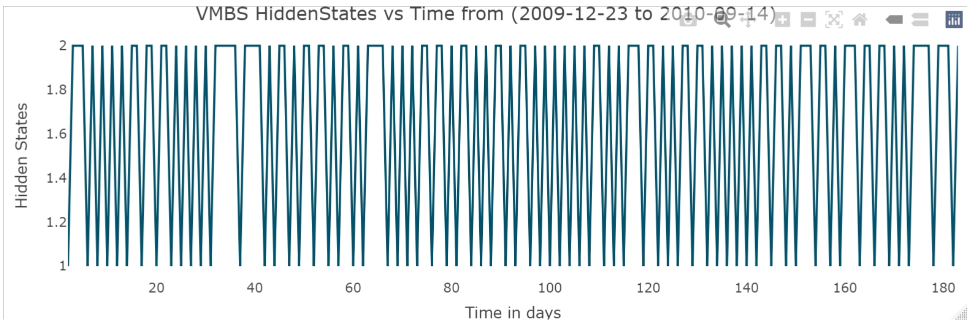

Figure 3: Two Hidden states, two symbols plot for VMBS dataset from 23/12/2009 to 14/09/2010

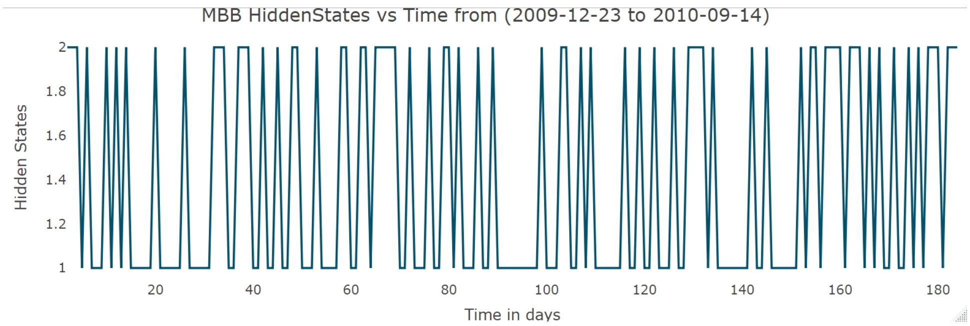

Figure 4: Two Hidden states, two symbols plot for MBB dataset from 23/12/2009 to 14/09/2010

5 APPENDIX

The following R code was used to determine the Hidden States in MBSs

```
library(HMM)
library(TSstudio)

data <- read.table("C:/Users/sryad/Downloads/VMBS.csv",header=TRUE,sep=",",as.is=TRUE)
close_price_data <- as.double(data$Close[1:183])
close_price_date <- data$Date[2:183]
timeDate <- as.Date.character(close_price_date)
close_price_data

A <- matrix(1:4,2,2)
A <- A/rowSums(A)

Time_len <- as.numeric(length(close_price_data))
B <- matrix(1:4, 2, 2)
B <- B/rowSums(B)
init <- c(0.3, 0.7)

data_close <- vector("numeric",Time_len-1)
for(i in 2:Time_len){
  v <- ((close_price_data[i]-close_price_data[i-1])/close_price_data[i-1])*100
  if(v<0)
  {
    data_close[i-1] <- 1
  }
  else if (v>=0){
    data_close[i-1] <- 2
  }
}
statesHMM <- c('re1', 'gr1')
symb <- c(1,2)
hmm <- initHMM(statesHMM, symb, init, A, B)
print(hmm)

start.time <- Sys.time()
bw <- baumWelch(hmm, data_close, maxIterations=500, delta = 1E-6 )
print(bw$hmm)
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken

vb <- viterbi(bw$hmm, data_close)
print(vb)

len <- length(vb)
a <- vb
b <- as.matrix(1,len-1)
time1 <- as.matrix(1:len-1,len-1)
for(i in 1:len){
  if(a[i]=="re1")
    b[i] <- 1
  else if (a[i]=="gr1")
    b[i] <- 2
}
```

```
myts <- ts(b, 1:182)
ts_plot(myts, title = 'VMBS Hidden States vs Time from (2009-12-23 to 2010-09-14)',
Xtitle = 'Time in days', Ytitle = 'Hidden States')
```

6 Acknowledgement

Finally we would like to convey special thanks to our supervisor Prof. Dr.Avijit Khanra who has guided us thoroughly and pointed out whenever we made an error or had any doubts and helped us resolve key concepts.

7 References

1)Hidden Markov Model for Portfolio Management with Mortgage-Backed Securities Exchange-Traded Fund,Nguyet (Moon) Nguyen,2017
2)Baum welch Algorithm,A Developer Diary,Abhishek Jana,2019
3)HMM, CRAN packages