UNIVERSITY OF CALIFORNIA

Los Angeles

Application of Hidden Markov Models in Financial Time Series: Inspection of the Capital Asset Pricing Model

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Minxuan Xu

2018

ABSTRACT OF THE THESIS

Application of Hidden Markov Models in Financial Time Series:
Inspection of the Capital Asset Pricing Model

by

Minxuan Xu
Master of Science in Statistics
University of California, Los Angeles, 2018
Professor Qing Zhou, Chair

In this thesis, we propose two Gaussian hidden Markov models: univariate Gaussian hidden Markov models with covariate and bivariate Gaussian hidden Markov models. After that they are applied to stock market returns to inspect the return-beta relationship stated in the capital asset pricing model (CAPM). The relationship is examined under 3 definitions of regimes: market regimes, idiosyncratic regimes and co-regimes. Results show that betas are larger under bullish market regime compared to bearish. Although no consistent patterns in beta are discovered under different idiosyncratic regimes and co-regimes, for each stock the betas do seem to vary considerably across regimes. Our model is also able to capture volatility clustering exhibited in return series.

The thesis of Minxuan Xu is approved.

Hongquan Xu

Nicolas Christou

Qing Zhou, Committee Chair

University of California, Los Angeles

2018

To my dad and grandma

who helped me through all the difficulties,

and those Mexican and Peruvian fiestas that

enriched my life in the past year

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

The stock market involves great uncertainty, and it is always necessary for investors to detect the signals and patterns in order to make wise decisions. Aside from using numerous technical indicators built from historical data and trading volume, many statistical techniques have been utilized to analyze financial time series and depict a more realistic representation of the underlying mechanism, including autoregressive integrated moving average models (ARIMA), generalized autoregressive conditional heteroscedastic models (GARCH) and hidden Markov models (HMM).

In this thesis, we propose two Gaussian hidden Markov models and make use of them to inspect the return-beta relationship stated in the capital asset pricing model (CAPM). We start by providing a brief overview of HMMs in Chapter 2. After that, a univariate Gaussian HMM with covariate and a bivariate Gaussian HMM are proposed in Chapter 3. Details about parameter learning and underlying state decoding are presented, followed by simulation studies to examine the effectiveness of the estimation procedures. Chapter 4 outlines some background knowledge about stock returns. In Chapter 5, we investigate the CAPM relationship under three definitions of regimes: market regimes, idiosyncratic regimes and co-regimes. After applying the proposed models on real-world stock data, results are interpreted to provide some insights. Chapter 6 concludes the thesis and suggests areas that need future work.

CHAPTER 2

Overview of Hidden Markov Models

Since its introduction in the 1970s to facilitate speech recognition [BP66], Hidden Markov Models (HMMs) have been extensively employed in various fields, including shape[BM04] and handwriting recognition [KHB89][SGH95], bioinformatics [KBM94] and financial modelling [HN05][BLD01]. This demonstrated power of HMM can be attributed to its flexibility to integrate two processes into one model.

In this section, we provide a brief overview of discrete-state HMMs. Most of the materials come from [Rab89]. HMMs with continuous hidden states or continuous time are beyond the scope of our discussion.

2.1 Basic Definitions

An HMM is formulated based on a stochastic process of non-observable states, with each state emitting an observable symbol. It is characterized by five key components:

- The sequence of hidden states $\{Z_t\}$, $t \in \mathbb{N}^+$, a first-order Markov process.
  Assume that there are N possible values for each state, i.e. $Z_t \in \{1, 2, \dots, N\}$. The memoryless property states that the conditional distribution of future states only depends on the current state, not on the past history,

$$\mathbb{P}\left(Z_{t+1}|Z_t, Z_{t-1}, \dots, Z_1\right) = \mathbb{P}\left(Z_{t+1}|Z_t\right).$$

- The sequence of observed symbols $\{Y_t\}$, $t \in \mathbb{N}^+$.
  Y_t could either be a discrete or continuous random variable. In the discrete case, assume that $Y_t \in \{1, 2, \dots, M\}$.

- The transition matrix of hidden states $A = (a_{ij})_{N \times N}$, where

$$a_{ij} = \mathbb{P}(Z_{t+1} = j | Z_t = i) \text{ and } \sum_{j=1}^{N} a_{ij} = 1. \tag{2.1}$$

- The conditional probability distribution of observed symbols given hidden states $\mathbb{P}(Y_t|Z_t)$, a.k.a. the emission probabilities. When Y_t is discrete, let $B = (b_j(k))_{N \times M}$, where

$$b_j(k) = \mathbb{P}(Y_t = k | Z_t = j) \text{ and } \sum_{k=1}^{M} b_j(k) = 1.$$

- The initial distribution of hidden states $\pi = (\pi_1, \ldots, \pi_N)$, where

$$\pi_i = \mathbb{P}(Z_1 = i), 1 \leq i \leq N. \tag{2.2}$$

Also note that the state transition probabilities and the emission probabilities do not vary with time. An example of HMM over the period $[1, n]$ is shown in Figure 2.1. Nodes in gray represent hidden variables, while nodes in white represent observed symbols. Arrows illustrate the dependency relationship between variables.

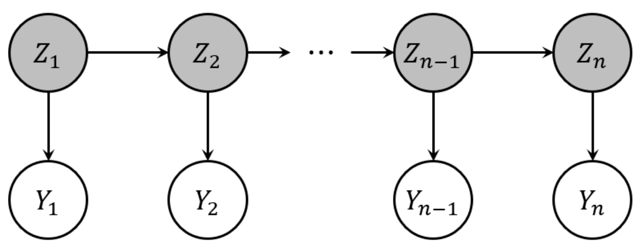

Figure 2.1: Graphical representation of an HMM

Let $\theta = (A, B)$ denote the model parameters. The coming sections address two tasks for an hidden Markov model:

1. **The Learning Problem:** Given $\{Y_t\}$, how to get the parameter estimate $\hat{\theta} = \arg\max_\theta \mathbb{P}(Y; \theta)$?

2. **The Decoding Problem:** Given $\{Y_t\}$ and $\hat{\theta}$, how to infer the hidden sequence $\{Z_t\}$?

Details about implementations are skipped in this chapter, and readers can feel free to refer to [Rab89]. It is in the next chapter that we are going to elaborate on the implementation of two proposed Gaussian HMMs.

2.2 Learning Model Parameters

Combining the conditional dependencies, the joint probability of the model is

$$\mathbb{P}\left(Y,Z;\theta\right)=\mathbb{P}(Z_1)\,\mathbb{P}\left(Y_1|Z_1\right)\prod_{t=2}^{n}\mathbb{P}\left(Z_t|Z_{t-1}\right)\mathbb{P}\left(Y_t|Z_t\right).$$

Since $\{Z_t\}$ is unobserved, the parameter estimate cannot be directly calculated by maximizing $\mathbb{P}\left(Y;\theta\right)=\sum_Z \mathbb{P}\left(Y,Z;\theta\right)$. Baum-Welch algorithm[BPS70], a variation of the more general Expectation-Maximization (EM) procedure[DLR77], is employed to deal with this issue. It iterates between an expectation step (E-step) and a maximization step (M-step).

In the E-step, the conditional expectation of the log-likelihood is computed given observed variables and current parameter estimate, i.e. $\mathbb{E}\left[\log\mathbb{P}\left(Y,Z;\theta\right)\middle|Y,\theta^{(m)}\right]$. In the M-step, parameter estimates are updated by maximizing this conditional expectation. We get $\theta^{(m+1)}=\arg\max_\theta \mathbb{E}\left[\log\mathbb{P}\left(Y,Z;\theta\right)\middle|Y,\theta^{(m)}\right]$. It has been shown that each iteration increases the log-likelihood unless a critical point has been reached. The two steps are repeated until after l iterations $\left|\theta^{(l)}-\theta^{(l-1)}\right|<\delta$. Here δ is the convergence tolerance. The final output of the method is then $\hat{\theta}=\theta^{(l)}$.

2.3 Decoding Hidden States

With the help of the Viterbi algorithm, one is able to find the hidden state sequence that is most likely to have generated the observations. The method produces the maximum a posteriori estimate

$$\hat{Z}=\arg\max_z \mathbb{P}\left(Z=z|Y\right)=\arg\max_z \mathbb{P}\left(Y,Z=z\right).$$

For more details, see [Vit67].

2.4 Model Selection

When fitting HMMs, it is possible to increase the likelihood by increasing the number of states N, but doing so may result in overfitting. Akaike Information Criterion (AIC) [Aka74] and Bayesian Information Criterion (BIC) [Sch78] provide a trade-off between parsimony and goodness-of-fit. They are defined as follows:

$$AIC = -2l\left(\hat{\theta}\middle|Y, \hat{Z}\right) + 2d$$
$$BIC = -2l\left(\hat{\theta}\middle|Y, \hat{Z}\right) + (\log n)d$$

where $l\left(\hat{\theta}\middle|Y, \hat{Z}\right)$ is the complete data log-likelihood evaluated at the parameter estimate from Section 2.2 and the inferred hidden sequence from Section 2.3, d is the number of parameters estimated by the model and n is the total number of observations.

Compared with AIC, BIC adds more penalty to the number of parameters. Models with different number of states are fitted, and the one with the lowest AIC or BIC is usually preferred.

CHAPTER 3

Two Gaussian Hidden Markov Models

Two Gaussian hidden Markov models are proposed in this chapter. We show their definitions and how they could be implemented step-by-step. After that, simulation studies are conducted to investigate the effectiveness of the model-estimation technique. These two models are employed in Chapter 5 to identify patterns in real-world data.

3.1 Univariate Gaussian HMMs with Covariate

3.1.1 Definition

Consider a hidden Markov model with discrete hidden states $\{Z_t\}$ and continuous observations $\{Y_t\}$. The space for the hidden states is $\{1, 2, \ldots, N\}$, and the space for the observations is $\mathbb{R}$. The state transition matrix and the initial state distribution are defined the same way as in (2.1) and (2.2). Let $\gamma_j = (\theta_{0,j}, \theta_{1,j}, \sigma_j^2)$ denote the parameter set for the jth state's emission, and $\theta = (A, \gamma)$ represent all the parameters. We assume that the emission density is Gaussian conditioning on the hidden state and the observable covariate $X_t \in \mathbb{R}$,

$$Y_t | X_t, Z_t = j \sim N(\theta_{0,j} + \theta_{1,j} X_t, \sigma_j^2)$$
$$f\left(Y_t | X_t; \gamma_j\right) = \left(2\pi\sigma_j^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\left(Y_t - \theta_{0,j} - \theta_{1,j} X_t\right)^2}{2\sigma_j^2}\right\}.$$

The mean and variance of the distribution are allowed to vary across different states.

The model is illustrated in Figure 3.1. Note that while the hidden states influence the observed symbols, it do not have any effects on the values taken by the covariates. In other words, we assume $\{X_t\}$ and $\{Z_t\}$ to be uncorrelated.

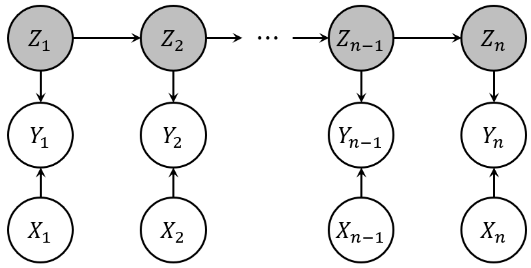

Figure 3.1: Graphical representation of an univariate HMM with covariate

Let $Z_{tj} = \mathbb{I}(Z_t = j)$, by (2.3) the joint probability of the model can be calculated by

$$
\begin{aligned}
\mathbb{P}(Y, Z|X;\theta) &= \mathbb{P}(Z_1)\,\mathbb{P}(Y_1|X_1, Z_1)\prod_{t=2}^{n}\mathbb{P}(Z_t|Z_{t-1})\,\mathbb{P}(Y_t|X_t, Z_t) \\
&\propto \prod_{j=1}^{N}\prod_{t=1}^{n}\left(f\left(Y_t|X_t;\gamma_j\right)\right)^{Z_{tj}} \times \prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{t=2}^{n}(a_{ij})^{Z_{(t-1)i}Z_{tj}}.
\end{aligned}
$$

The complete data log-likelihood is

$$
\begin{aligned}
l(\theta|Y, Z) &= \log\mathbb{P}(Y, Z|X;\theta) \\
&= \sum_{j=1}^{N}\sum_{t=1}^{n} Z_{tj}\log f\left(Y_t|X_t;\gamma_j\right) + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{n} Z_{(t-1)j}Z_{tj}\log a_{ij} + const \\
&= \sum_{j=1}^{N}\sum_{t=1}^{n} Z_{tj}\left(-\frac{1}{2}\log\sigma_j^2 - \frac{\left(Y_t - \theta_{0,j} - \theta_{1,j}X_t\right)^2}{2\sigma_j^2}\right) \\
&\quad + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{n} Z_{(t-1)j}Z_{tj}\log a_{ij} + const.
\end{aligned}
$$

For simplicity, the constant term is dropped in subsequent sections.

3.1.2 Implementation

Learning Parameters

The E-step:

First write down the objective function

$$\begin{aligned} Q\left(\theta\middle|\theta^{(m)}\right) &\triangleq \mathbb{E}\left[l\left(\theta|Y,Z\right)\middle|Y,X;\theta^{(m)}\right] \\ &= \sum_{j=1}^{N}\sum_{t=1}^{n}\mathbb{E}\left[Z_{tj}\middle|Y,X;\theta^{(m)}\right]\left(-\frac{1}{2}\log\sigma_j^2 - \frac{\left(Y_t-\theta_{0,j}-\theta_{1,j}X_t\right)^2}{2\sigma_j^2}\right) \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{n}\mathbb{E}\left[Z_{(t-1)i}Z_{tj}\middle|Y,X;\theta^{(m)}\right]\log a_{ij}. \end{aligned}$$

Let $\alpha_t^{(m)}(j) = \mathbb{P}\left(Y_{1:t}, Z_t = j\middle|X;\theta^{(m)}\right)$, $\beta_t^{(m)}(j) = \mathbb{P}\left(Y_{(t+1):n}\middle|Z_t = j, X;\theta^{(m)}\right)$. Note that $\forall j$,

$$\mathbb{E}\left[Z_{tj}\middle|Y,X;\theta^{(m)}\right] = \mathbb{P}\left(Z_t=j\middle|Y,X;\theta^{(m)}\right) \propto \mathbb{P}\left(Y, Z_t=j\middle|X;\theta^{(m)}\right) = \alpha_t^{(m)}(j)\beta_t^{(m)}(j)$$

$$\Rightarrow \mathbb{E}\left[Z_{tj}\middle|Y,X;\theta^{(m)}\right] = \frac{\alpha_t^{(m)}(j)\beta_t^{(m)}(j)}{\sum_{i=1}^{N}\alpha_t^{(m)}(i)\beta_t^{(m)}(i)} \triangleq u_t^{(m)}(j).$$

Similarly, $\forall i, j$,

$$\begin{aligned} \mathbb{E}\left[Z_{(t-1)i}Z_{tj}\middle|Y,X;\theta^{(m)}\right] &= \mathbb{P}\left(Z_{t-1}=i, Z_t=j\middle|Y,X;\theta^{(m)}\right) \\ &\propto \mathbb{P}\left(Y, Z_{t-1}=i, Z_t=j\middle|X;\theta^{(m)}\right) \\ &= a_{ij}^{(m)} f\left(Y_t\middle|X_t;\gamma_j^{(m)}\right)\alpha_{t-1}^{(m)}(j)\beta_t^{(m)}(j) \end{aligned}$$

$$\Rightarrow \mathbb{E}\left[Z_{(t-1)i}Z_{tj}\middle|Y,X;\theta^{(m)}\right] = \frac{a_{ij}^{(m)} f\left(Y_t\middle|X_t;\gamma_j^{(m)}\right)\alpha_{t-1}^{(m)}(j)\beta_t^{(m)}(j)}{\sum_{k=1}^{N}\sum_{l=1}^{N} a_{kl}^{(m)} f\left(Y_t\middle|X_t;\gamma_l^{(m)}\right)\alpha_{t-1}^{(m)}(m)\beta_t^{(m)}(l)} \triangleq w_t^{(m)}(i,j).$$

To calculate α's and β's, we make use of their recursion relationships. Forward summation is used to compute $\alpha_t^{(m)}(j)$:

1. Initialization: $\alpha_1^{(m)}(i) = \pi_i f\left(Y_1\middle|X_1;\gamma_i^{(m)}\right), i = 1, 2, \ldots, N.$
2. Recursion: For $t = 1, 2, \ldots, n-1$,
   $\alpha_{t+1}^{(m)}(j) = f\left(Y_{t+1}\middle|X_{t+1};\gamma_j^{(m)}\right)\sum_{i=1}^{N} a_{ij}^{(m)}\alpha_t^{(m)}(i), j = 1, 2, \ldots, N.$

Backward summation is used to compute $\beta_t^{(m)}(i)$:

1. Initialization: $\beta_n^{(m)}(i) = 1, i = 1, 2, \ldots, N$.

2. Recursion: For $t = n-1, n-2, \ldots, 1$,
   $\beta_t^{(m)}(i) = \sum_{j=1}^{N} a_{ij}^{(m)} f\left(Y_{t+1}\middle| Xt+1; \gamma_j^{(m)}\right) \beta_{t+1}^{(m)}(j), j = 1, 2, \ldots, N.$

The M-step:

Already know how to calculate the conditional expectations of Z_{tj} and $Z_{(t-1)j}Z_{tj}$, we plug them back into the objective function. Maximizing the objective function w.r.t. $a_{ij}, \theta_{0,j}, \theta_{1,j}$ and σ_j^2 (see Appendix A.1), we get:

$$\begin{aligned}
a_{ij}^{(m+1)} &= \frac{\sum_{t=2}^{n} w_t^{(m)}(i,j)}{\sum_{k=1}^{N}\sum_{t=2}^{n} w_t^{(m)}(i,k)} \\
\theta_{1,j}^{(m+1)} &= \frac{\sum_{t=1}^{n} u_t^{(m)}(j) X_t Y_t - n_j \overline{X}_j \overline{Y}_j}{\sum_{t=1}^{n} u_t^{(m)}(j) X_t^2 - n_j \left(\overline{X}_j\right)^2} \\
\theta_{0,j}^{(m+1)} &= \overline{Y}_j - \theta_{1,j}^{(m+1)} \overline{X}_j \\
\left(\sigma_j^2\right)^{(m+1)} &= \frac{\sum_{t=1}^{n} u_t^{(m)}(j)\left(Y_t - \theta_{0,j}^{(m+1)} - \theta_{1,j}^{(m+1)} X_t\right)^2}{n_j}.
\end{aligned}$$

Here $n_j = \sum_{t=1}^{n} u_t^{(m)}(j), \overline{X}_j = \frac{\sum_{t=1}^{n} u_t^{(m)}(j) X_t}{n_j}, \overline{Y}_j = \frac{\sum_{t=1}^{n} u_t^{(m)}(j) Y_t}{n_j}$.

Decoding Hidden States

1. Initialization: $\delta_1(i) = \pi_i f\left(Y_1|X_1; \hat{\gamma}_i\right), i = 1, 2, \ldots, N$.

2. Forward maximization: For $t = 1, 2, \ldots, n-1,\ j = 1, 2, \ldots, N$
   $\delta_{t+1}(j) = \max_{1 \le i \le N}\left\{\delta_t(i)\hat{a}_{ij}\right\} f\left(Y_{t+1}|X_{t+1}, \hat{\gamma}_j\right)$
   $\eta_{t+1}(j) = \arg\max_{1 \le i \le N}\left\{\delta_t(i)\hat{a}_{ij}\right\}.$

3. Backward tracking: Let $\hat{Z}_n = \arg\max_i \delta_n(i)$; for $t = n-1, n-2, \ldots, 1$, $\hat{Z}_t = \eta_{t+1}(\hat{Z}_{t+1})$.

Summary of the Algorithm

- Step 0: Initialize $\theta^{(0)}$, set $m = 0$;

- Step 1: Calculate $\alpha_t(i)^{(m)}$ and $\beta_t(i)^{(m)}$;
- Step 2: Calculate $u_t(i)^{(m)}$ and $w_t(i,j)^{(m)}$;
- Step 3: Calculate $a_{ij}^{(m+1)}$, $\theta_{1,j}^{(m+1)}$, $\theta_{0,j}^{(m+1)}$ and $\left(\sigma_j^2\right)^{(m+1)}$.
  Add m by 1, repeat Steps 1 to 3 until convergence;
- Step 4: Calculate $\hat{Z}$.

3.1.3 Simulation Study

Data Generation

Suppose that there are $N = 3$ distinct states and we want $n = 400$ observations. The first "hidden" state, Z_1, is drawn from the initial state distribution $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. It is in fact visible to us in this study, so that we can later judge the performance of the Viterbi algorithm by comparing the inferred states with the true values. Following the state transition probabilities, all the hidden states are drawn one-by-one. After that, covariates are randomly generated from $N(0.01, 0.01^2)$. Given $\{Z_t\}, \{X_t\}$ and the γ_j's, observed symbols $\{Y_t\}$ are drawn from $N\left(\theta_{0,Z_t} + \theta_{1,Z_t}X_t, \sigma^2_{Z_t}\right)$.

Model Estimation

Models with $N = 2, 3, 4$ hidden states are fitted using the "depmixS4" package[VS10] in R. According to its manual, it is believed that this package implements our proposed method to do model-fitting. A simple linear model without hidden states is also fitted. From Figure 3.2, the model with 3 hidden states is selected over the other threes, which is in line with the data-generating process.

As the parameters of our model are non-identifiable (you can feel free to re-label the states), we try to match the fitted parameters to the true ones by switching the order. From Table 3.1 one can see that the fitted transition probabilities are quite close to the true ones except for $\hat{a}_{31}$ and $\hat{a}_{33}$. Table 3.2 shows the true emission density parameters and the fitted

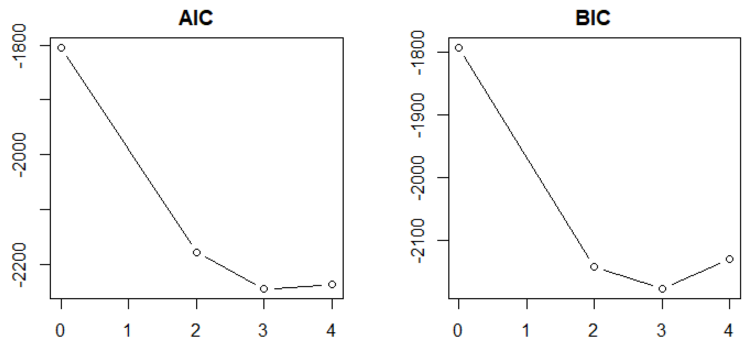

Figure 3.2: Information criteria of models fitted on the simulated data

ones. Standard deviations seem to be estimated really well, while the estimated slope $\hat{\theta}_{1,3}$ differ a little bit too much from the true value.

True	1	2	3	Fitted	1	2	3
1	0.4	0.5	0.1	1	0.358	0.563	0.079
2	0.2	0.7	0.1	2	0.218	0.670	0.112
3	0.5	0.3	0.2	3	0.676	0.324	0.000

Table 3.1: True and fitted state transition matrices (3-state model)

True	θ_0	θ_1	σ	Fitted	θ_0	θ_1	σ
1	0.03	1.20	0.010	1	0.032	1.106	0.010
2	-0.01	0.85	0.005	2	-0.012	0.921	0.005
3	0.01	1.50	0.050	3	0.006	1.776	0.043

Table 3.2: True and fitted emission density parameters (3-state model)

In terms of hidden states, the prediction error is $26/400 = 6.50\%$, a very exceptional performance. The following plot also illustrates that the true and inferred states series are highly similar to each other. Overall, the package works well to fit univariate Gaussian HMMs with covariate.

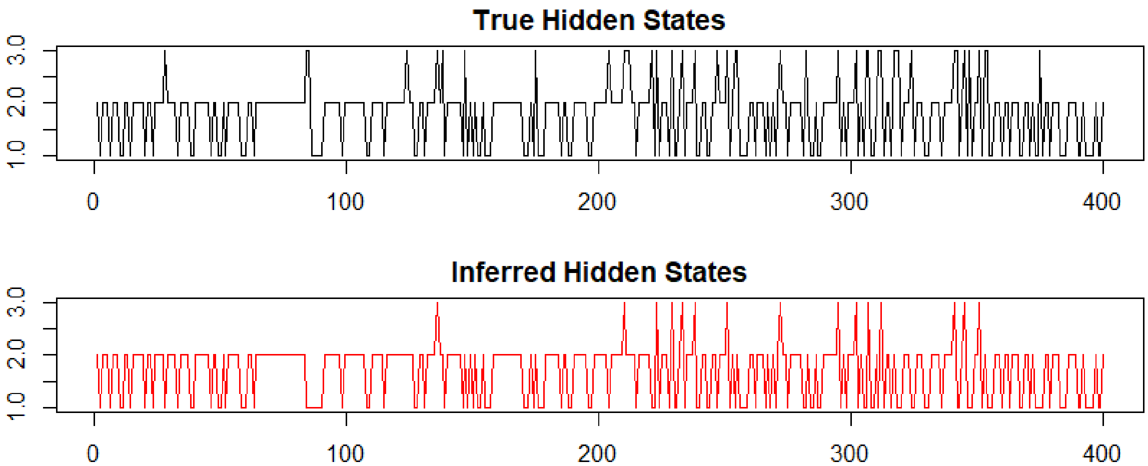

Figure 3.3: True and inferred hidden states of the simulated data

3.2 Bivariate Gaussian HMMs

3.2.1 Definition

Consider another hidden Markov model with discrete hidden states $\{Z_t : Z_t = 1, 2, \ldots, N\}$ and continuous observations $\{Y_t : Y_t \in \mathbb{R}^2\}$. The state transition matrix and the initial state distribution are defined the same way as in (2.1) and (2.2). Let $\gamma_j = (\mu_j, \Sigma_j)$ denote the parameter set for the jth state's emission, and $\theta = (A, \gamma)$ represent all the parameters, where

$$\mu_j = \begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix}, \Sigma_j = \begin{pmatrix} \sigma^2_{1,j} & \sigma_{12,j} \\ \sigma_{12,j} & \sigma^2_{2,j} \end{pmatrix}.$$

Suppose that the emission density is bivariate Gaussian conditioning on the hidden state, we have

$$\begin{aligned} Y_t | Z_t = j &\sim N(\mu_j, \Sigma_j) \\ f(Y_t; \gamma_j) &= (2\pi)^{-1} |\Sigma_j|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (Y_t - \mu_j)^\intercal \Sigma_j^{-1} (Y_t - \mu_j) \right\}. \end{aligned}$$

The mean vector and covariance matrix of the distribution are allowed to vary across different states. The model is illustrated in Figure 3.4.

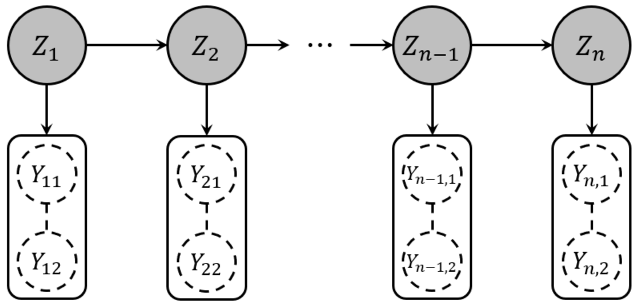

Figure 3.4: Graphical representation of an bivariate HMM

Let $Z_{tj} = \mathbb{I}(Z_t = j)$, by (2.3) the joint probability of the model can be calculated by

$$\mathbb{P}(Y, Z; \theta) \propto \prod_{j=1}^{N} \prod_{t=1}^{n} (f(Y_t; \gamma_j))^{Z_{tj}} \times \prod_{i=1}^{N} \prod_{j=1}^{N} \prod_{t=2}^{n} (a_{ij})^{Z_{(t-1)i} Z_{tj}}.$$

The complete data log-likelihood is

$$\begin{aligned} l(\theta | Y, Z) &= \log \mathbb{P}(Y, Z; \theta) \\ &= \sum_{j=1}^{N} \sum_{t=1}^{n} Z_{tj} \left(-\frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (Y_t - \mu_j)^{\intercal} \Sigma_j^{-1} (Y_t - \mu_j) \right) \\ &\quad + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{n} Z_{(t-1)j} Z_{tj} \log a_{ij} + const. \end{aligned}$$

For simplicity, the constant term is dropped in subsequent sections.

3.2.2 Implementation

Learning & Decoding

In the E-step, the quantities $\alpha_t^{(m)}(j), \beta_t^{(m)}(j), u_t^{(m)}(j), w_t^{(m)}(i,j)$ can be computed as in Section 3.1.2 by slightly modifying the emission density. In the M-step, after maximization (See Appendix A.2) we get

$$\begin{aligned} a_{ij}^{(m+1)} &= \frac{\sum_{t=2}^{n} w_t^{(m)}(i,j)}{\sum_{k=1}^{N} \sum_{t=2}^{n} w_t^{(m)}(i,k)} \\ \mu_j^{(m+1)} &= \overline{Y}_j \\ \Sigma_j^{(m+1)} &= \frac{\sum_{t=1}^{n} u_t^{(m)}(j) (Y_t - \mu_j)(Y_t - \mu_j)^{\intercal}}{n_j}. \end{aligned}$$

Where $n_j = \sum_{t=1}^{n} u_t^{(m)}(j), \overline{Y}_j = \frac{\sum_{t=1}^{n} u_t^{(m)}(j) Y_t}{n_j}$. Decoding of hidden states also follows the procedures in Section 3.1.2.

Summary of the Algorithm

- Step 0: Initialize $\theta^{(0)}$, set $m = 0$;
- Step 1: Calculate $\alpha_t(i)^{(m)}$ and $\beta_t(i)^{(m)}$;
- Step 2: Calculate $u_t(i)^{(m)}$ and $w_t(i,j)^{(m)}$;
- Step 3: Calculate $a_{ij}^{(m+1)}$, $\mu_j^{(m+1)}$ and $\Sigma_j^{(m+1)}$.

  Add m by 1, repeat Steps 1 to 3 until convergence;
- Step 4: Calculate $\hat{Z}$.

3.2.3 Simulation Study

Similar to the previous case, we first generate the hidden state sequence $\{Z_t\}$. After that, each Y_t is drawn from the bivariate normal distribution $N(\mu_{Z_t}, \Sigma_{Z_t})$. The R package "mhsmm"[OH11] is adopted to fit the simulated data assuming there are 2, 3 and 4 hidden states. A bivariate normal model with no hidden state is also fitted.

Figure 3.5 presents the information criteria of the fitted 4 models. Both AIC and BIC are in favor of the 2-state model, which agrees with the true data-generating process. The closeness between true and fitted transition probabilities are demonstrated in Table 3.3. In Table 3.4, we can see that the fitted parameter also do not differ much from the true values, except for the estimate of μ_1 and σ_{12} in state 2.

True	1	2	Fitted	1	2
1	0.800	0.200	1	0.802	0.198
2	0.300	0.700	2	0.203	0.797

Table 3.3: True and fitted state transition matrices (2-state model)

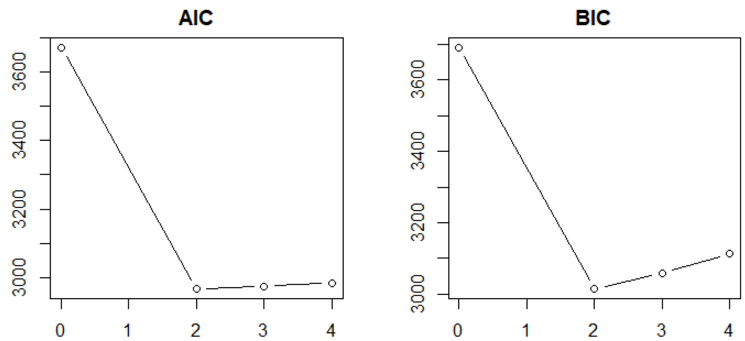

Figure 3.5: Information criteria of models fitted on the simulated data

True	μ_1	μ_2	σ_1^2	σ_2^2	σ_{12}
1	-3.000	0.000	1.000	1.000	0.000
2	1.000	2.000	4.000	3.000	2.000
Fitted	μ_1	μ_2	σ_1^2	σ_2^2	σ_{12}
1	-2.979	0.005	1.068	0.854	-0.085
2	0.562	1.840	5.203	3.163	2.709

Table 3.4: True and fitted emission density parameters (2-state model)

The prediction error for hidden states is as low as $32/400 = 8.00\%$, as is depicted in Figure 3.6. In general, the package seems to have a satisfactory performance in fitting bivariate HMMs.

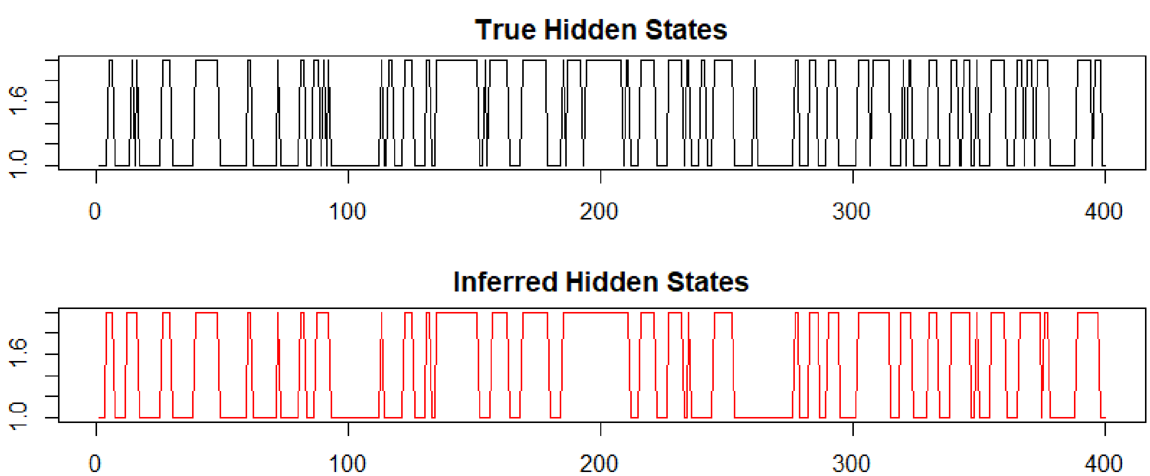

Figure 3.6: True and inferred hidden states of the simulated data

CHAPTER 4

Stock Returns

This chapter provides some background knowledge about stock returns. Concepts and notations defined in this chapter are extensively used in the next chapter.

4.1 Basic Definitions

4.1.1 Returns

Let $\{S_t\}$ denote a discrete-time stochastic process of asset price. Each S_t stands for the asset price at time t. The simple return (arithmetic return) of the asset over the period $[t-1, t]$ is defined by

$$r_t^{simple} = \frac{S_t - S_{t-1}}{S_{t-1}}.$$

The unit time period could be a day, a week, etc. In real-world applications, the gaps in trading days are often ignored. On each trading day, the stock exchange quotes 5 prices for each stock listed: open, high, low, closing and adjusted closing. An adjusted closing price is calculated by removing the effects of stock splits and dividends from the closing price, and thus is a better measure of capital gains. Without special mentioning, 'price' refers to adjusted closing price in this thesis.

Note that the range of simple return is asymmetric: it can never fall below -1. For this reason, another definition of return is preferred when doing analyses. The continuously compounded return (log return) over the period $[t-1, t]$ is

$$r_t^{cc} = \log\left(\frac{S_t}{S_{t-1}}\right) = \log S_t - \log S_{t-1}. \tag{4.1}$$

For simplicity of notation, by default r_t is equivalent to r_t^{cc} throughout the thesis.

To convert simple returns to log returns, notice the following relationship:

$$\begin{aligned} S_{t-1}(1+r_t^{simple}) &= S_t = S_{t-1}e^{r_t} \\ 1+r_t^{simple} &= e^{r_t} \\ \Rightarrow r_t &= \log\left(1+r_t^{simple}\right). \end{aligned} \tag{4.2}$$

Consider a portfolio made up of all risky assets in the market, held in proportion to their value. This is the so-called market portfolio. It is not tractable, and a market index is good enough to serve as a proxy. Let M_t denote the value of the market index at time t, the log return on the market index (market return) over the period $[t-1, t]$ is

$$r_{m,t} = \log\left(\frac{M_t}{M_{t-1}}\right) = \log M_t - \log M_{t-1}. \tag{4.3}$$

Let $r_{f,t}$ denote the log return on a risk-free asset (risk-free return) over the period $[t-1, t]$. As there is no absolutely zero risk assets in the world, the interest rate on a three-month U.S. Treasury Bill is usually used as a proxy.

4.1.2 Excess Returns

Excess return is the difference between asset return and risk-free return. Its expectation form is known as risk premium, since investors are only compensated for taking extra risks compared to holding a risk-free investment. The asset excess return and market excess return are defined as

$$R_t = r_t - r_{f,t}, R_{m,t} = r_{m,t} - r_{f,t}. \tag{4.4}$$

And the corresponding risk premiums are

$$\mathbb{E}\left(R_t\right) = \mathbb{E}\left(r_t\right) - r_{f,t}, \mathbb{E}\left(R_{m,t}\right) = \mathbb{E}\left(r_{m,t}\right) - r_{f,t}.$$

4.2 Volatility of Returns

Volatility refers to the standard deviation of asset log returns, i.e.

$$Volatility\ \left(r_t\right) = \sqrt{Var\left(r_t\right)}.$$

Previous studies [Con01] has revealed that the return series of stock exhibit volatility clustering, when periods of high volatility are followed by high volatility and periods of lower volatility are followed by low volatility. In Chapter 5, readers are going to see how this stylized fact is captured by the hidden Markov models.

4.3 Normality of Stock Returns

Consider the following asset dynamics (Geometric Brownian Motion) for stock price used by Black and Scholes in their options pricing model [BS73]:

$$dS_t = \mu S_t dt + \tau S_t dW_t$$

where μ represents drift, τ represents volatility, W_t is a Wiener process (standard Brownian motion) that satisfies $W_t \sim N(0, t)$. Using Ito's lemma we have

$$d \log S_t = \left(\mu - \frac{1}{2}\tau^2\right) dt + \tau dW_t.$$

In discrete time,

$$r_t = \log S_t - \log S_{t-1} = \left(\mu - \frac{1}{2}\tau^2\right) + \tau(W_t - W_{t-1}).$$

As $W_t - W_{t-1} \sim N\left(0, t - (t-1)\right) = N(0, 1)$,

$$r_t \sim N\left(\left(\mu - \frac{1}{2}\tau^2\right), \tau^2\right).$$

So the log return of the stock is approximately normal. This argument can be extended to market return as well. In terms of the risk-free return, theoretically it should not exhibit any volatility. Empirically, the proxy for risk-free rate does fluctuate as time goes, but the variance is negligible. For this reason, we could also regard the stock excess return and market excess return to be normal.

CHAPTER 5

Inspection of the Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) is widely accepted and used in the financial industry to compute benchmark returns for securities and portfolios. Developed by William Sharpe [Sha64] and John Lintner [Lin69], it connects the risk premium of an asset to the market risk premium by a linear relationship. This relationship seems to vary with different underlying non-observable states. In the field of econometrics, a state is often referred to as an regime. We use state and regime interchangeably.

This chapter starts with an outline of CAPM, followed by the definitions of regimes from three perspectives and the methodologies to investigate the CAPM relationship under various regimes. Finally, the proposed procedures are applied into the analysis of real-world financial time series. Results are shown with interpretations.

5.1 The Capital Asset Pricing Model (CAPM)

According to the Capital Asset Pricing Model (CAPM), in an efficient market there exists a linear relationship between the expectation of a stock's excess return (called the stock's risk premium) and the market risk premium:

$$\mathbb{E}\left(R_t\right) = \beta\, \mathbb{E}\left(R_{m,t}\right).$$

This is also known as the return-beta relationship. Now consider a regression model with a Gaussian noise that is independent to the predictor:

$$R_t = \alpha + \beta R_{m,t} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2). \tag{5.1}$$

Here β describes the sensitivity of the stock excess returns to the market excess returns. It is also a measure of the systematic risk. α usually plays a role in stock selection: a positive α means the stock is underpriced, while a negative α means it is overpriced. If CAPM holds (which is too good to be true in reality), after taking expectation on both sides one can see that $\alpha = 0$. Readers may have noticed that the α and β symbols used here carry meanings different from those in Chapter 3.

Take another look at (5.1), taking covariance with $R_{m,t}$ on both sides, we have

$$Cov(R_t, R_{m,t}) = \beta Var(R_{m,t}) \Rightarrow \beta = \frac{Cov(R_t, R_{m,t})}{Var(R_{m,t})} \tag{5.2}$$

Traditional CAPM assumes that betas are fixed in time, or at most slowly evolving as the company grows. However, some findings in the literature argue that the beta coefficients may be time-varying [GF99][Rey99][AF09]. If it is really the case, fund managers and option traders need to frequently adjust their model inputs to avoid receiving distorted predictions and losing money. In the coming sections we are going to focus on β only, α is not our concern.

5.2 Regimes from Three Perspectives

Before examining the return-beta relationship under different regimes, it is necessary to clarify how we define regimes and what dynamics they are supposed to capture. Three kinds of regimes are considered: market, idiosyncratic and co-regimes. While market regimes are frequently mentioned by economists, the latter two are named by us. We also elaborate on how to identify regimes with the data, which is made up of observed excess return pairs $(R_t, R_{m,t})$.

5.2.1 Market Regimes

Market regimes are the underlying states of the stock market, to be more specific, the market index return or market index excess return. Previous studies have found evidence of regimes in market returns, usually a high-mean, low variance regime and a low-mean high variance regime[HL96][SN97]. [TSN89] has suggested negative correlation in movements between volatility and market excess returns. Market regimes capture market-wide news and events. Possible driving force of regime-switching are fluctuations in macroeconomic factors, for instance, exchange rate[WCM11], inflation rate[Che09] and level of industrial productions[HL96] .

To identify market regimes, consider a univariate Gaussian HMM on the excess return of market index. This is a simplified version of both models proposed in Chapter 3. We assume that

$$R_{m,t}|Z_{m,t} = j \sim N(\mu_{m,j}, \sigma^2_{m,j}),$$

where $Z_{m,t}$ represents the market regime. As one can see, the mean and the volatility are allowed to differ from one regime to another.

After regimes are discovered by the Viterbi algorithm, we cut our data pairs into blocks accordingly. Linear regressions of R_t on $R_{m,t}$ are then implemented for blocks characterized by the same regime.

5.2.2 Idiosyncratic Regimes

The return of each stock has two sources of variation: the market factor and the firm-specific factor. Suppose that news comes out, say, that the CEO of the company is detained by law enforcement authorities due to embezzlements. This would hardly affect the overall market returns, but it would definitely lower investors' confidence in the future earnings of the firm, and thus reduces the stock price. In this scenario, the sensitivity of stock excess returns to the market excess returns (beta) may also alter.

To identify idiosyncratic regimes, consider a univariate Gaussian HMM with covariate. We assume that

$$R_t|R_{m,t}, Z_t = j \sim N(\alpha_j + \beta_j R_{m,t}, \sigma^2_j),$$

where Z_t stands for the idiosyncratic regime. This conditional distribution could easily follow from (5.1). Recall from Section 3.1.1 that the model requires the hidden states and the covariates to be uncorrelated. This needs to be examined before fitting the model.

5.2.3 Co-Regimes

Co-regimes captures the phenomena in stock excess returns and market excess returns that tend to emerge simultaneously. A typical example is co-movement in returns or volatilities.

To identify co-regimes, consider a bivariate Gaussian HMM . We assume that

$$\begin{pmatrix} R_t \\ R_{m,t} \end{pmatrix} | Z_{co,t} = j \sim N\left(\mu_j, \Sigma_j = \begin{pmatrix} Var(R_t)_j & Cov(R_t, R_{m,t})_j \\ Cov(R_t, R_{m,t})_j & Var(R_{m,t})_j \end{pmatrix}\right).$$

Here $Z_{co,t}$ represents the co-regime. After model-fitting, betas can be easily deduced from the covariance matrix by (5.2).

5.3 Data Description & Preparation

Five stocks from various industries in the U.S. stock market are chosen for our study. They are McDonald's (MCD, food), Nike (NKE, apparel), Twitter (TWTR, information technology), Verizon (VZ, telecommunication) and Walgreens Boots Alliance (WBA, retail pharmacy). The market index S&P500 serves as the proxy for market portfolio.

Daily data is not used for analysis as it tends to contain too many noises, which may distort the underlying pattern. The weekly adjusted closing prices from Jan 6, 2014 to Oct 22, 2018 are downloaded from Yahoo Finance, from which log returns are calculated according to (4.1) and (4.3). Risk-free simple returns are retrieved from Dr. Kenneth R. French's research website[1]. After that, risk-free log returns are computed by (4.2). Finally, subtract risk-free returns from stock/market index log returns to get excess returns, as in (4.4). We have 250 excess return observations for each stock and the market index.

5.4 CAPM under Different Market Regimes

The univariate Gaussian HMM in Section 5.2.1 is fitted with $2, 3, 4$ hidden states. A simple linear regression without regressors is also fitted to serve as the case with no hidden states.

[1] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

The information criteria of these models are compared. As one can see from the figure below, the AIC plot prefers the 3-state model (yet the AICs of the 2-state and the 3-state models are quite close, taking values of -1372.779 and -1378.294 respectively), while the BIC plot suggests the use of 2 states. Thus, the 2-state model is selected.

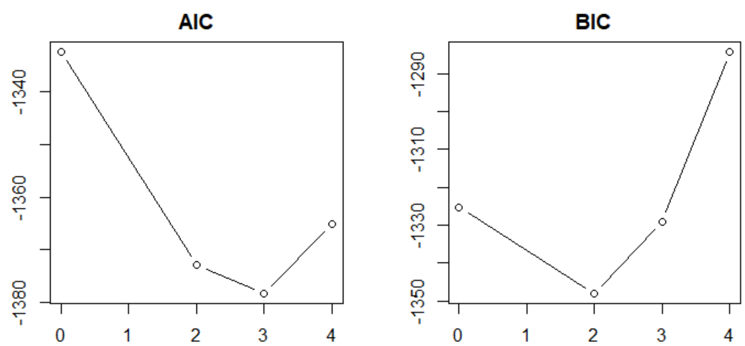

Figure 5.1: Information criteria of models fitted on S&P500 excess return

From Table 5.1 we can recognize a state with negative mean and high volatility, and a state with positive mean and low volatility. The first state can be interpreted as "bearish" market regime, while the second state stands for "bullish" market regime. This parallels with the existing studies [HL96][SN97][TSN89]. Both states are strongly persistent, as indicated by p_{11} and p_{22}. The duration for each state is $1/0.1 = 10$ weeks and $1/0.072 = 13.89$ weeks respectively. Also note that the model is able to capture the volatility clustering effect in the time series (see Figure 5.2).

$\mu_{m,1}$	$\mu_{m,2}$	$\sigma_{m,1}$	$\sigma_{m,2}$	p_{11}	p_{12}	p_{21}	p_{22}
-0.001	0.003	0.023	0.009	0.900	0.100	0.072	0.928

Table 5.1: Fitted parameters of the 2-state model on S&P500 excess return

Now inspect the CAPM relationship and the distribution of stock excess returns under each market regime (Table 5.2). Except for NKE, the betas under the bearish regime are lower or at least no bigger than the betas under the bullish regime. This asymmetry indicates that stocks are more sensitive to market growth than shrinkage. Market regimes do not seem to have a distinguishable effect on stock excess returns in terms of mean and variance, though.

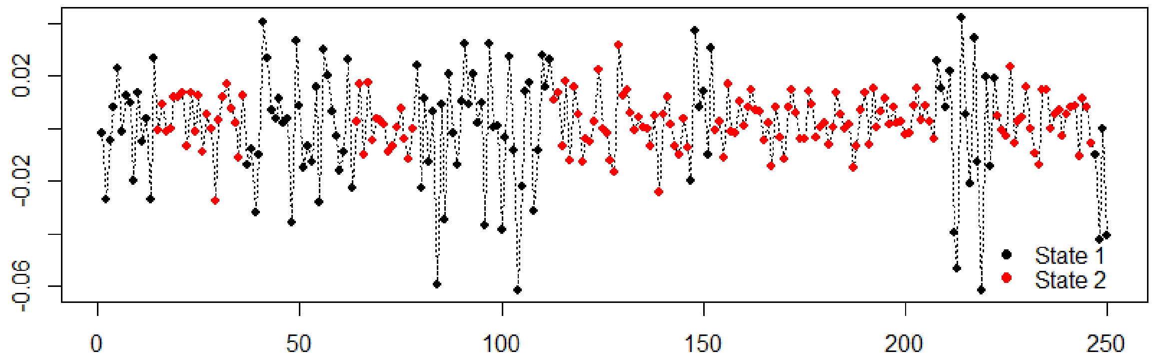

Figure 5.2: Weekly S&P500 excess returns with market regime marked

It is counterintuitive to notice that the excess returns are slightly lower for MCD and WBA under bullish regime when compared with bearish regime.

Stock	β_1	β_2	μ_1	μ_2	σ_1	σ_1
MCD	0.567	0.592	0.004	0.002	0.026	0.019
NKE	0.858	0.813	0.002	0.003	0.033	0.027
TWTR	0.904	1.751	-0.005	0.001	0.072	0.073
VZ	0.666	0.660	0.001	0.001	0.026	0.023
WBA	0.734	1.189	0.003	0.000	0.031	0.031

Table 5.2: Betas and distribution statistics under different market regimes

5.5 CAPM under Different Idiosyncratic Regimes

To examine the underlying assumption, 2,3 and 4-state univariate Gaussian HMMs (no covariates) are fitted on stock excess returns. Their AIC and BIC plots are supportive of the 2-state model (omitted in this thesis). Then, we calculate the correlation between the inferred hidden states of this model and the hidden states of the market calculated in the previous section. Results are displayed in the table below, suggesting negligible or week correlation.

MCD	NKE	TWTR	VZ	WBA
0.120	-0.081	-0.003	0.053	-0.010

Table 5.3: Correlation between market hidden states and stock hidden states

After the preliminary step, for each pair the univariate Gaussian HMM with covariate in Section 5.2.2 is fitted with $2, 3, 4$ hidden states. A simple linear regression on market excess returns is also fitted to serve as the case with no hidden states. Almost all the plots in Figure 5.3 are in favor of the 2-state model. Hence it is chosen. Figure 5.4 has demonstrated the power of idiosyncratic regimes to mark volatility clustering in individual stock excess returns.

Just like market excess return, which experiences high volatility and low means at the same time, stock excess returns show similar patterns in 5.4. The first state can be interpreted as "individually bearish" regime, and the second can be considered as "individually bullish" regime. The reason that NKE's β_1 is negative could be attributed to the fact that its mean is higher in state 1, an exception to our identified pattern. For other stocks, there is no consistent patterns in beta, yet their betas do vary under different regimes.

Figure 5.5 indicates that a stock's duration in the individually bullish state is longer compared with that in the individually bearish one.

Stock	β_1	β_2	μ_1	μ_2	σ_1	σ_2	$\mu_{m,1}$	$\mu_{m,2}$	$\sigma_{m,1}$	$\sigma_{m,2}$
MCD	0.776	0.409	0.000	0.004	0.041	0.014	0.000	0.002	0.021	0.015
NKE	-2.01	0.936	0.065	0.001	0.056	0.027	-0.007	0.002	0.011	0.017
TWTR	0.218	1.558	-0.016	0.002	0.131	0.043	0.000	0.002	0.017	0.017
VZ	0.668	0.653	0.002	0.001	0.038	0.017	-0.001	0.002	0.017	0.016
WBA	0.126	0.938	-0.006	0.002	0.073	0.027	-0.004	0.002	0.019	0.017

Table 5.4: Betas and distribution statistics under different idiosyncratic regimes

Stock	p_{11}	p_{12}	p_{21}	p_{22}	Dur_1	Dur_2
MCD	0.697	0.303	0.149	0.851	3.30	6.71
NKE	0.118	0.882	0.059	0.941	1.134	16.95
TWTR	0.346	0.654	0.365	0.635	1.53	2.74
VZ	0.693	0.307	0.229	0.771	3.26	4.37
WBA	0.465	0.535	0.105	0.895	1.87	9.52

Table 5.5: Transition probabilities between different idiosyncratic regimes

5.6 CAPM under Different Co-Regimes

For each pair, the bivariate Gaussian HMM is fitted with $2, 3, 4$ hidden states. Bivariate normals without hidden states are also fitted. Again we compare the information criteria of these models. In Figure 5.5, while AICs do not seem to give a consistent suggestion, for all pairs BICs are lowest when the number of states is 2. Thus, for the sake of parsimony we choose the 2-state model.

Stock	β_1	β_2	μ_1	μ_2	σ_1	σ_2	$\mu_{m,1}$	$\mu_{m,2}$	$\sigma_{m,1}$	$\sigma_{m,2}$
MCD	0.670	0.377	-0.002	0.004	0.041	0.015	-0.008	0.003	0.029	0.011
NKE	0.944	0.158	-0.002	0.007	0.030	0.028	0.000	0.003	0.022	0.008
TWTR	0.663	1.374	-0.033	0.005	0.135	0.046	-0.013	0.005	0.025	0.012
VZ	0.688	0.597	0.002	0.001	0.025	0.024	0.000	0.003	0.023	0.008
WBA	0.483	0.909	-0.030	0.004	0.062	0.026	-0.027	0.004	0.026	0.013

Table 5.6: Betas and distribution statistics under different co-regimes

Stock	p_{11}	p_{12}	p_{21}	p_{22}	Dur_1	Dur_2
MCD	0.582	0.418	0.159	0.841	2.40	6.29
NKE	0.881	0.119	0.115	0.885	8.40	8.70
TWTR	0.248	0.752	0.286	0.714	1.33	3.50
VZ	0.897	0.103	0.088	0.912	9.71	11.36
WBA	0.286	0.714	0.131	0.869	1.40	7.63

Table 5.7: Transition probabilities between different co-regimes

Now take a look at Table 5.6 and 5.7. The means for stock excess returns and market excess returns are non-positive in state 1 except for VZ, while the means are positive in state

2. Speaking of volatilities, those in state 1 are considerably higher than those in state 2. This motivates us to regard the first state as "co-bearish" and the second one as "co-bullish". For each pair, co-regimes is able to account for volatility clustering in the market, but works poorly for the phenomenon in excess returns of certain stocks (NKE & VZ), as it is shown in Figure 5.6.

A similar study by Fridman[Fri94] on three oil stocks argues that when the volatility is high, the betas are consistently larger while the self-transition probability p_{11}'s are lower. In our case, however, betas do not seem to have a consistent pattern across the two co-regimes. A stock and the market could move in the same direction, but the sensitivity might be determined by both market and firm-specific factors. However, we can at least argue that for each stock the sensitivity is different under the two co-regimes. Fridman's conclusion about persistency does hold in our case, implying that stocks and market tend to stay longer period in co-bullish regime.

For investors, an ideal investment is those with long duration in co-bullish and short duration in co-bearish. A natural choice is MCD and WBA. VZ do have remarkably long duration in the co-bullish state, but its equally length stay in the co-bearish state could be a nightmare for investors, resulting in continuous losses.

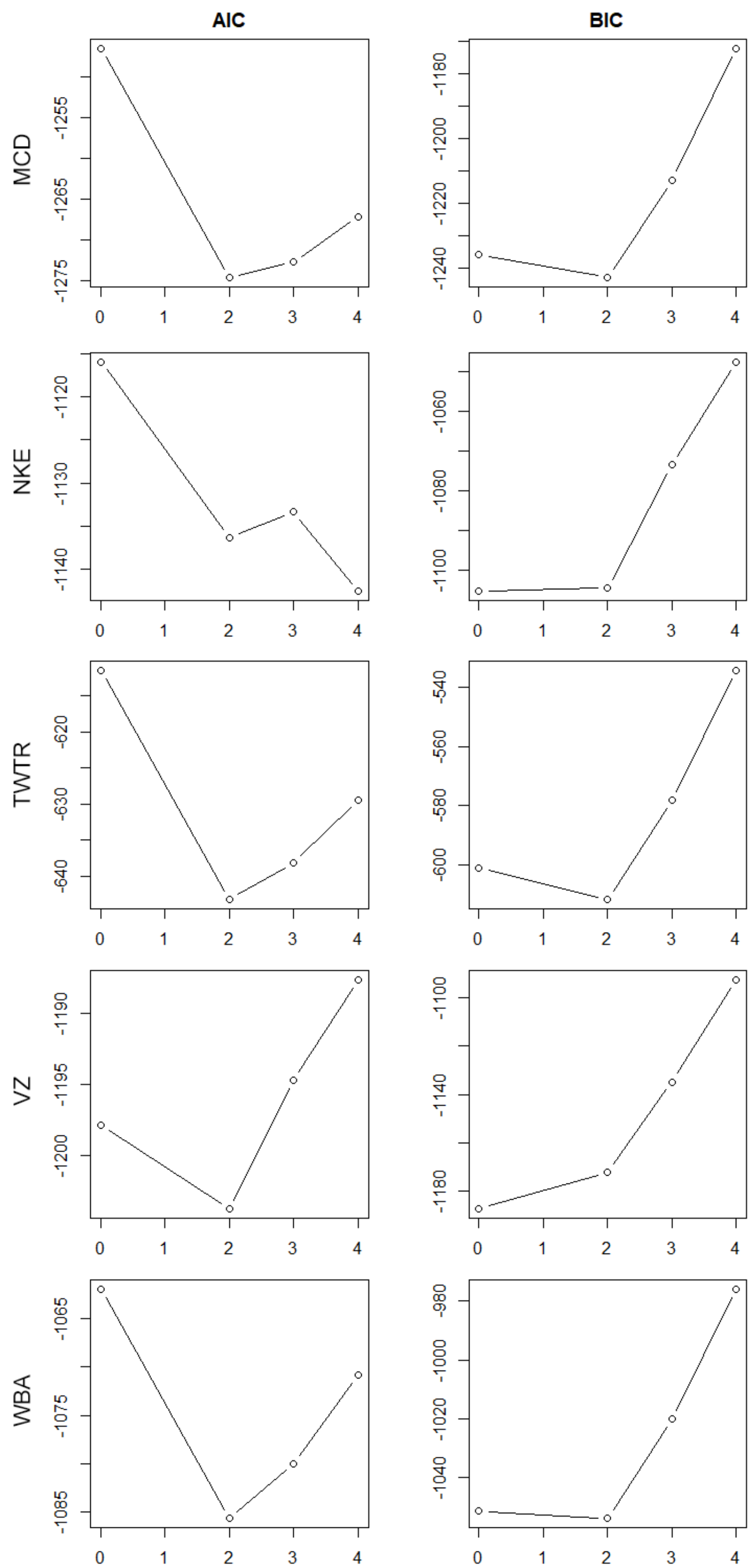

Figure 5.3: Information criteria of fitted univariate HMMs with covariate on each pair

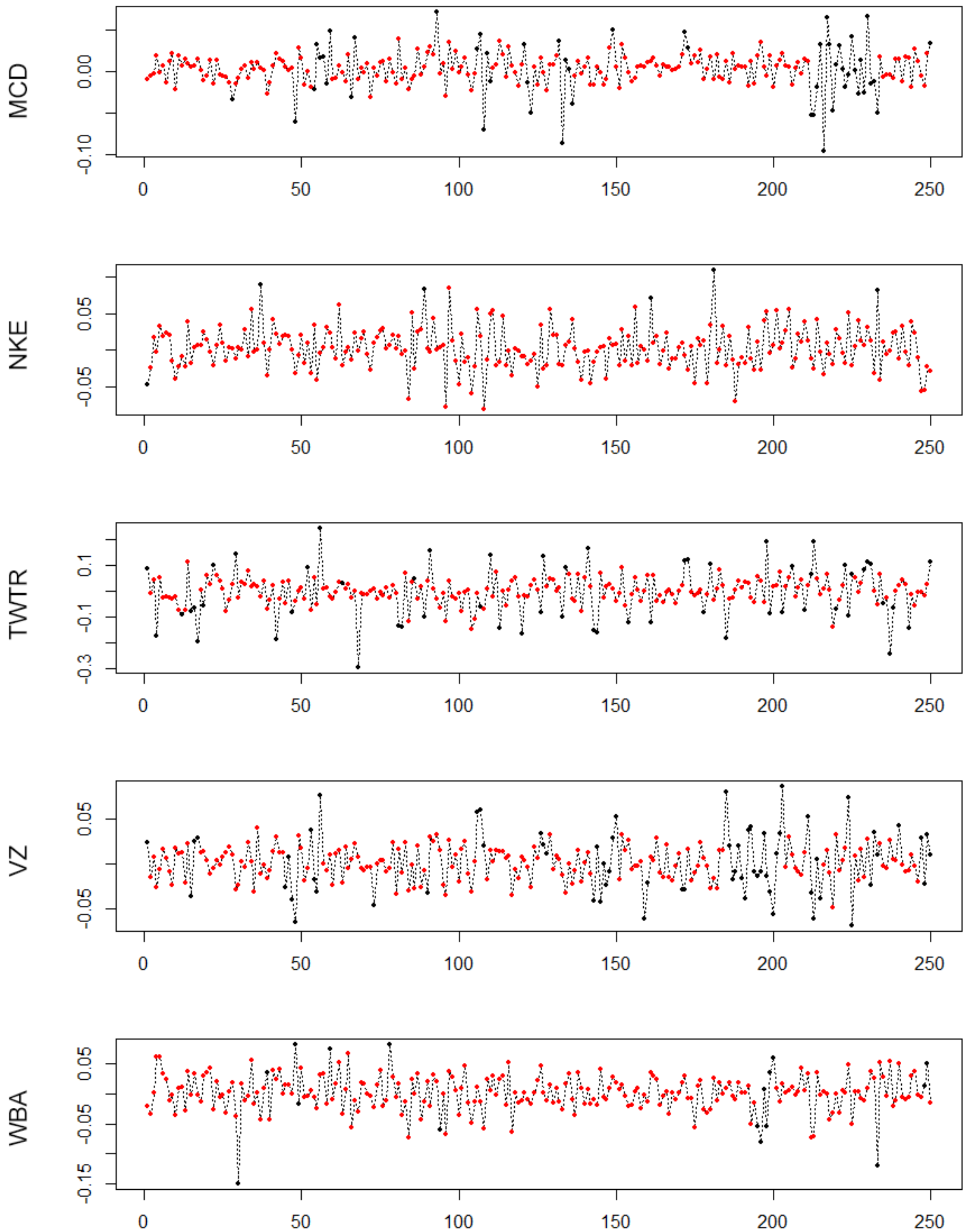

Figure 5.4: Weekly excess returns for each stock with idiosyncratic regime marked

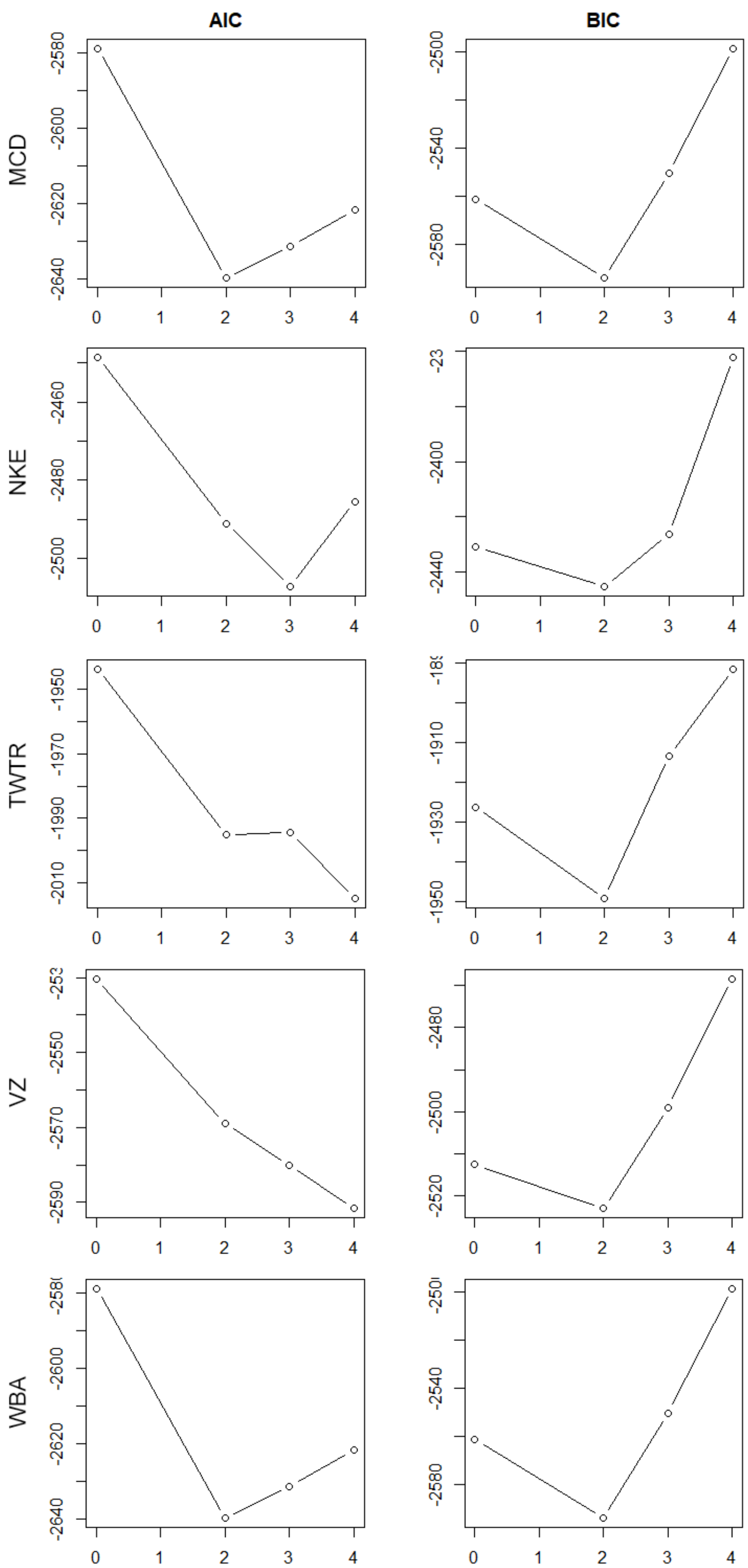

Figure 5.5: Information criteria of fitted bivariate HMMs on each pair

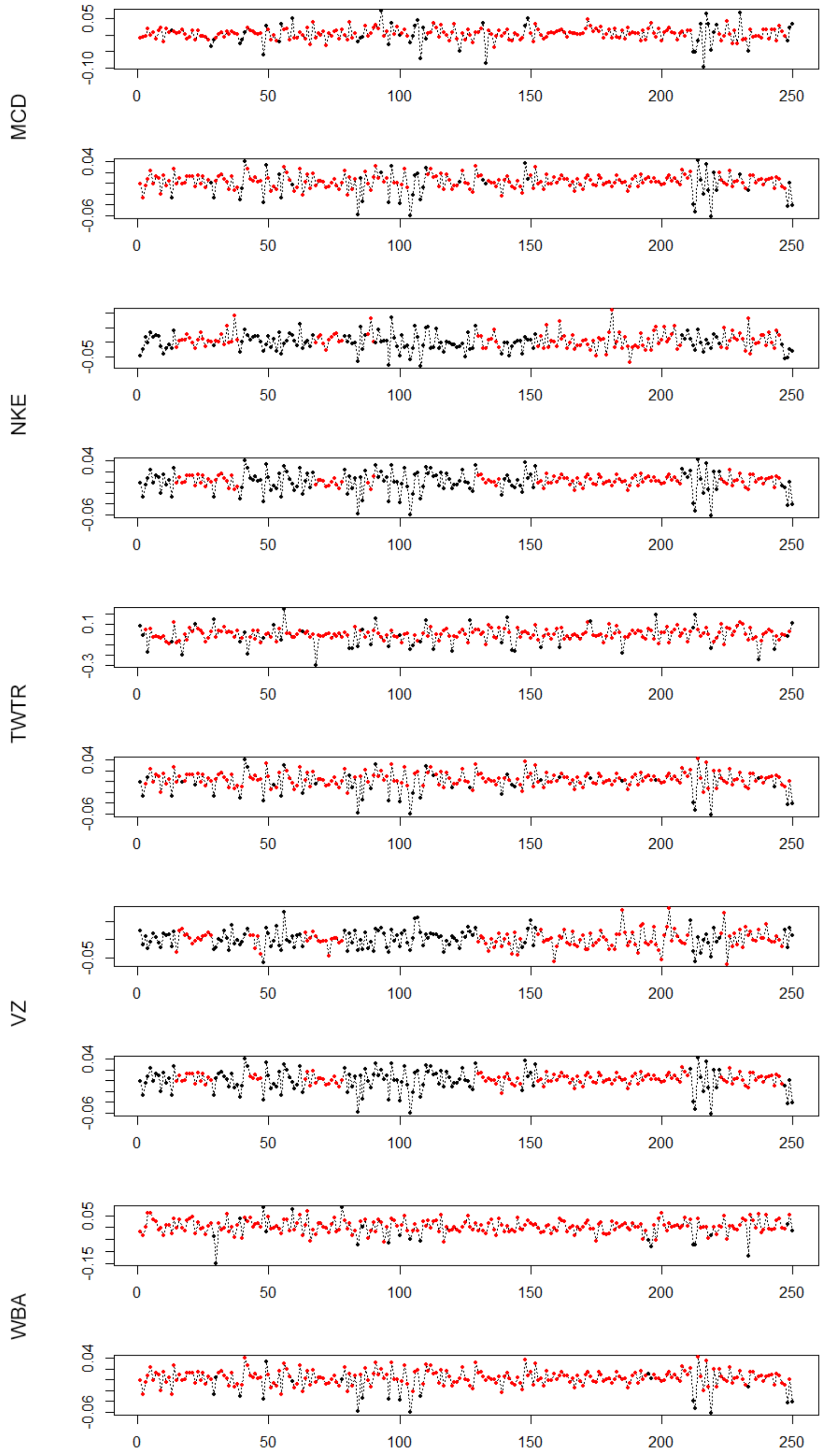

Figure 5.6: Weekly excess returns for each pair with co-regimes marked

CHAPTER 6

Discussion

In this thesis we have proposed two Gaussian hidden Markov models: univariate HMMs with covariate and bivariate HMMs. The methods of parameter estimation, hidden states decoding are also demonstrated, followed by simulation studies to show that the procedures are effective.

Proposed models are then employed to investigate the return-beta relationship of the capital asset pricing model under 3 types of regimes. Results show that betas are larger under bullish market regime compared to bearish. Although no consistent patterns in beta are discovered under different idiosyncratic regimes and co-regimes, for each stock the betas do seem to vary considerably across regimes. The duration in each regime as well as the captured volatility clustering effect shed light on investor's strategies.

For future work, we may consider examine the relationship between the inferred market, idiosyncratic and co-regimes. Instead of picking stocks from different industries, using sector data may provide more insights in the patterns of beta. Our model can also be extended to the well -established Fama French three factor model to see how the loadings of the factors vary across states. Normality assumption of returns can be lifted to take skewness and fat tails into consideration.

APPENDIX A

Appendix

A.1 Univariate Gaussian HMMs with Covariate: The M-step

We start from the objective function

$$Q\left(\theta \middle| \theta^{(m)}\right) = \sum_{j=1}^{N}\sum_{t=1}^{n} u_t^{(m)}(j)\left(-\frac{1}{2}\log\sigma_j^2 - \frac{\left(Y_t-\theta_{0,j}-\theta_{1,j}X_t\right)^2}{2\sigma_j^2}\right) + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{n} w_t^{(m)}(i,j)\log a_{ij}.$$

Easy to see that the first term contains γ only, while the second term contains a_{ij}'s only. To estimate $\gamma_j = \left(\theta_{0,j}, \theta_{1,j}, \sigma_j^2\right)$, differentiate the first term w.r.t. $\theta_{0,j}$ and $\theta_{1,j}$ and set it zero,

$$\begin{cases} -\frac{1}{2\sigma_j^2}\sum_{t=1}^{n} u_t^{(m)}(j)\times(-2X_t)(Y_t-\theta_{0,j}-\theta_{1,j}X_t) & = 0 \\ -\frac{1}{2\sigma_j^2}\sum_{t=1}^{n} u_t^{(m)}(j)\times(-2)(Y_t-\theta_{0,j}-\theta_{1,j}X_t) & = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \sum_{t=1}^{n} u_t^{(m)}(j)X_tY_t - \theta_{0,j}\sum_{t=1}^{n} u_t^{(m)}(j)X_t - \theta_{1,j}\sum_{t=1}^{n} u_t^{(m)}(j)X_t^2 & = 0 \\ \sum_{t=1}^{n} u_t^{(m)}(j)Y_t - \theta_{0,j}\sum_{t=1}^{n} u_t^{(m)}(j) - \theta_{1,j}\sum_{t=1}^{n} u_t^{(m)}(j)X_t & = 0. \end{cases}$$

Let $n_j = \sum_{t=1}^{n} u_t^{(m)}(j)$, $\overline{X}_j = \frac{\sum_{t=1}^{n} u_t^{(m)}(j)X_t}{n_j}$, $\overline{Y}_j = \frac{\sum_{t=1}^{n} u_t^{(m)}(j)Y_t}{n_j}$, then

$$\begin{cases} \sum_{t=1}^{n} u_t^{(m)}(j)X_tY_t - \theta_{0,j}n_j\overline{X}_j - \theta_{1,j}\sum_{t=1}^{n} u_t^{(m)}(j)X_t^2 & = 0 \\ n_j\overline{Y}_j - \theta_{0,j}n_j - \theta_{1,j}n_j\overline{X}_j & = 0. \end{cases}$$

Multiply the second equation by $\overline{X}_j$ and subtract it from the first equation, and divide the second equation by n_j,

$$\begin{cases} \sum_{t=1}^{n} u_t^{(m)}(j)X_tY_t - n_j^{(m)}\overline{X}_j\overline{Y}_j - \theta_{1,j}\left(\sum_{t=1}^{n} u_t^{(m)}(j)X_t^2 - n_j\left(\overline{X}_j\right)^2\right) & = 0 \\ \overline{Y}_j - \theta_{0,j} - \theta_{1,j}\overline{X}_j & = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \theta_{1,j}^{(m+1)} & = \frac{\sum_{t=1}^{n} u_t^{(m)}(j) X_t Y_t - n_j \overline{X}_j \overline{Y}_j}{\sum_{t=1}^{n} u_t^{(m)}(j) X_t^2 - n_j \left(\overline{X}_j\right)^2} \\ \theta_{0,j}^{(m+1)} & = \overline{Y}_j - \theta_{1,j}^{(m+1)} \overline{X}_j. \end{cases}$$

For the variance, take derivative w.r.t. σ_j^2,

$$\sum_{t=1}^{n} u_t^{(m)}(j) \left(-\frac{1}{2\sigma_j^2} + \frac{(Y_t - \theta_{0,j} - \theta_{1,j} X_t)^2}{2\sigma_j^4} \right) = 0$$

$$\sum_{t=1}^{n} u_t^{(m)}(j) \left(Y_t - \theta_{0,j} - \theta_{1,j} X_t\right)^2 - \sigma_j^2 n_j = 0$$

$$\Rightarrow \left(\sigma_j^2\right)^{(m+1)} = \frac{\sum_{t=1}^{n} u_t^{(m)}(j) \left(Y_t - \theta_{0,j}^{(m+1)} - \theta_{1,j}^{(m+1)} X_t\right)^2}{n_j}.$$

For a given i, to estimate $a_{ij} \forall j$ is equivalent to maximizing $\sum_{j=1}^{N} \sum_{t=2}^{n} w_t^{(m)}(i,j) \log a_{ij}$ subject to the constraint $\sum_{j=1}^{N} a_{ij} = 1$. Since this is an analogy of the log-likelihood of a multinomial distribution with cell probabilities $(a_{i1}, \ldots, a_{iN})$ and cell counts $\left(\sum_{t=2}^{n} w_t^{(m)}(i,1), \ldots, \sum_{t=2}^{n} w_t^{(m)}(i,N)\right)$, we have

$$a_{ij}^{(m+1)} = \frac{\sum_{t=2}^{n} w_t^{(m)}(i,j)}{\sum_{k=1}^{N} \sum_{t=2}^{n} w_t^{(m)}(i,k)}.$$

A.2 Bivariate Gaussian HMMs: The M-step

We start from the objective function

$$Q\left(\theta \middle| \theta^{(m)}\right) = \sum_{j=1}^{N} \sum_{t=1}^{n} u_t^{(m)}(j) \left(-\frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (Y_t - \mu_j)^\intercal \Sigma_j^{-1} (Y_t - \mu_j) \right)$$
$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{n} w_t^{(m)}(i,j) \log a_{ij}.$$

Easy to see that the first term contains γ only, while the second term contains a_{ij}'s only. Differentiate the first term w.r.t. μ_j and set it zero,

$$\sum_{t=1}^{n} u_t^{(m)}(j) \left(-\frac{1}{2} \left(-2\Sigma_j^{-1} Y_t + 2\Sigma_j^{-1} \mu_j\right) \right) = 0$$

$$\Rightarrow \Sigma_j^{-1} \sum_{t=1}^{n} u_t^{(m)}(j) Y_t - \Sigma_j^{-1} \mu_j \sum_{t=1}^{n} u_t^{(m)}(j) = 0.$$

Let $n_j = \sum_{t=1}^{n} u_t^{(m)}(j), \overline{Y}_j = \frac{\sum_{t=1}^{n} u_t^{(m)}(j) Y_t}{n_j}$, and multiply the equation by Σ_j,

$$\sum_{t=1}^{n} u_t^{(m)}(j) Y_t - n_j \mu_j = 0 \Rightarrow \mu_j^{(m+1)} = \overline{Y}_j.$$

To estimate the covariance matrix, rewrite the objective function,

$$Q\left(\theta \middle| \theta^{(m)}\right) = \sum_{j=1}^{N} \sum_{t=1}^{n} u_t^{(m)}(j) \left(\frac{1}{2} \log |\Sigma_j^{-1}| - \frac{1}{2} tr \left[\Sigma_j^{-1} \left(Y_t - \mu_j\right) \left(Y_t - \mu_j\right)^\intercal \right] \right)$$
$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{n} w_t^{(m)}(i, j) \log a_{ij}.$$

Differentiate w.r.t. Σ_j^{-1},

$$\sum_{t=1}^{n} u_t^{(m)}(j) \left(\frac{1}{2} \Sigma_j^\intercal - \frac{1}{2} \left(Y_t - \mu_j\right) \left(Y_t - \mu_j\right)^\intercal \right) = 0$$
$$n_j \Sigma_j - \sum_{t=1}^{n} u_t^{(m)}(j) \left(Y_t - \mu_j\right) \left(Y_t - \mu_j\right)^\intercal = 0$$
$$\Rightarrow \Sigma_j^{(m+1)} = \frac{\sum_{t=1}^{n} u_t^{(m)}(j) \left(Y_t - \mu_j\right) \left(Y_t - \mu_j\right)^\intercal}{n_j}.$$

For the estimation of $a_i j$'s, see Appendix A.1.

REFERENCES

[AF09] Tobias Adrian and Francesco Franzoni. "Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM." *Journal of Empirical Finance*, **16**(4):537 – 556, 2009.

[Aka74] H. Akaike. "A new look at the statistical model identification." *IEEE Transactions on Automatic Control*, **19**(6):716–723, December 1974.

[BLD01] Y. Bengio, V. . Lauzon, and R. Ducharme. "Experiments on the application of IOHMMs to model financial returns series." *IEEE Transactions on Neural Networks*, **12**(1):113–123, Jan 2001.

[BM04] M. Bicego and V. Murino. "Investigating hidden Markov models' capabilities in 2D shape classification." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**(2):281–286, Feb 2004.

[BP66] Leonard E. Baum and Ted Petrie. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains." *The Annals of Mathematical Statistics*, **37**(6):1554–1563, 1966.

[BPS70] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." *The Annals of Mathematical Statistics*, **41**(1):164–171, 1970.

[BS73] Fischer Black and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, **81**(3):637–654, 1973.

[Che09] Shiu-Sheng Chen. "Predicting the bear stock market: Macroeconomic variables as leading indicators." *Journal of Banking & Finance*, **33**(2):211 – 223, 2009.

[Con01] R. Cont. "Empirical properties of asset returns: stylized facts and statistical issues." *Quantitative Finance*, **1**(2):223–236, 2001.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1):1–38, 1977.

[Fri94] Moshe Fridman. "A two state capital asset pricing mode." 1994.

[GF99] Nicolaas Groenewold and Patricia Fraser. "Time-varying estimates of CAPM betas." *Mathematics and Computers in Simulation*, **48**(4):531 – 539, 1999.

[HL96] James D. Hamilton and Gang Lin. "Stock market volatility and the business cycle." *Journal of Applied Econometrics*, **11**(5):573–593, 1996.

[HN05] M. R. Hassan and B. Nath. "Stock market forecasting using hidden Markov model: a new approach." In *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, pp. 192–196, Sept 2005.

[KBM94] Anders Krogh, Michael Brown, I.Saira Mian, Kimmen Sjölander, and David Haussler. "Hidden Markov Models in Computational Biology: Applications to Protein Modeling." *Journal of Molecular Biology*, **235**(5):1501 – 1531, 1994.

[KHB89] Amlan Kundu, Yang He, and Paramvir Bahl. "Recognition of handwritten word: First and second order hidden Markov model based approach." *Pattern Recognition*, **22**(3):283 – 297, 1989.

[Lin69] John Lintner. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets: A Reply." *The Review of Economics and Statistics*, **51**(2):222–224, 1969.

[OH11] Jared O'Connell and Søren Højsgaard. "Hidden Semi Markov Models for Multiple Observation Sequences: The mhsmm Package for R." *Journal of Statistical Software*, **39**(4):1–22, 2011.

[Rab89] L. R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE*, **77**(2):257–286, Feb 1989.

[Rey99] Mario G. Reyes. "Size, time-varying beta, and conditional heteroscedasticity in UK stock returns." *Review of Financial Economics*, **8**(1):1 – 10, 1999.

[Sch78] Gideon Schwarz. "Estimating the Dimension of a Model." *Ann. Statist.*, **6**(2):461–464, 03 1978.

[SGH95] M. Schenkel, I. Guyon, and D. Henderson. "On-line cursive script recognition using time-delay neural networks and hidden Markov models." *Machine Vision and Applications*, **8**(4):215–223, Jul 1995.

[Sha64] William F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions Of Risk." *The Journal of Finance*, **19**(3):425–442, 1964.

[SN97] Huntley Schaller and Simon Van Norden. "Regime switching in stock market returns." *Applied Financial Economics*, **7**(2):177–191, 1997.

[TSN89] Christopher M. Turner, Richard Startz, and Charles R. Nelson. "A Markov model of heteroskedasticity, risk, and learning in the stock market." *Journal of Financial Economics*, **25**(1):3 – 22, 1989.

[Vit67] A. Viterbi. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm." *IEEE Transactions on Information Theory*, **13**(2):260–269, April 1967.

[VS10] Ingmar Visser and Maarten Speekenbrink. "depmixS4: An R Package for Hidden Markov Models." *Journal of Statistical Software*, **36**(7):1–21, 2010.

[WCM11] Chkili Walid, Aloui Chaker, Omar Masood, and John Fry. "Stock market volatility and exchange rates in emerging countries: A Markov-state switching approach." *Emerging Markets Review*, **12**(3):272 – 292, 2011.