VOLATILITY ESTIMATION AND PRICE PREDICTION USING A HIDDEN MARKOV MODEL WITH EMPIRICAL STUDY

A Dissertation
presented to
the Faculty of the Graduate School
University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree

Doctor of Philosophy

by
PEI YIN

Dr. Allanus H. Tsoi, Dissertation Supervisor

AUGUST 2007

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

VOLATILITY ESTIMATION AND PRICE PREDICTION
USING A HIDDEN MARKOV MODEL
WITH EMPIRICAL STUDY

Presented by Pei Yin

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance.

Professor Allanus Tsoi

Professor Michael Taksar

Professor Carlo Morpurgo

Professor Carmen Chicone

Professor Tony Sun

Acknowledgments

My first thank goes to the Department of Mathematics, University of Missouri at Columbia for its financial support and for facilitating my work.

Next I wish to express my sincere gratitude to my teacher, Professor Allanus Tsoi, for introducing me to the subject and for his continuous advice, support and encouragement during this work.

I am also grateful to Professors Michael Taksar, Professors Carlo Morpurgo, Professors Carmen Chicone, Professor Zhenbo Qin and Dr. Eric Zeng for their interest and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction and Preliminaries

In this chapter, we first give a brief introduction of this dissertation and then provide some preliminary background of our study. We will outline the basic concepts of hidden Markov models, Kalman Filter algorithm, EM algorithm, and Monte Carlo simulation method. Finally, we shall question the Efficient Market Hypothesis and claim the effectiveness of technique analysis as there indeed exist certain trends that financial time series follow to a large extent.

1.1 Introduction

A number of recent studies have sought to characterize the nature of financial market return process, which has always been described as a combination of drift and volatility. More models are established to focus on the volatility. Early studies simply assume constant volatility, while nowadays it is widely believed that volatility itself is volatile. In this work we suggest a hidden Markov model(HMM) where both drift and volatility are stochastic and they are driven by some sort of underlying economic forces which evolves as a finite-state, time-invariant Markov chain. Through stochastic filtering techniques in the same flavor of Kalman Filter algorithm we formulate EM estimates of the parameters by iterations in the same spirit as the EM algorithm. Then we apply the HMM estimates to a price model and develop the prediction formula. On an empirical level, we perform Monte Carlo simulation analysis and apply our model to 73 cautiously selected data sets of historical security prices. The results suggest great applicability of our HMM. Moreover, we compare HMM and the well established GARCH(1,1) with the same data sets, as far as the prediction performance is concerned, our results suggest that HMM outperforms GARCH(1,1). Lastly, we provide some further avenues for the future study.

1.2 Hidden Markov Models

A hidden Markov model (HMM) is characterized by the following elements:

1. $\{X\} = \{X_1, X_2, ...\}$: the unobserved Markov chain, called the signal sequence

2. $\{Y\} = \{Y_1, Y_2, ...\}$: the observation sequence

3. N: the dimension of the state space of the Markov chain

4. M: the dimension of the state space of the observations sequence

5. $S_X = \{s_1, s_2, ..., s_N\}$: the state space for the Markov chain

6. $O_Y = \{o_1, o_2, ..., o_M\}$: the state space for the observation sequence

7. $A_{N\times N} = (a_{ij})$ where $i, j = 1,, N$ the transition probability matrix: $a_{ij} = P(X_{n+1} = s_i|X_n = s_j), \forall n$

8. $B_{M\times N} = (b_{ij})$ where $i = 1,, M$ and $j = 1, N$ the conditional distribution of the observation given the signal: $b_{ij} = P(Y_n = o_i|X_n = s_j), \forall n$

9. $\Pi_{N\times 1} = (\pi_i)$ where $i = 1,, N$ the initial state distribution

Let us use a compact notation: $\lambda = \{A, B, \Pi\}$.

There are three basic problems of HMM when it is applied to the real world.

1. Given an observation sequence $\{Y\}$, and the model parameters $\lambda = \{A, B, \Pi\}$, how to compute $P(Y|\lambda)$. i.e. once the model is known, what is the probability of getting a specific observation sequence? This problem can be solved forward-backward algorithms (Rabiner 1989) .

2. Given the observation sequence $\{Y\}$, and the model specifications $\lambda = \{A, B, \Pi\}$, what is the underlying sequence $\{X\}$ that can best explain the outcome of the observation? This problem can be solved be Viterbi algorithms (Forney 1973).

3. Given the observation sequence $\{Y\}$ and the state space $\{S_X\}$, how do we estimate the model parameters $\lambda = \{A, B, \Pi\}$ which maximizes $P(Y|\lambda)$? This problem can be solved by Baum-Welch algorithms (Welch 2003).

In the world of finance, if we consider $\{Y\}$ as an observed financial time series, $\{X\}$ as some underlying economic forces, then the second problem is to find the hidden path of the economic forces, the third problem is to estimate the model parameters by using the observed information, the first problem is to predict the future by employing the model with the historical observation.

Formally, the following definitions are widely accepted in the literature of HMM. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X = \{X_t\}_{t \in T}$ be a stochastic process, T denotes the collection of positive integers.

Definition 1.2.1. *$S_X = \{s_1, s_2, ...\}$ is called the state space of X if for $\forall t \in T$, $\forall \omega \in \Omega$, $X_t(\omega) \in S_X$. Moreover, $X_t = s_j$ where $s_j \in S_X$ means the process is in state s_j at time t.*

Definition 1.2.2. *X is called a homogeneous Markov chain with state space S_X if: $P(X_{t+1} = s_i | X_t = s_j, ..., X_2, X_1) = P(X_{t+1} = s_i | X_t = s_j) = a_{ij}$*

for $\forall s_j, s_i \in S_X$ and $\forall t \in T$.

The probabilities a_{ij} are called the transition probabilities.

$A = \{a_{ij}\}_{i,j=1,2,...,N}$ is called the transition probability matrix (or transition matrix) of the Markov chain X.

Note: throughout this research we only discuss the homogeneous Markov Chains with finite state space. i.e. $P(X_{t+1} = s_i|X_t = s_j)$ does not depend on t, and $S_X = \{s_1, s_2, ...s_N\}$ is finite.

Definition 1.2.3. *A pair of stochastic processes $\{X, Y\}$ where $X = \{X_t\}_{t\in T}$ and $Y = \{Y_t\}_{t\in T}$ is said to be a hidden Markov chain, if X is a Markov chain which can not be observed directly, and $Y_t = f(X_t, \omega_t)$, where f is a deterministic Borel measurable function and $\{\omega_t\}_{t\in T}$ is a sequence of i.i.d. random variables that are independent of X. Here the process Y is called the observation process.*

1.3 Kalman Filter Algorithm

To have a general idea of what filtering, smoothing and prediction are, we consider the following example. We are given a signal process $\{x_t\}$ and a noise process $\{u_t\}$, neither of them is observable. However we can observe a sequence $y_t = f(x_t, u_t)$. Suppose we have observed the values of $\{y_1, y_2, ..., y_T\}$, what can we infer from this knowledge in regard to the underlying process $\{x_1, x_2, ..., x_t\}$?

if $t = T$, this is called a filtering problem,

if $t < T$, this is called a smoothing problem,

if $t > T$, this is called a prediction problem.

The following system is suggested by Kalman(1960):

$$x_t = Fx_{t-1} + Gu_t + \epsilon_t$$

$$y_t = Hx_t + \eta_t$$

$$\epsilon_t \sim N(0, R_t) \quad i.i.d \qquad \eta_t \sim N(0, Q_t) \quad i.i.d$$

Thus,

$$(y_t|x_t, H, Q_t) \sim N(Hx_t, Q_t)$$

$$(x_t|x_{t-1}, F, G, R_t) \sim N(Fx_{t-1} + Gu_t, R_t)$$

The underlying variable x_t is an $n \times 1$ vector, F is an $n \times n$ matrix. u_t is an $m \times 1$ vector, and $m \leq n$, G is an $n \times m$ matrix. We know u_t and G. The observation y_t is a $p \times 1$ vector, H is a $p \times n$ matrix. Let $\mathcal{Y}_t$ be the $\sigma - field$ generated by $\{y_1, y_2, ..., y_t\}$. The noise: ϵ_t is $n \times 1$ and η_t is $p \times 1$, therefore, R_t is $n \times n$, and Q_t is $p \times p$. We take the following notation: (G. Welch and G. Bishop 2006)

$$x_{t|t} = E(x_t|\mathcal{Y}_t)$$

$$x_{t|t-1} = E(x_t|\mathcal{Y}_{t-1})$$

$$\Sigma_{t|t} = E((x_t - x_{t|t})(x_t - x_{t|t})'|\mathcal{Y}_t)$$

$$\Sigma_{t|t-1} = E((x_t - x_{t|t-1})(x_t - x_{t|t-1})'|\mathcal{Y}_{t-1})$$

By taking expectations on the original system we can have the following equations:

(1) $x_{t|t-1} = Fx_{t-1|t-1} + Gu_t$

(2) $\Sigma_{t|t-1} = R_t + F\Sigma_{t-1|t-1}F'$

(3) Define J_t as: $J_t = \Sigma_{t|t-1}H'(H\Sigma_{t|t-1}H' + Q_t)^{-1}$ so the dimension of J_t is $n \times p$

(4) $x_{t|t} = x_{t|t-1} + J_t(y_t - Hx_{t|t-1})$

(5) $\Sigma_{t|t} = (I - J_tH)\Sigma_{t|t-1}$ where I is the $n \times n$ unit matrix.

Given F, G, H, Q, R and suppose we know the initial conditions: $x_{0|0}$ and $\Sigma_{0|0}$, the Kalman Filter algorithm works as the following:

step 1. Obtain $x_{1|0}$ and $\Sigma_{1|0}$ from equations (1) and (2).

step 2. Obtain J_1 from (3)

step 3. Obtain $x_{1|1}$ and $\Sigma_{1|1}$ from equations (4) and (5).

step 4. Now we know $x_{1|1}$ and $\Sigma_{1|1}$, through the same procedures as in steps 1-3, we obtain $x_{2|2}$ and $\Sigma_{2|2}$.

....

Therefore, we get the filters $x_{t|t}$, $\Sigma_{t|t}$ and the predictions $x_{t+1|t}$, $\Sigma_{t+1|t}$, $t = 1, 2, ...$

As for the smoothers after obtaining T observations, $x_{t-1|T}$, $\Sigma_{t-1|T}$, next we run a backwards algorithm for $t = T, T-1, ..., 1$ to get the following equations:

(6) $K_{t-1} = \Sigma_{t-1|t-1} F' \Sigma_{t|t-1}$ so the dimension of K_t is $n \times n$

(7) $x_{t-1|T} = x_{t-1|t-1} + K_{t-1}(x_{t|T} - x_{t|t-1})$

(8) $\Sigma_{t-1|T} = \Sigma_{t-1|t-1} + K_{t-1}(\Sigma_{t|T} - \Sigma_{t|t-1})K'_{t-1}$

Take the initial conditions from the filters $x_{T|T}$, $\Sigma_{T|T}$ and follow the steps below we shall be able to get the smoothers $x_{t|T}$ and $\Sigma_{t|T}$ where $t = T-1, T-2, ..., 0$.

step 1. Incorporating the filter $\Sigma_{T-1|T-1}$ and the prediction $\Sigma_{T|T-1}$ into equation (6), we get K_{T-1}.

step 2. Obtain $x_{T-1|T}$ by (7).

step 3. Obtain $\Sigma_{T-1|T}$ by (8).

step 4. Repeat the above three steps to get $X_{T-2|T}$, $\Sigma_{T-2|T}$, ... , $X_{0|T}$, $\Sigma_{0|T}$.

1.4 EM Algorithm

Let us first review what is a maximum likelihood problem. Suppose we have a probability distribution function $p(z|\Theta)$ that is determined by a set of parameters Θ, and N independent random draw from this distribution to form a data set $\mathcal{Z} = \{z_1, z_2, ...z_N\}$. The resulting joint probability density function for this sample is $p(\mathcal{Z}|\Theta) = \prod_{i=1}^{N} p(z_i|\Theta) = L(\Theta|\mathcal{Z})$ The function $L(\Theta|\mathcal{Z})$ is called the likelihood function. A maximizing likelihood problem is to estimate Θ under which the observed data are most likely, i.e. the likelihood function has the greatest value. To achieve this goal, we let the derivative of $L(\Theta|\mathcal{Z})$ with respect to each component of Θ to be zero. In practice, the log likelihood function is more often used for easier computation.

However if we are not able to obtain such a complete data set $\mathcal{Z}$, but rather $\mathcal{Z}$ is composed by two parts: $\mathcal{Z} = \{\mathcal{X}, \mathcal{Y}\}$, where $\mathcal{Y}$ are the values of the observed variable, $\mathcal{X}$ are the missing data (can be either actual missing measurements or hidden variables). In such a case, the likelihood function $p(\mathcal{Z}|\Theta) = p(\mathcal{X}, \mathcal{Y}|\Theta) = L(\Theta|\mathcal{X}, \mathcal{Y})$ is actually a random variable since the missing information $\mathcal{X}$ is unknown, random, and presumably governed by an underlying distribution. Therefore, we can not maximize it by choosing the best Θ, i.e. the regular maximum likelihood method will not give the estimate of Θ. An alternative way is to maximize the expected likelihood function, and this is the spirit of the EM algorithm.

The EM (Expectation-Maximization) algorithm will iteratively improve an initial estimate Θ_0 and construct new estimates Θ_1,..., Θ_k... in this way:

To derive Θ_{k+1} from Θ_k, we maximize the following function $Q(\Theta)$ with respect to Θ.

$$Q(\Theta) = E_x\left(ln\ p(\mathcal{X}, \mathcal{Y}|\Theta)\middle|\mathcal{Y}\right) = \int_{-\infty}^{\infty} ln\ p(\mathcal{X}, \mathcal{Y}|\Theta) \cdot p(\mathcal{X}|\mathcal{Y}, \Theta_k)dx$$

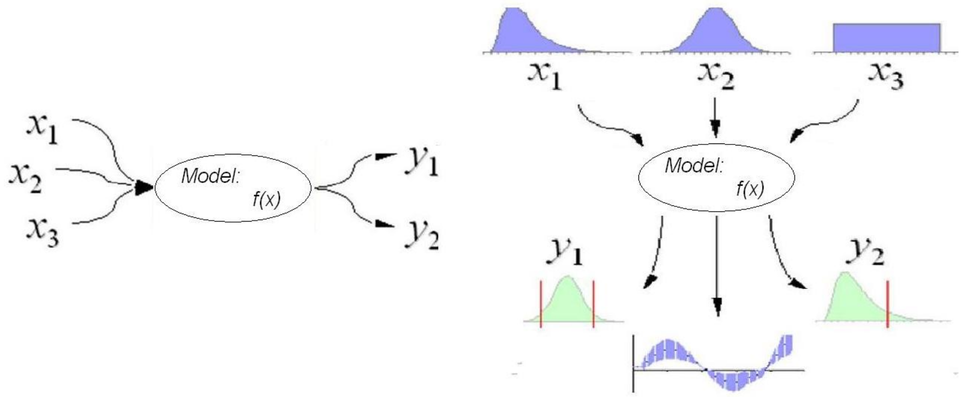
There are two major steps repeated in the EM algorithm: the evaluation of the above expectation is called the "E-step", to maximize this expectation is called the "M-step". Each iteration is guaranteed to increase the log likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function, depending on starting values. However, there is no guarantee that the sequence converges to a maximum likelihood estimator.

Note that Expectation-Maximization is a concept for a class of related algorithms, not a specific one. The EM estimates we shall develop in our primary model is in line with the concept of Expectation-Maximization, but we modify the way of iteration in the above EM algorithm.

1.5 Monte Carlo Simulation

Monte Carlo simulation method was originated as a reference to casino games. Its use of randomness in the process are analogous to the activities conducted at a casino. Stanislaw Marcin Ulam, an early pioneer in this field tells in his autobiography “Adventures of a Mathematician” that the method was named in honor of his uncle, who was a gambler.

Monte Carlo simulation method is a technique that involves using simulated random numbers as input to estimate stochastic models. It is useful especially for modeling phenomena with significant uncertainty in inputs, such as estimating the risk in financial time series. Therefore, what is a stochastic model? While creating a model, we usually have a certain number of data as inputs and a few equations that use those inputs to give you a set of outputs. This type of model is called a deterministic model. For example, to model a risk free investment, let the principle, annual interest rate, length of a financial period, number of years be the inputs, you will then have a certain future value as the output no matter how many times you re-calculate. However, if we replace some of the inputs in a deterministic model by random variables, it turns into a stochastic model. For instance, to model the stock returns, we let the risk element evolve as a random draw from a normal distribution. Even if the mean and variance of such distribution is known, when you re-calculate the future value of an investment, each time you may get a different result. The following diagram shows the difference between a deterministic model and a stochastic model.



a deterministic model a stochastic model

The Monte Carlo simulation method is categorized as a sampling method because the inputs are randomly generated from some probability distributions to simulate a sample from an actual population. So we try to choose a distribution for the inputs that most closely matches data we already have. It also bears mentioning that the use Monte Carlo method requires large amounts of random numbers, which has spurred the development of programmatic random number generators.

The use of Monte Carlo simulation method usually involves the following steps:(Gilks, Richardson, Spiegelhalter 1996; Gamerman 1997)

step 1: Generate a set of random inputs, $x^i = \{x_1^i, x_2^i, ..., x_n^i\}$.

step 2: Create a parametric model, $y = f(x_1, x_2, ..., x_n)$.

step 3: Evaluate $y^i = f(x^i)$ and store the results as $\{y^1, y^2, ...y^t\}$.

step 4: Analyze the results by comparing $\{y^1, y^2, ...y^t\}$ with the observed data and adjust the model specifications. What we mean by "analyze" depends on the application, typically they should pass a series of statistical tests.

1.6 The Paradox of Efficient Market Hypothesis

The concept of Efficient Market Hypothesis (EMH) was originally from Eugene Fama's influential PhD dissertation (1965) "Random Walks in Stock Market Prices", in which he persuasively made the argument that when information arrives at the financial market, it is immediately incorporated into the security prices without any delay.

> Eugene F. Fama (1965) *"An 'efficient' market is defined as a market where there are large numbers of rational, profit-maximizers actively competing, with each trying to predict future market values of individual securities, and where important current information is almost freely available to all participants. In an efficient market, competition among the many intelligent participants leads to a situation where, at any point in time, actual prices of individual securities already reflect the effects of information based both on events that have already occurred and on events which, as of now, the market expects to take place in the future. In other words, in an efficient market at any point in time the actual price of a security will be a good estimate of its intrinsic value."*

If market is indeed efficient, any technical analysis, the study of historical price processes in an attempt to predict the future price is of no use since the past prices are fully reflected in the current prices. In addition, any fundamental analysis, the study of company's asset value, financial statements, credit risk, etc., is of no use since any public information is fully reflected in the security prices.

As a result of the efficient market, the price movement is a "random walk" which will not follow any patterns. Actually, this "random walk" theory can be traced back to 1900 when a French mathematician Louis Bachelier asserted in his PhD dissertation “The Theory of Speculation” that “the mathematical expectation of the speculation is zero”. However his theory was too profound to be accepted by then and it finally became noticed after 1964 when the English translation was published.

EMH has been an issue of debate among both the market practitioners and academic researchers. For example, the investment media asserts that the market is not efficient as they make profit by supplying information to the investors. Similarly, active fund managers[1] are ambivalent toward EMH, otherwise their well paid job is nothing but speculation. However passive managers[2], who support EMH, argue with the fact that majority of the active managers in a given market will under perform an appropriate benchmark index[3] in the long run.

In academics, EMH is first of all challenged by economists who take human behavior and psychological factors as dominating elements in their equilibrium price models. Moreover, another group of proponents of EMH are the econometricians, statisticians and financial mathematicians. They have built numerous models under the assumption that the return processes are to a large extent predictable.

[1]Active fund management is an investment strategy of making specific portfolio selections in an attempt to beat the market.

[2]Passive fund management is the strategy of investing in broad sectors of the market without making any attempt to distinguish attractive or unattractive securities.

[3]For example, a small-cap stock fund is only comparable with indices of small-cap stocks, therefore indices such as S&P 600 or Russell 2000 could be appropriate benchmarks. Or, a growth stock fund can be compared with growth indices such as Vangard Growth Index.

As we quoted above, the realization of "Efficient Market" is based on the hypothesis that all market participants believe that the market is not efficient so they can use certain strategies to outperform the benchmarks. In another word, if every investor believes that the market is efficient, then the market will not be efficient. This is the paradox of Efficient Market Hypothesis that makes the hypothesis itself doubtful. In this thesis we shall discuss various models that capture the stylized facts about return processes and their applicability put EMH questionable .

Chapter 2

Constant Volatility

In this chapter we start with the first milestone in modeling the security price processes, the Geometric Brownian Motion Model, which has introduced the concept of drift and volatility in the return process. Then we talk about the role of volatility estimation in option pricing given that the underlying stock price evolves as the Geometric Brownian Motion.

2.1 Geometric Brownian Motion Model

The modeling for security price processes has been addressed for nearly a century. Especially in the recent two decades, with the explosion of mathematic finance, econometrics, financial engineering, there have been several well recognized theories regarding this issue. Moreover, with the emergence of computer based numerical experiments and historical data providers, most of the theories are challenged by and supported with empirical results. Nevertheless, the problem of security price processes is far from being solved. As Robert Almgren noted in an issue of The

American Mathematical Monthly *"Construct improved model for asset price motion is a subject of active research"*.

Among the numerous models, the first one with great recognition is the Brownian Motion Model (Bachelier, 1900). $dS = \mu dt + \sigma dz$, where z is a Wiener Process. In this model, it is assumed that the magnitude of the stock price variation is unrelated to the price itself. However, historical observation tells that the variation is larger at a higher price level. Furthermore, take an extreme case where $\sigma = 0$, this stock becomes a risk-free bond with an instantaneous rate of return μ. i.e. $S_t = S_0 e^{\mu t}$. However Brownian Motion Model suggests $S_t = \mu t$.

Three decades later, in their landmark paper about option pricing, Black and Scholes adopted Geometric Brownian Motion to describe the underlying stock price process

$$\frac{dS_t}{S_t} = \mu dt + \sigma dz \tag{2.1}$$

i.e. the instantaneous rate of return rather than the price itself is a Brownian Motion. In this equation, μ is the drift term that gives direction to the movement of the instantaneous rate of return. σ is the volatility term, which describes its tendency to undergo price changes, i.e. more volatile stocks undergo larger or more frequent price changes. Notice that z is a Wiener increment, $dz = \epsilon d\sqrt{t}$, where ϵ is the standard normal distribution. And this specification of can lead to a "jumpy" movement in the S process. That is because, for a small time interval

Δt, $\sqrt{\Delta t}$ is much larger than Δt, as a result, the standard deviation movement will be much larger than the mean of movement. In the discrete form of this equation, $ln\frac{S(t+\Delta t)}{S(t)}$ is a normal random variable with mean $\mu\Delta t$ and variance $\sigma^2\Delta t$. The attractiveness of Geometric Brownian Motion includes:

1. In case where $\sigma = 0$, it implies $S_t = S_0 e^{\mu t}$, which is consistent with the fact the price of a risk-free stock will grow over time just like a risk-free Treasury Bill.

2. In agreement with the empirical finding, this model indicates that the magnitude of the price change is positively correlated with the price level.

3. As it can be solved: $E(S_t) = S_0 \cdot exp\left\{t \cdot \left(\mu + \frac{\sigma^2}{2}\right)\right\}$, the prices governed by a Geometric Brownian Motion are unlikely to fall below zero, which is in accord with the feature of "Limited Liability" of any equity.

4. The change of the price is independent of the past price history and this independence makes the price series a Markov Process, which is a powerful mathematic tool.

5. From an analytic point of view, by using Geometric Brownian Motion, the Black-Scholes Differential Equation can be transformed to a Heat Equation and solved in a closed form.

Even as a milestone, the Geometric Brownian motion is still far from being "accurate" in describing the stock processes. It violates a lot of stylized facts about the stock returns. We name a few below:

1. Geometric Brownian Motion suggests: $ln\frac{S(t+\Delta t)}{S(t)} \sim N(\mu\Delta t, \sigma^2\Delta t)$, but empirical distribution is usually more "peaked" with a "fatter tail" than the normal distribution.

2. Historically, the mean and variance change over time, which can not be the case given constant μ and σ.

3. As the information arrives in chunks rather than a continuous stream, there are surprises in the sample path of return: a rapid jumping up/down rather than a continuous moving up/down.

2.2 Volatility in Option Pricing

For option traders, volatility is the essence of trading as options derive their theoretic prices in part from it. In this section, we first let the underlying stock price evolve as a Geometric Brownian Motion, and derive the Black-Scholes formula for European Call option by using Heat Equation and solve it in a closed form.[1] Then we give an example to show how the theoretic option price changes with the

[1]Summarized from **A Course in Financial Calculus**, *Cambridge University Press*, **Option Pricing, Mathematical Models and Computation**,*Oxford Financial Press*

volatility of the underlying stock, thus it is important for professional traders to have knowledge of theoretical option prices.

First we construct a portfolio with one unit of call option and $-\Delta$ units of the underlying stock, then its value is given by: $\Pi_t = C_t - \Delta S_t$, where S_t is value of the underlying stock, C_t is value of the value of the option.

$$dS_t = \mu S_t dt + \sigma S_t dz$$

$$dC_t = \sigma S_t \frac{\partial C_t}{\partial S_t} dz + \left(\mu S_t \frac{\partial C_t}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C_t}{\partial S_t^2} + \frac{\partial C_t}{\partial t}\right) dt$$

Therefore, $d\Pi_t = \sigma S_t \left(\frac{\partial C_t}{\partial S_t} - \Delta\right) dz + \left(\mu S_t \frac{\partial C_t}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C_t}{\partial S_t^2} + \frac{\partial C_t}{\partial t} - \Delta \mu S_t\right) dt$

We eliminate the random component by choosing $\Delta = \frac{\partial C_t}{\partial S_t}$ i.e. let the coefficient of dz term be 0.

From now on, we shall use $\Pi, C, S,$ in stead of $\Pi_t, C_t, S_t,$ for the simplicity of notation. Thus, $d\Pi = \left(\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + \frac{\partial C}{\partial t}\right) dt$

On the other hand, the return on an amount of Π invested in riskless assets would see a growth of $r\Pi dt$ in a time period dt. Then the assumption of no arbitrage opportunity in the market requires:

$$d\Pi = \left(\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + \frac{\partial C}{\partial t}\right) dt = r\Pi dt = r(C - \Delta S) dt$$

$$\Rightarrow \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0$$

The above equation is called the Black-Scholes Partial Differential Equation.

The first boundary condition comes from the definition of the European Call Option, $C(S,T) = max(S - E, 0)$. Then, if S=0 on the expiration day, the payoff is

0, this call option is worthless at anytime, so $C(0,t) = 0$. Finally, if $S \to \infty$, this option will be exercised for sure, and the magnitude of the exercise price becomes unimportant, so $C(S,t) \simeq S$ as $S \to \infty$. Now we obtain a system about $C(S,t)$:

$$\begin{aligned} &\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rc = 0 \\ &C(S,T) = max(S-E,0) \quad C(0,t) = 0 \quad C(S,t) \simeq S \text{ as } S \to \infty \end{aligned} \tag{2.2}$$

To solve this system, we do the following substitutions: let $S = Ee^x$, $t = T - \frac{\tau}{\sigma^2/2}$, and let $C(S,t) = Ev(x,\tau)$. Then the Black-Scholes Partial Differential Equation system becomes a system about $v(x,\tau)$:

$$\begin{aligned} &\frac{\partial v}{\partial \tau} = \frac{\partial^2 v}{\partial x^2} + (k_1 - 1)\frac{\partial v}{\partial x} - k_1 v \quad \text{where } k_1 = \frac{r}{\sigma^2/2} \\ &v(x,0) = max(e^x - 1, 0) \end{aligned} \tag{2.3}$$

Next we do another substitution, write $v(x,\tau)$ in the form of $e^{ax+b\tau}u(x,\tau)$. Choose constants a and b that make $u(x,\tau)$ a Heat Equation of the form $\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2}$, i.e. the coefficients of $\frac{\partial u}{\partial x}$ and u in the equation of $\frac{\partial u}{\partial \tau}$ are 0. So the previous system about $v(x,\tau)$ becomes a system about $u(x,\tau)$:

$$\begin{aligned} &\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2} \\ &u_0(x) \hat{=} u(x,0) = max(e^{\frac{k_1+1}{2}x} - e^{\frac{k_1-1}{2}x}, 0) \end{aligned} \tag{2.4}$$

Solving the above system about $u(x,\tau)$, we get:

$$u(x,\tau) = e^{\frac{(k_1+1)x}{2} + \frac{(k_1+1)^2\tau}{4}} \cdot N(d_1) - e^{\frac{(k_1-1)x}{2} + \frac{(k_1-1)^2\tau}{4}} \cdot N(d_2)$$

where $d_1 = \frac{x}{\sqrt{2\tau}} + \frac{k_1+1}{2}\sqrt{2\tau} \quad d_2 = \frac{x}{\sqrt{2\tau}} + \frac{k_1-1}{2}\sqrt{2\tau} \quad N(s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{s} e^{-\frac{\rho^2}{2}}\, d\rho$

By retracing the steps from $C(S,t)$ to $v(x,\tau)$ to $u(x,\tau)$, we obtain the Black-Scholes formula for pricing the European call option:

$$\begin{aligned} C(S,t) &= sN(d_1) - Ee^{-r(T-t)}N(d_2) \\ d_1 &= \frac{\ln\frac{S}{E} + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= \frac{\ln\frac{S}{E} + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \\ N(s) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{s} e^{-\frac{\rho^2}{2}}\, d\rho \end{aligned} \tag{2.5}$$

Therefore to theoretically price an option, we would like to know the volatility (σ) that the underlying stock is going to have from the time the option is purchased (or sold) until the expiration day. Unfortunately that volatility can never be known, since the time frame is the future. Thus, we use a volatility estimate that is based upon the historical volatility[2] of the stock instead.

In practice, volatility is measured in percentages per annum, and price changes are measured from one days closing price to the next. For example, when a stock is described as having a volatility of 30, it means the stock moves either up or down by 30% annually.

Consider an European call option with 9 months until expiration, the strike price is 55, the underlying stock is priced at 50, zero dividend, and the riskfree interest

[2]It also takes into account any events that may have a significant impact on the price of the stock and are known to be occurring during the lifetime of the option. For instance, the quarterly announcement of the company's earnings. Another factor to be taken into consideration is the general condition of the market. With calm markets, all volatility estimates are reduced. But sometimes world events have a great impact on stock prices then volatility estimates are raised.

rate is 2%. By the Black-Scholes formula: [3]

if the volatility estimate is 60, the theoretic price of the option is 8.7947

if the volatility estimate is 40, the theoretic price of the option is 5.2886

if the volatility estimate is 20, the theoretic price of the option is 1.906

if the volatility estimate is 10, the theoretic price of the option is 0.4158

Obviously the larger the volatility of the underlying stock is, the higher the theoretic price of the option is. Intuitively, this is because option buyers make profits when the underlying stocks undergo significant price changes in the correct direction. Volatile stocks are much more likely to undergo large price changes, therefore option buyers pay a much higher price for options of volatile stocks.

Knowing how to use theoretical option prices can help the trader to select the particular option to buy or sell after he chooses a particular stock. The observation that one series (a specific option, with a specific strike and expiration) is much more overpriced or underpriced than others can help a trader decide a long or short position.

It bears mentioning that except for option pricing, volatility estimation is also a tool of quantitative risk management. Moreover, volatility, which is considered the most accurate measure of risk, reflects underlying problems with the overall financial market. For instance: lack of transparency, bad loans, default rates, uncertainty, illiquidity, external shocks, and other negative externalities.

[3] The results below are calculated from an Equity Options Calculator by Dr. Robert Lum

Chapter 3

Time Varying Volatility

After discovering the deficiency in the Geometric Brownian Motion Model as we mentioned in the last chapter, there have been various models which can better describe the statistic features of financial time series emerged in this literature, enough to fill a library of textbooks. According to Bob Jarrow, a professor at the Cornell School of Business Management and research consultant for Kamakura Software, *"many of these models come out of the academic community, where disagreements over the latest and greatest models have become something of an armchair sport".* The most popular ones are stochastic volatility models including the ARCH and GARCH models, and the Regime-Switching models. In this chapter, we first talk about these two families of models as we shall compare them with our primary hidden Markov model both theoretically and empirically. Then we talk about the implied volatility, which is a whole different approach in a sense that these models derive the volatilities of the underlying securities from the life prices of their derivatives, namely the options. Lastly, we talk bout the realized volatility which is a "model-free" measure of volatility but its computation depends on high frequency intra-daily data such as an observation for every five minutes.

3.1 ARCH and GARCH Models

Let $y_{t+1} = ln \frac{S_{t+1}}{S_t}$. According to the discrete version of Geometric Brownian Motion model $ln \frac{S(t+\Delta t)}{S(t)} \sim N(c\Delta t, \sigma^2 \Delta t)$, we have $y_t \sim N(c, \sigma^2)$. i.e. $y_t = c + \epsilon_t$ where $\epsilon = \sigma \cdot z_t$ and z_t is a Wiener Process. In this model, both the drift tern c and the volatility term σ are constants. It is important not to misunderstand ϵ_t as the volatility. Actually, ϵ_t is called the error term. As we discussed earlier, empirical findings suggest that volatility is time varying. That is to say, we should model the error as $\epsilon_t = \sigma_t \cdot z_t$ instead of $\epsilon_t = \sigma \cdot z_t$.

The ARCH and GARCH class of models have been very popular in the area of modeling σ_t, the time varying volatility. The ARCH model was first introduced by Robert F. Engle in 1982 and he won the Nobel prize in 2003 for his contribution in modeling volatility in the financial time series. ARCH means Autoregressive Conditional Heteroscedasticity. Heteroscedasticity refers to a variable whose volatility changes over time.

In ARCH(1), it is assumed that $\sigma_t^2 = K + A \cdot \epsilon_{t-1}^2$ where K and A are constants with $|A| < 1$. Therefore, $\sigma_t^2 = K + A \cdot \sigma_{t-1}^2 \cdot z_{t-1}^2$. This reliance of σ_t^2 on σ_{t-1}^2 is what we mean by "autoregressive". Actually, although the return process y_t

usually has little or no serial correlation, but the squared precess y_t^2 often exhibit significant autocorrelation, which indicates ARCH as a good candidate of modeling y_t. In addition, "conditional" implies given the past sequence of observation. Let ψ_t be the $\sigma - field$ generated by $\{y_t, y_{t-1}, ..., \epsilon_t, \epsilon_{t-1}, ...\}$, then "conditional heteroscedastic" means that $Var(\epsilon_t|\psi_{t-1})$ is time varying.

All models in the ARCH and GARCH family has the following four properties and we shall take ARCH(1) as an example to prove them. For convention, we take $E_t(.) = E(.|\psi_t)$

Proposition 3.1.1. *ϵ_t has zero mean.*

Proof.

$$\epsilon_t = \sqrt{K + A \cdot \epsilon_{t-1}^2} \cdot z_t$$

$$E_{t-1}(\epsilon_t) = \sqrt{K + A \cdot \epsilon_{t-1}^2} \cdot E_{t-1}(z_t) = \sqrt{K + A \cdot \epsilon_{t-1}^2} \cdot 0 = 0$$

$$E_{t-2}(\epsilon_t) = E(\epsilon_t|\psi_{t-2}) = E(E(\epsilon_t|\psi_{t-1})|\psi_{t-2}) \quad \text{since: } \psi_{t-2} \subseteq \psi_{t-1}$$

Therefore, $E_{t-2}(\epsilon_t) = 0$

......

$$E(\epsilon_t) = E_0(\epsilon_t) = 0$$

□

Proposition 3.1.2. $E_{t-1}(\epsilon_t^2) = K + A \cdot \epsilon_{t-1}^2,$ *Conditional Heteroscedastic*

Proof.

$$E_{t-1}(\epsilon_t^2) = E_{t-1}[(K + A\epsilon_{t-1}^2) \cdot z_t^2] = K + A\epsilon_{t-1}^2$$

□

Proposition 3.1.3. *$E(\epsilon_t^2)$ is a constant. Unconditional Homoscedastic*

Proof.

$$\begin{aligned}
E_{t-2}(\epsilon_t^2) &= E(\epsilon_t^2|\psi_{t-2}) \\
&= E(E(\epsilon_t^2|\psi_{t-1})|\psi_{t-2}) \\
&= K + A \cdot E(\epsilon_{t-1}^2|\psi_{t-2}) \\
&= K + K \cdot A + A^2 \cdot \epsilon_{t-2}^2 \\
E_{t-3}(\epsilon_t^2) &= K + K \cdot A + K \cdot A^2 + A^3 \cdot \epsilon_{t-3}^2 \\
&\ldots\ldots \\
E(\epsilon_t^2) &= E_0(\epsilon_t^2) = K + K \cdot A + ... + K \cdot A^{t-1} + A^t \cdot \epsilon_0^2 = \frac{K}{1-A}
\end{aligned}$$

□

Proposition 3.1.4. *$E_{t-1}(\epsilon_t \epsilon_{t-1}) = 0$, zero autocovariance*

Proof.

$$E_{t-1}(\epsilon_t \epsilon_{t-1}) = \epsilon_{t-1} E_{t-1}(\epsilon_t) = 0$$

□

An ARCH(Q) model is defined as:

$$\begin{aligned}
&\sigma_t^2 = K + \sum_{j=1}^{Q} A_j \cdot \epsilon_{t-j}^2 \\
&s.t. \sum A_j < 1, K > 0, A_j \geq 0
\end{aligned} \tag{3.1}$$

The volatility depends on the errors of the last Q periods. It may have a long memory as Q might be greater than 1. The ARCH models became so popular as they take care of changes in the econometrician's ability to forecast. In the history of this literature, interesting interpretations to the conditional heteroscedasticity can be found. For instance, Lamoureux and Lastrapes(1990) explain it as a result of the time dependence in the rate of information arrival to the market. They use the daily trading volume as a proxy for such information arrival and show its significance.

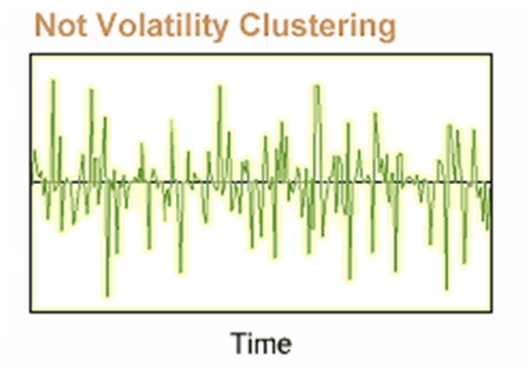
In 1986, Bollerslev improved the ARCH models by inventing the Generalized ARCH models, the GARCH models, where the current volatility depends not only on the past errors, but also on the past volatilities.

$$\begin{aligned}
&\sigma_t^2 = K + \sum_{i=1}^{P} G_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} A_j \cdot \epsilon_{t-j}^2 \\
&s.t. \sum G_i + \sum A_j < 1, K > 0, G_i \geq 0, A_j \geq 0
\end{aligned} \tag{3.2}$$

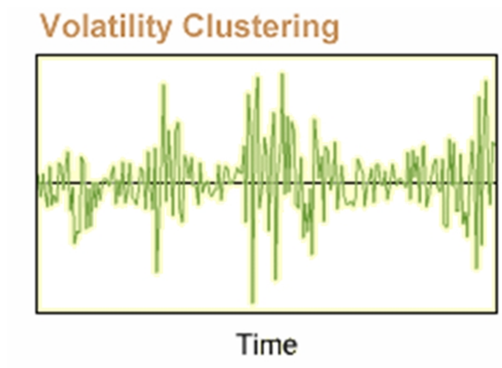
$\sum_{i=1}^{P} G_i \cdot \sigma_{t-i}^2$ is the autoregressive part, and $\sum_{j=1}^{Q} A_j \cdot \epsilon_{t-j}^2$ is the moving average part. It should be noted that when $P = 0$, a GARCH(P,Q) model is an ARCH(Q) model. The GARCH models are designed to capture the following three characteristics associated with the returned process.

1. Volatility Clustering: while plotting the daily returns over a long term period, it is found that large changes tend to be followed by large changes and small changes are tend to be followed by small changes. It suggests that successive

volatilities are serially dependent although uncorrelated.



the fluctuation is independent from time to time



large (small) changes cluster together

2. Fat Tail: the observation of asset return series y_t often exhibit a fatter tail than a standard normal distribution. Statistically, this is known as excess kurtosis, i.e the sample kurtosis $\dfrac{T\sum_{t=1}^{T}(y_t-\bar{y})^4}{\left(\sum_{t=1}^{T}(y_t-\bar{y})^2\right)^2}-3>0.$

3. Leverage Effect: the changes in security returns are often found negatively correlated with the changes in volatility. This is due to the fact that a lower rate of return is always associated with a higher risk.

The error term $\epsilon_t = y_t - c$ is also known as the shock, and a negative shock is always associated with bad news. As $\epsilon_t < 0$ implies $y_t < c$, the return falls below its mean because of the bad news. Empirical studies on financial time series have shown that conditional variance $E_{t-1}(\sigma_t^2)$ often increases after negative shocks, i.e., after bad news being released to the market, the risk is higher. However, as suggested by the equation $\sigma_t^2 = K + \sum_{i=1}^{P} G_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} A_j \cdot \epsilon_{t-j}^2$, the GARCH models are symmetric models, where the sign of ϵ_t is ignored. Some asymmetric version of GARCH models have been invented to capture this characteristic.

AGARCH(P,Q)-type 1

$$\begin{aligned} &\sigma_t^2 = K + \sum_{i=1}^{P} G_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} A_j \cdot (\epsilon_{t-j} + \gamma)^2 \\ &s.t. \sum G_i + \sum A_j < 1, K > 0, G_i \geq 0, A_j \geq 0, \gamma < 0 \end{aligned} \tag{3.3}$$

AGARCH(P,Q)-type 2

$$\begin{aligned} &\sigma_t^2 = K + \sum_{i=1}^{P} G_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} A_j \cdot (|\epsilon_{t-j}| + \gamma \cdot \epsilon_{t-j})^2 \\ &s.t. \sum G_i + \sum A_j < 1, K > 0, G_i \geq 0, A_j \geq 0, \gamma < 0 \end{aligned} \tag{3.4}$$

GJR-GARCH(P,Q)

$$\begin{aligned} &\sigma_t^2 = K + \sum_{i=1}^{P} G_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} (A_j + \gamma \cdot I_{t-j}) \epsilon_{t-j}^2 \\ &s.t. \sum G_i + \sum A_j < 1, K > 0, G_i \geq 0, A_j \geq 0, \gamma < 0 \\ &I_t = 1 \quad \text{if } \sigma_t < 0 \\ &I_t = 0 \quad \text{if } \sigma_t \geq 0 \end{aligned} \tag{3.5}$$

EGARCH(P,Q)

$$\begin{aligned} &ln(\sigma_t^2) = K + \sum_{i=1}^{P} G_i ln(\sigma_{t-i}^2) + \sum_{j=1}^{Q} A_j z_{t-j} + \sum_{j=1}^{Q} \gamma_j \cdot (|z_{t-j}| - E(|z_{t-j}|)) \\ &s.t. \sum G_i + \sum A_j < 1, K > 0, G_i \geq 0, A_j \geq 0 \\ &z_t = \frac{\epsilon_t}{\sigma_t} \end{aligned} \tag{3.6}$$

In a standard GARCH(1,1) model $\sigma_t^2 = K + G\sigma_{t-1}^2 + A\epsilon_{t-1}^2$, conditional variance $E_{t-1}(\sigma_t^2)$ is minimized when there was no shock in the last period ($\epsilon_{t-1} = 0$). In an AGARCH(1,1)-type 1 model, $E_{t-1}(\sigma_t^2)$ is minimized when ($\epsilon_{t-1} = -\gamma > 0$), which means the risk decreases after the release of good news.

In an AGARCH(P,Q)-type 2 model, $\epsilon_{t-j} < 0 \Rightarrow |\epsilon_{t-j}| + \gamma\epsilon_{t-j} = (\gamma - 1)\epsilon_{t-j}$, thus a decrease in ϵ_{t-j} shall cause an increase in $E_{t-1}(\sigma_t^2)$. That is to say, given there was bad news, then the worse it was, the higher the risk will be.$\epsilon_{t-j} > 0 \Rightarrow |\epsilon_{t-j}| + \gamma\epsilon_{t-j} = (\gamma + 1)\epsilon_{t-j}$, thus a decrease in ϵ_{t-j} shall cause a decrease in $E_{t-1}(\sigma_t^2)$. That is to say, given there was good news, then the better it was, the lower the risk will be.

After the release of good news, a GJR-GARCH(P,Q) model is the same as a GARCH(P,Q) model since $\epsilon_{t-j} \geq 0 \Rightarrow I_{t-j} = 0$. However, GJR-GARCH models magnify the impact of bad news, or say, negative shocks. This is because $\epsilon_{t-j} < 0 \Rightarrow [E_{t-1}(\sigma_t^2)]_{GJR-GARCH} > [E_{t-1}(\sigma_t^2)]_{GARCH}$.

All the GARCH models are uniquely described by the parameters
$\Theta = \{K, G_1, G_2, ..., G_P, A_1, A_2, ..., A_Q\}$. The most widely accepted method of estimating GARCH models is to maximize the conditional log likelihood function.

$$p(\epsilon_t|\sigma_t, \Theta) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{\epsilon_t^2}{2\sigma_t^2}}$$

$$L(\epsilon_t|\sigma_t, \Theta) = \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{\epsilon_t^2}{2\sigma_t^2}} \right)$$

$$ln\ L(\epsilon_t|\sigma_t, \Theta) = -\sum_{t=1}^{T} \left(ln\sigma_t + \frac{\epsilon_t^2}{2\sigma_t^2} \right)$$

The last one is our objective function, where T is the number of terms in the sequence.

The technique difficulty of ARCH, GARCH modeling lies in its dependence on P (the length of its autoregressive part) and Q (the length of its moving average part). The higher the P is, the longer volatility memory the process has and therefore the less accurate the model is. Meanwhile, when P and Q are increased, there are more parameters to be estimated.

There are some limitations in the design of GARCH models. For instance, they can not fully explain the fat tail phenomenon. Most importantly, they often fail to capture the highly irregular phenomena, including wild market fluctuations such as crashes and subsequent rebounds, and other highly unanticipated events that can lead to significant structural change. In the next section, we shall talk about another popular type of models that can describe the relatively unstable market conditions.

3.2 Regime-Switching Models

In his influential paper, Hamilton (1989) suggested Regime-Switching model for non-stationary time series (log of GDP). The parameters of an autoregression are viewed as the outcome of a discrete state stationary Markov process $\{X_t\}$, which could be explained as the underlying economic forces.

$$R_t = \phi R_{t-1} + [c_0(1 - X_t) + c_1 X_t] + [\sigma_0(1 - X_t) + \sigma_1 X_t]\epsilon_t$$

where, R_t is the log of GDP, $\epsilon_t \sim N(0, 1)$ *i.i.d.* X_t is a random variable that capture the changes in the underlying economic forces. $\{X_t\}$ evolves as a two-state markov chain with the state space $\{0, 1\}$, and the transition probabilities: $p(X_t = 0|X_{t-1} = 0) = q$, $p(S_t = 1|S_{t-1} = 1) = p$.

Later on, many papers have discussed the regime switching phenomena in stock market returns under the framework of Hamilton's Regime-Switching model. They drop the autoregressive term by letting $\phi = 0$. Turner, Startz, and Nelson (1989) consider a Markov switching model in which either the drift, the volatility, or both of them may differ between two regimes. They use a two-state Markov process with constant transition probabilities. Schwert (1989) considers a model in which returns may have either a high or a low volatility. He claims that a two-state Markov process determines the switches of return distributions. Hamilton and Susmel (1993) combine the ARCH model with the Markov switching method and propose a Switching- ARCH (SWARCH) model. They address that there exist sudden discrete changes in the process that determine volatility. These three papers use the same technique provided by Hamilton (1989) to estimates the transition probabilities in the two-state Markov processes.

Basically it is assumed in those papers that the return process can be described as the following:

$$R_t = [c_0(1 - X_t) + c_1 X_t] + [\sigma_0(1 - X_t) + \sigma_1 X_t]\epsilon_t \tag{3.7}$$

where, $\epsilon_t \sim N(0,1)$ $i.i.d.$ $\{S_t\}$ is a random variable that capture the changes in the underlying economic forces nad it evolves as a two-state markov chain with the state space $\{0,1\}$, and the transition probabilities: $pr(X_t = 0|X_{t-1} = 0) = q$, $pr(X_t = 1|X_{t-1} = 1) = p$. That is:

$$R_t|_{\{X_t=0\}} = c_0 + \sigma_0\epsilon_t$$

$$R_t|_{\{X_t=1\}} = c_1 + \sigma_1\epsilon_t$$

They use the technique provided by Hamilton(1989) to estimate the parameters $\Theta = \{c_1, c_2, \sigma_1, \sigma_2, p, q\}$, which is to maximize the log likelihood function of $(R_1, R_2, ..., R_T)$ given observation up to $T - 1$. Let Ω_t be the $\sigma - field$ generated by $\{R_t, R_{t-1}, ..., R_1, R_0\}$

$$max \qquad ln\ L(R_1, R_2, ..., R_T|R_0, \Theta) = \sum_{t=1}^{T} ln\ p(R_t|\Omega_{t-1}; \Theta) \qquad w.r.t. \quad \Theta$$

An obvious deficiency in this technique lies in the fact that it is not easy to extend it analytically to the case where the state space of the Markov chain has a higher dimension.

3.3 Implied Volatility

Unlike the previous models, the implied volatility of the underlying security is derived from the life prices of its derivatives, namely the options.

In an option pricing model, such as Black-Scholes, a variety of variables are needed as inputs to derive a theoretical value for an option. These inputs may vary depending on the type of option being priced and the pricing model used. However, the theoretic value of any option in any pricing model would depend on an estimate of the volatility of the underlying security, σ. To express this idea mathematically: $C = f(\sigma, \Upsilon)$, where C is the theoretic price of the option, Υ stands for all the other inputs except for the volatility of the underlier, and $f(.)$ is the option pricing model. Assume there exists an inverse function $h(.) = f^{-1}(.)$, then $\sigma_{\hat{C}} = h(\Upsilon, \hat{C})$, where $\hat{C}$ is the realized option price. Therefore, we say $\sigma_{\hat{c}}$ is the volatility implied by the market price $\hat{C}$ of the derivative (the option), thus $\sigma_{\hat{c}}$ is called the implied volatility. The most technique difficulties in this approach lie in the fact that there is usually not a closed form of such h function. In fact, for most of the derivative securities, even the exact formulas of $C = f(\sigma, \Upsilon)$ are not available. In practice, an iterative search procedure is often used to find the implied volatility. [1]

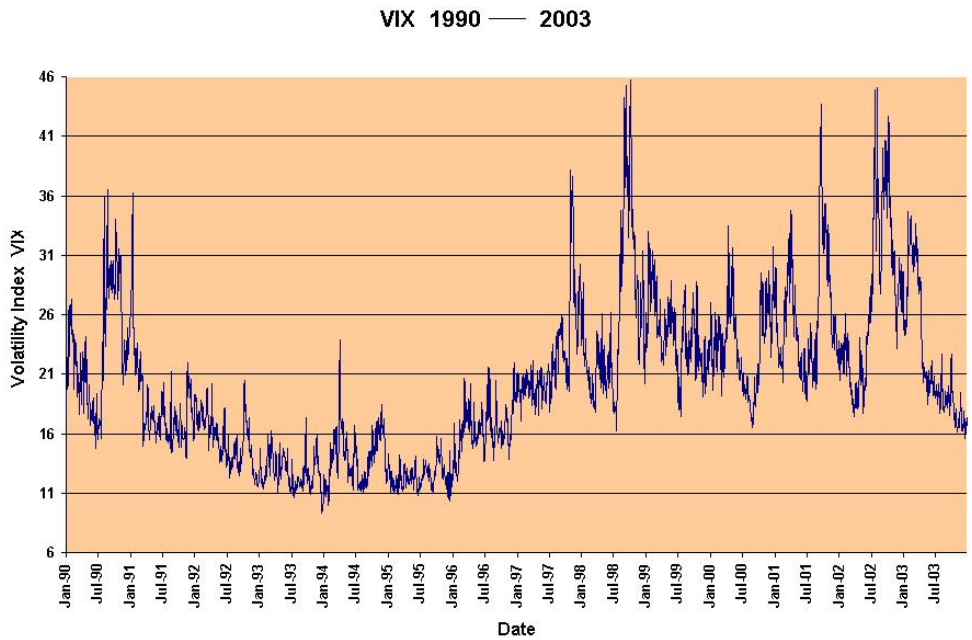
Implied volatilities are of great interest to option traders as they reflect the market's opinion toward the volatility of a particular stock. They can also be used to estimate the price of one option from the price of another option with the same underlying. Very often, the implied volatilities are obtained simultaneously from

[1]There are generally two types of derivatives for which no exact pricing formulas available. The first type is when the payoff of the derivative security is dependent on the history of the underlying variable or where there are several underlying variables, in this case a Monte Carlo simulation method is always useful; the second type is when the holder of the security has early exercise decisions or other types of decisions to make prior to maturity, in this case, trees or finite difference methods are useful.

different options on the same stock and then a composite implied volatility for the stock is calculated by taking suitable weighted average of the individual implied volatilities. By "suitable", we mean the weight should reflect the the sensitivity of the option price to the volatility. To illustrate this point, suppose there are two implied volatility estimates available, the first one is 0.32, based on an at-the-money option, the second one is 0.26, based one an out-of-money option. The price of the at-the-money option is far more sensitive to the volatility then the out-of-the-money option, so we give the first one a weight of 80%, and give the second one a weight of 20%, thus the composite volatility is $0.8 \times 0.32 + 0.2 \times 0.26 = 0.308$, or 30.8% per annum. The most popular composite implied volatility is the VIX(Volatility Index)on Chicago Board Options Exchange (CBOE), which is calculated using a weighted average of implied volatilities in options on the S&P 500 Index futures. There also exists the Nasdaq 100 index futures volatility measure (VXN) and the QQQQ volatility measure (QQV). The next graph[2] displays the hisotical VIX from Jan, 1990. [3]

[2]The resource of this plot is MDWoptions - Options Education for the Public Investor

[3]CBOE changed its methodology for calculating the VIX in 1990



Moreover, implied volatilities is used more and more in conjunction with statistic volatilities (volatilities derived by measuring the actual price movements that the underlying stock has made in the past.) to identify new profit-making opportunities. If implied volatility is substantially higher than statistic volatility, it is time to sell volatility. Likewise, if implied volatility is much lower than statistic volatility, it is time to buy volatility.

There has been disagreement among scholars and traders whether it is better off to use historical security prices or current option prices to estimate volatility.

> John Campbell, Andrew Lo, and Craig MacKinlay, (1997): *"Consider the argument that implied volatilities are better forecasts of future volatility because changing market conditions cause volatilities to vary through time stochastically, and historical volatilities cannot adjust to changing market conditions as rapidly. The folly of this argument lies in the fact that stochastic volatility contradicts the assumption required by the B-S model - if volatilities do change stochastically through time, the Black-Scholes formula is no longer the correct pricing formula and an implied volatility derived from the Black-Scholes formula provides no new information."*

Meanwhile, it is found in practice that options based on the same underlying securities but with expiration date and different strike value yield different implied volatilities. This phenomenon is generally believed as evidence that volatility is stochastic.

Whether or not you receive a valid implied volatility depends on the option model you are using to solve for volatility. As Bob Jarrow noted *"Any mis- specifications in the model will affect the resulting implied volatility."*

3.4 Realized Volatility

Many difficulties in evaluating volatility models arise from the fact that volatility is not observable, since one can not compared the forecasted volatility with a "benchmark" that physically exists. Anderson and Bollerslev (1998) introduced the concept of realized volatility, a "model-free" measure of volatility from which evaluation of volatility models can be made. Realized volatility is calculated from high frequency intra-daily data, rather than inter-daily data.

Anderson and Bollerslev collected data on the Deutsche Mark - U.S. Dollar and Japanese Yen - U.S. Dollar spot exchange rates for every five minutes resulting in 288 observations per day. Then these 288 observations were used to compute the variance of the exchange rate on that particular day. Their methodology can be summarized as the following:

Let p_t^n denote the security price at the n^{th} time interval on the t^{th} day. $n =$

$1, 2, ..., N$ and $t = 1, 2, ...$ For example, $N = 288$ if we collect data on a five-minute interval basis. Note that when N equals 1, which mean n could only be 1, and p_t^1 is simply the inter-daily price rather than intra-daily prices. In such a case, we just write it as p_t. Define $r_t^n = log(p_t^n) - log(p_t^{n-1}$, thus r_t^n is the observed intra-daily returns on the t^{th} day. The inter-daily return on the t^{th} day can be obtained by $r_t = \sum_{n-1}^{N} r_t^n$. Since

$$r_t^2 = \left(\sum_{n=1}^{N} r_t^n \right)^2 = \sum_{n=1}^{N} (r_t^n)^2 + 2 \sum_{n=1}^{N} \sum_{m=n+1}^{N} r_t^n r_t^{m-n}$$

$$\sigma_t^2 = E(r_t^2) = E\bigg(\sum_{n=1}^{N} (r_t^n)^2 \bigg) + 2 \sum_{n=1}^{N} \sum_{m=n+1}^{N} E(r_t^n r_t^{m-n}) = E\bigg(\sum_{n=1}^{N} (r_t^n)^2 \bigg)$$

Note that $E(r_t^n r_t^{m-n})$ equals zero as it is believed that the intra-daily returns are uncorrelated.

Let $s_t^2 = \sum_{n=1}^{N} (r_t^n)^2$, then s_t is called the realized volatility.

The properties of the realized volatility are discussed by Anderson, Bollerslev, Diebold and Labys (2001). While the concept of realized volatility does provide an efficient way of estimating volatility, the problem of collecting information on the security price every minute or so is immense.

Chapter 4

HMM with Stochastic Drift and Volatility

In this chapter we start with the model setup and a general guideline of the methodology in section 4.1. Then we explain the four major estimation steps one by one in section 4.2 - 4.5. Finally, in section 4.6, we will introduce a price model which can predict the next price by using the current price and the estimates we obtained in section 4.5.

4.1 Model Setup

Assume the daily stock (could be any risky security) return process $\left\{ln\frac{S_{t+1}}{S_t}\right\}_{t=1,2,3,......}$ is governed by:

$$ln\frac{S_{t+1}}{S_t} = G_t + V_t \cdot b_{t+1}, \qquad b_t \sim N(0,1), i.i.d. \tag{4.1}$$

This model is basically under the framework of the discrete version of the Geometric Brownian Motion Model except that we allow the drift $\{G_t\}$ and the volatility

$\{V_t\}$ to be time varying. Unlike the Efficient Market Hypothesis, which claims the current price reflect all the information, we consider there is a one day delay in the information spread, and for investors, the information they gain from today's price is actually the information that should be available from the last trading day. As a result, the evolving of today's information: G_t and V_t depends on the information of the last trading day: G_{t-1} and V_{t-1}. Therefore, we model this pair $\{(G_t, V_t)\}$ to evolve as a first order Markov chain. As this Markov chain is unobservable(hidden), this model belongs to the category of Hidden Markov Models.

The attractiveness of this model includes:

1. It maintained all the charms of the Geometric Brownian Motion Model with loosened hypothesis. i.e., the drift and volatility can be time varying.

2. Unlike most of the time varying volatility models which force the drift to be constant basically for analytic purpose (such as a typical GARCH model) , this model allows the drift to be time varying. However, as we admit there are certain return processes do show a constant trend over certain periods, we do not force the drift to be time varying. For instance, G_t can be some constant. In our empirical implementation with 73 different data sets, we did find some security returns have relatively steady drift.

3. Unlike the estimation technique employed in a typical Regime-Switching model, we suggest a technique that can be easily extended to the cases where

there are more than 2 states.

4. Unlike the Efficient Market Hypothesis, it admits the inefficiency: there exists some delay in the information spread. As a result, the current price has some sort reliance on the past price history. Meanwhile it capture of feature of modern financial market, i.e. the information is quickly incorporated in the security prices so the reliance on the past does not have a long memory. In another word, what is reflected by today's price depends only on what is reflected by yesterday's price.

5. Such dependence on the last trading day is a stochastic rather than a deterministic function.

6. As it is shown by our paper, the implementation of this model is manageable and the result is meaningful.

Suppose the chain $\{(G_t, V_t)\}_{t=1,2,3,......}$ has an N dimensional state space $B = \{(g_1, v_1), (g_2, v_2), ..., (g_N, v_N)\}$ and stationary transition probabilities $P((G_{t+1}, V_{t+1}) = (g_s, v_s)|(G_t, V_t) = (g_r, v_r)) = a_{sr}$. Thus, $A = (a_{sr})_{s,r=1,2,...,N}$ is the transition probability matrix.

Let R_t be the return process, i.e. $R_t = \left\{ ln\frac{S_t}{S_{t-1}} \right\}_{t=1,2,3,......}$

Obviously, the return process is observable. What we would like is to "see" the unobservable Markov process through the return process so that we can have an expected next stage of this Markov cahin and use it to predict the next return and thus make the investment decision. To achieve this goal, we need to estimate the specifications of this Markov Chain, namely the state space B and the transition probability matrix A.

The methodology can be summarized as the following:

First of all we transform the model by using a $1 - to - 1$ mapping ϕ from B to $\Sigma = \{e_1, e_2, ..., e_N\}$ where e_r is an $N \times 1$ unit vector with the r^{th} component equals 1 and the others equal zero. i.e. $\phi((g_r, v_r)) = e_r$ for $\forall r = 1, 2, ..., N$. Obviously Σ is numerically more manageable than B as each of its entry is a unit vector rather than a pair of numbers. More importantly, we use this $1-to-1$ mapping to define a process $\{X_t\}$ which has a profound economic meaning. We shall discuss this step in section 4.2.

Next, consider the probability measure p in the real world. Under p, the drift and volatility changes from one state to another with certain probabilities and the return process has a noise term evolving as a random draw from a sequence of i.i.d. random variables with Gaussian distribution. What we do in this step is to define a probability ρ in a fictitious world under which the hidden Markov chain is still the Markov chain with the same state space and transition probability matrix, however the return process $\{R_t\}$ becomes a sequence of i.i.d. random variables with uniform distributions. The purpose of defining such a probability measure is to make the next two steps possible: obtaining recursive filters, obtaining EM estimates. We call this step “change of probability measure”. The theoretic development of such a probability measure and its properties are discussed in section 4.3.

Mathematically a recursive filter is a recursive equation where only the estimate of the previous state and the current measurement are needed to compute the estimate of the current state. In section 4.4, we first develop a general filter in the fictitious world we have just defined in 4.3 (under the new probability measure) using a similar idea to the Kalman Filter algorithm then use this general filter to obtain the following four filters: the underlying Markov chain, occupation time, number of jumps, and functionals of the observation.

Finally by maximizing the expectation of the likelihood functions, we obtain the estimates of the transition probability matrix, the state space of the drift and volatility, and some other model specifications as by-products. Our iteration is similar to the EM algorithm but different in a sense that we update the observation in each pass so that more emphasis is given to the recent information as it is more relevant in regard to predicting the near future. We shall discuss how these EM estimates are achieved in section 4.5

4.2 The Underlying Process

The state space B for drift and volatility is N dimensional, and so is $\Sigma = \{e_1, e_2, ..., e_N\}$. If we can define a 1-to-1 mapping between these two spaces then Σ could help us to capture some features of B since after all, the shape of Σ is a lot easier.

Proposition 4.2.1. *There exists* ϕ*:* $B \rightarrow \Sigma$*, s.t.* $\phi((g_r, v_r)) = e_r \quad \forall r = 1, 2, ..., N$

The existence of ϕ is ensured by the fact that $dim(B) = dim(\Sigma)$. We actually do not have to worry about the exact formula of ϕ. Then we define a process $\{X_t\}$ by: $X_t = \phi((G_t, V_t))$

Proposition 4.2.2. $\{X_t\}$ *evolves as a Markov Chain with state space* Σ*.*

Proof.

$$
\begin{aligned}
P(X_{t+1}|X_t, X_{t-1}, ..., X_0) &= P(\phi((G_{t+1}, V_{t+1}))|\phi((G_t, V_t)), \phi((G_{t-1}, V_{t1})), ..., \phi((G_0, V_0))) \\
&= P((G_{t+1}, V_{t+1})|(G_t, v_t), (G_{t-1}, V_{t-1}), ..., (G_0, V_0)) \\
&= P((G_{t+1}, V_{t+1})|(G_t, V_t)) \\
&= P(\phi((G_{t+1}, V_{t+1}))|\phi((G_t, V_t))) \\
&= P(X_{t+1}|X_t)
\end{aligned}
$$

$(G_t, V_t) \in B$ and ϕ: $B \rightarrow \Sigma$, so $X_t = \phi((G_t, V_t)) \in \Sigma$

i.e. the state space of $\{X_t\}$ is Σ. □

Proposition 4.2.3. $P(X_{t+1} = e_s|X_t = e_r) = a_{sr}$*, therefore the chain* $\{(G_t, V_t)\}$ *and* $\{X_t\}$ *have the same transition probability matrix.*

Proof.

$$\begin{aligned} P(X_{t+1} = e_s | X_t = e_r) &= P(\phi((G_{t+1}, V_{t+1})) = \phi((g_s, v_s)) | \phi((G_t, V_t)) = \phi((g_r, v_r))) \\ &= P((G_{t+1}, V_{t+1}) = (g_s, v_s) | (G_t, V_t) = (g_r, v_r)) \\ &= a_{sr} \end{aligned}$$

□

Next we denote the following filtrations: $\{\mathcal{X}_k\} = \sigma(X_0, X_1, ..., X_k)$,

$\{\mathcal{F}_k\} = \sigma(X_0, X_1, ..., X_k, R_1, ..., R_k)$, and $\{\mathcal{Y}_k\} = \sigma(R_1, ..., R_k)$.

Let $g = (g_1, g_2, ..., g_N)'$ and $v = (v_1, v_2, ..., v_N)$, then $(g_r, v_r) = (< g, e_r >, < v, e_r >)$ for $\forall r = 1, 2, ..., N$. Therefore (4.1) is transformed to:

$$\begin{aligned} &R_{t+1} =< g, X_t > + < v, X_t > b_{t+1} \\ &X_{t+1} = AX_t + M_{t+1} \\ &A \text{ is the transition probability matrix} \\ &\{M_{t+1}\} \text{ is a martingale increment process with respect to the filtration } \{\mathcal{F}_t\} \end{aligned} \tag{4.2}$$

This transforming not only makes the model easier analytically but also gives rise to an economically meaningful process $\{X_t\}$. As we all know, forces that move the security prices including the information about market sentiment[1]; fundamen-

[1]Market sentiment refers to the psychology of market participants, individually and collectively.

tal factors such as earning base[2], valuation multiple[3]; technique factors such as inflation[4], economic strength of market and peers, substitutions, incidental transactions, demographics[5], trends[6], liquidity,... Their overall impact varies on different securities at different time period, so we can not have a clean equation to put all the forces together. Instead we consider the combination of them as a magic "manipulator" of the financial market. For example, this manipulator makes a daily choice on any specific security from the following strategy set: { high rise, low rise, stay, low drop, high drop }, and his choice is based on his observation from the previous trading day. For instance, if it was a 'low rise' on Monday, then the probability for a 'high rise' on Tuesday is 0.2, for a 'low rise' again is 0.2, for 'stay' is 0.4, for a 'low drop' is 0.1, and for a 'high drop' is 0.1. As a result, the process of his every day strategy evolves as a first order Markov Chain. Why does he only take yesterday's observation into account? First of all, all the information we just mentioned are embedded in the observation. Secondly, the market adjusts so fast upon the information arrival that only previous day's observation could reflect all

[2]There are different measures of earning base, for example the earnings per share (EPS) is the owner's return on his or her investment.

[3]The valuation multiple is a way of representing the discounted present value of the anticipated future earnings stream.

[4]Inflation is generally considered as good for stocks because it signifies a gain in pricing power for companies.

[5]It is believed that the greater the proportion of middle-aged investors among the investing population, the greater the demand for equities and the higher the valuation multiples.

[6]Trend is a two-sided impact, a stock that is moving up can gather momentum, as "success breeds success", however it also suggests to move in opposite way in a trend and does what is called "reverting to the mean."

the historical information until yesterday. In fact, the underlying process $\{X_t\}$ that we defined earlier can be viewed as this market manipulator who works behind the scene. When $\{X_t\}$ is in state e_r, it drives the drift to g_r and the volatility to v_r. We have introduced the hidden Markov models by a situation of playing bridge in section 1.2, we estimate the opponent's dealing habit and his evaluation of each round. Here investing is like playing against this market manipulator and we try to estimate his strategy of each trading day.

While we consider the underlying Markov chain as a market "manipulator", the combination of all the forces that move the security prices, which differs from stock to stock, and time to time, some other papers have suggested that it is a specific economic process. Blanchard and Watson (1982) provide one example of such processes, the stochastic bubbles. They stress that a bubble may either survive or collapse in each period; in such a world, returns would be drawn from one of two distributions - surviving bubbles or collapsing bubbles. Cecchetti, Lam, and Mark (1990) provide another example. They consider a Lucas asset pricing model in which the economy's endowment switches between high economic growth and low economic growth. They show that such switching in fundamentals accounts for patterns in the stock market.

4.3 Change of Probability Measure

Let $\varphi(x)$ denote the probability density function of a standard normal distribution. i.e. $\varphi(x) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-x^2}{2}\right)$.

Let $\psi_t = \frac{\varphi\left(\frac{R_t - < g, X_{t-1} >}{< v, X_{t-1} >}\right)}{< v, X_{t-1} > \varphi(R_t)}$ and $\lambda_t = \prod_{k=1}^{t} \psi_k$

Define a new probability measure ρ by the Radon-Nikodym derivative $\frac{dp}{d\rho}\Big|\mathcal{F}_t = \lambda_t$

i.e. for $\forall \Lambda \in \mathcal{F}_t \quad \int_\Lambda \lambda_t d\rho = p(\Lambda)$

In the sections 4.4 and 4.5, we shall derive recursive filters that describe the dynamics of $E_\rho(\cdot)$, the expectations under ρ, and then we use those filters to obtain the estimates, therefore it is necessary for us to be able to change them back to $E(\cdot)$, the expectations under p. In plain English, it is just like we ship the raw materials to a fictitious world where manufacturing is easier, then we ship the products back to the real world. Sections 4.4 and 4.5 shall talk about the manufacturing in the fictitious world, and in this section, we talk about the shipping devices.

Lemma 4.3.1.

$$E(\omega_t) = E_\rho(\lambda_t \omega_t) \quad \textit{For } \forall \textit{ random variable } \omega_t \textit{ that is } \mathcal{F}_t \textit{ measurable} \tag{4.3}$$

Proof. $E(\omega_t) = \int \omega_t dp = \int \omega_t \frac{dp}{d\rho} d\rho = \int \omega_t \lambda_t d\rho = E_\rho(\omega_t \lambda_t)$ □

Lemma 4.3.2. *The abstract Bayes' Theorem:*

$$E(H_t|\mathcal{Y}_t) = \frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E_\rho(\lambda_t|\mathcal{Y}_t)} \quad \mathcal{Y}_t \subseteq \mathcal{F}_t \tag{4.4}$$

For $\forall$ random variable H_t that is integrable and $\mathcal{F}_t$ measurable

Proof. For $\forall B \in \mathcal{Y}_t$, Let I_B be the indicator function of B

$$\int_B \frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E_\rho(\lambda_t|\mathcal{Y}_t)} dp = E\left[I_B \frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E_\rho(\lambda_t|\mathcal{Y}_t)}\right]$$

$$= E_\rho\left[\lambda_t I_B \frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E_\rho(\lambda_t|\mathcal{Y}_t)}\right]$$

$$= E_\rho\left\{E_\rho\left[\lambda_t I_B \frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E_\rho(\lambda_t|\mathcal{Y}_t)}\right]\middle|\mathcal{Y}_t\right\}$$

$$= E_\rho\left[I_B \cdot E_\rho(\lambda_t|\mathcal{Y}_t) \frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E_\rho(\lambda_t|\mathcal{Y}_t)}\right]$$

$$= E_\rho[I_B E_\rho(\lambda_t H_t|\mathcal{Y}_t)] = E_\rho[I_B \lambda_t H_t] = E(I_B H_t)$$

$$= \int_B E(H_t|\mathcal{Y}_t) dp$$

the 2nd and the 7th equal signs are due to the lemma 4.3.1

therefore: $$\frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E_\rho(\lambda_t|\mathcal{Y}_t)} = E(H_t|\mathcal{Y}_t)$$

□

For the convenience of notation, we define: $\sigma_t(H_t) = E_\rho(\lambda_t H_t|\mathcal{Y}_t)$.

The abstract Bayes Theorem gives,

$$\sigma_t(1) = E_\rho(\lambda_t|\mathcal{Y}_t) = \frac{E_\rho(\lambda_t H_t|\mathcal{Y}_t)}{E(H_t|\mathcal{Y}_t)} \qquad \forall H_t \text{ that is } \mathcal{F}_t \text{ measurable}$$

We call $\sigma_t(1)$ the normalizer.

When recursive filters are achieved in the fictitious world, we obtain estimates with these filters then we take the estimates back to the real world by this normalizer.

4.4 Recursive Filters

In this section, we shall first develop a general unnormalized filter then use it to four specific unnormalized filters that will be needed in the EM estimation.

Let $\Gamma^i(R_t) = \dfrac{\varphi\left(\dfrac{R_t - g_i}{v_i}\right)}{v_i \cdot \varphi(R_t)}$ Note that $\psi_t|_{X_{t-1}=e_i} = \Gamma^i(R_t)$, and it is $\mathcal{Y}_t$ measurable.

Let $a_i = A \cdot e_i$

$diag(a_i)$ is a diagonal matrix with its entries on the diagonal to be entries in a_i.

As a remark, there are two types of index sets involved: $\{r\}$, $\{s\}$, and $\{i\}$ are index sets for states, which means, $r, s, i = 1, ..., N$. While $\{k\}$ is an index set for time, i.e., $k = 1, 2, ..., t$.

Consider the following process: $H_t = H_{t-1} + \alpha_t + < \beta_t, M_t > + \delta_t f(R_)$

where $M_t = X_t - AX_{t-1}$, α_t is a scalar, β_t is an $N \times 1$ vector, δ_t is a scalar. Moreover, α_t, β_t, and δ_t are all $\mathcal{F}_{t-1}$ measurable. i.e. H_t, as a scalar variable, is a function of X_t, R_t and other variables that are all $\mathcal{F}_t$ predictable.

Theorem 4.4.1. *For any $\{H_t\}$ processes defined as above,*

$$\begin{aligned}
\sigma_t(H_t X_t) = \sum_{i=1}^{N} \Big[& < \sigma_{t-1}(H_{t-1}X_{t-1}), e_i > \Gamma^i(R_t) a_i \\
& + \sigma_{t-1}(\alpha_t < X_{t-1}, e_i >) \Gamma^i(R_t) a_i \\
& + \sigma_{t-1}(\delta_t < X_{t-1}, e_i >) \Gamma^i(R_t) f(R_t) a_i \\
& + (diag(a_i) - a_i a_i') \sigma_{t-1}(\beta_t < X_{t-1}, e_i >) \Gamma^i(R_t) \Big]
\end{aligned} \tag{4.5}$$

Proof.

$$\sigma_t(H_tX_t) = E_\rho(\lambda_t H_t X_t|\mathcal{Y}_t)$$

$$= E_\rho[(\lambda_{t-1}\psi_t)(H_{t-1} + \alpha_t + <\beta_t, M_t> + \delta_t f(R_t))(AX_{t-1} + M_t)|\mathcal{Y}_t]$$

this is to break H_t into parts with zero and nonzero conditional expectations.

$$= E_\rho[(\lambda_{t-1}\psi_t)(H_{t-1} + \alpha_t + \delta_t f(R_t))AX_{t-1}|\mathcal{Y}_t]$$

$$+ E_\rho[(\lambda_{t-1}\psi_t)(H_{t-1} + \alpha_t + \delta_t f(R_t))M_t|\mathcal{Y}_t]$$

$$+ E_\rho[(\lambda_{t-1}\psi_t) < \beta_t, M_t > AX_{t-1}|\mathcal{Y}_t]$$

$$+ E_\rho[(\lambda_{t-1}\psi_t) < \beta_t, M_t > M_t|\mathcal{Y}_t]$$

since $E_\rho(M_t|\mathcal{Y}_t) = E_\rho[E_\rho(M_t|\mathcal{Y}_t, \mathcal{F}_{t-1})|\mathcal{Y}_t] = E_\rho[E_\rho(M_t|\mathcal{F}_{t-1})|\mathcal{Y}_t]$

and $E_\rho(M_t|\mathcal{F}_{t-1}) = 0$

we have $E_\rho(M_t|\mathcal{Y}_t) = 0$

therefore, the second and the third terms in the above equation are zero.

Thus, we get the following equation, call it ($\diamondsuit$):

$$\sigma_t(H_tX_t)$$

$$= E_\rho[(\lambda_{t-1}\psi_t)(H_{t-1} + \alpha_t + \delta_t f(R_t))AX_{t-1}|\mathcal{Y}_t] + E_\rho[(\lambda_{t-1}\psi_t) < \beta_t, M_t > M_t|\mathcal{Y}_t]$$

$$= E_\rho[(\lambda_{t-1}\psi_t)H_{t-1}AX_{t-1}|\mathcal{Y}_t] \text{ .... } {}_{1^\circ}$$

$$+ E_\rho[(\lambda_{t-1}\psi_t)\alpha_t AX_{t-1}|\mathcal{Y}_t] \text{ .... } {}_{2^\circ}$$

$$+ E_\rho[(\lambda_{t-1}\psi_t)\delta_t f(R_t)AX_{t-1}|\mathcal{Y}_t] \text{ .... } {}_{3^\circ}$$

$$+ E_\rho[(\lambda_{t-1}\psi_t) < \beta_t, M_t > M_t|\mathcal{Y}_t] \text{ .... } {}_{4^\circ}$$

and

$$\psi_t A X_{t-1}$$
$$= \frac{\varphi\left(\dfrac{R_t - <g, X_{t-1}>}{<v, X_{t-1}>}\right)}{<v, X_{t-1}> \varphi(R_t)} \sum_{i=1}^{N} a_i <X_{t-1}, e_i>$$
$$= \sum_{i=1}^{N} a_i \frac{\varphi\left(\dfrac{R_t - <g, X_{t-1}>}{<v, X_{t-1}>}\right)}{<v, X_{t-1}> \varphi(R_t)} <X_{t-1}, e_i>$$

$$= \begin{cases} \sum_{i=1}^{N} a_i 0 = 0 \quad X_{t-1} \neq e_i \\ \sum_{i=1}^{N} a_i \frac{\varphi\left(\dfrac{R_t - <g, e_i>}{<v, e_i>}\right)}{<v, e_i> \varphi(R_t)} = \sum_{i=1}^{N} a_i \frac{\varphi\left(\dfrac{R_t - g_i}{v_i}\right)}{v_i \varphi(R_t)} = \sum_{i=1}^{N} a_i \Gamma^i(R_t) \quad X_{t-1} = e_i \end{cases}$$

$$\text{i.e.} \quad \psi_t A X_{t-1} = \sum_{i=1}^{N} a_i \Gamma^i(R_t) <X_{t-1}, e_i> \;\; \ldots\ldots\ldots\ldots \;(*)$$

Plug $(*)$ into 1° :

$E_\rho[(\lambda_{t-1}\psi_t) H_{t-1} A X_{t-1} | \mathcal{Y}_t]$ note that H_t is a scalar

$$= E_\rho\left[\lambda_{t-1} H_{t-1} \left(\sum_{i=1}^{N} a_i \Gamma^i(R_t) <X_{t-1}, e_i> \right) \middle| \mathcal{Y}_t \right]$$

$= \sum_{i=1}^{N} E_\rho[\lambda_{t-1} H_{t-1} <X_{t-1}, e_i> | \mathcal{Y}_t] \; a_i \Gamma^i(R_t)$ note that $\Gamma^i(R_t)$ is $\mathcal{Y}_t$ measurable

$$= \sum_{i=1}^{N} E_\rho[<\lambda_{t-1} H_{t-1} X_{t-1}, e_i> | \mathcal{Y}_t] \; a_i \Gamma^i(R_t)$$
$$= \sum_{i=1}^{N} <E_\rho[\lambda_{t-1} H_{t-1} X_{t-1} | \mathcal{Y}_t], e_i> \; a_i \Gamma^i(R_t)$$
$$= \sum_{i=1}^{N} <\sigma_{t-1}(H_{t-1} X_{t-1}), e_i> \; \Gamma^i(R_t) a_i$$

Plug $(*)$ into $2°$:

$$E_\rho[(\lambda_{t-1}\psi_t)\alpha_t AX_{t-1}|\mathcal{Y}_t]$$
$$= E_\rho\left[\lambda_{t-1}\alpha_t\left(\sum_{i=1}^{N} a_i\Gamma^i(R_t) < X_{t-1}, e_i >\right)\middle|\mathcal{Y}_t\right]$$
$$= \sum_{i=1}^{N} E_\rho[\lambda_{t-1}\alpha_t < X_{t-1}, e_i > |\mathcal{Y}_t]\ a_i\Gamma^i(R_t)$$
$$= \sum_{i=1}^{N} E_\rho[< \lambda_{t-1}\alpha_t X_{t-1}, e_i > |\mathcal{Y}_t]\ a_i\Gamma^i(R_t)$$
$$= \sum_{i=1}^{N} E_\rho[< \lambda_{t-1}\alpha_t X_{t-1}, e_i > |\mathcal{Y}_{t-1}]\ a_i\Gamma^i(R_t) \quad \text{note that } \alpha_t \text{ is } \mathcal{F}_{t-1} \text{ measurable}$$
$$= \sum_{i=1}^{N} < \sigma_{t-1}(\alpha_t < X_{t-1}, e_i >)\ \Gamma^i(R_t)a_i$$

Plug $(*)$ into $3°$:

$$E_\rho[(\lambda_{t-1}\psi_t)\delta_t f(R_t)AX_{t-1}|\mathcal{Y}_t]$$
$$= E_\rho\left[\lambda_{t-1}\delta_t f(R_t)\left(\sum_{i=1}^{N} a_i\Gamma^i(R_t) < X_{t-1}, e_i >\right)\middle|\mathcal{Y}_t\right]$$
$$= \sum_{i=1}^{N} E_\rho[\lambda_{t-1}\delta_t < X_{t-1}, e_i > |\mathcal{Y}_t]\ a_i\Gamma^i(R_t)f(R_t) \quad \text{note that } f(R_t) \text{ is } \mathcal{Y}_t \text{ measurable}$$
$$= \sum_{i=1}^{N} E_\rho[< \lambda_{t-1}\delta_t X_{t-1}, e_i > |\mathcal{Y}_t]\ a_i\Gamma^i(R_t)f(R_t)$$
$$= \sum_{i=1}^{N} E_\rho[< \lambda_{t-1}\delta_t X_{t-1}, e_i > |\mathcal{Y}_{t-1}]\ a_i\Gamma^i(R_t)f(R_t) \quad \text{note that } \delta_t \text{ is } \mathcal{F}_{t-1} \text{ measurable}$$
$$= \sum_{i=1}^{N} < \sigma_{t-1}(\delta_t < X_{t-1}, e_i >)\ \Gamma^i(R_t)f(R_t)a_i$$

To proceed with $4°$, we need to work on $M_t M_t'$ first:

since $X_t = AX_{t-1} + M_t$

$$\Longrightarrow X_t X_t' = AX_{t-1}X_{t-1}'A' + M_t(AX_{t-1})' + (AX_{t-1})M_t' + M_t M_t'$$

meanwhile, $X_tX_t' = diag(X_t) = diag(AX_{t-1} + M_t) = diag(AX_{t-1}) + diag(M_t)$

therefore:

$$M_tM_t' = diag(AX_{t-1}) + diag(M_t) - AX_{t-1}X_{t-1}'A' - M_t(AX_{t-1})' - (AX_{t-1})M_t' \qquad \cdots\cdots\cdots\cdots(**)$$

and sine:

$$\begin{aligned}
&diag(AX_{t-1}) - AX_{t-1}X_{t-1}'A' \\
&= diag\left(\sum_{i=1}^{N} a_i < X_{t-1}, e_i >\right) - \left(\sum_{i=1}^{N} a_i < X_{t-1}, e_i >\right)\left(\sum_{i=1}^{N} a_i < X_{t-1}, e_i >\right)' \\
&= diag\left(\sum_{i=1}^{N} a_i < X_{t-1}, e_i >\right) - \sum_{i=1}^{N}\left(\sum_{i=1}^{N} a_ia_i' < X_{t-1}, e_i >^2\right) \\
&= diag\left(\sum_{i=1}^{N} a_i < X_{t-1}, e_i >\right) - \sum_{i=1}^{N}\left(\sum_{i=1}^{N} a_ia_i' < X_{t-1}, e_i >\right) \\
&= \sum_{i=1}^{N} (diag(a_i) - a_ia_i') < X_{t-1}, e_i > \qquad \cdots\cdots\cdots\cdots (***)
\end{aligned}$$

take $(**)$ and $(***)$ into $4°$

$$\begin{aligned}
&E_\rho[(\lambda_{t-1}\psi_t) < \beta_t, M_t > M_t|\mathcal{Y}_t] \\
&= E_\rho[(\lambda_{t-1}\psi_t)(M_tM_t')\beta_t|\mathcal{Y}_t] \\
&= E_\rho[(\lambda_{t-1}\psi_t) \\
&\qquad \cdot [diag(AX_{t-1}) + diag(M_t) - AX_{t-1}X_{t-1}'A' - M_t(AX_{t-1})' - (AX_{t-1})M_t']\beta_t|\mathcal{Y}_t] \\
&= E_\rho[(\lambda_{t-1}\psi_t)[diag(AX_{t-1}) - AX_{t-1}X_{t-1}'A']\beta_t|\mathcal{Y}_t] \\
&= E_\rho\left[(\lambda_{t-1}\psi_t) \sum_{i=1}^{N} (diag(a_i) - a_ia_i') < X_{t-1}, e_i > \beta_t \middle| \mathcal{Y}_t\right]
\end{aligned}$$

$$= \sum_{i=1}^{N} (diag(a_i) - a_i a_i') \; E_\rho[(\lambda_{t-1}\psi_t) < X_{t-1}, e_i > \beta_t | \mathcal{Y}_t]$$

$$= \sum_{i=1}^{N} (diag(a_i) - a_i a_i') \; E_\rho[\lambda_{t-1} < X_{t-1}, e_i > \beta_t | \mathcal{Y}_t] \; \Gamma^i(R_t) \qquad \text{since } \psi_t|_{(X_{t-1}=e_i)} = \Gamma^i(R_t)$$

$$= \sum_{i=1}^{N} (diag(a_i) - a_i a_i') \sigma_{t-1}(\beta_t < X_{t-1}, e_i >) \Gamma^i(R_t)$$

Finally, take these four terms back to equation ($\diamondsuit$), we finish the proof. □

Next we develop an unnormalized filter for the conditional expectation of the underlaying Markov chain $\{X_t\}$. By conditional, we mean given the observation up to time t.

Theorem 4.4.2.

$$\sigma_t(X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(X_{t-1}), e_i > \Gamma^i(R_t) a_i \tag{4.6}$$

Proof. Let $H_0 = 1$, $\alpha_t = 0, \beta_t = (0, ..., 0)'_{N\times 1}, \delta_t = 0$,

thus $H_t = H_{t-1} + \alpha_t + < \beta_t, M_t > + \delta_t f(R_) = 1$.

Note that $\alpha_t, \beta_t, \delta_t$ are all $\mathcal{F}_{t-1}$ measurable, H_t is $\mathcal{F}_t$ measurable, so we can plug them into (4.5), then this theorem is proved. □

Define $N_t^{rs} = \sum_{k=1}^{t} < X_{k-1}, e_r >< X_k, e_s >$.

We know that $< X_{k-1}, e_r >< X_k, e_s >= 1 \quad iff \quad X_{k-1} = e_r, X_k = e_s$, otherwise $< X_{k-1}, e_r >< X_k, e_s >= 0$. In another word, during the time interval $[0, t]$, we count 1 whenever X_t jumps from state e_r to state e_s, otherwise we count 0. Therefore N_t^{rs} represents the number of jumps from state e_r to state e_s during $[0, t]$. Thus we call N_t^{rs} the number of jumps. Next we develop an unnormalized filter for the conditional expectation of the number of jumps. By conditional, we mean given the observation up to time t.

Theorem 4.4.3.

$$\sigma_t(N_t^{rs}X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(N_{t-1}^{rs}X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > a_{sr}\Gamma^r(R_t)e_s \tag{4.7}$$

Proof. Let $H_0 = 0$, $\alpha_t =< X_{t-1}, e_r > a_{sr}, \beta_t =< X_{t-1}, e_r > e_s, \delta_t = 0$, thus

$$H_t = H_{t-1} + \alpha_t + < \beta_t, M_t > + \delta_t f(R_)$$

$$= H_{t-1} + < X_{t-1}, e_r > a_{sr} + < X_{t-1}, e_r >< e_s, M_t >$$

$$= H_{t-1} + < X_{t-1}, e_r >< AX_{t-1}, e_s > + < X_{t-1}, e_r >< M_t, e_s >$$

$$= H_{t-1} + + < X_{t-1}e_r >< X_t, e_s >$$

therefore, $\quad H_t = \sum_{k=1}^{t} < X_{k-1}, e_r >< X_k, e_s >= N_t^{rs}$

Note that $\alpha_t, \beta_t, \delta_t$ are all $\mathcal{F}_{t-1}$ measurable, H_t is $\mathcal{F}_t$ measurable,so we can plug them into (4.5),

$$\sigma_t(N_t^{rs}X_t) = \sum_{i=1}^{N} \left[< \sigma_{t-1}(N_{t-1}^{rs}X_{t-1}), e_i > \Gamma^i(R_t)a_i \right] \text{............ } 1^\circ$$

$$+ \sum_{i=1}^{N} \left[\sigma_{t-1}(< X_{t-1}, e_r > a_{sr} < X_{t-1}, e_i >)\Gamma^i(R_t)a_i \right] \text{............ } 2^\circ$$

$$+ \sum_{i=1}^{N} \left[(diag(a_i) - a_i a_i')\sigma_{t-1}(< X_{t-1}, e_r > e_s < X_{t-1}, e_i >)\Gamma^i(R_t) \right] \text{............ } 3^\circ$$

For $2°$, only the term with $i = r$ is left in the summation.

$$\sum_{i=1}^{N}\left[\sigma_{t-1}(< X_{t-1}, e_r > a_{sr} < X_{t-1}, e_i >)\Gamma^i(R_t)a_i\right]$$
$$= \sigma_{t-1}(< X_{t-1}, e_r > a_{sr})\Gamma^r(R_t)a_r$$
$$=< \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)a_{sr}a_r$$

For $3°$, only the term with $i = r$ is left in the summation.

$$\sum_{i=1}^{N}\left[(diag(a_i) - a_i a_i')\sigma_{t-1}(< X_{t-1}, e_r > e_s < X_{t-1}, e_i >)\Gamma^i(R_t)\right]$$
$$= (diag(a_r) - a_r a_r')\sigma_{t-1}(< X_{t-1}, e_r > e_s)\Gamma^r(R_t)$$
$$= (diag(a_r) - a_r a_r')e_s < \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)$$
$$=< \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)(a_{sr}e_s - a_{sr}a_r)$$

the last equal sign is because:

$$(diag(a_r) - a_r a_r')e_s$$
$$= \begin{pmatrix} a_{1r} & & & & \\ & \ddots & & & \\ & & a_{sr} & & \\ & & & \ddots & \\ & & & & a_{Nr} \end{pmatrix}\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} a_{1r}^2 & \cdots & a_{sr}a_{1r} & \cdots & a_{Nr}a_{1r} \\ & & \cdots & & \\ a_{1r}a_{sr} & \cdots & a_{sr}^2 & \cdots & a_{Nr}a_{sr} \\ & & \cdots & & \\ \cdots & \cdots & a_{sr}a_{Nr} & \cdots & a_{Nr}^2 \end{pmatrix}\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} 0 \\ \vdots \\ a_{sr} \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} a_{sr}a_{1r} \\ \vdots \\ a_{sr}^2 \\ \vdots \\ a_{sr}a_{Nr} \end{pmatrix}$$
$$= a_{sr}e_s - a_{sr}a_r$$

therefore

$$\sigma_t(N_t^{rs}X_t)$$
$$= \sum_{i=1}^{N} < \sigma_{t-1}(N_{t-1}^{rs}X_{t-1}), e_i > \Gamma^i(R_t)a_i$$
$$+ < \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)a_{sr}a_r + < \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)(a_{sr}e_s - a_{sr}a_r)$$
$$= \sum_{i=1}^{N} < \sigma_{t-1}(N_{t-1}^{rs}X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)a_{sr}e_s$$

□

Define $J_t^r = \sum_{k=1}^{t} < X_{k-1}, e_r >$.

We know that $< X_{k-1}, e_r >= 1 \quad iff \quad X_{k-1} = e_r$, otherwise $< X_{k-1}, e_r >= 0$. In another word, during the time interval $[0, t]$, we count 1 whenever X occupies the state e_r once, otherwise we count 0. Therefore J_t^r represents the occupation time on state e_r during $[0, t]$. Thus we call J_t^r the occupation time. Next we develop an unnormalized filter for the conditional expectation of the occupation time. By conditional, we mean given the observation up to time t.

Theorem 4.4.4.

$$\sigma_t(J_t^r X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(J_{t-1}^r X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > a_r\Gamma^r(R_t) \tag{4.8}$$

Proof. Let $H_0 = 0$, $\alpha_t =< X_{t-1}, e_r >, \beta_t = 0, \delta_t = 0$, thus

$$H_t = H_{t-1} + \alpha_t + < \beta_t, M_t > + \delta_t f(R_)) = H_{t-1} + < X_{t-1}, e_r >$$

$$\text{therefore,} \quad H_t = \sum_{k=1}^{t} < X_{k-1}, e_r >= J_t^r$$

Note that $\alpha_t, \beta_t, \delta_t$ are all $\mathcal{F}_{t-1}$ measurable, H_t is $\mathcal{F}_t$ measurable,so we can plug them into (4.5),

$$\sigma_t(J_t^r X_t)$$
$$= \sum_{i=1}^{N} \left[< \sigma_{t-1}(J_{t-1}^r X_{t-1}), e_i > \Gamma^i(R_t)a_i \right] + \sum_{i=1}^{N} \left[\sigma_{t-1}(< X_{t-1}, e_r >< X_{t-1}, e_i >)\Gamma^i(R_t)a_i \right]$$

note that for the second term, only the case where $i = r$ is left. Therefore:

$$\sigma_t(J_t^r X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(J_{t-1}^r X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)a_r$$

□

Define $G_t^r(f) = \sum_{k=1}^{t} < X_{k-1}, e_r > f_k$, where $f = (f_1, f_2, ..., f_t)$ and $f_k = f(R_k)$ is any functional of R_k.

For example, $f(R_k) = R_k$, then $G_t^r(f) = \sum_{k=1}^{t} < X_{k-1}, e_r > R_k$, denote it as $G_t^r(R)$ or $f(R_k) = R_k^2$, then $G_t^r(f) = \sum_{k=1}^{t} < X_{k-1}, e_r > R_k^2$, denote it as $G_t^r(R^2)$

Next we develop an unnormalized filter for the conditional expectation of $G_t^r(f)$, and we shall need this filter for two cases: $f(R_k) = R_k$ and $f(R_k) = R_k^2$. By conditional, we mean given the observation up to time t.

Theorem 4.4.5.

$$\sigma_t(G_t^r(f)X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(G_{t-1}^r(f)X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > a_r\Gamma^r(R_t)f(R_t) \tag{4.9}$$

Proof. Let $H_0 = 0$, $\alpha_t = 0, \beta_t = 0, \delta_t =< X_{t-1}, e_r >$, thus

$$H_t = H_{t-1} + \alpha_t + < \beta_t, M_t > + \delta_t f(R_) = H_{t-1} + < X_{t-1}, e_r > f(R_t)$$

therefore, $$H_t = \sum_{k=1}^{t} < X_{k-1}, e_r > f(R_t) = G_t^r(f)$$

Note that $\alpha_t, \beta_t, \delta_t$ are all $\mathcal{F}_{t-1}$ measurable, H_t is $\mathcal{F}_t$ measurable,so we can plug them into (4.5),

$$\sigma_t(G_t^r(f)X_t) = \sum_{i=1}^{N} \Big[< \sigma_{t-1}(G_{t-1}^r(f(R_{t-1}))X_{t-1}), e_i > \Gamma^i(R_t)a_i \Big]$$
$$+ \sum_{i=1}^{N} \Big[\sigma_{t-1}(< X_{t-1}, e_r >< X_{t-1}, e_i >)\Gamma^i(R_t)f(R_t)a_i \Big]$$

note that for the second term, only the case where $i = r$ is left. Therefore:

$$\sigma_t(G_t^r(f)X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(G_{t-1}^r(f)X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > \Gamma^r(R_t)f(R_t)a_r$$

□

4.5 EM Estimates

In the previous sections, we developed recursive filters given that the model parameters are known. In this section, we estimate these parameters by maximizing the expectation of likelihood functions.

The following lemma is prepared for developing the likelihood function of $(X_0, X_1, ..., X_t)$.

Lemma 4.5.1. *Given* $a_{sr} \quad s, r = 1, ..., N$

$$p(X_k|X_{k-1}) = \prod_{r,s=1}^{N} a_{sr}^{<X_{k-1}, e_r><X_k, e_s>}$$

Proof.

$$p(X_k|X_{k-1}) = \begin{cases} a_{11} \quad if \quad X_{k-1} = e_1, X_k = e_1 \\ ...... \\ a_{1N} \quad if \quad X_{k-1} = e_N, X_k = e_1 \\ ...... \\ a_{sr} \quad if \quad X_{k-1} = e_r, X_k = e_s \\ ...... \\ a_{N1} \quad if \quad X_{k-1} = e_1, X_k = e_N \\ ...... \\ a_{NN} \quad if \quad X_{k-1} = e_N, X_k = e_N \end{cases} = \prod_{r,s=1}^{N} a_{sr}^{<X_{k-1},e_r><X_k,e_s>}$$

□

Assume the initial probability $p(X_0)$ is π_0, and it is known.

Thus, the likelihood function of $(X_0, X_1, ..., X_t)$ given π_0 and a_{sr} is:

$$\begin{aligned} p(X_0, X_1, ..., X_t|\pi_0, a_{sr}) &= p(X_0)p(X_1|X_0) \cdot \cdots \cdot p(X_t|X_{t-1}) \\ &= \pi_0 \prod_{k=1}^{t} p(X_t|X_{t-1}) \\ &= \pi_0 \prod_{k=1}^{t} \prod_{r,s=1}^{N} a_{sr}^{<X_{k-1},e_r><X_k,e_s>} \end{aligned}$$

Similarly, given the estimates of the transition probabilities after t observations $\widehat{a_{sr}(t)}$, the likelihood function of $(X_0, X_1, ..., X_t)$ is:

$$p(X_0, X_1, ..., X_t|\pi_0, \widehat{a_{sr}(t)}) = \pi_0 \prod_{k=1}^{t} \prod_{r,s=1}^{N} \widehat{a_{sr}(t)}^{<X_{k-1},e_r><X_k,e_s>}$$

Then we can write the likelihood ratio and its natural logarithm as:

$$\xi_t = \frac{\pi_0 \prod_{k=1}^{t} \prod_{r,s=1}^{N} \widehat{a_{sr}(t)}^{<X_{k-1},e_r><X_k,e_s>}}{\pi_0 \prod_{k=1}^{t} \prod_{r,s=1}^{N} a_{sr}^{<X_{k-1},e_r><X_k,e_s>}} = \prod_{k=1}^{t} \prod_{r,s=1}^{N} \left(\frac{\widehat{a_{sr}(t)}}{a_{sr}} \right)^{<X_{k-1},e_r><X_k,e_s>}$$

$$
\begin{aligned}
ln\xi_t &= \sum_{k=1}^{t}\sum_{r,s=1}^{N} < X_{k-1}, e_r >< X_k, e_s > (ln\widehat{a_{sr}(t)}(t) - lna_{sr}) \\
&= \sum_{r,s=1}^{N} N_t^{rs}(ln\widehat{a_{sr}(t)} - lna_{sr})
\end{aligned}
$$

From the equation of the log likelihood ratio above we find it is impossible to maximize it directly since the time series of $\{X_t\}$ is not observable, so we can not treat N_t^{rs} as given information. Fortunately we know the time series of $\sigma_t(N_t^{rs}X_t)$ from proposition 4.4.3 and we can always use a normalizer to change it back to $E(N_t^{rs}|\mathcal{Y}_t)$. Therefore, instead of maximizing $ln\xi_t$, which would result in unsolvable equations, we choose to maximize the expected log likelihood ratio $E(ln\xi_t|\mathcal{Y}_t)$.

Before proceeding to first order, second order conditions, we consider the constrain $\sum_{s=1}^{N}\widehat{a_{sr}(t)} = 1$. In the next lemma, we convert this condition in another form. The purpose of this step is to pave the way for the coming maximization problem.

Lemma 4.5.2.

$$
\sum_{s=1}^{N}\widehat{a_{sr}(t)} = 1 \Longleftrightarrow \sum_{s=1}^{N} J_t^r\widehat{a_{sr}(t)} = t
$$

Thus $\sum_{s=1}^{N} E(J_t^r|\mathcal{Y}_t)\widehat{a_{sr}(t)} = t$ *is the constrain.*

The following theorem gives the estimates of transition probabilities.

Theorem 4.5.1.

$$
\widehat{a_{sr}(t)} = \frac{\sigma_t(N_t^{rs})}{\sigma_t(J_t^r)} \tag{4.10}
$$

Proof. The maximization problem is:

$$max_{\widehat{a_{sr}(t)}} \quad E(ln\xi_t|\mathcal{Y}_t)$$

$$s.t. \quad \sum_{s=1}^{N} E(J_t^r|\mathcal{Y}_t)\widehat{a_{sr}(t)} = t$$

$$\begin{aligned}
&E(ln\xi_t|\mathcal{Y}_t)\\
&= E\bigg(\sum_{r,s=1}^{N} N_t^{rs}(ln\widehat{a_{sr}(t)} - lna_{sr})\bigg|\mathcal{Y}_t\bigg)\\
&= E\bigg(\sum_{r,s=1}^{N} N_t^{rs}ln\widehat{a_{sr}(t)}\bigg|\mathcal{Y}_t\bigg) - Z(a)\\
&= \sum_{r,s=1}^{N} ln\widehat{a_{sr}(t)}E(N_t^{rs}|\mathcal{Y}_t) - Z(a)\\
&\text{where } Z(a) = E\bigg(\sum_{r,s=1}^{N} N_t^{rs}lna_{sr}(t)\bigg|\mathcal{Y}_t\bigg) \text{ is irrelevant to } \widehat{a_{sr}(t)}
\end{aligned}$$

The lagrange function is:

$$L(\lambda, \widehat{a_{sr}}) = \sum_{r,s=1}^{N} ln\widehat{a_{sr}(t)}E(N_t^{rs}|\mathcal{Y}_t) - Z(a) + \lambda\bigg(\sum_{r,s=1}^{N} \widehat{a_{sr}(t)}E(J_t^r|\mathcal{Y}_t) - t\bigg)$$

$$\begin{aligned}
f.o.c. \quad &\frac{E(N_t^{rs}|\mathcal{Y}_t)}{\widehat{a_{sr}(t)}} + \lambda E(J_t^r|\mathcal{Y}_t) = 0 \quad \forall r, s = 1, ..., N\\
&E(N_t^{rs}|\mathcal{Y}_t) = (-\lambda)E(J_t^r|\mathcal{Y}_t)\widehat{a_{sr}(t)} = (-\lambda)t\\
&\because \sum_{r,s=1}^{N} N_t^{rs} = \sum_{r,s=1}^{N}\bigg(\sum_{k=1}^{t} < X_{k-1}, e_r >< X_k, e_s >\bigg)\\
&= \sum_{k=1}^{t}\bigg(\sum_{r,s=1}^{N} < X_{k-1}, e_r >< X_k, e_s >\bigg) = \sum_{k=1}^{t} 1 = t\\
&\therefore \sum_{r,s=1}^{N} E(N_t^{rs}|\mathcal{Y}_t) = t
\end{aligned}$$

Therefore $\lambda = -1$ plug it into the $f.o.c.$

$$\widehat{a_{sr}(t)} = \frac{E(N_t^{rs}|\mathcal{Y}_t)}{E(J_t^r|\mathcal{Y}_t)} = \frac{\dfrac{\sigma_t(N_t^{rs})}{\sigma_t(1)}}{\dfrac{\sigma_t(J_t^r)}{\sigma_t(1)}} = \frac{\sigma_t(N_t^{rs})}{\sigma_t(J_t^r)}$$

□

The above theorem could be intuitively explained as: given observation up to time t, the estimated transition probability from state r to state s equals the expected number of jumps from state r to states$s$ divided by the time length of occupation on state r. Moreover, this conditional expectation could be either in the real world (under probability p) or in the fictitious world (under probability ρ).

Notice that we do not have the filters for $\sigma_t(N_t^{rs})$ or $\sigma_t(J_t^r)$, instead we have them for $\sigma_t(N_t^{rs}X_t)$ and $\sigma_t(J_t^rX_t)$, the next lemma provides the formulas to connect them.

Lemma 4.5.3.

$$\begin{aligned}
&\sigma_t(1) =< \sigma_t(X_t), I > \\
&\sigma_t(N_t^{rs}) =< \sigma_t(N_t^{rs}X_t), I > \\
&\sigma_t(J_t^r) =< \sigma_t(J_t^rX_t), I > \\
&\textit{where} \quad I_{N\times 1} = (1, 1, ..., 1)'
\end{aligned} \tag{4.11}$$

Proof.

since $\quad < X_t, I >= (0, ..., 1, ..., 0)(1, 1, ..., 1) = 1$

we have $\quad \sigma_t(1) = \sigma_t(< X_t, I >) =< \sigma_t(X_t), I >$

moreover $\quad \sigma_t(N_t^{rs}) = \sigma_t(N_t^{rs} < X_t, I >) = \sigma_t(< N_t^{rs}X_t, I >) =< \sigma_t(N_t^{rs}X_t), I >$

similarly $\quad \sigma_t(J_t^r) = \sigma_t(J_t^rs < X_t, I >) = \sigma_t(< J_t^rX_t, I >) =< \sigma_t(J_t^rX_t), I >$

□

To estimate the transition probabilities, we consider the likelihood function of $(X_0, X_1, ..., X_t)$. To estimate $g = (g_1, ..., g_N)'$ and $v = (v_1, ..., v_N)'$, we use the EM algorithm on the likelihood function of $(R_1, R_2, ..., R_t)$.

Since $R_t =< g, X_{t-1} > + < v, X_{t-1} > b_t \quad \sim \quad N(< g, X_{k-1}, < v, X_{t-1} >^2) \quad i.i.d.$

For $\forall k = 1, ..., t \quad p(R_k|X_{k-1}, g, v) = \frac{1}{\sqrt{2\pi} < v, X_{k-1} >} exp\Big(-\frac{(R_k - < g, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big).$

Therefore, the likelihood function of $(R_1, ..., R_t)$ given $\mathcal{X}_{t-1}$, g, and v is:

$$\begin{aligned} p(R_1, ..., R_t|\mathcal{X}_{t-1}, g, v) &= p(R_1, ..., R_t|X_0, ..., X_{t-1}, g, v) \\ &= p(R_1|X_0, g, v) \cdot \cdots \cdot p(R_t|X_{t-1}, g, v) \\ &= \prod_{k=1}^{t} \frac{1}{\sqrt{2\pi} < v, X_{k-1} >} exp\Big(-\frac{(R_k - < g, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big) \end{aligned}$$

In order to estimate g, we fix v first, the likelihood function of $(R_1, ..., R_t)$ given $\mathcal{X}_{t-1}$, $\widehat{g(t)}$, and v is:

$$p(R_1, ..., R_t|\mathcal{X}_{t-1}, \widehat{g(t)}, v) = \prod_{k=1}^{t} \frac{1}{\sqrt{2\pi} < v, X_{k-1} >} exp\Big(-\frac{(R_k - < \widehat{g(t)}, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big)$$

So we can write the likelihood ratio and its natural logarithm as:

$$\xi_t = \prod_{k=1}^{t} exp\Big[-\frac{(R_k - < \widehat{g(t)}, X_{k-1} >)^2}{2 < v, X_{k-1} >^2} + \frac{(R_k - < g, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big]$$

$$\begin{aligned} ln\xi_t &= \sum_{k=1}^{t} \Big[-\frac{(R_k - < \widehat{g(t)}, X_{k-1} >)^2}{2 < v, X_{k-1} >^2} + \frac{(R_k - < g, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big] \\ &= \sum_{k=1}^{t} \frac{2R_k < \widehat{g(t)}, X_{k-1} > - < \widehat{g(t)}, X_{t-1} >^2 - 2R_k < g, X_{k-1} > + < g, X_{t-1} >^2}{2 < v, X_{k-1} >^2} \end{aligned}$$

$$= \sum_{k=1}^{t} \Bigg(\sum_{r=1}^{N} < X_{k-1}, e_r > \Bigg)$$

$$\cdot \frac{2R_k < \widehat{g(t)}, X_{k-1} > - < \widehat{g(t)}, X_{t-1} >^2 - 2R_k < g, X_{k-1} > + < g, X_{t-1} >^2}{2 < v, X_{k-1} >^2}$$

$$= \sum_{k=1}^{t} \Bigg(\sum_{r=1}^{N} < X_{k-1}, e_r > \Bigg)$$

$$\cdot \frac{2R_k < \widehat{g(t)}, e_r > - < \widehat{g(t)}, e_r >^2 - 2R_k < g, e_r > + < g, e_r >^2}{2 < v, e_r >^2}$$

$$= \sum_{k=1}^{t} \Bigg(\sum_{r=1}^{N} < X_{k-1}, e_r > \Bigg) \cdot \frac{2R_k \widehat{g_r(t)} - \widehat{g_r(t)}^2 - 2R_k g_r + g_r^2}{2v_r^2}$$

For the same reason, we can not maximize the $ln\xi_t$, since $\{X_t\}$ is not observable, instead we shall maximize the expectation of $ln\xi_t$. The following theorem gives the estimates of g.

Theorem 4.5.2.

$$\widehat{g_r(t)} = \frac{\sigma_t(G_t^r(R))}{\sigma_t(J_t^r)} \tag{4.12}$$

Proof. The maximization problem is:

$$max_{\widehat{g_r(t)}} \quad E(ln\xi_t | \mathcal{Y}_t)$$

$$max_{\widehat{g_r(t)}} \quad E\Bigg[\sum_{k=1}^{t} \Bigg(\sum_{r=1}^{N} < X_{k-1}, e_r > \Bigg) \cdot \frac{2R_k \widehat{g_r(t)} - \widehat{g_r(t)}^2 - 2R_k g_r + g_r^2}{2v_r^2} \Bigg| \mathcal{Y}_t \Bigg]$$

$$f.o.c. \quad E\Bigg[\sum_{k=1}^{t} < X_{k-1}, e_r > \cdot \frac{2R_k - 2\widehat{g_r(t)}}{2v_r^2} \Bigg| \mathcal{Y}_t \Bigg] = 0$$

therefore,

$$E\Bigg(\sum_{k=1}^{t} < X_{k-1}, e_r > R_k \Bigg| \mathcal{Y}_t \Bigg) = E\Bigg(\sum_{k=1}^{t} < X_{k-1}, e_r > \widehat{g_r(t)} \Bigg| \mathcal{Y}_t \Bigg)$$

$$= \widehat{g_r(t)} E\Bigg(\sum_{k=1}^{t} < X_{k-1}, e_r > \Bigg| \mathcal{Y}_t \Bigg)$$

$$E(G_t^r(R) | \mathcal{Y}_t) = \widehat{g_r(t)} E(J_t^r | \mathcal{Y}_t)$$

thus,

$$\widehat{g_r(t)} = \frac{E(G_t^r(R)|\mathcal{Y}_t)}{E(J_t^r|\mathcal{Y}_t)} = \frac{\dfrac{\sigma_t(G_t^r(R))}{\sigma_t(1)}}{\dfrac{\sigma_t(J_t^r)}{\sigma_t(1)}} = \frac{\sigma_t(G_t^r(R))}{\sigma_t(J_t^r)}$$

□

Where,

$$\sigma_t(G_t^r(R)) =< \sigma_t(G_t^r(R)X_t), I > \quad \text{and } I_{N\times 1} = (1, 1, ..., 1)' \tag{4.13}$$

In order to estimate v, we fix g at this time, take $g_r = \widehat{g_r(t)}$, the likelihood function of $(R_1, ..., R_t)$ given $\mathcal{X}_{t-1}$, $\widehat{v(t)}$, and g is:

$$p(R_1, ..., R_t|\mathcal{X}_{t-1}, \widehat{v(t)}, g) = \prod_{k=1}^{t} \frac{1}{\sqrt{2\pi} < \widehat{v(t)}, X_{k-1} >} exp\Big(- \frac{(R_k - < g, X_{k-1} >)^2}{2 < \widehat{v(t)}, X_{k-1} >^2}\Big)$$

So we can write the likelihood ratio and its natural logarithm as:

$$\xi_t = \prod_{k=1}^{t} \frac{< v, X_{t-1} >}{< \widehat{v(t)}, X_{t-1} >} exp\Big[- \frac{(R_k - < g, X_{k-1} >)^2}{2 < \widehat{v(t)}, X_{k-1} >^2} + \frac{(R_k - < g, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big]$$

$$\begin{aligned}
ln\xi_t &= \sum_{k=1}^{t} \Big[ln < v, X_{k-1} > - ln < \widehat{v(t)}, X_{k-1} > \\
&\quad - \frac{(R_k - < g, X_{k-1} >)^2}{2 < \widehat{v(t)}, X_{k-1} >^2} + \frac{(R_k - < g, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big] \\
&= \sum_{k=1}^{t} \Big(\sum_{r=1}^{N} < X_{k-1}, e_r > \Big) \\
&\quad \cdot \Big[ln < v, X_{k-1} > - ln < \widehat{v(t)}, X_{k-1} > - \frac{(R_k - < g, X_{k-1} >)^2}{2 < \widehat{v(t)}, X_{k-1} >^2} + \frac{(R_k - < g, X_{k-1} >)^2}{2 < v, X_{k-1} >^2}\Big] \\
&= \sum_{k=1}^{t} \Big(\sum_{r=1}^{N} < X_{k-1}, e_r > \Big) \\
&\quad \cdot \Big[ln < v, e_r > - ln < \widehat{v(t)}, e_r > - \frac{(R_k - < g, e_r >)^2}{2 < \widehat{v(t)}, e_r >^2} + \frac{(R_k - < g, e_r >)^2}{2 < v, e_r >^2}\Big]
\end{aligned}$$

$$= \sum_{k=1}^{t} \left(\sum_{r=1}^{N} < X_{k-1}, e_r > \right) \cdot \left[ln v_r - ln \widehat{v_r(t)} - \frac{(R_k - g_r)^2}{2\widehat{v_r(t)}^2} + \frac{(R_k - g_r)^2}{2v_r^2} \right]$$

We obviously can not maximize the $ln\xi_t$ directly, since $\{X_t\}$ is not observable, instead we shall maximize the expectation of $ln\xi_t$. The following theorem gives the estimates of v.

Theorem 4.5.3.

$$\widehat{v_r(t)}^2 = \frac{\sigma_t(G_t^r(R^2)) - 2\widehat{g_r(t)}\sigma_t(G_t^r(R)) + \widehat{g_r(t)}^2 \sigma_t(J_t^r)}{\sigma_t(J_t^r)} \tag{4.14}$$

Proof. The maximization problem is:

$$max_{\widehat{v_r(t)}} \quad E(ln\xi_t | \mathcal{Y}_t)$$

$$max_{\widehat{v_r(t)}} \quad E\left[\sum_{k=1}^{t} \left(\sum_{r=1}^{N} < X_{k-1}, e_r > \right) \cdot \left(ln v_r - ln \widehat{v_r(t)} - \frac{(R_k - g_r)^2}{2\widehat{v_r(t)}^2} + \frac{(R_k - g_r)^2}{2v_r^2} \right) \Big| \mathcal{Y}_t \right]$$

$$f.o.c. \quad E\left[\sum_{k=1}^{t} < X_{k-1}, e_r > \cdot \left(\frac{-1}{\widehat{v_r(t)}} + \frac{(R_k - g_r)^2}{\widehat{v_r(t)}^3} \right) \Big| \mathcal{Y}_t \right] = 0$$

therefore,

$$\frac{1}{\widehat{v_t(t)}^3} E\left(\sum_{k=1}^{t} < X_{k-1}, e_r > (R_k^2 - 2R_k g_r + g_r^2) \Big| \mathcal{Y}_t \right) = \frac{1}{\widehat{v_t(t)}} E\left(\sum_{k=1}^{t} < X_{k-1}, e_r > \Big| \mathcal{Y}_t \right)$$

$$\frac{1}{\widehat{v_t(t)}^3} \left(E(G_t^r(R^2)|\mathcal{Y}_t) - 2g_r E(G_t^r(R)|\mathcal{Y}_t) + g_r^2 E(J_t^r|\mathcal{Y}_t) \right) = \frac{1}{\widehat{v_t(t)}} E(J_t^r|\mathcal{Y}_t)$$

$$\widehat{v_r(t)}^2 = \frac{E(G_t^r(R^2)|\mathcal{Y}_t) - 2g_r E(G_t^r(R)|\mathcal{Y}_t) + g_r^2 E(J_t^r|\mathcal{Y}_t)}{E(J_t^r|\mathcal{Y}_t)}$$

$$= \frac{\sigma_t(G_t^r(R^2)) - 2g_r \sigma_t(G_t^r(R)) + g_r^2 \sigma_t(J_t^r)}{\sigma_t(J_t^r)}$$

Note that we fix $g_r = \widehat{g_r(t)}$ at the begin when we try to estimate v, therefore:

$$\widehat{v_r(t)}^2 = \frac{\sigma_t(G_t^r(R^2)) - 2\widehat{g_r(t)}\sigma_t(G_t^r(R)) + \widehat{g_r(t)}^2 \sigma_t(J_t^r)}{\sigma_t(J_t^r)}$$

$\square$

Where,

$$\sigma_t(G_t^r(R^2)) =< \sigma_t(G_t^r(R^2)X_t), I > \quad \text{and } I_{N\times 1} = (1, 1, ..., 1)' \tag{4.15}$$

Up to here, with the estimated model specifications, A, g, and σ, we can simulate the underlying Markov chain $\{X_t\}$ by: $E(X_t|\mathcal{Y}_t) = \dfrac{\sigma_t(X_t)}{\sigma_t(1)}$

Moreover, we can estimate:

(1) the number of jumps from the r^{th} state to the s^{th} state within the time interval $[0\ ,\ t]$:

$$E(N_t^{rs}|\mathcal{Y}_t) = \frac{\sigma_t(N_t^{rs})}{\sigma_t(1)}$$

(2) the occupation time on the r^{th} state within the time interval $[0\ ,\ t]$:

$$E(J_t^r|\mathcal{Y}_t) = \frac{\sigma_t(J_t^r)}{\sigma_t(1)}$$

Suppose we have observation up to time T and we are going to have a data length of n for each pass of iteration. We start with any initial guess $\Theta^0 = \{A^0, g^0, v^0\}$. The first pass is done after having n observation $\{R_1, ..., R_n\}$ and we get $\Theta^1 = \{A^1, g^1, v^1\}$. They become the initials of the second pass where we have observation $\{R_{n+1}, ..., R_{2n}\}$, and we get $\Theta^2 = \{A^2, g^2, v^2\}$ ...... Our iteration is essentially similar to the spirit of the EM algorithm we presented in chapter one, namely, estimate Θ^{k+1} from Θ^k, where Θ stands for $\{A, g, v\}$ in our case, and Θ^k means the estimates from the k^{th} pass. However, they are not exactly the same, instead of using the same observation sequence $\{R_1, R_2, ..., R_T\}$ for all the passes, we update the observation sequence as well, i.e., $\{R_1, ..., R_n\}$ for the first pass, $\{R_{n+1}, ..., R_{2n}\}$ for the second pass...... The idea is to seek the quality of information rather than just the quantity, as the recent data are more relevant in terms of predicting the near future. In our iteration, not only the parameters but also the observation get updated in each pass.

4.6 Price Prediction

In this section we develop a formula for price prediction.

First let us review a lemma.

Lemma 4.6.1.

$$\textit{if } b \sim N(0,1), \textit{ we have } E(e^{v \cdot b}) = e^{v^2/2}$$

Proof.

$$\begin{aligned} E(e^{vz}) &= \int_{-\infty}^{\infty} e^{vz} \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z-v)^2}{2}} e^{\frac{v^2}{2}} dz \\ &= e^{\frac{v^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \quad \text{(change variable, } y = z - v\text{)} \\ &= e^{\frac{v^2}{2}} \end{aligned}$$

□

Based on the information up to the current trading day, we obtain estimates of A, g, and v by (4.10) to (4.14). Then we can predict the price of the next trading day by using the estimates of A, g, v and the price of the current trading day. The formula is given by the theorem below.

Theorem 4.6.1.

$$E(S_{n+1}|\mathcal{Y}_n, g, v, A) = S_n \sum_{i=1}^{N} e^{g_i} e^{\sigma_i^2/2} \left\langle \frac{\sigma_n(X_n)}{\sigma_n(1)} \ , \ e_i \right\rangle \tag{4.16}$$

$$\textit{where} \quad \sigma_n(X_n) = \sum_{i=1}^{N} < \sigma_{n-1}(X_{n-1}) \ , \ e_i > \cdot \frac{\varphi(\dfrac{R_n - g_i}{v_i})}{v_i \cdot \varphi(R_n)} \cdot A \cdot e_i$$

φ is the probability density function of a standard normal distribution

$$\textit{and} \qquad \sigma_n(1) =< \sigma_n(X_n), I >$$

Proof. since $S_{n+1} = S_n e^{R_{n+1}}$

$$E(S_{n+1}|\mathcal{Y}_n, g, v, A)$$

$$= S_n \cdot E\left(e^{R_{n+1}} \middle| \mathcal{Y}_n, g, v, A \right)$$

$$= S_n \cdot E\left(\sum_{i=1}^{N} < X_n, e_i > e^{<g,X_n>+<v,X_n>b_{n+1}} \middle| \mathcal{Y}_n, g, v, A \right)$$

$$= S_n \cdot E\left(\sum_{i=1}^{N} < X_n, e_i > e^{<g,e_i>+<v,e_i>b_{n+1}} \middle| \mathcal{Y}_n, g, v, A \right)$$

$$= S_n \cdot \sum_{i=1}^{N} E\left(< X_n, e_i > \middle| \mathcal{Y}_n, g, v, A \right) E\left(e^{g_i} \middle| \mathcal{Y}_n, g, v, A \right) E\left(e^{v_i b_{n+1}} \middle| \mathcal{Y}_n, g, v, A \right)$$

since g_i, v_i are independent of $\{\mathcal{Y}_n, g, v, A\}$, and $b_i \sim N(0,1)$ *i.i.d.*

$$E\left(e^{g_i} \middle| \mathcal{Y}_n, g, v, A \right) = e^{g_i}$$

$$E\left(e^{v_i b_{n+1}} \middle| \mathcal{Y}_n, g, v, A \right) = e^{v_i b_{n+1}} = e^{v_i^2/2}$$

as to the first expectation

$$E\Big(< X_n, e_i > \Big| \mathcal{Y}_n, g, v, A \Big) = \Big\langle E(X_n \Big| \mathcal{Y}_n, g, v, A) \ , \ e_i \Big\rangle$$
$$= \Big\langle \frac{\sigma_n(X_n)}{\sigma_n(1)} \ , \ e_i \Big\rangle$$

thus, we get

$$E(S_{n+1}|\mathcal{Y}_n, g, v, A) = S_n \sum_{i=1}^{N} e^{g_i} e^{\sigma_i^2/2} \Big\langle \frac{\sigma_n(X_n)}{\sigma_n(1)} \ , \ e_i \Big\rangle$$

from (4.6), we get

$$\sigma_n(X_n) = \sum_{i=1}^{N} < \sigma_{n-1}(X_{n-1}) \ , \ e_i > \cdot \frac{\varphi(\frac{R_n - g_i}{v_i})}{v_i \cdot \varphi(R_n)} \cdot A \cdot e_i$$

form (4.11), we get

$$\sigma_n(1) = < \sigma_n(X_n), I >$$

□

After obtaining the estimates of $A = (a_{sr})_{s,r=1,\ldots,N}$ from (4.10), $g = \{g_r\}_{r=1,\ldots,N}$ from (4.12), and $v = \{v_r\}_{r=1,\ldots,N}$ from (4.14), we plug them in (4.17) to predict the price of the next trading day.

Later on in section 5.5, we shall compare the prediction performance of HMM and GARCH(1,1). The predicted price from GARCH(1,1) is obtained in this way: $E(S_{n+1}|\mathcal{Y}_n, g, v) = S_n e^g e^{\sigma^2/2}$ where, g and v are the drift and volatility forecasts by GARCH(1,1).

Chapter 5

Model Implementation

In this chapter we start with the Monte Carlo Simulation Analysis followed by the statistic analysis of the prediction performance with real data sets. Then we talk about how the data sets are cautiously selectly and some finding within the data sets. Next we give a general outline of our empirical finds with HMM. Finally, we compare the prediction performance of HMM with GARCH(1,1).

5.1 Monte Carlo Simulation Analysis

In the Monte Carlo simulation analysis, we assume the length of observation process is 1398[1] and the dimension for the state space is 2, thus the state space for the underlying Markov chain $\{X_t\}$ is $\Sigma = \{e_1, e_2\}$, where $e_1 = (1,0)'$, $e_2 = (0,1)'$. Suppose we know the initial distribution π_0. Then we fix the transition probability matrix $A_{2\times 2}$, the state space of the drift $g = (g_1, g_2)'$, and the state space of the volatility $v = (v_1, v_2)'$ as the "true" values.

[1]We select such a length because the length of each real data set we shall use later is 1398

First of all, we generate a sequence of *i.i.d.* random variables with Gaussian distribution $\{b_t\}_{t=1}^{1398}$, then by using the "true" value of A we generate a Markov chain $\{X_t\}_{t=0}^{1397}$. Next, with this generated $\{X_t\}$ and the "true" values of g and v, we simulate a sequence of 1398 daily returns by: $R_{t+1} =< g, X_t > + < v, X_t > b_{t+1}$

Finally, we take this simulated sequence as the observation process and use it to estimate the model, namely, A, g, v. Ideally, the estimates should resemble the "true" values we fixed at the beginning.

Let us summarize all the equations that will be used in the model implementation.

$$\Gamma^i(R_t) = \frac{\varphi\left(\dfrac{R_t - g_i}{v_i}\right)}{v_i \cdot \varphi(R_t)} \quad a_i = A \cdot e_i \quad I_{N\times 1} = (1, 1, ..., 1)' \tag{5.1}$$

$$\sigma_t(X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(X_{t-1}), e_i > \Gamma^i(R_t) a_i \qquad \sigma_t(1) =< \sigma_t(X_t), I > \tag{5.2}$$

$$\begin{aligned} &\sigma_t(N_t^{rs} X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(N_{t-1}^{rs} X_{t-1}), e_i > \Gamma^i(R_t) a_i + < \sigma_{t-1}(X_{t-1}), e_r > a_{sr} \Gamma^r(R_t) e_s \\ &\sigma_t(N_t^{rs}) =< \sigma_t(N_t^{rs} X_t), I > \end{aligned} \tag{5.3}$$

$$\begin{aligned} &\sigma_t(J_t^r X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(J_{t-1}^r X_{t-1}), e_i > \Gamma^i(R_t) a_i + < \sigma_{t-1}(X_{t-1}), e_r > a_r \Gamma^r(R_t) \\ &\sigma_t(J_t^r) =< \sigma_t(J_t^r X_t), I > \end{aligned} \tag{5.4}$$

$$\sigma_t(G_t^r(R)X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(G_{t-1}^r(R)X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > a_r\Gamma^r(R_t)R_t$$

$$\sigma_t(G_t^r(R)) =< \sigma_t(G_t^r(R)X_t), I >$$

$$\sigma_t(G_t^r(R^2)X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(G_{t-1}^r(R^2)X_{t-1}), e_i > \Gamma^i(R_t)a_i + < \sigma_{t-1}(X_{t-1}), e_r > a_r\Gamma^r(R_t)R_t^2$$

$$\sigma_t(G_t^r(R^2)) =< \sigma_t(G_t^r(R^2)X_t), I > \tag{5.5}$$

$$\widehat{a_{sr}(t)} = \frac{\sigma_t(N_t^{rs})}{\sigma_t(J_t^r)} \tag{5.6}$$

$$\widehat{g_r(t)} = \frac{\sigma_t(G_t^r(R))}{\sigma_t(J_t^r)} \tag{5.7}$$

$$\widehat{v_r(t)}^2 = \frac{\sigma_t(G_t^r(R^2)) - 2\widehat{g_r(t)}\sigma_t(G_t^r(R)) + \widehat{g_r(t)}^2\sigma_t(J_t^r)}{\sigma_t(J_t^r)} \tag{5.8}$$

$$\begin{aligned} E(N_t^{rs}|\mathcal{Y}_t) &= \frac{\sigma_t(N_t^{rs})}{\sigma_t(1)} \\ E(J_t^r|\mathcal{Y}_t) &= \frac{\sigma_t(J_t^r)}{\sigma_t(1)} \end{aligned} \tag{5.9}$$

The first pass of estimation is done after 56 observation $\{R_1, ..., R_{56}\}$. We start with some randomly picked A^0, g^0, and v^0. And Let $\sigma_0(X_0) = X_0 = \pi_0^1 e_1 + \pi_0^2 e_2$, $\sigma_0(N_0^{rs}X_0) = \sigma_0(J_0^r X_0) = \sigma_0(G_0^r(R)X_0) = \sigma_0(G_0^r(R^2)X_0) = 1$. Next we simulate the filters by (5.1) – (5.5), then plug them into (5.6) – (5.8). In this way we obtain the estimates after the first pass: A^1, g^1, and v^1. They becomes the initials of the second pass where we have observation $\{R_{57}, ..., R_{112}\}$, and we perform the same estimation as in the first pass to obtain the estimates after the second pass: A^2, g^2, and v^2. ...... Since the length of observation is 1398, and the length of each pass is 56, we have repeated 24 passes and the final estimates are: A^{24}, g^{24}, and v^{24}.

After obtaining estimates of $\{A, g, v\}$, we use (5.9) to estimate the number of jumps within the last pass and the occupation time within the last pass. In addition, we compare them with the real number of jumps and occupation time.

The following result shows that our estimates closely resemble the true values.

Simulation Result
Data Length=1398
Data length of each pass=56
24 Passes

Drift		
True:	-0.07	0.07
Estimate:	-0.08	0.0699

Volatility		
True:	0.06	0.02
Estimate:	0.0611	0.0204

Transition Prob		
True:	0.1	0.6
	0.9	0.4
Estimate:	0.163	0.6702
	0.837	0.3298

Number of Jumps within a pass		
True:	3	20
	21	11
Estimate:	3.0412	20.7447
	20.9212	10.2929

Occupation Time with a Pass		
True:	24	32
Estimate:	24.7859	31.2141

5.2 Statistic Analysis of the Prediction Performance

We implement our hidden Markov model on 73 data sets of historical price processes. The Predicted Prices are achieved by (4.16).

$$E(S_{t+1}|\mathcal{Y}_t, g, v, A) = S_n \sum_{i=1}^{N} e^{g_i} e^{\sigma_i^2/2} \left\langle \frac{\sigma_t(X_t)}{\sigma_t(1)} \ , \ e_i \right\rangle$$

where e_i is the unit vector

$$\sigma_t(X_t) = \sum_{i=1}^{N} < \sigma_{t-1}(X_{t-1}) \ , \ e_i > \cdot \frac{\varphi(\dfrac{R_t - g_i}{v_i})}{v_i \cdot \varphi(R_t)} \cdot A \cdot e_i$$

$$\sigma_t(1) =< \sigma_t(X_t), I >$$

The prediction performance of our model is judged by the following regression:

$$\text{Actual Price}_i = \alpha + \beta \cdot \text{Predicted Price}_i + \varepsilon_i \tag{5.10}$$

We report the following variables in the output table: estimates of Drift, Volatility, Transition Probability (Appendix A); estimates of the regression coefficients α and β; as well as the relative α, standard error of regression (s), as well as the relative standard error, Durbin-Watson statistics (DW), in addition, t-statistics and p-values for the coefficients are also included (Appendix B).

First of all, the regression result is assessed on a basis of the 3 criteria for a good model proposed by Fama and Gibbons (1984):

(1) serially uncorrelated residuals

(2) a low standard error of regression

(3) conditional unbiasedness, i.e., the intercept α should be close to zero, and the regression coefficient β should be close to one.

We use Durbin-Watson (DW) Statistics to check whether the residuals are serially uncorrelated. The DW stat is defined as:

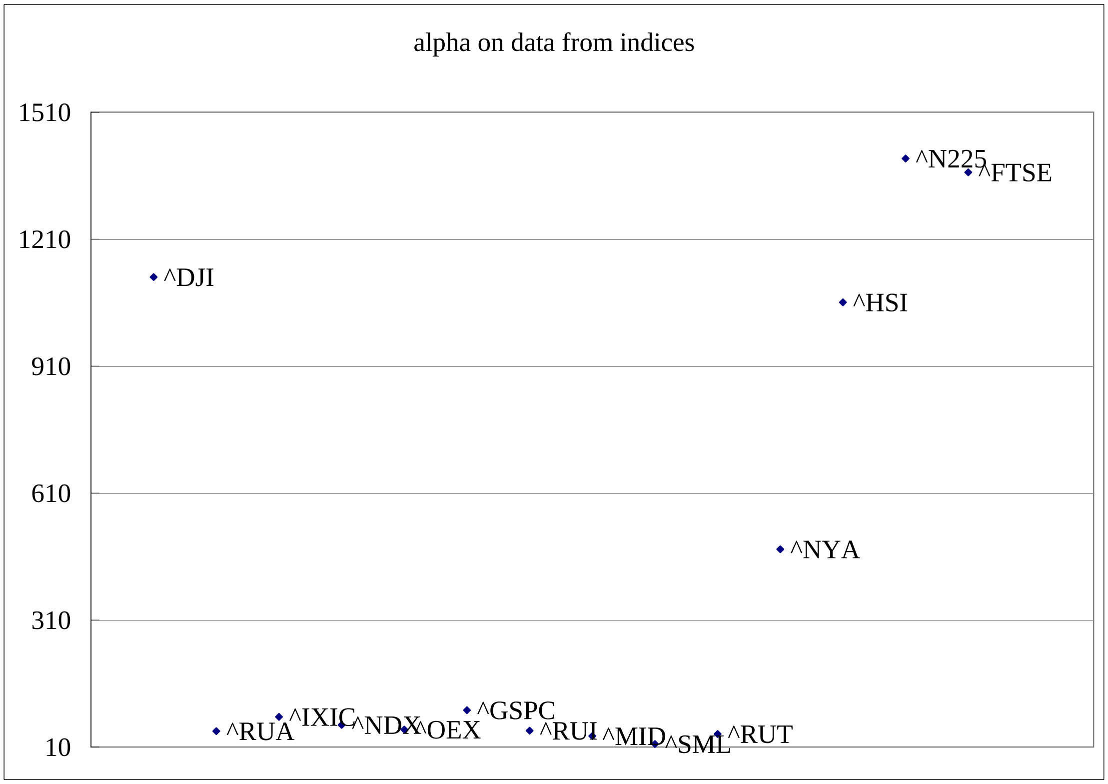
$$DW = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2} = \frac{\sum_{i=2}^{n}\varepsilon_i^2 + \sum_{i=2}^{n}\varepsilon_{i-1}^2 - 2\sum_{i=2}^{n}\varepsilon_i\varepsilon_{i-1}}{\sum_{i=1}^{n}\varepsilon_i^2}$$

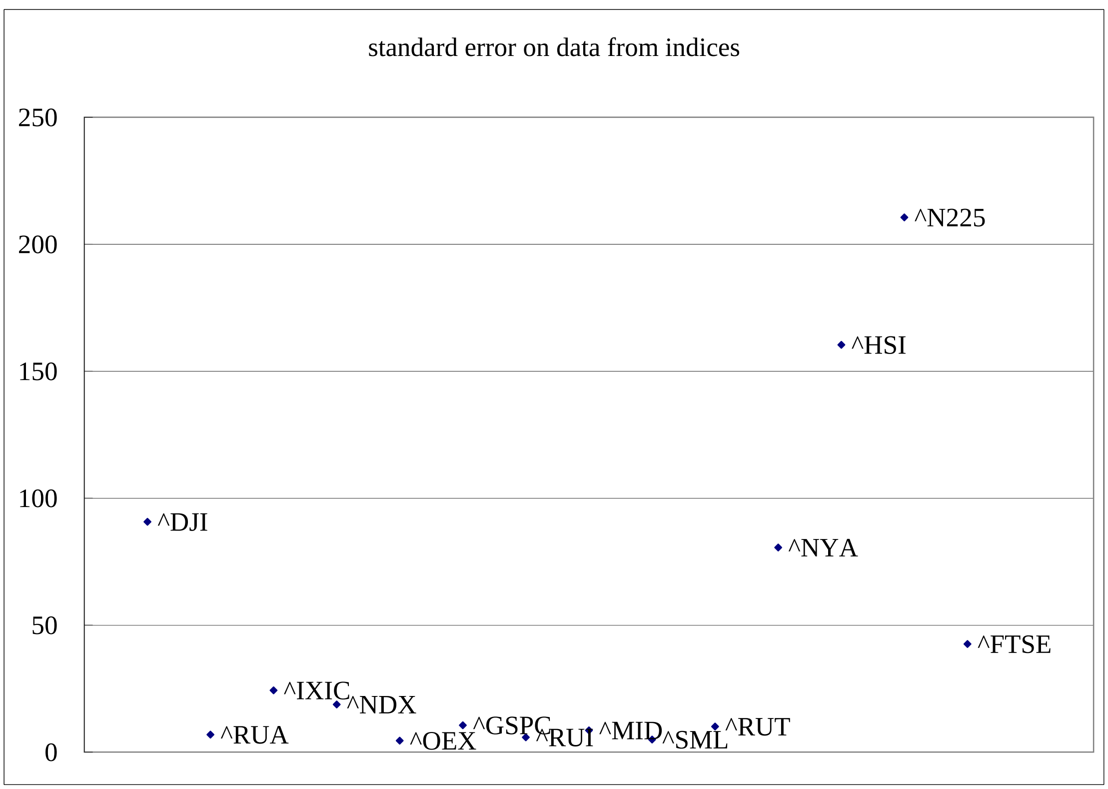
When n is large enough, both $\sum_{i=2}^{n}\varepsilon_i^2$ and $\sum_{i=1}^{n-1}\varepsilon_i^2 \to \sum_{i=1}^{n}\varepsilon_i^2$, in addition, if residuals are serially uncorrelated, $\sum_{i=2}^{n} \varepsilon_i \varepsilon_{i-1} \rightarrow 0$. Therefore $DW \rightarrow 2$ indicates that residuals are serially uncorrelated.

A regression is considered as a good fit if the standard error of the regression (s) is close to zero. We obtain s in this way: $s = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{df}}$ where ε_i = data - fit $= Price_i - \widehat{Price}_i$; and df is the degree of freedom, computed by subtracting the number of terms included in the regression (which is 2 in our case) from the number of data points.

Hoping that the predicted price closely resembles the actual price, it is quite nature for us to expect that α is close to zero and β is close to 1. The problem is "how close is close enough". For instance, 0.01 is close to zero compared with 1, but compared with 10,000 why can we think 100 is close to zero? It is a problem of relatively close to zero.

A quick glance at the plots of α and the standard error of regression (s), we see a few points that are way above zero. Therefore, special attention is paid to those points and their similarities are drawn to the next table. We found that α and s are extraordinarily high with the indices and the actual prices for indices are substantially larger than the others.


alpha on data from indices
1510
1210
910
610
310
10
^N225
^FTSE
^DJI
^HSI
^NYA
^GSPC
^IXIC
^NDX
^OEX
^RUA
^RUI
^MID
^SML
^RUT


standard error on data from indices
250
200
150
100
50
0
^N225
^HSI
^DJI
^NYA
^FTSE
^IXIC
^NDX
^RUA
^OEX
^GSPC
^RUI
^MID
^SML
^RUT

Index		Average Price	alpha	standard error
^DJI	DOW JONES INDUSTRIAL AVERAGE IN	11146.27	1120.2	90.71440
^RUA	RUSSELL 3000 INDEX	738.1819596	47.4651	6.85640
^IXIC	NASDAQ COMPOSITE	2179.66756	81.3016	24.31200
^NDX	NASDAQ-100 (DRM)	1589.556122	62.3106	18.77580
^OEX	S&P 100 INDEX,RTH	581.2536806	51.3247	4.53120
^GSPC	S&P 500 INDEX,RTH	1271.124672	96.8597	10.53510
^RUI	RUSSELL 1000 INDEX	690.9786325	48.5067	5.80320
^MID	S&P 400 MIDCAP INDEX	761.7627119	35.9552	8.57450
^SML	S&P 600 SMALLCAP INDEX	373.6832972	17.2268	4.95650
^RUT	RUSSELL 2000 INDEX	719.9188713	40.8194	10.04050
^NYA	NYSE COMPOSITE INDEX (NEW METHO	8175.071918	477.42420	80.59000
^HSI	HANG SENG INDEX	16316.92308	1060.60000	160.38850
^N225	NIKKEI 225	15257.08061	1400.60000	210.53720
^FTSE	FTSE 100	5864.922813	1367.70000	42.56620

Go back to the original regression: $\text{Actual Price}_i = \alpha + \beta \cdot \text{Predicted Price}_i + \varepsilon_i$, compared with these large actual prices, their corresponding α, although still large numerically, can be viewed as "close to zero". Prompted by this observation, we divide both side of (5.10) by the mean of the actual prices ($m.ac.$),

$$\frac{\text{Actual Price}_i}{m.ac.} = \frac{\alpha}{m.ac.} + \beta \cdot \frac{\text{Predicted Price}_i}{m.ac.} + \frac{\varepsilon_i}{m.ac.} \tag{5.11}$$

We shall call $\dfrac{\alpha}{m.ac.}$, the new constant term, 'relative α'.

Hoping that the $\dfrac{\text{Predicted Price}_i}{m.ac.}$ can closely resemble the $\dfrac{\text{Actual Price}_i}{m.ac.}$, we expect that relative alpha is close to zero.

Meanwhile, for (5.11), a brief computation suggests that the standard error of this regression is actually $\dfrac{s}{m.ac.}$, which we shall call it 'relative standard error'.

$$\sqrt{\frac{\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{m.ac.}\right)^2}{df}} = \frac{1}{m.ac.}\sqrt{\frac{\sum_{i=1}^{n}\varepsilon_i^2}{df}} = \frac{s}{m.ac.}$$

As a modification of Fama and Gibbons (1984), we think the relative α is a better criterion than α as it can eliminate the effect caused by the magnitude of the actual prices and the same applies to the relative standard error. In another word, as long as the relative α (relative standard error) is close to zero numerically, we consider α (standard error) as close enough to zero.

In addition to the estimates of α and β, we also seek information about the precision of the estimates. Are there any variables removable from the regression? So the t-stats and P-values for α and β under the null hypothesis that the corresponding coefficient equals zero are also reported in the output table. Obviously if the hypothesis that a coefficient is zero can not be rejected statically, we consider this variable as removable.

For a regression model $Y = \alpha + \beta \cdot X$, under the null hypothesis $H_0 : \beta = 0$, we know that $t_{\beta=0} = \dfrac{\hat{\beta} - 0}{\text{standard error of } \beta}$ is a value of a random variable having the t-distribution with n-2 (as we have 2 terms, thus 2 coefficients to be estimated) degrees of freedom. Similarly, under the null hypothesis $H_0 : \alpha = 0$, the t-stat for α is $t_{\alpha=0} = \dfrac{\hat{\alpha} - 0}{\text{standard error of } \alpha}$. These t-statistics, coefficient estimates divided by their respective standard errors, are often used to test the hypothesis that the true value of the coefficient is zero, in another word, if the variable is significant. The t-distribution resembles the standard normal distribution, with a somehow “fatter tail”, i.e., relatively more extreme values. However, the difference between the t

and the standard normal is negligible if the number of degrees of freedom is more than 30, which is indeed satisfied in our regression, where the number of degrees of freedom is 54. In a standard normal distribution, only 5% of the values fall outside the plus-or-minus 2 range, i.e., a t-statistics larger than 2 in magnitude would only have a 5% or smaller change of happening under the assumption that the true coefficient is zero. Hence, there is the commonest rule-of-thumb in this regard: if its t-stat is greater than 2 in absolute value, we reject the null hypothesis and the corresponding variable could be removed without seriously affecting the standard error of the model. But this rule does not serve as a basis for deciding whether or not to include the constant term, as the constant term is usually included for priori reason(s).

The results in appendix B (4)-(6) suggest no doubt that β is significant, actually we are more concerned about whether β is significantly 1. So here comes another null hypothesis to test: $H_0 : \beta = 1$. The corresponding t-stat is

$$t_{\beta=1} = \frac{\hat{\beta} - 1}{\text{standard error of } \beta}.$$

The P-value is the probability that a t-distributed variable is larger than the corresponding t-statistics given the null hypothesis is true. Specifically in our case, $P_{\alpha=0} = Prob(t > |t_{\alpha=0}|) + Prob(t < -|t_{\alpha=0}|)$ under the hypothesis $H_0 : \alpha = 0$, where the variable t in this equation has a t distribution with the degrees of freedom equals the degrees of freedom in the regression. We have included $P_{\alpha=0}$, $P_{\beta=0}$,and $P_{\beta=1}$ in appendix B (4)-(6). Any coefficient θ may be only "accidentally" significant if the corresponding $P_{\theta=0}$ is greater than 0.05. Basically the P-value is a measure of how much evidence you have against the null hypothesis. The smaller the P-value, the more evidence you have. The following table provides a widely accepted interpretation of the P-values:

$P \leq 0.01$ very strong evidence against H_0

$0.01 < P \leq 0.05$ moderate evidence against H_0

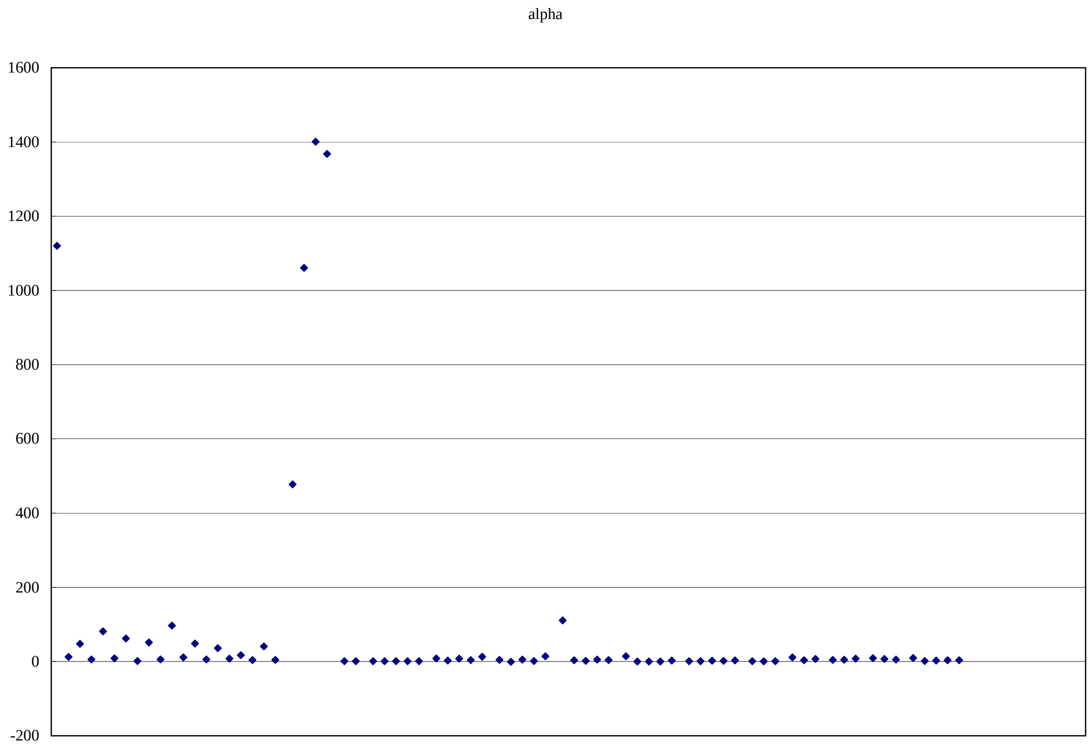
$0.05 < P \leq 0.1$ suggestive evidence against H_0

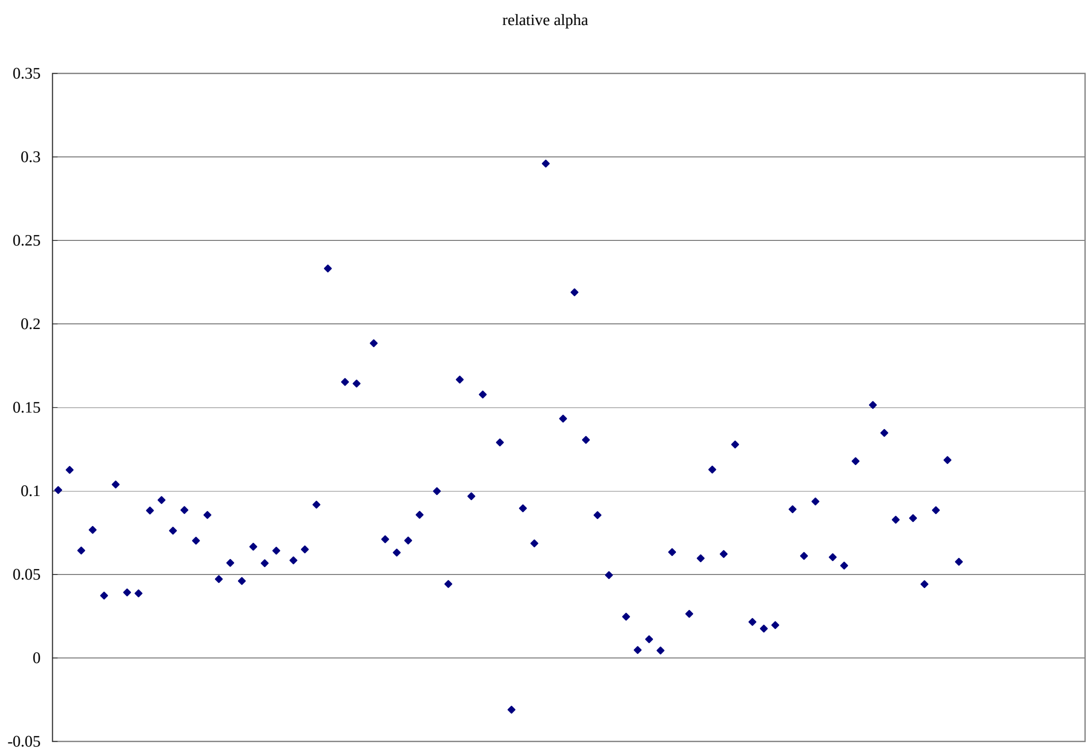
$0.1 < P$ little or no real evidence against H_0

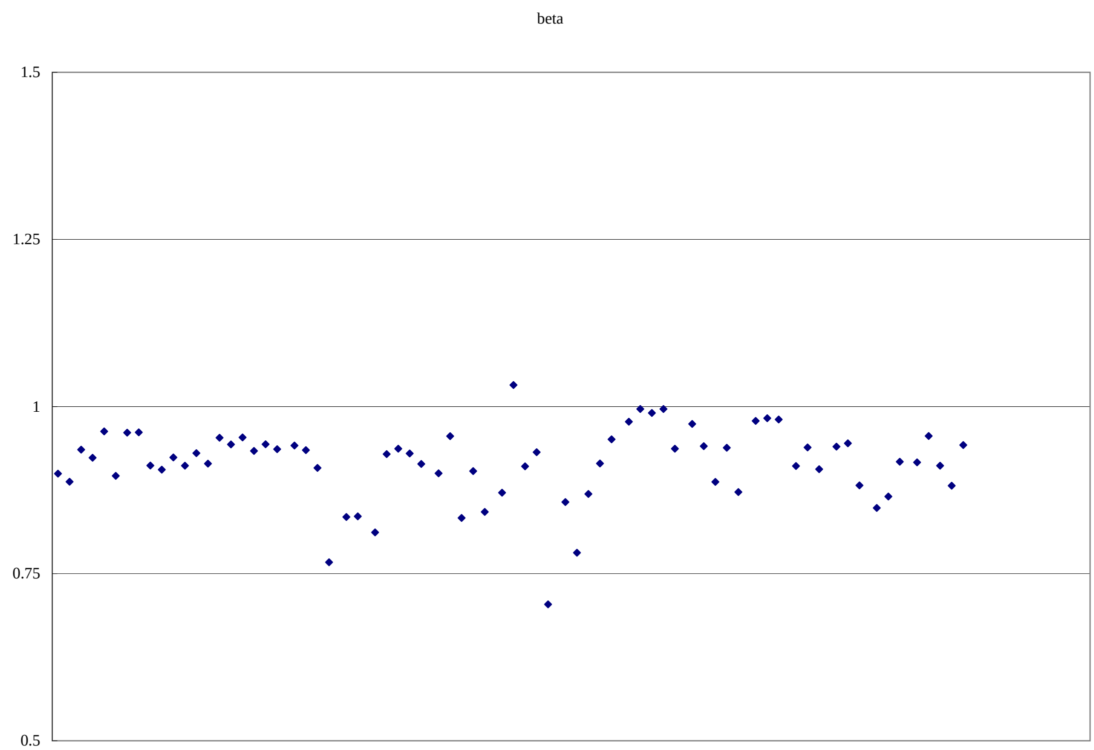
First of all, we plot all the estimates of α hoping the points could be close to zero, which is fairly the case with only a few exceptions jumping too high. Since those anomies are all from the indices with very high actual prices, we plot the relative alpha next and it brings all the points to $[-0.05, 3]$, moreover, they are clustered around $[0, 0.2]$.

Next, we plot the estimates of β. Almost all the points are nicely gathered between 0.75 and 1.

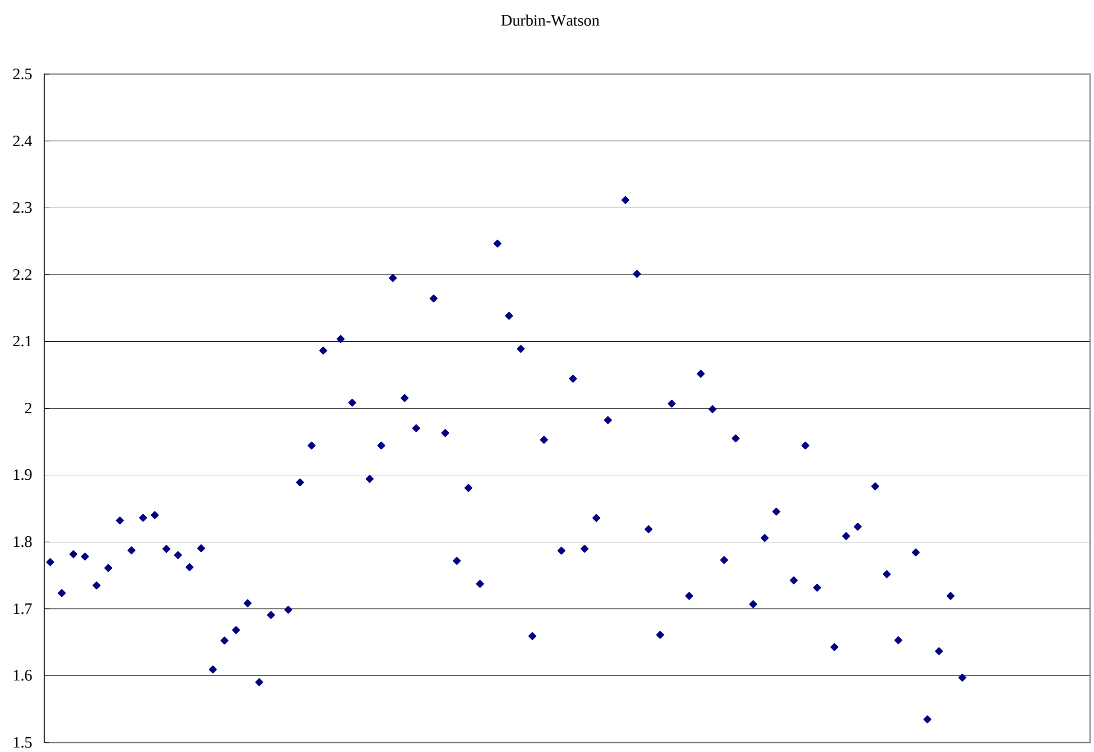
We know that $E[\text{Actual Price} \mid \text{Predicted Price}] = \alpha + \beta \cdot \text{Predicted Price}$. As the next three figures suggested, the predicted price is quite an unbiased estimate.


alpha
1600
1400
1200
1000
800
600
400
200
0
-200

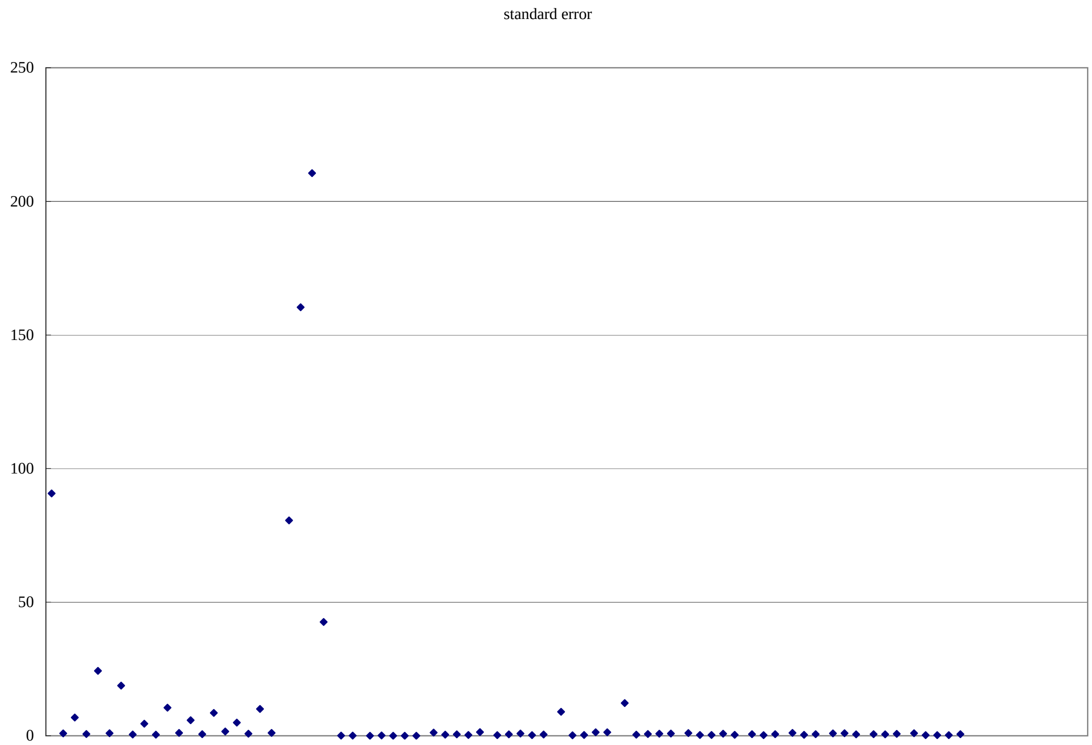

relative alpha
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05

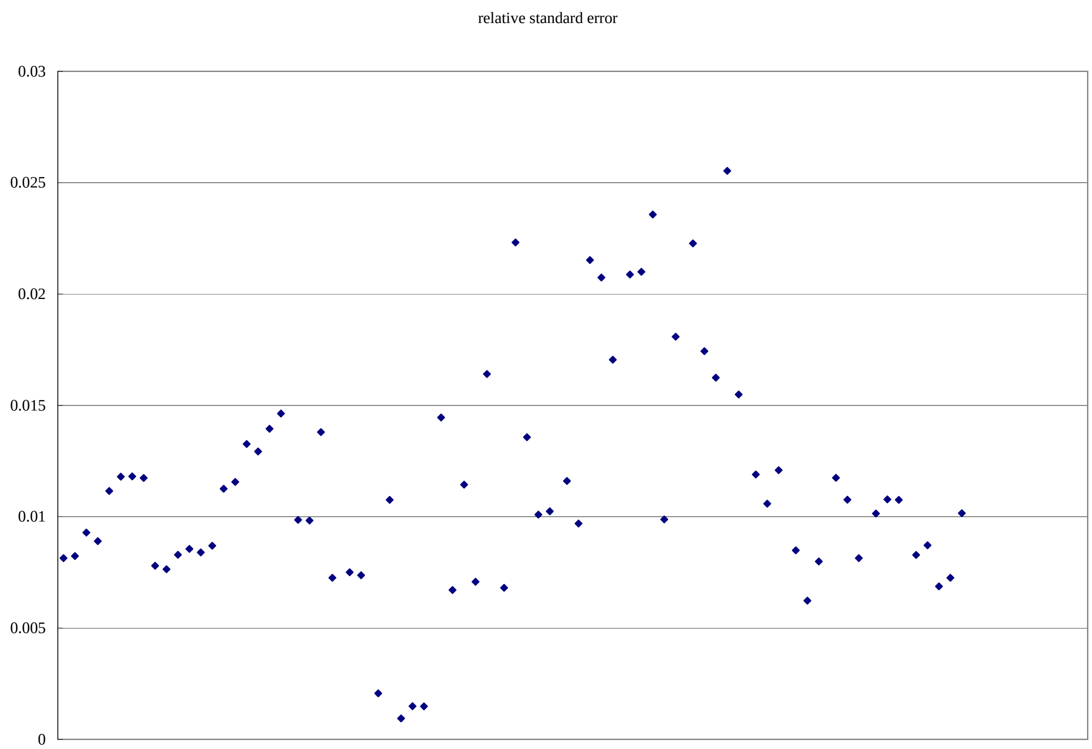

beta
1.5
1.25
1
0.75
0.5

As we can see from the next plot of the Durbin-Watson statistics, the points are scattered within $[1.6, 2.3]$. This is indicative of a great extent to which residuals are serially uncorrelated.

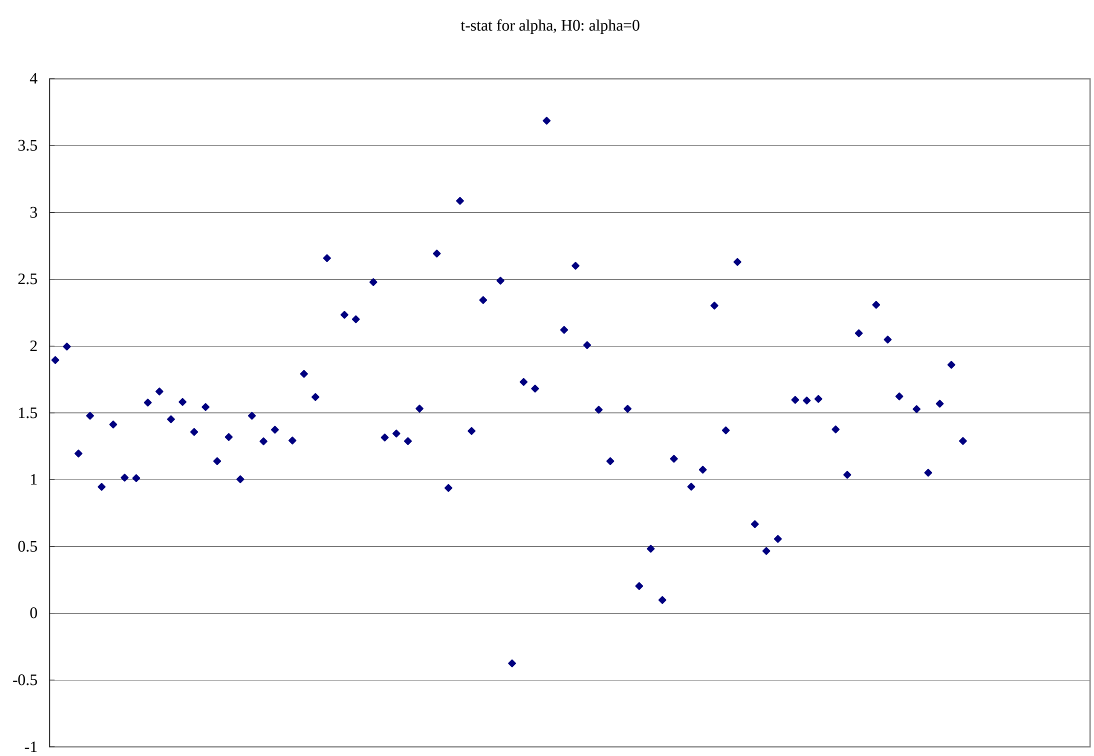


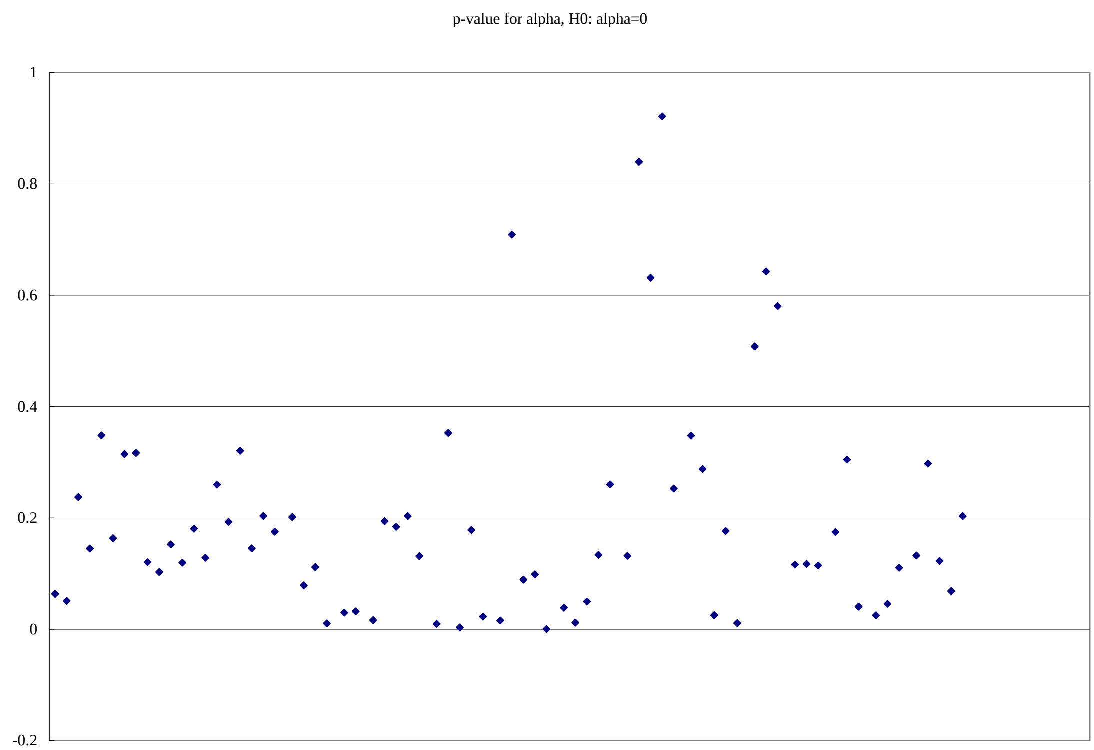
To check the overall fit of the regression, we plot the standard errors and observe a few points climbing too high as the price level is high. Therefore we plot the relative standard errors as well, and gladly find most of the points are under 0.025.


standard error
250
200
150
100
50
0


relative standard error
0.03
0.025
0.02
0.015
0.01
0.005
0

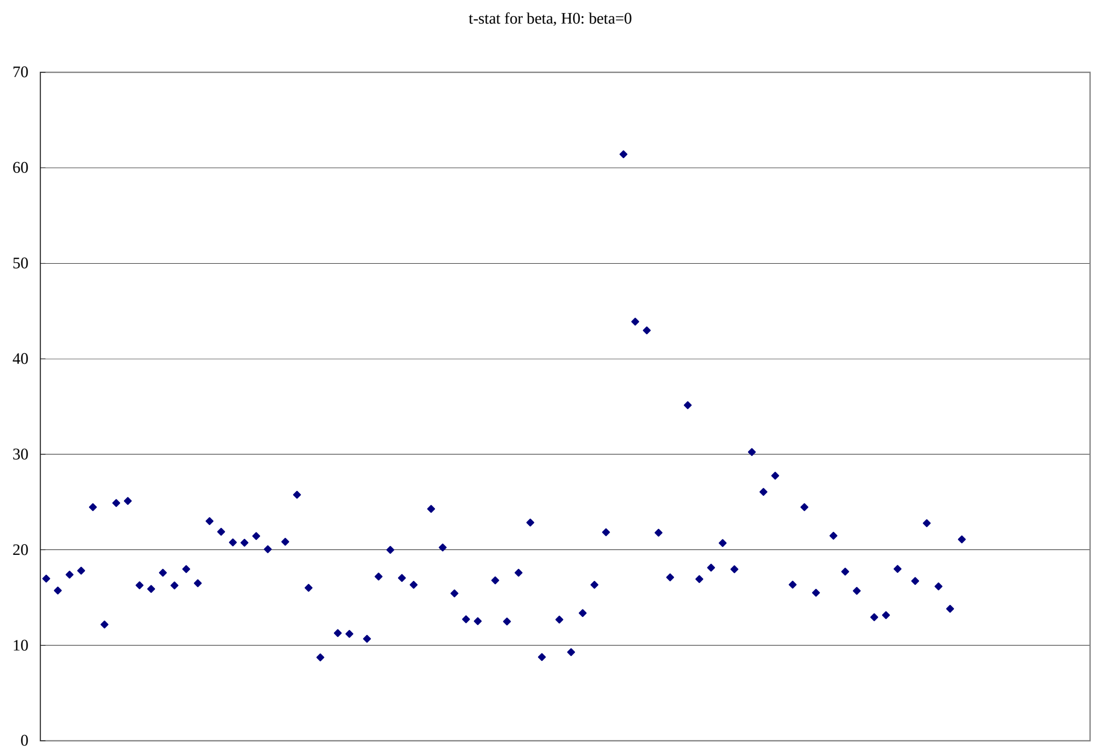
While plotting the t-statistics for α under $H_0 : \alpha = 0$, we see about one fourth of them slightly above 2, but still controlled by 3, which says no significant evidence to reject the null hypothesis. As to the P-values for α, the figure presents it clearly that the majority of the points are above 0.05, which means there is only a little evidence against H_0. Keep in mind that alpha is the constant term and whether to include the constant term or not is usually determined by priori reason(s) instead of the t-stat or P-value. x

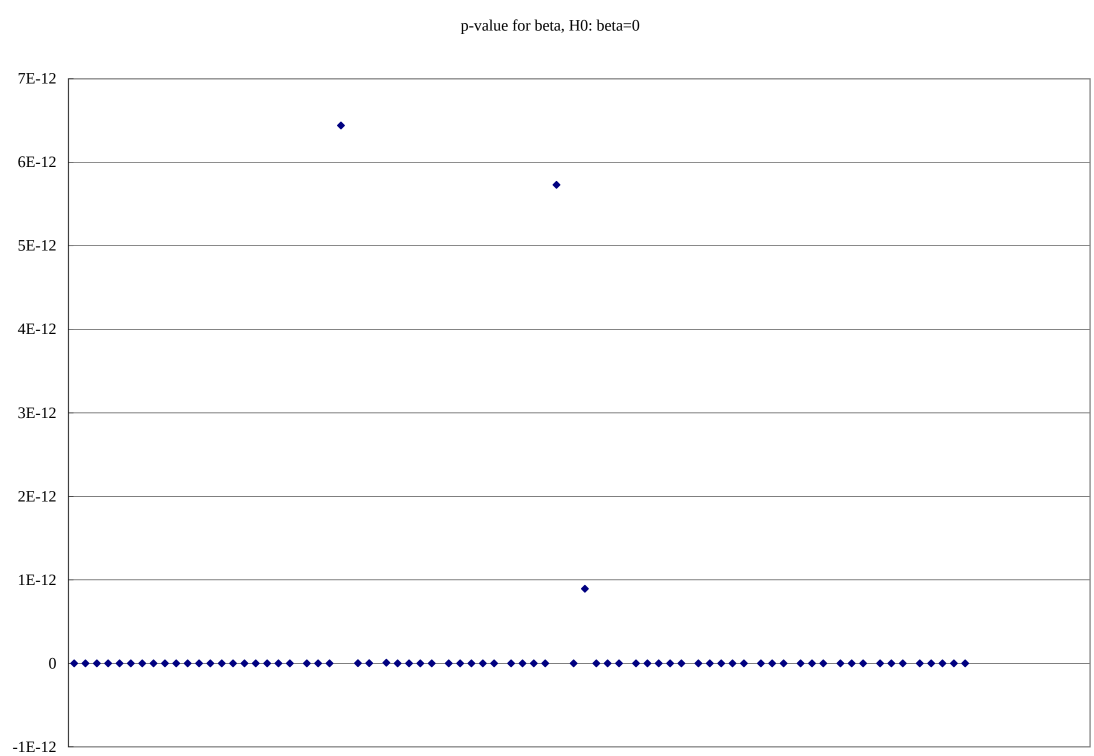


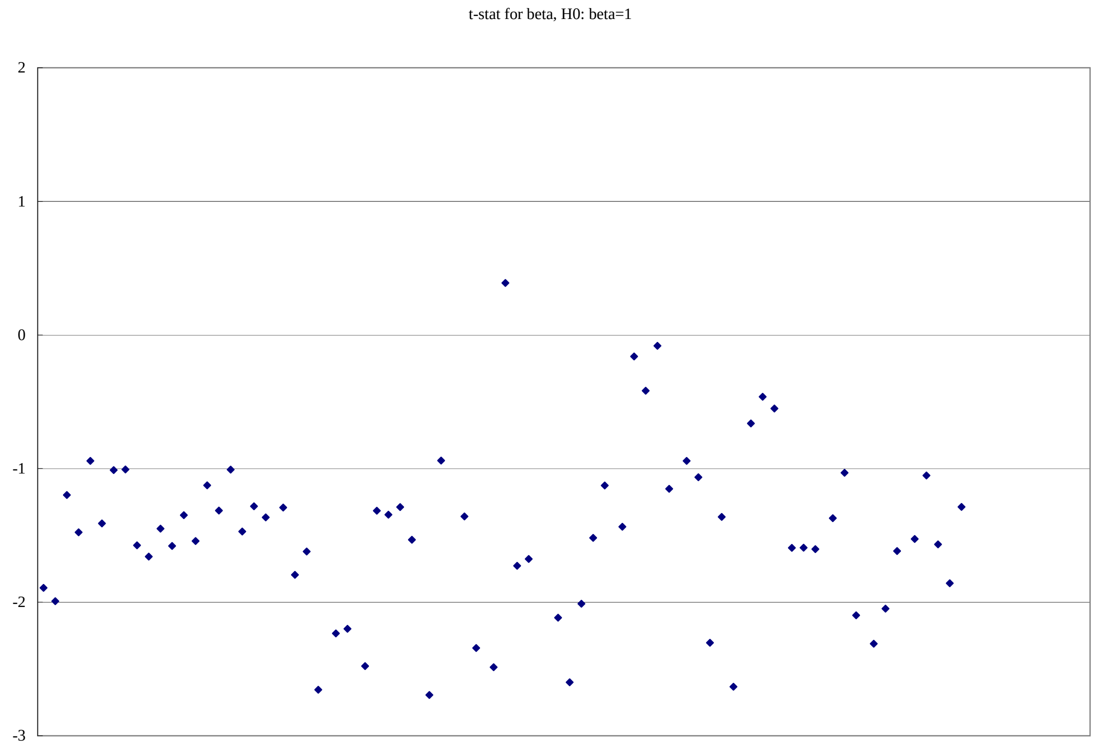


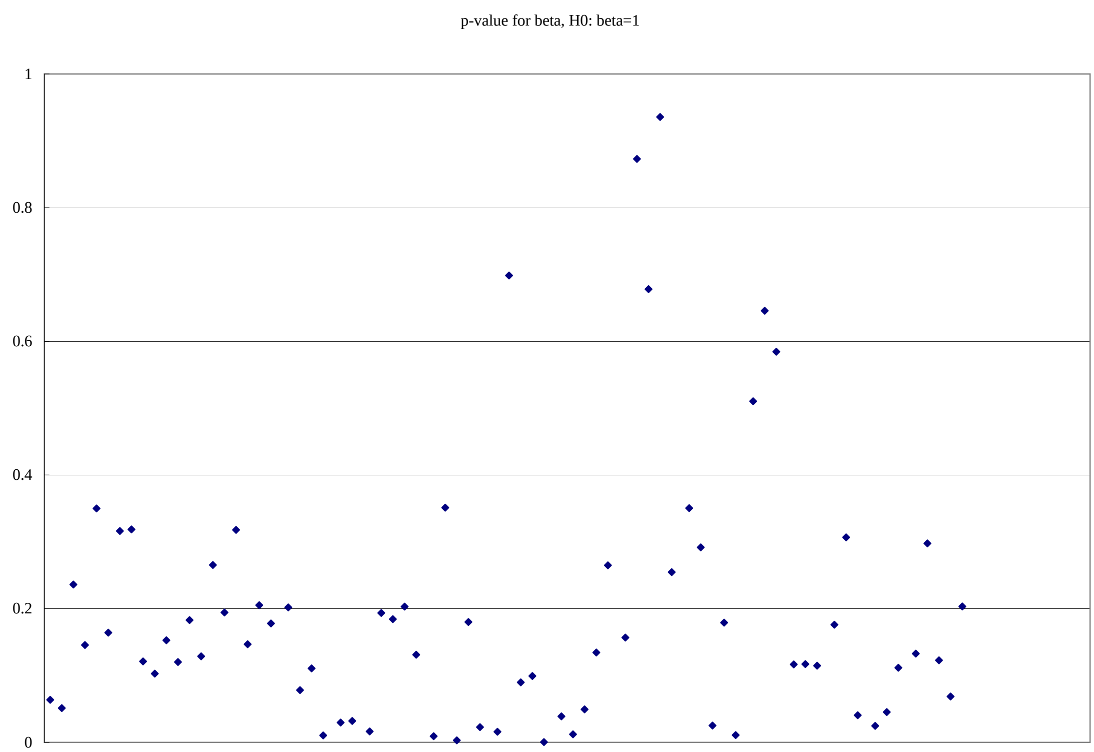
As to the t-statistics for β under $H_0 : \beta = 0$, the first plot below shows that all points are way higher than 2. This, in addition with the fact that the P-values for β under $H_0 : \beta = 0$ are much lower than 0.05, which is displayed in the second figure below, suggests a great significance of the regression.

Finally, we plot the t-statistics and P-values for β under $H_0 : \beta = 1$ in the third and the fourth figures below. They both suggest no significant evidence against this null hypothesis $H_0 : \beta = 1$. Up to now, we should have no doubt about the precision of this regression.


t-stat for beta, H0: beta=0
70
60
50
40
30
20
10
0


p-value for beta, H0: beta=0
7E-12
6E-12
5E-12
4E-12
3E-12
2E-12
1E-12
0
-1E-12


t-stat for beta, H0: beta=1
2
1
0
-1
-2
-3


p-value for beta, H0: beta=1
1
0.8
0.6
0.4
0.2
0

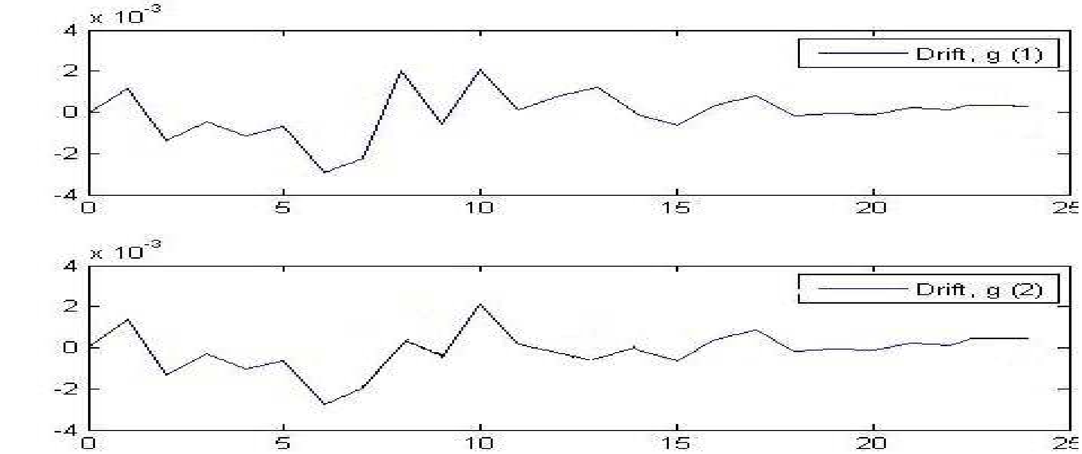
In summary, according to the 3 criteria proposed by Fama and Gibbons (1984) with our modification, i.e. replace α by relative α, replace standard error by relative standard error, our prediction results indicate a great applicability of our prediction methodology based on the hidden Markov model.

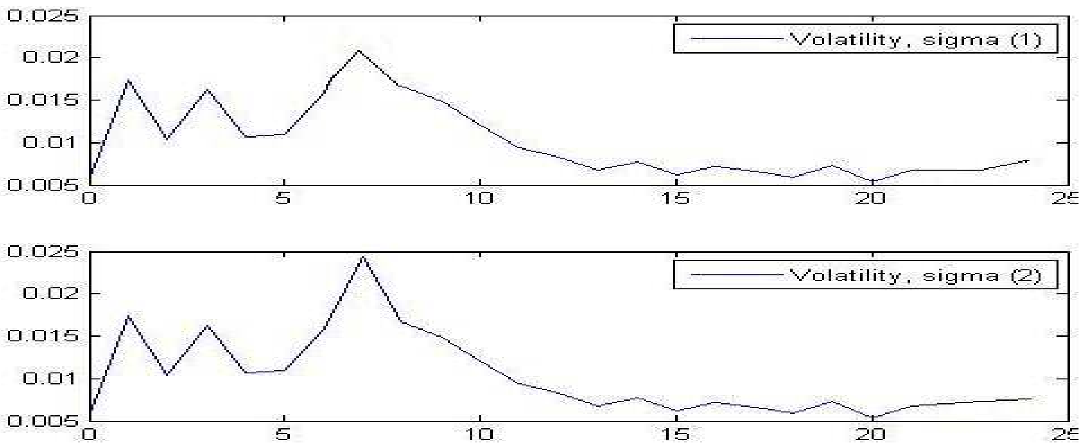
5.3 The Selection of Data Sets

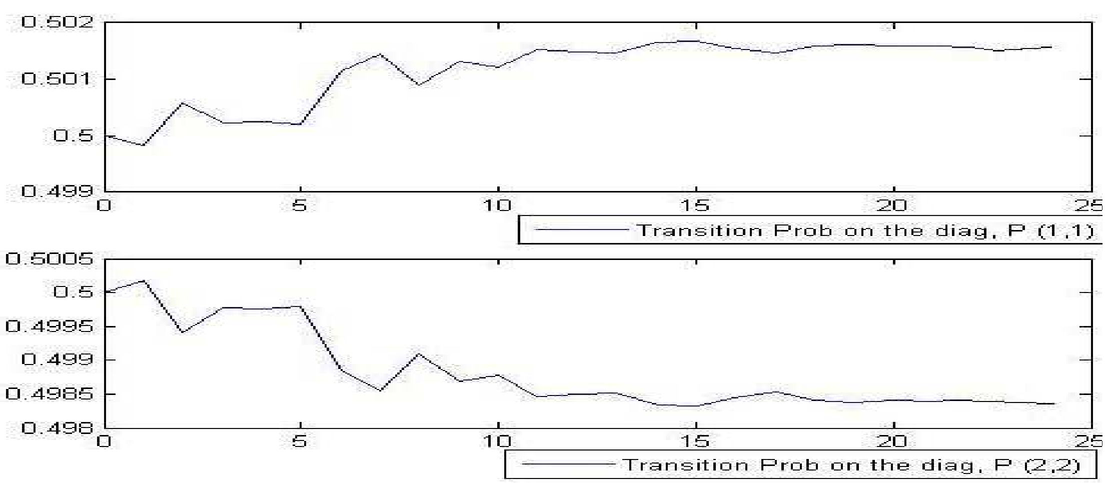
To thoroughly check the applicability of our prediction methodology, we try to achieve the highest completeness and diversification of the data sets. Meanwhile, after the methodology has been testified, it also provides us a tool to draw the similarities and distinctions among those data sets. Each of the data set contains historical prices from 3/12/2001 to 10/3/2006, so the data length for each one is about 1,938 except that it is 1,377 for Heng Seng Index, 1,369 for Nikkei-225, 1,404 for FTSE-100, this slight difference is due to the fact that different countries have different holidays (no trading, no historical price). It also bears mentioning that we start with 2001 just to make sure that all the historical prices are available. The rest of this chapter explains how and why we choose such a portfolio for this experiment.

Our data sources are CRSP (Center for Research in Security Prices, maintained by the University of Chicago) and Yahoo Finance(maintained by Yahoo INC) . From any historical price providers we can get 5 kinds of daily prices: open, low, high, close, and adjusted prices. The classical models on daily prices are all implemented with close-to-close or adjusted-to-adjusted prices. We use adjusted prices in our research. The adjusted price is made from closing price adjusted according to the most common corporate actions: cash dividend, stock dividend and stock split. For example, assume the closing price for one share of Company X is \$20 on March 9, 2006 (Thursday). After close on that day, Company X announced a dividend distribution of \$1.50 per share. The adjusted closing price for the stock would then be \$18.50. If Company X announced a 2:1 stock dividend instead, which means for any investor, she would receive two more shares for any share she owned. In this case, the adjusted price would be \$20/3=\$6.67, rounded to penny. If Company X announced a 2:1 stock split, then any investor would receive an extra share for every share she owned. As a result, the adjusted price would be \$20/2=\$10.

We use a length of 56 for each pass, so there are 24 passes (except that 25 passes for Nikkei-225) for each data set. The following three figures plot the estimates of drift, volatility and the diagonal of the transition matrix from the historical data of OEX (The S&P 100 Index). They show a trend of convergency from about the 17th pass. As a matter of fact, the results from all the data sets are convergent given such a 56×24 structure.


x 10^-3
4
2
0
-2
-4
0 5 10 15 20 25
Drift, g (1)
x 10^-3
4
2
0
-2
-4
0 5 10 15 20 25
Drift, g (2)


0.025
0.02
0.015
0.01
0.005
0 5 10 15 20 25
Volatility, sigma (1)
0.025
0.02
0.015
0.01
0.005
0 5 10 15 20 25
Volatility, sigma (2)


0.502
0.501
0.5
0.499
0 5 10 15 20 25
Transition Prob on the diag, P (1,1)
0.5005
0.5
0.4995
0.499
0.4985
0.498
0 5 10 15 20 25
Transition Prob on the diag, P (2,2)

Names and Symbols of selected data sets

Table 1	Indices and the corresponding ETFs
^DJI	DOW JONES INDUSTRIAL AVERAGE
DIA	DIAMONDS TRUST SER 1
^RUA	RUSSELL 3000 INDEX
IWV	ISHARE RUS 3000 INDX
^IXIC	NASDAQ COMPOSITE
ONEQ	NASDAQ COMP NDX FUND
^NDX	NASDAQ-100 (DRM)
QQQQ	NASDAQ 100 TR SER I
^OEX	S&P 100 INDEX,RTH
OEF	ISHARE SP 100 INDEX
^GSPC	S&P 500 INDEX,RTH
SPY	S&P DEP RECEIPTS
^RUI	RUSSELL 1000 INDEX
IWB	ISHARE RUS 1000 INDX
^MID	S&P 400 MIDCAP INDEX
MDY	S&P MID DEPOSIT RCPT
^SML	S&P 600 SMALLCAP INDEX
IJR	ISHARE SP SC 600 INX
^RUT	RUSSELL 2000 INDEX
IWM	ISHARE RUS 2000 INDX
^NYA	NYSE COMPOSITE INDEX
^HSI	HANG SENG INDEX
^N225	NIKKEI 225
^FTSE	FTSE 100

Tabe 2	Bonds
Gov Bonds	
^TNX	10-YEAR TREASURY NOTE
^TYX	30-YEAR TREASURY BOND
Bond Funds	
INBNX	RIVERSOURCE DIVERSIFIED BOND
AFTEX	AMERICAN FDS TAX-EX BOND FUND
PTHYX	PUTNAM TAX-FREE HIGH YIELD FUND
ELFTX	ELFUN TAX EXEMPT INCOME FUND
MDXBX	MARYLAND TAX-FREE BOND FUND

Table3	Stocks
Stocks in S&P500(large-cap)	
GE	GEN ELECTRIC CO
MSFT	MICROSOFT CP
XOM	EXXON MOBIL CP
PFE	PFIZER INC
C	CITIGROUP INC
Stocks in S&P400(middle-cap)	
WPO	WASHINGTN POST CO B
NYB	NEW YORK CMMTY BNC##
TSN	TYSON FOODS INC CL A
VLO	VALERO ENERGY CP
LLL	L-3 COMM HLDGS INC
Stocks in S&P600(small-cap)	
NVR	N V R L P
URBN	URBAN OUTFITTERS I
MRX	MEDICIS PHARMA CP
IDXX	IDEXX LABS
ROP	ROPER INDUST INC
Stocks in NASDAQ 100(tech)	
QCOM	QUALCOMM INC
INTC	INTEL CP
CSCO	CISCO SYS INC
EBAY	EBAY INC
DELL	DELL INC

Table 4	ETFs
ETF, Specialty-Technology	
MTK	MORGAN STANLEY TECHN
XLK	TECHNOLOGY SPDR
IYW	ISHARE DJ TCH SC INX
ETF, Large Blend	
IVV	ISHARE S&P 500 INDX
VV	VANGUARD LG-CAP ETF
ELV	STREETTRACKS SERIES
ETF, Mid-Cap Blend	
IJH	ISHARE SP MC 400 INX
IWR	ISHARE RUS MC INDX
VO	VANGUARD MID-CAP ETF
ETF, Small Blend	
VB	VANGUARD SM-CAP ETF
IWC	ISHARES RUSSELL MICR
DSV	SPDR DJ WILSHIRE SMA
BA	BOEING CO

Table 5	Mutual Funds
VFINX	VANGUARD INDEX TRUST 500 INDEX
AGTHX	AMERICAN FDS GROWTH FUND
AIVSX	AMERICAN FDS INVESTMENT CO
AWSHX	AMERICAN FDS WASHINGTON MUTUAL
FCNTX	FIDELITY CONTRA FUND

The above 5 tables present the names and symbols of the 73 data sets that were explored in this experiment including: 14 major Indices; 10 corresponding Index Based ETFs; 2 government bonds and 5 bond funds; 5 mutual funds; 25 common stocks including 5 from the composition of Dow Johns Industrial Average, 5 from S&P LargeCap 500, 5 from S&P MidCap 400, 5 from S&P SmallCap 600, and 5 from Nasdaq-100; finally another 12 ETFs including 3 'Specialty-Technology', 3 'Large Blend', 3 'Mid-Cap Blend', 3 'Small Blend'.

Next, we shall explain how and why we choose such a portfolio to meet our goal of completeness and diversification.

Table 1 includes major indices (Arnott, Hsu, and Moore, 2005; Amenc, Goltz, and Sourd, 2006) and their most well known corresponding ETFs. We start with the Dow Jones Industrial Average (DJI), the best-known and most widely followed index among the world. Even though it only contains 30 industrial companies, it is highly correlated to more diverse indices like the S&P 500.

Next, S&P 500 LargeCap (GSPC), the most commonly referenced U.S. equity benchmark.. It consists 500 leading companies from a wide variety of 100 economic sectors and thus they respond to every important factor in the overall economy. However, the S&P 500 does not provide investors with exposure to some of the smaller, yet in many cases faster growing, companies on the market because they are unlikely to qualify due to the index's high market cap requirements. For the completeness of our experiment, we also include its siblings: S&P 400 MidCap

(MID), S&P 600 SmallCap (SML). A mid-cap company is broadly defined as one with a market capitalization ranging from about $2 billion to $10 billion. S&P 400 MidCap contains solid companies with good track records that are simply not large enough to be included in the much larger S&P 500. A small-cap company is generally defined as one with a market capitalization between $300 million and $2 billion. S&P SmallCap 600 index was introduced in an effort to represent a smaller segment of the market than the S&P MidCap 400 Index. In order to compare the prediction performance of our model within different markets, we then include another member of this family, S&P 100 (OEX), a subset of the S&P 500 tracking the largest, most liquid stocks in it. S&P 100 is made up of 100 major, blue chip stocks across diverse industry groups.

In addition to the S&P family, we also exam the three members in the Russell family. Russell 3000 (RUA), the index comprised of the 3000 largest and most liquid stocks based and traded in the U.S. It can be subdivided into two segments: the Russell 1000 large-cap (RUI) and Russell 2000 small-cap (RUT). The delineation is clear enough–the Russell 1000 represents the 1000 largest stocks in the index (based on market cap), while the remaining 2000 are placed in the Russell 2000. Because of its broad diversification and large number of holdings, Russell 3000 often makes a good capture of the overall market.

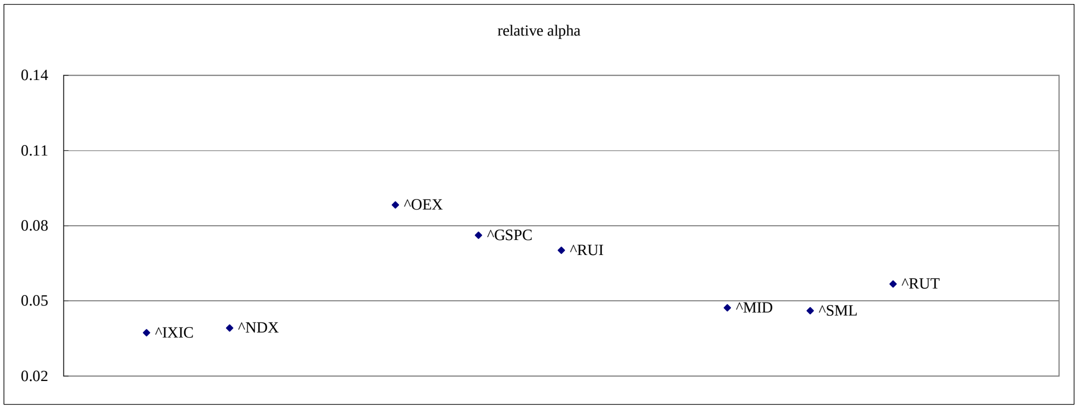
The next one is Nasdaq Composite (IXIC), a broad market index that encompasses about 4,000 issues traded on the Nasdaq National Market. Although it is not as actively traded as its much smaller cousin, the Nasdaq-100, this index is more commonly referred to by investors and the financial press. When the question "How did the Nasdaq (National Association of Securities Dealers Automated Quotations) do today?" is asked, the answer is usually the value of this index. Because technology firms account for roughly 2/3 of the index, investors often use it as a guide to help them determine the strength of technology stocks. We also include Nasdaq-100 (NDX), which gives investors a quick snapshot of how some of the nation's largest technology firms are faring. It includes almost all of the country's top technology stocks, so it is a better proxy for this sector than most other indices. In addition to the issues traded on the Nasdaq Exchange, we are also interested in those traded on the New York Stock Exchange, so the NYSE Composite (NYA) is included in our experiment as well. This index measures all common stocks listed on the New York Stock Exchange and four subgroup indexes: Industrial, Transportation, Utility, and Finance.

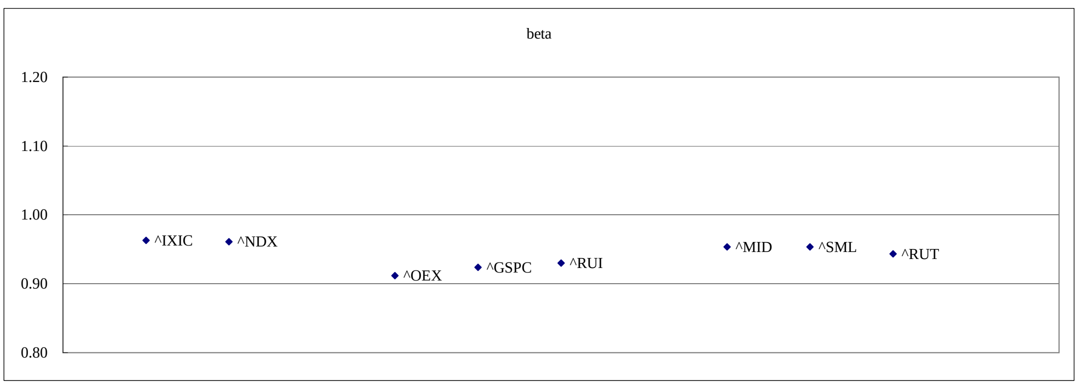
Finally, for a taste of some major oversea Exchanges, we add three more major foreign indices: Hang Seng Index (HIS), which consists of the 33 largest companies traded on the Hong Kong Stock Exchange and represents approximately 70% of the value of all stocks traded on the exchange; Nikkei-225 (N225), the index comprised of 225 top-rated Japanese companies listed in the First Section of the Tokyo Stock Exchange; and Financial Times 100 Index (FTSE), the most widely used benchmark for the performance of equities traded on the London Stock Exchange, representing about 80% of the value of all issues traded on the exchange.

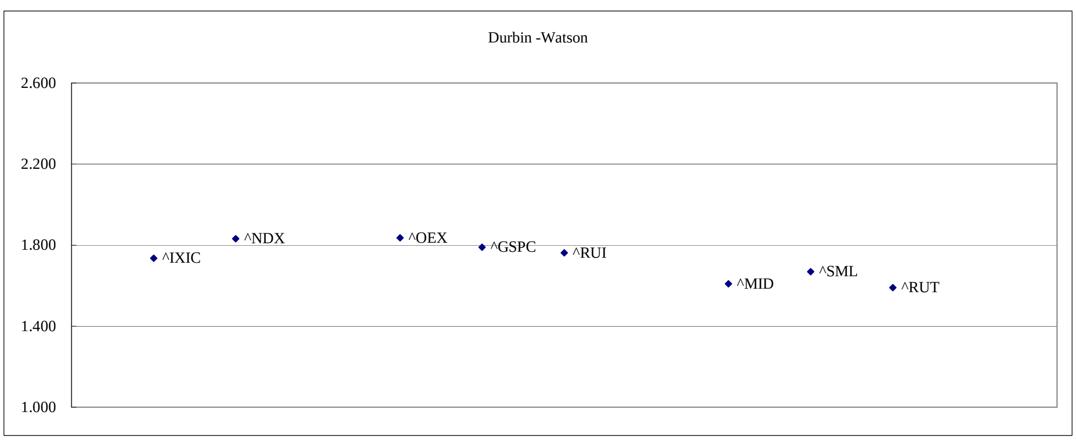
To meet the goal of a highly complete and diversified portfolio, after the indices, we select the following two types of equities that should be very different in their price processes: the bonds and the common stocks. The first difference between them is that bonds are less risky, i.e. less volatile in prices, especially for government bonds, they can never default, as the government can always print more money to pay you back. In addition, stocks and bonds move in the opposite directions. Because money flows into and out of assets on a regular basis as investors try to find the hot spot in the world of asset allocation. This means that as stock prices rise, bonds fall. It bears mentioning that this was not the case historically, especially before 2000, back to then, stocks and bonds periodically moved in the same direction, but at other times have moved in opposite directions (depending on where we were in the business cycle). This is another reason we start our data sets in 2001. As you can see from table 2, there are 2 Treasury Bills included. In addition, there are 5 bond funds studied in this experiment for the following two reasons: historical data for corporate bonds are hard to achieve while the bond funds we choose are portfolios of corporate bonds, secondly, they are from different categories of bond funds so such a selection meet our goal of diversification.

Table 3 lists the 25 common stocks that are studied in this experiment. There are thousands of stocks traded in the U.S. market, as suggested by Wilshire 5000, the “total market index”, which encompasses about 6,700 stocks. Which ones should be included in our portfolio? Thanks to the analysis on indices at the beginning. If we put technology based indices in one category, large cap indices in another category, and small/mid cap indices in the third category, the table and plots below show the similarities within each category and the differences between categories in the prediction performance. We achieve the lowest relative α (< 0.04) and the most close-to-1 β (> 0.96)with Nasdaq cousins. For large cap indices, relative α is within $[0.07, 0.09]$ and β is within $[0.91, 0.93]$, while relative α is within $[0.046, 0.057]$ and β is within $[0.94, 0.95]$ for the small/mid cap indices. i.e. we achieve the best prediction on technology based indices, second best on small/mid cap indices, the third on large cap indices.

		relative alpha	beta	Durbin-Watson
^IXIC	NASDAQ COMPOSITE+B49	0.03730	0.96290	1.73500
^NDX	NASDAQ-100 (DRM)	0.03920	0.96090	1.83210
^OEX	S&P 100 INDEX,RTH	0.08830	0.91180	1.83630
^GSPC	S&P 500 INDEX,RTH	0.07620	0.92390	1.78970
^RUI	RUSSELL 1000 INDEX	0.07020	0.93020	1.76220
^MID	S&P 400 MIDCAP INDEX	0.04720	0.95340	1.60920
^SML	S&P 600 SMALLCAP INDEX	0.04610	0.95370	1.66830
^RUT	RUSSELL 2000 INDEX	0.05670	0.94360	1.59020







Suggested by the similarities within each category and the differences between categories in the prediction performance, we take 5 stocks from Nadaq-100; 5 stocks form the large cap index, S&P 500; 5 from a middle cap index, S&P 400; 5 from a small cap index, S&P 600; and of course, 5 from the Dow.

Basically we seek stocks with highest weight thus the greatest influence in the corresponding indices. What does weight mean? (Fernholz, Garvy, and Hannon, 1998; Haugen and Baker, 1991) The Efficient Market Inefficiency of Capitalization-Weighted Stock Portfolios,) Among all the major U.S. indices, the Dow is the only price-weighted index. In other words, stocks with higher prices are given a greater weighting in the index than lower-priced stocks regardless of each company's actual size. The calculation is quite complex, but essentially it is summing up the prices of all 30 member stocks and then dividing that figure by a "magic number." In an effort to maintain the index's continuity, this divisor changes over time to reflect changes in the Dow's 30 component stocks. The other indices are calculated based on a market cap weighting, i.e., the weight of any given stock holds in the index is determined by this stock's market capitalization (its price multiplied by the number of shares outstanding). Therefore, the largest firms have the greatest impact on the value of a market cap weighted index. There is a slight difference in Nasdq-100 though. It is computed using a modified market weighting. Although firms with the largest market caps still have the largest influence on the index, its value is modified to keep any issues from having an overwhelming effect on the index value.

For example, Microsoft has a market cap 5X larger than that of Qualcomm, yet it only boasts a 50% greater weighting in the index. This keeps the index from being dominated by a handful of stocks. The actual computation methods are proprietary to the Nasdaq. The next table lists all the stock we have selected with their weighting in the indices.[2]

Note: data as of July 2004			Companies in S&P 500		
Company	Symbol	Weight	General Electric	GE	3.20%
			Microsoft	MSFT	2.90%
			ExxonMobil	XOM	2.70%
			Pfizer	PFE	2.50%
			Citigroup	C	2.30%
Companies in Dow			Companies in S&P 400		
3M	MMM	6.30%	Washington Post	WPO	0.90%
Johnson & Johnson	JNJ	4.00%	NY Comm. Bancorp	NYB	0.80%
Wal-Mart Stores	WMT	3.80%	Valero Energy	VLO	0.80%
Coca-Cola	KO	3.70%	Tyson Foods	TSN	0.70%
Boeing	BA	3.60%	L3 Communications	LLL	0.70%
Companies in Nasdaq-100			Companies in S&P 600		
Qualcomm	QCOM	5.40%	NVR Inc.	NVR	1.30%
Intel	INTC	5.10%	Urban Outfitters	URBN	1.00%
Cisco Systems	CSCO	4.80%	Medicis Pharma.	MRX	1.00%
eBay	EBAY	3.10%	IDEXX Labs	IDXX	0.90%
Dell	DELL	2.80%	Roper Ind.	ROP	0.70%

An other type of securities that has grown increasingly popular in recent years is

[2]Major resource: Yahoo Finance, Street Authority

the exchange-traded fund (ETF). ETFs are securities that closely resemble index funds, but can be bought and sold throughout the day, purchased on margin, or even sold short, just like common stocks. These investment vehicles allow investors a convenient way to purchase a broad basket of securities in a single transaction. Essentially, ETFs offer the convenience of a stock along with the diversification of a mutual fund. Whenever an investor purchases an ETF, he or she is basically investing in the performance of an underlying bundle of securities – usually those representing a particular index or sector. ETFs are very liquid compared with the traditional mutual funds, as they can be bought and sold at any time throughout the trading day, many have average daily trading volumes in the hundreds of thousands or even millions per day. For instance, the average daily volume of QQQQ (NASDAQ 100 TR SER I) from March 12, 2001 to October 3, 2006 is 88,967,298 and it is 44,009,298 for SPY(S&P DEP RECEIPTS). Compared with some relatively liquid stocks, say, WMT (WAL MART STORES), the average daily volume during the same period is 10,087,066 and it is14,539,203 for C (CITIGROUP INC). Given its popularity, we definitely need to include this type of securities in our study. Meanwhile, ETFs have been very attractive to market players ever since the first one (SPDR) was born in 1993, however most of the theoretic and empirical research has not paid enough attention to it. In this paper, we showed that ETFs and common stocks do share a lot of statistic properties and our HMM works on ETFS just as well as it does to the other securities.

The selection of ETFs started from QQQQ, the best-known ETF in existence. It tracks the Nasdaq-100 Trust Index (NDX). The experiment result is astounding, we found QQQQ resembles NDX very closely, as the next table suggested. Although the price level of NDX is almost 60 times that of the QQQQ, the difference in the estimates of drift, volatility, transition probability can almost be neglected. Moreover they can be predicted at the same level of accuracy as it is shown in the estimates of the regression coefficients and statistics for testing significance.

	Drift 1	Drift 2	Volatility 1	Volatility 2	Transition	Prob
^NDX	-0.0028	-0.0027	0.0118	0.0118	0.5008	0.5008
					0.4992	0.4992
QQQQ	-0.0028	-0.0028	0.0118	0.0118	0.5007	0.5007
					0.4993	0.4993

	relative alpha	beta	relative standard error	D-W
^NDX	0.0392	0.9609	0.0118	1.8321
QQQQ	0.0387	0.9615	0.0117	1.7874

	t for alpha H0: alpha=0	t for beta H0: beta=0	t for beta H0: beta=1	p for alpha H0: alpha=0	p for beta H0: beta=0	p for beta H0: beta=1
^NDX	1.0153	24.8931	-1.0118	0.3145	0.0000	0.3161
QQQQ	1.0106	25.1208	-1.0068	0.3167	0.0000	0.3185

	Opem	High	Low	Close	Volume	Adjusted
^NDX	1427.64	1442.58	1412.11	1427.00	1759978005.72	1427.00
QQQQ	35.39	35.79	34.96	35.37	88967298.08	34.99

Average Mar 12, 2001 --- Oct 3, 2006

Impelled by such an amazing discovery, we shall go further to compare each major index with its most well known ETF as you shall find in table 1. The outcome is again promising, most of the index based ETFs do a good job to mimic the corresponding indices. According to *ETF Investment Outlook*, there are 4 index based ETFs among the 10 most active ETFs,[3] and they are Nasdaq 100 Trust (QQQQ), S&P 500 SPDR (SPY), iShare SP SC 600 INX (IJR), Dow Diamonds Trust (DIA). We have the result from SPY, IJR, and DIA reported in the table below along with the indices we are trying to resemble.

As far as the estimates of drift and volatility are concerned, there is no difference between the index and its corresponding ETF up to the third decimal place. Even with the prediction performance of our methodology, the differences in relative α, and β between the index and its corresponding ETF is basically under 0.02.

[3]The others are: Energy SPDR (XLE), Semiconductor HOLDRS (SMH), Oil Service HOLDRS (OIH), Financial SPDR (XLF), Retail HOLDRS (RTH), Utilities SPDR (XLU)

	D 1	D 2	V 1	V 2	Transition	Prob
^DJI	-0.0011	-0.0008	0.0083	0.0083	0.5104	0.5113
					0.4896	0.4887
DIA	-0.0012	-0.0005	0.0088	0.0080	0.5228	0.5238
					0.4772	0.4762
^GSPC	-0.0011	-0.0009	0.0084	0.0083	0.5068	0.5070
					0.4932	0.4930
SPY	-0.0010	-0.0008	0.0087	0.0086	0.5119	0.5116
					0.4881	0.4884
^SML	-0.0042	0.0006	0.0134	0.0136	0.524	0.5261
					0.476	0.4739
IJR	-0.0041	0.0007	0.0131	0.0125	0.5193	0.5120
					0.4807	0.4880

	relative alpha	beta	relative S. E. for regression	D-W	t for alpha H0: alpha=0	t for beta H0: beta=0	t for beta H0: beta=1	p for alpha H0: alpha=0	p for beta H0: beta=0	p for beta H0: beta=1
^DJI	0.1005	0.8996	0.0081	1.7699	1.8955	16.9745	-1.8934	0.0634	2.64E-23	0.0637
DIA	0.1126	0.8875	0.0082	1.7235	1.9958	15.7297	-1.9932	0.0510	8.16E-22	0.0513
^GSPC	0.0762	0.9239	0.0083	1.7897	1.4516	17.6010	-1.4503	0.1524	5.00E-24	0.1528
SPY	0.0886	0.9116	0.0086	1.7804	1.5810	16.2712	-1.5787	0.1197	1.79E-22	0.1202
^SML	0.0461	0.9537	0.0133	1.6683	1.0026	20.7732	-1.0084	0.3206	2.02E-27	0.3178
IJR	0.0666	0.9337	0.0129	1.7084	1.4776	20.7362	-1.4716	0.1453	2.20E-27	0.1469

ETFs are often categorized by the types of their underlying securities. For example, a 'Large Blend' ETF holds only large cap stocks, the same rule applies to 'Mid-Cap Blend' and 'Small Blend'. Some ETFs consist of stocks from a certain economic sector, such as technology, we call this category 'Specialty-Technology'. From each of the above four categories, we chose 3 ETFs with the highest NAV except for those listed in table 1 already. Selected ETFs are listed in table 4.

The Last investment vehicles that we shall discuss is the Mutual Fund. According to *ICI (the Investment Company Institute)*, as of April 1, 2006, the number of registered mutual funds has grown to 8606 ever since the first one was launched 65 years ago, and they manage about $9.2 trillion dollars. Over 91 million Americans, from 47% of the nation's households, invest in mutual funds. In our experiment, we choose the five biggest mutual funds available to individuals. The ranking below is from *Cox Newspaper* by Hank Ezell(2005).

AMERICA'S BIGGEST MUTUAL FUNDS

Fund	Net assets	Total return*	Sales charge	Annual expenses	Minimum purchase	Ticker symbol
Vanguard 500 Index	$70.9 billion	9.3%	0	0.18%	$3,000	VFINX
Growth Fund of America A	$68.6 billion	12.8%	5.75%	0.68%	$250	AGTHX
Investment Company of America A	$66.3 billion	11.0%	5.75%	0.57%	$250	AIVSX
Washington Mutual Investors Fund A	$62.8 billion	10.7%	5.75%	0.61%	$250	AWSHX
Fidelity Contrafund	$55.0 billion	11.8%	0	0.94%	$2,500	FCNTX
Fidelity Magellan	$50.7 billion	7.0%	0	0.57%	Closed	FMAGX
Income Fund of America A	$48.1 billion	9.9%	5.75%	0.55%	$250	AMECX
Dodge & Cox Stock	$46.3 billion	14.8%	0	0.53%	Closed	DODGX
Capital Income Builder A	$42.5 billion	11.0%	5.75%	0.60%	$250	CAIBX
EuroPacific Growth Fund A	$41.9 billion	10.2%	5.75%	0.83%	$250	AEPGX

*Annual average over the last 10 years.
Sources: Morningstar, fund prospectuses

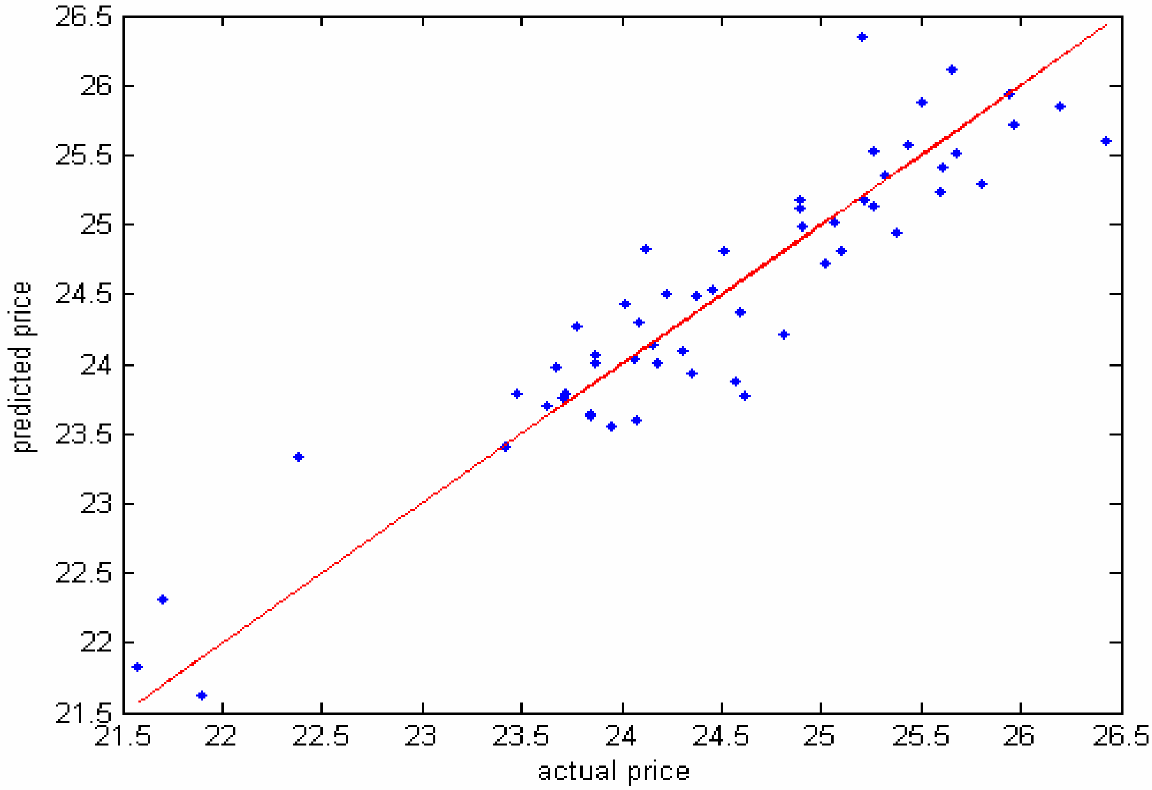
The biggest mutual fund of all is the Vanguard 500 Index fund, which is built to match the performance of the Standard & Poor's 500 index. From the second to the fourth are managed by the American Funds. The fifth one is managed by Fidelity.

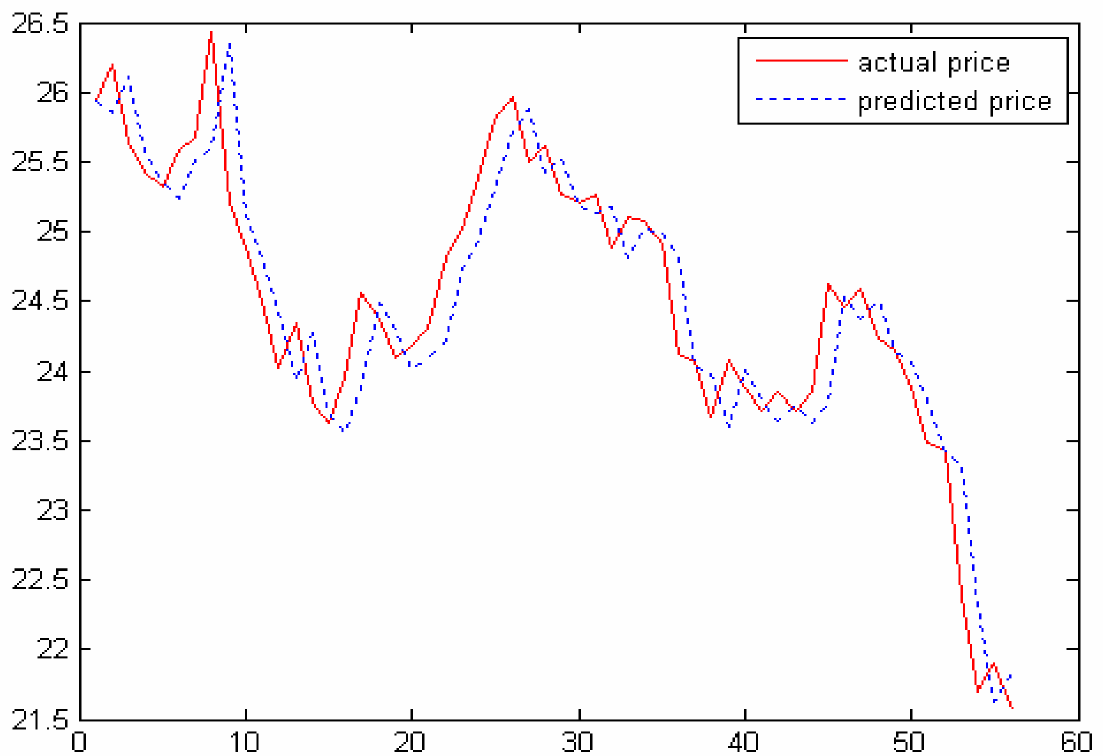
5.4 Empirical Findings

The most significant advantage of this hidden Markov model while implemented with daily data is its predictability. The three figures below show the prediction performance on DEll (DELL INC.), MMM (3M COMPANY), and QQQQ (NASDAQ 100 TRUST, SERIES I) .

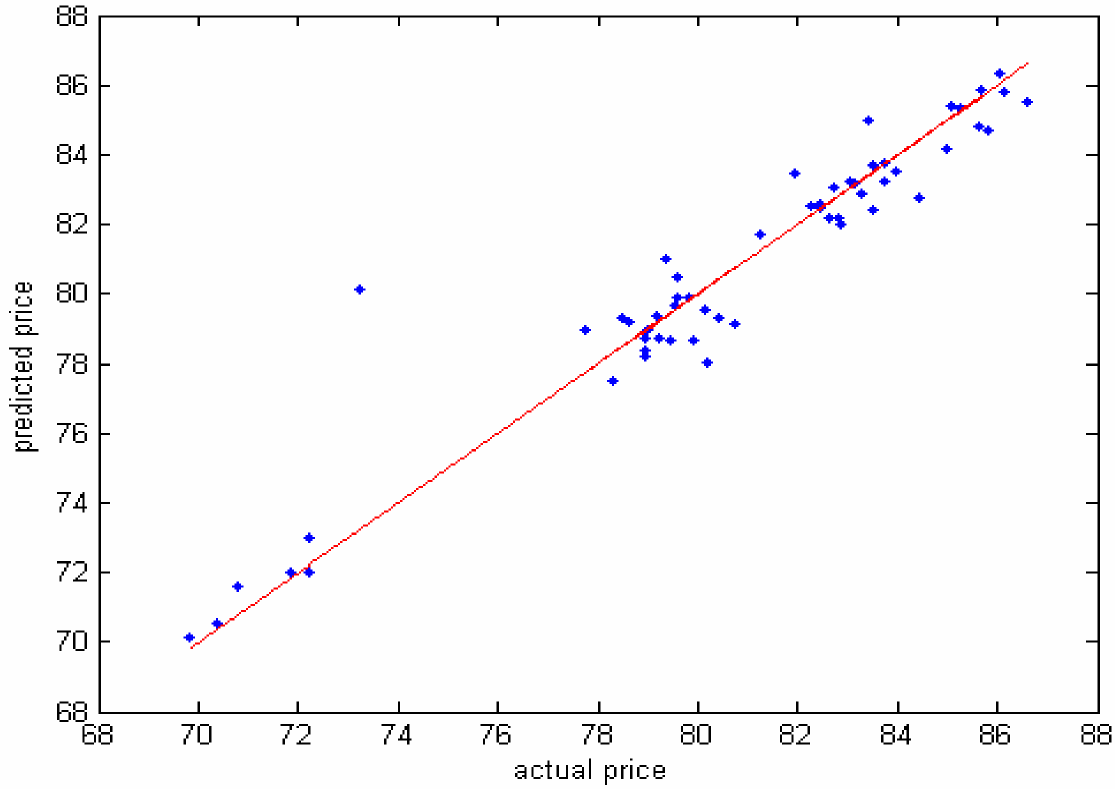
Take the prediction on the stock prices of DELL for example. First we plot the predicted prices on the y-axis and the actual prices on the x-axis. The first plot displays the prediction of prices on a range of 56 days. For each point, the x coordinate represents the actual price and the y coordinate represents the predicated price given information of the actual prices up to the last trading day. We can see that all points are closely gathered around the diagonal $y = x$. Generally it indicates a good prediction. To see how good it is, the second plot shows the whole paths of actual price (solid line) and predicted price (dashed line). Not only we find the predicted path closely resembles the actual path, but also we discover an interesting phenomenon. When the prices are increasing, we see under prediction, conversely when the prices are decreasing, we see over estimate. This phenomenon can be found with quite a lot of data sets. One might take this into account when using our program to predict the prices. And of course more work has to done to study the cause of this phenomenon.

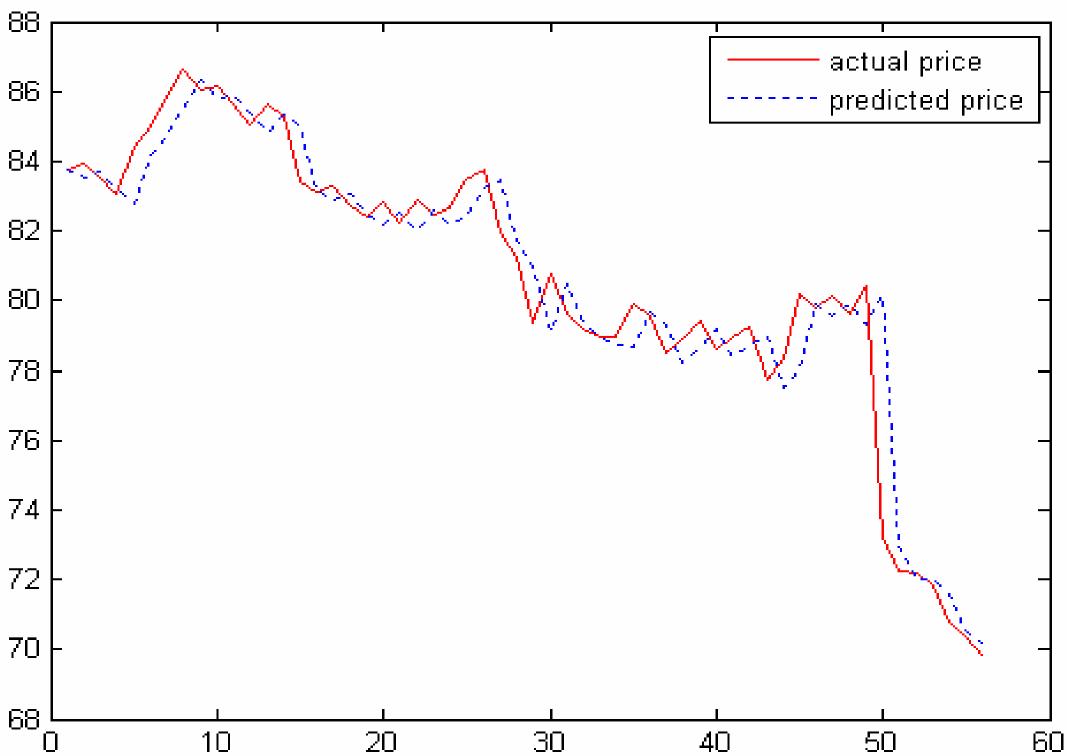
Prediction on DELL



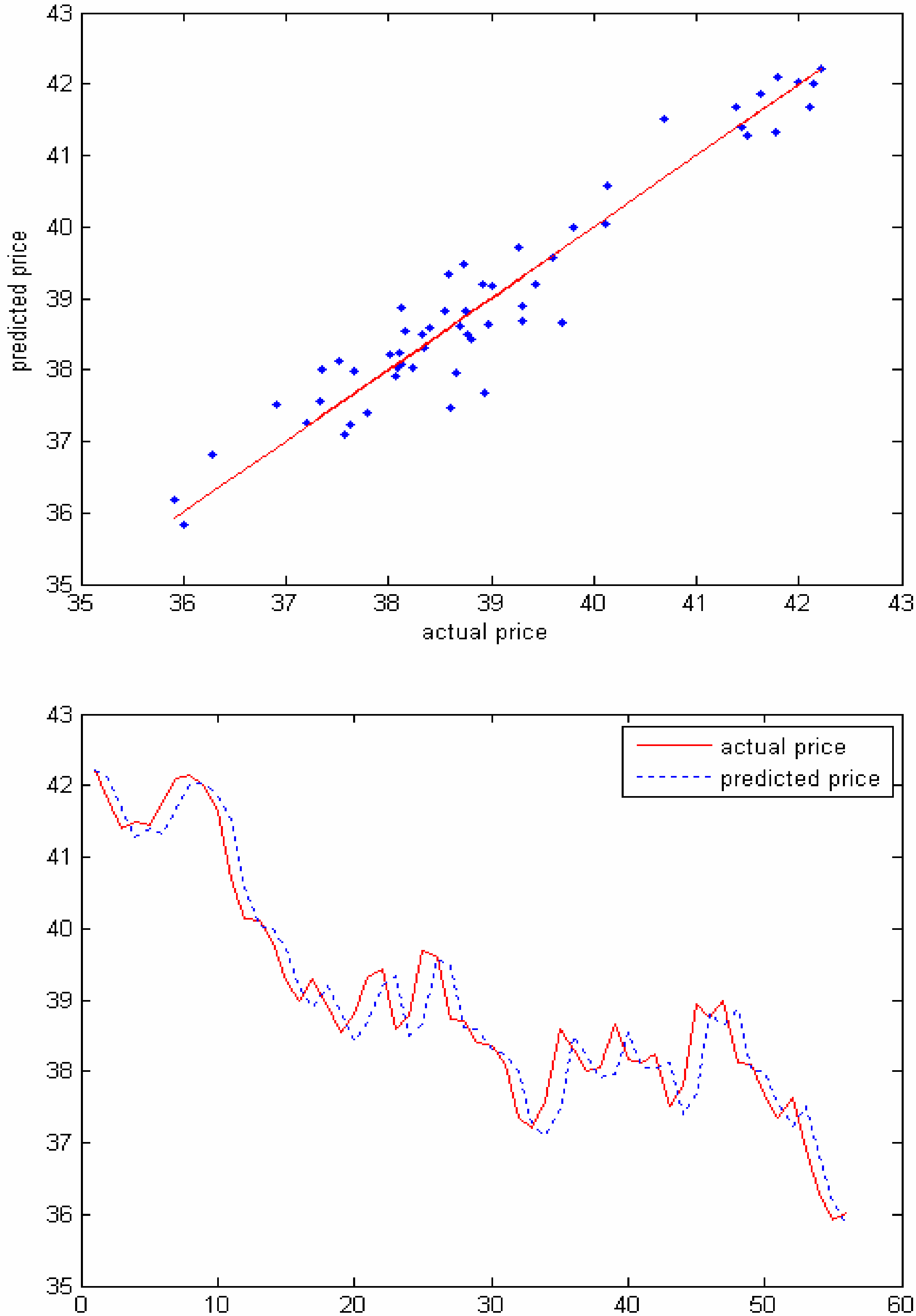


Prediction on MMM

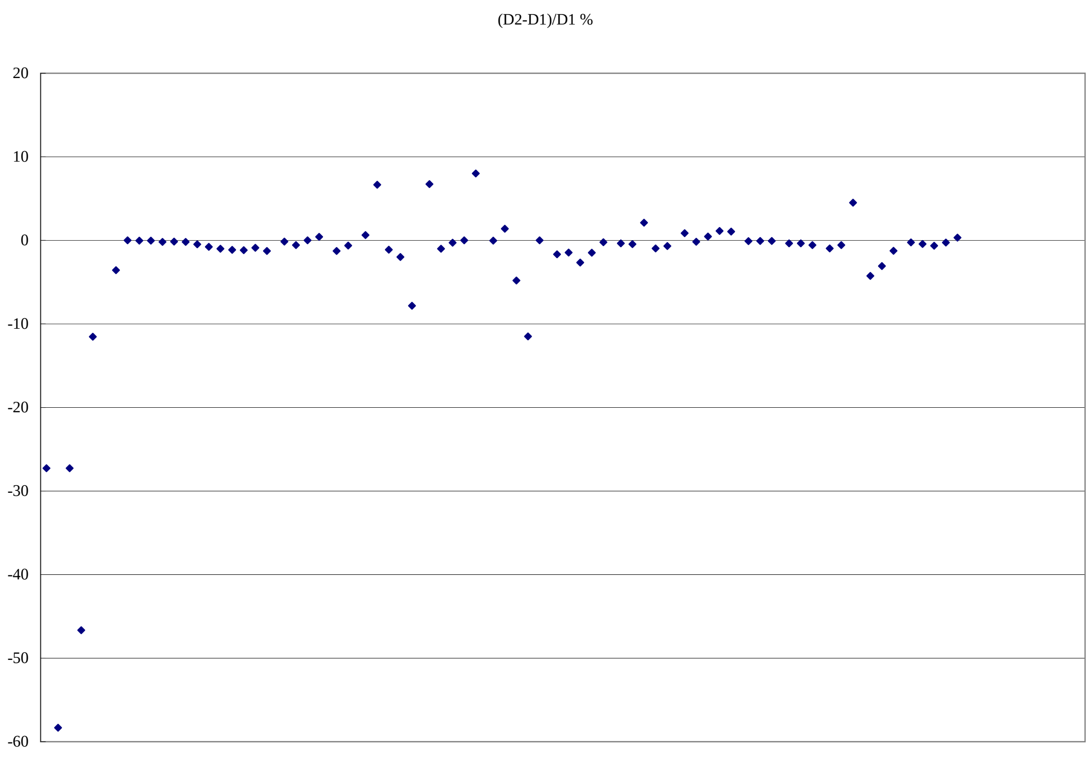




Prediction on QQQQ

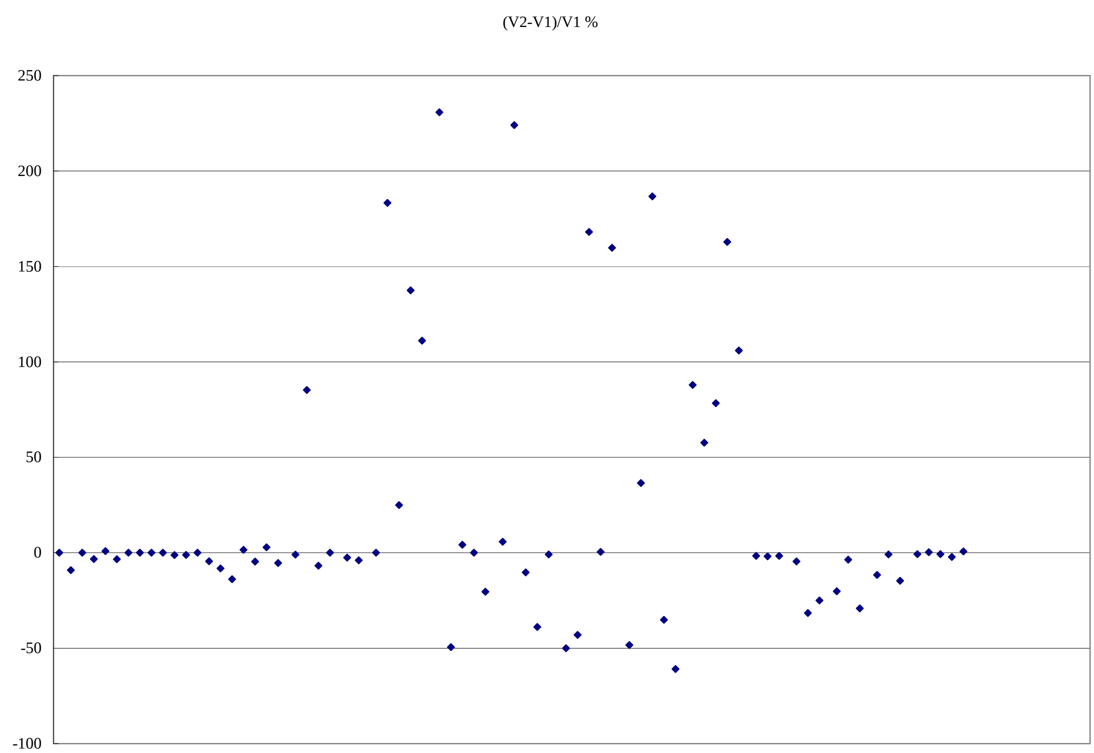


Historically most of the models assume constant drift. If that is indeed the case we should not see much differences in the estimates of two states. The figure below plots $\dfrac{D_1 - D_2}{D_1}\,\%$ where $\{D_1, D_2\}$ is the estimate of the state space of the drift. The 73 points representing the results from 73 data sets. We find there are indeed two different states for the drift (except for a few points lie on the line zero). Moreover, the difference in the two states is bigger among technology stocks and small-cap stocks.

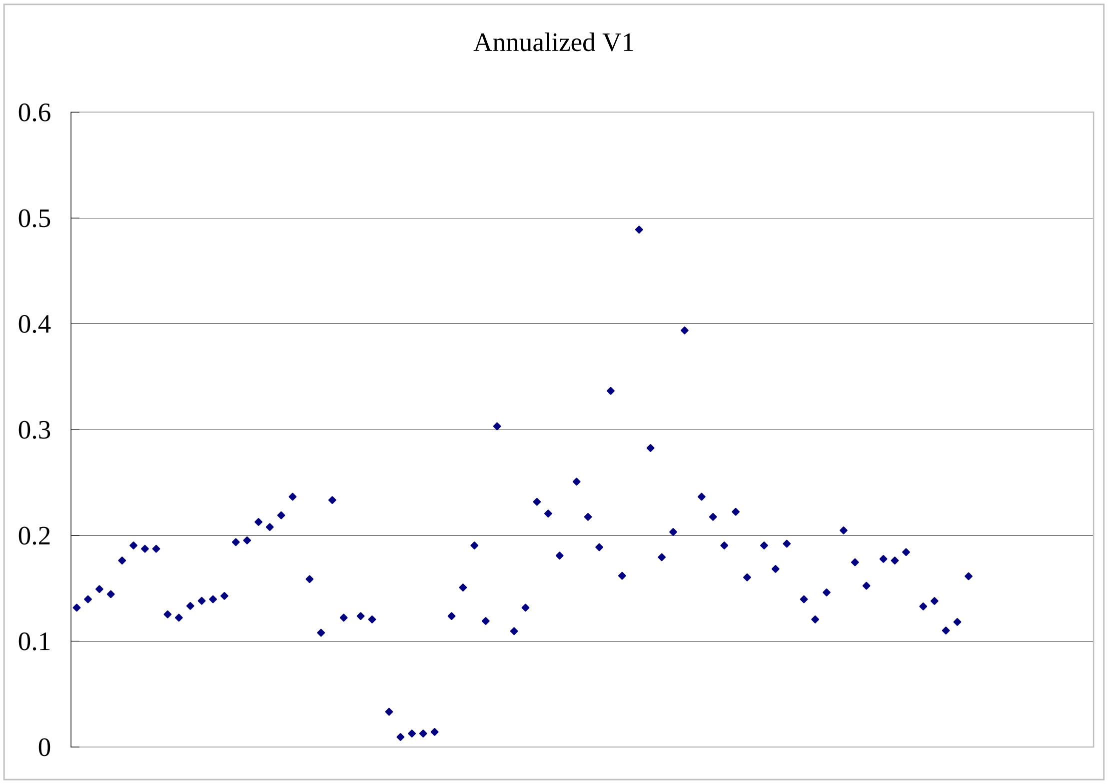


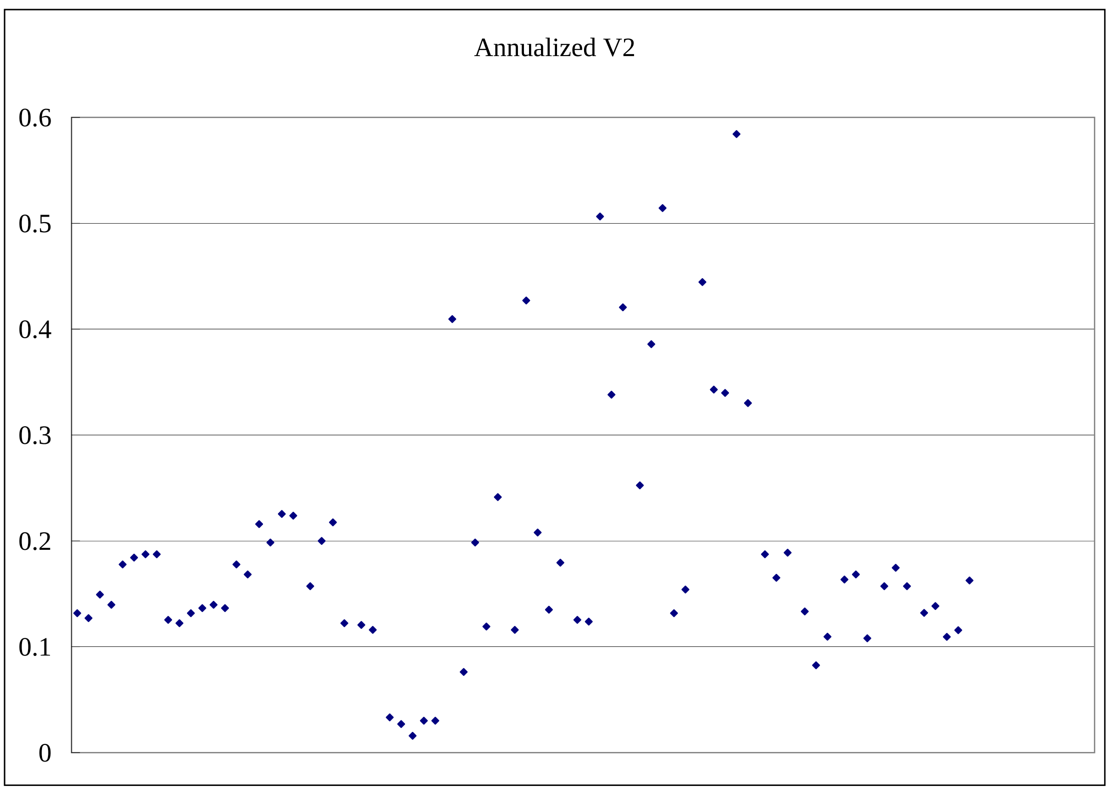
There is another interesting finding, during the period from Mar 2001 to Oct 2006, when most of the U.S. indices suggest negative drifts, the estimate of the drifts for the foreign indices (Hang Seng, Nikkei, FTSE) are positive for both states.

Next we plot $\dfrac{V_1 - V_2}{V_1}$ % where $\{V_1, V_2\}$ is the estimate of the state space of the volatility. Comparatively, the difference in the two states of volatilities (mostly from -50% to $+100\%$) is a lot more significant than the difference in the two states of drifts (mostly from -10% to $+10\%$). Therefore, for analytic reason, some practices can assume constant mean.



According to our estimates, the typical values of annualized volatility are within 10% to 30%, as indicated from the next two plots.


Annualized V1
0.6
0.5
0.4
0.3
0.2
0.1
0


Annualized V2
0.6
0.5
0.4
0.3
0.2
0.1
0

More over, when we look at the transition probabilities, the overall financial market (indicated by the indices) has pretty close probabilities of staying in each state which is indicative to a steady movement of the price process. However, some of the individual stocks such as XOM, PFE have big differences in the transition probabilities, which is an evidence that one state dominant the other over that period.

5.5 Empirical Comparison of HMM with GARCH(1,1)

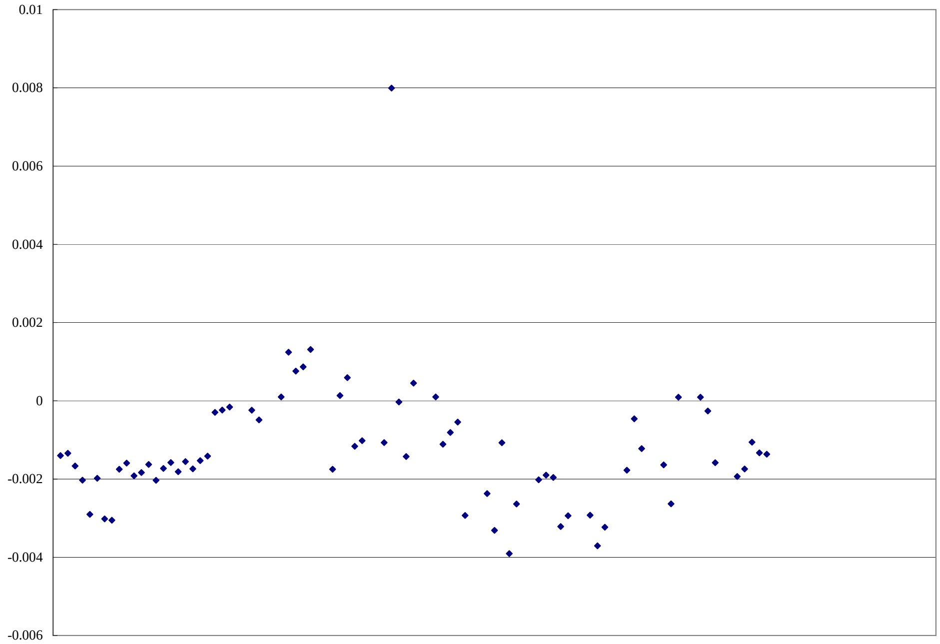
The classical GARCH(1,1) model and Our primary HMM model both assume that the value of the current volatility depends only on the information from previous adjacent period. In this section, we compare the prediction performance of these two models.

First of all, we achieve the predicted prices by GARCH(1,1) on the same data sets as we worked with HMM, then we run the same regression:

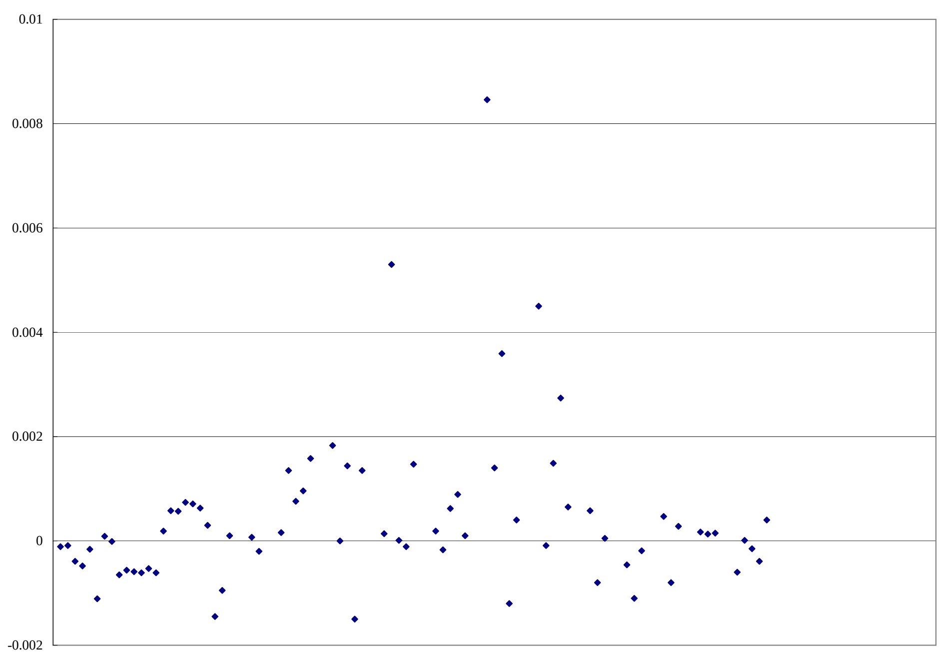
$$\frac{\text{Actual Price}_i}{m.a.c.} = \text{ relative } \alpha + \beta \cdot \frac{\text{Predicted Price}_i}{m.a.c.} + \frac{\varepsilon_i}{m.a.c.}$$

The prediction performance of GARCH(1,1) is also assessed by relative α, β, relative standard error, the t-statistics, and the P-values (Appendix C). Then we compare the relative α, β, relative standard error, and Durbin-Watson statistics achieved from GARCH(1,1) and from HMM (Appendix D).

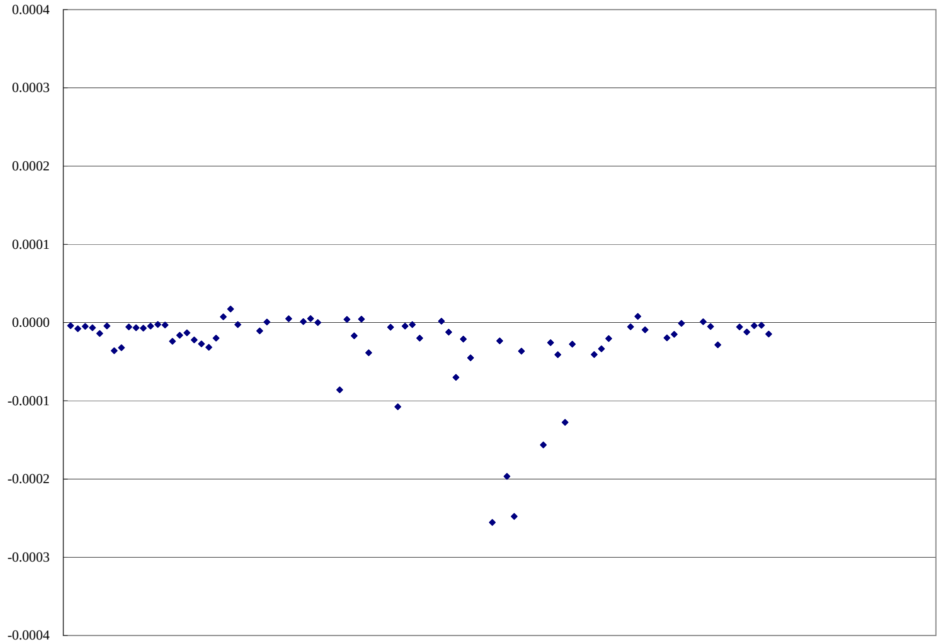
For the four figures below, each point is corresponding to a data set, so there are 73 points in each figure. Take the leftmost (first) point in the first figure for example, it corresponds to DJI, the first data set (Appendix D). We run regression for the actual prices of DJI on the HMM predicted price, and we obtain relative α_{HMM} then we run regression for the actual prices of DJI on the GARCH(1,1) predicted price, and we obtain relative $\alpha_{\text{GARCH(1,1)}}$. The y-coordinate of the first point is $|\text{relative } \alpha_{\text{HMM}} - 0| - |\text{relative } \alpha_{\text{GARCH(1,1)}} - 0|$ on data set DJI.



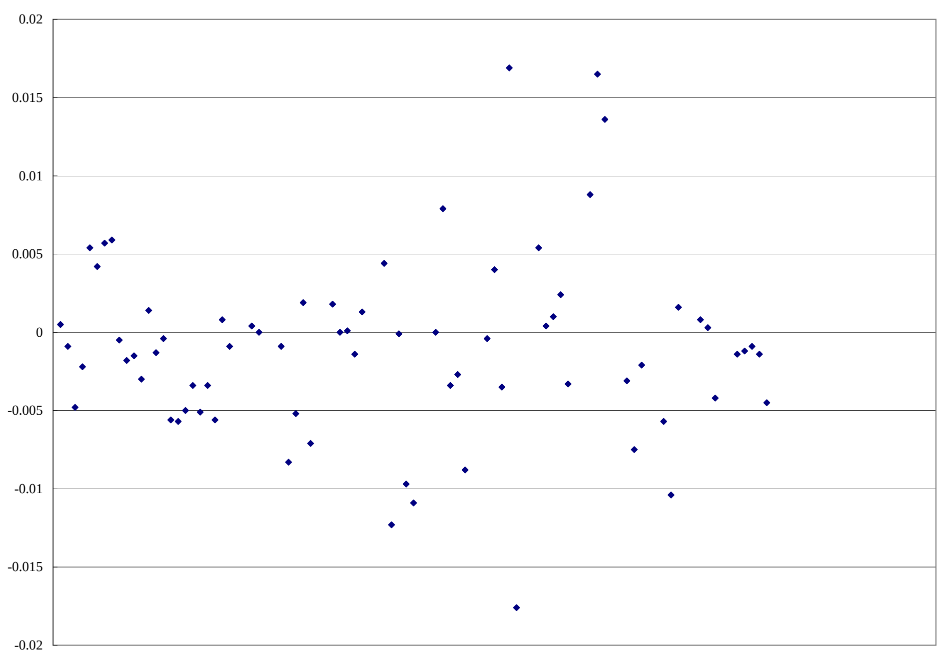
|relative $\alpha_{\text{HMM}} - 0| - |$relative $\alpha_{\text{GARCH}(1,1)} - 0|$
Most of the points are below zero $\Rightarrow$ relative α_{HMM} is closer to 0.



$|\beta_{\text{HMM}} - 1| - |\beta_{\text{GARCH}(1,1)} - 1|$
Majority of the points are below zero $\Rightarrow$ β_{HMM} is closer to 1.



| relative standard error$_{\text{HMM}} - 0| - |$ relative standard error$_{\text{GARCH(1,1)}} - 0|$
Most of the points are below zero $\Rightarrow$ relative standard error$_{\text{HMM}}$ is smaller.



| Durbin-Watson$_{\text{HMM}} - 2| - |$ Durbin-Watson$_{\text{GARCH(1,1)}} - 2|$
Most of the points are below zero $\Rightarrow$ Durbin-Watson Sat$_{\text{HMM}}$ is closer to 2.

The first figure displays $|$ relative $\alpha_{\text{HMM}} - 0| - |$ relative $\alpha_{\text{GARCH(1,1)}} - 0|$. It is obvious that most of the points are below zero, which suggests that relative α_{HMM} is closer to zero. The second figure displays $|\beta_{\text{HMM}} - 1| - |\beta_{\text{GARCH(1,1)}} - 1|$. We still have majority of the points below zero, i.e. β_{HMM} is closer to one. So far we can conclude that HMM provides a more unbiased prediction than GARCH(1,1).

Next we plot $|$ relative standard $\text{error}_{\text{HMM}} - 0| - |$ relative standard $\text{error}_{\text{GARCH(1,1)}} - 0|$. Again we find that most of the points are below zero. This suggests that relative standard $\text{error}_{\text{HMM}}$ is smaller, i.e. the regression of actual price on HMM predicted price has less error than the regression on the GARCH(1,1) predicted price.

The last figure above displays $|$ $\text{Durbin-Watson}_{\text{HMM}} - 2| - |$ $\text{Durbin-Watson}_{\text{GARCH(1,1)}} - 2|$. Since most of the points are below zero, we have Durbin-Watson Sat_{HMM} closer to 2, i.e. the residuals in the regressions on HMM predicted prices are less serially correlated.

According to the 3 criteria for a good model proposed by Fama and Gibbons (1984), we conclude that HMM outperforms GARCH(1,1).

Chapter 6

Conclusion and Future Work

In this chapter we conclude our work and provide some further avenues for the future work.

6.1 Conclusion

This work starts with summarizing and comparing all the popular models for financial market return process on both a theoretic level and an empirical level. It ranges from the ancient Geometric Brownian Motion model to the cutting edge stochastic models, along with the implied volatility and the "model free" realized volatility. It provides a solid development of a hidden Markov model (HMM), from the economic insight of the hypothesis to the mathematic formulation of the estimation and prediction.

In addition, enormous empirical work is done with HMM on our cautiously selected data sets including: 14 major Indices; 10 corresponding Index Based ETFs; 2 government bonds and 5 bond funds; 5 mutual funds; 25 common stocks including 5 from the composition of Dow Johns Industrial Average, 5 from S&P LargeCap 500, 5 from S&P MidCap 400, 5 from S&P SmallCap 600, and 5 from Nasdaq-100; finally another 12 ETFs including 3 'Specialty-Technology', 3 'Large Blend', 3 'Mid-Cap Blend', 3 'Small Blend'. The results can be incorporated with the phenomena observed from the real financial market. Moreover, we compare the applicability of our model with the well established GARCH(1,1) model. As far as the prediction performance is concerned, our results indicate that HMM outperforms GARCH(1,1).

There are several originalities in this work, from the mathematic inferences to the economic interpretations, and from the theoretic foundation to the empirical implementation. We name a few below:

1. In our EM iteration, not only the parameters but also the observation get updated in each pass. This is similar to but different from the classical EM algorithm. The idea is to seek quality rather than quantity of the information, as the recent data are more relevant in terms of predicting the near future. (pp. 69)

2. While judging the prediction performance, we employ the concepts of "relative α" and "relative standard error" as modification to the three criteria for a good model proposed by Fama and Gibbons (1984). This modification has been proved to be more appropriate. (pp. 84)

3. As far as the underlying Markov process is concerned, previous work intended to explain it as a specific economic process. We suggest that it is a combination of all the forces that can move the stock price, including the information about market sentiment, the psychology of market participants; fundamental factors such as earning base, evaluation multiple; and technical factors such as inflation, economic strength of market and peers, substitutions, incidental transactions, demographics, trends, liquidity. Our discovery is that we do not need a clean equation to describe them, which is impossible anyway, and it is sufficient to estimate their overall impact on any price process, i.e. to estimate the hidden Markov chain. (pp. 45)

4. It is interesting to consider how the behaviors of drift and volatility are related across states. However, most of the models force the drift to be constant for analytic reasons. We suggest a technique to model them together as a pair of random variables. Our results indicate that except for a few securities the drift does have different states. In addition, we find very little evidence of leverage effect. (pp. 39, pp. 132 - 134)

5. Our empirical findings include:

    - Let large cap, middle cap, small cap, and technology based indices be different categories, the prediction performance of our model is similar within each category and different between categories. (pp. 100)

- As far as the estimates of drift, volatility, and transition probabilities are concerned, Nasdaq 100 Trust (QQQQ) resembles Nasdaq-100 Index (NDX), S&P 500 SPDR (SPY) resembles S&P 500 Index (GSPC), iShare SP SC 600 (IJR) resembles S&P 600 Small Cap Index (SML), Dow Diamonds Trust (DIA) resembles Dow Jones Industrial Average Index (DJI). (pp. 105 - 106)

- The difference in the two states of volatilities (mostly from -50% to +100%) is a lot more significant than the difference in the two states of drifts (mostly from -10% to +10%). (pp. 114)

- Generally the difference in two states is higher among technology stocks and small-cap stocks. (pp. 114 - 115)

- The transition probabilities among indices are relatively closer compared with the transition probabilities among individual securities. (pp. 132 - 134)

6.2 Future Work

In addition to the improvements and developments we have made in this research, there are further avenues for future work that could be done.

1. Our model assumes that the probability of a change in the underlying economic forces depends on the past only through the value of the most recent

state but this hypothesis has not been tested against a more general case where it could evolve as a higher order Markov chain. This question is similar to the GARCH(P,Q) model, a longer autoregressive part(higher P) or a long moving average part(higher Q) will increase the number of parameters to be estimated thus decrease the accuracy. i.e. In order to loosen the hypothesis, we may sacrifice the accuracy in the estimates.

2. It is important to test the hypothesis of N states against $N+1$ states.

3. The simple time-invariant Markov chain is a good starting point but more work should be done to study the case where the transition probability matrix is A_t rather than the constant A. For example, the January effect[1], Mark Twain effect[2], Halloween indicator[3] would all suggest that A_t can be different in January, October, or November.

4. For the EM estimates, each iteration is guaranteed to increase the log likelihood and the algorithm is guaranteed to converge to a local maximum depending on starting values. However, there is no guarantee that the sequence converges to a maximum likelihood estimator. So the initial value has a great

[1]The January effect is the phenomenon that stocks, especially small-cap stocks, have historically tended to rise during the period starting on the last day of December and ending on the fifth trading day of January.

[2]The Mark Twain effect is the phenomenon of stock returns in October being lower than in other months.

[3]The Halloween indicator refers to the phenomenon that the period from November to April inclusive has significantly stronger stock market growth on average than the other months.

impact on the final estimates and even the computation time, thus the cost of computation in reality. Therefore, certain standard of choosing the initial value has to be established.

5. Some previous papers have suggested an autoregressive tern in the return process itself and it was shown to be relevant for Danish Data. We might consider adding an autoregressive term under the hidden Markov framwork.

6. As volatility estimation plays such an important role in option trading and portfolio risk management, the application of our model to these two fields should be further studied

7. On an empirical level, a lot more experiments can be done. For example: our empirical results are from a complete and diversified portfolio but they are restricted to a certain historical period, namely from Mar 12, 2001 to Oct 3, 2006 on a daily basis. It is quite natural to ask: would the monthly data has a stronger (or weaker) hidden Markov property? The daily data is more meaningful as far as the prediction is concerned, however it can be more valid to check the leverage effect by using monthly data. Is the lower drift always associated with a higher volatility while using monthly data? Even with the daily data, we could ask: are the estimates different from period to period? For example, during the period of financial crisis, business cycles, or fundamental changes in the monetary or fiscal policy, can we have $a_{kk} = 1$?

i.e. the underlying economic forces would stay in a certain state and will not go out. If so, how long would it stay? Can we predict the occupation time of such state? Our prediction generally has $\beta < 1$ which indicates over prediction generally. Moreover there is usually an under prediction when the price is on an increasing trend and over prediction when the price is on a decreasing trend. What could be the reason for it?

Appendix

A. Estimates of Drift, Volatility, Transition

Model: HMM

B. Prediction Performance

Model: HMM

C. Prediction Performance

Model: GARCH(1,1)

D. Model Comparison

Model: HMM and GARCH(1,1)

HMM	**Drift 1**	**Drift 2**	**Volatility 1**	**Volatility 2**	**Transition Prob**	
Data: adjusted price: 3/12/2001---10/3/2006, data length = 1938						
Indices, corresponding						
^DJI	-0.0011	-0.0008	0.0083	0.0083	0.5104	0.5113
					0.4896	0.4887
DIA	-0.0012	-0.0005	0.0088	0.008	0.5228	0.5238
					0.4772	0.4762
^RUA	-0.0011	-0.0008	0.0094	0.0094	0.5100	0.5103
					0.4900	0.4897
IWV	-0.0015	-0.0008	0.0091	0.0088	0.5165	0.5194
					0.4835	0.4806
^IXIC	-0.0026	-0.0023	0.0111	0.0112	0.5036	0.5040
					0.4964	0.4960
ONEQ	-0.0013	0.0003	0.012	0.0116	0.5121	0.5107
					0.4879	0.4893
^NDX	-0.0028	-0.0027	0.0118	0.0118	0.5008	0.5008
					0.4992	0.4992
QQQQ	-0.0028	-0.0028	0.0118	0.0118	0.5007	0.5007
					0.4993	0.4993
^OEX	-0.0008667	-0.0008412	0.0079	0.0079	0.5018	0.5018
					0.4982	0.4982
OEF	-0.0008187	-0.0007881	0.0077	0.0077	0.5021	0.5022
					0.4979	0.4978
^GSPC	-0.0011	-0.0009	0.0084	0.0083	0.5068	0.5070
					0.4932	0.4930
SPY	-0.000979	-0.0008368	0.0087	0.0086	0.5119	0.5116
					0.4881	0.4884
^RUI	-0.0008411	-0.0006916	0.0088	0.0088	0.5087	0.5089
					0.4913	0.4911
IWB	-0.0013	-0.0007	0.009	0.0086	0.5209	0.5224
					0.4791	0.4776
^MID	-0.0023	-0.0005	0.0122	0.0112	0.5300	0.5290
					0.4700	0.4710
MDY	-0.0032	0	0.0123	0.0106	0.5399	0.5362
					0.4601	0.4638
^SML	-0.0042	0.0006	0.0134	0.0136	0.5240	0.5261
					0.4760	0.4739
IJR	-0.0041	0.0007	0.0131	0.0125	0.5193	0.5120
					0.4807	0.4880
^RUT	-0.0036	-0.0004	0.0138	0.0142	0.4996	0.4993
					0.5004	0.5007
IWM	-0.0048	0.0013	0.0149	0.0141	0.5200	0.5137
					0.4800	0.4863
^NYA	-0.0013	-0.0011	0.01	0.0099	0.5160	0.5150
					0.4840	0.4850
^HSI	0.0023	0.001	0.0068	0.0126	0.4837	0.5157
					0.5163	0.4843
^N225	0.0013	0.0013	0.0147	0.0137	0.5203	0.5213
					0.4797	0.4787
^FTSE	0.0000828	0.0001178	0.0077	0.0077	0.5003	0.5003
					0.4997	0.4997

† Appendix A. Estimates of Drift, Volatility, Transition Matrix (1) Model: HMM

Gov Bonds						
^TNX	-0.0008309	0.0002196	0.0078	0.0076	0.5382	0.5486
					0.4618	0.4514
^TYX	0.0016	0.0006	0.0076	0.0073	0.4972	0.5002
					0.5028	0.4998
Bond Funds						
INBNX	0.0000882	0.0001445	0.0021	0.0021	0.4714	0.4837
					0.5286	0.5163
AFTEX	0.0000234	0.0001794	0.0006	0.0017	0.2553	0.3115
					0.7447	0.6885
PTHYX	0.0003311	-0.0000385	0.0008	0.001	0.3841	0.5129
					0.6159	0.4871
ELFTX	-0.00009709	0.00009709	0.0008	0.0019	0.4408	0.5428
					0.5592	0.4572
MDXBX	-0.0000284	0.0001941	0.0009	0.0019	0.5455	0.4329
					0.4545	0.5671
Stocks in DJI						
MMM	-0.0011	-0.0085	0.0078	0.0258	0.7296	0.6642
					0.2704	0.3358
JNJ	0.0022	0	0.0095	0.0048	0.3104	0.3162
					0.6896	0.6838
WMT	-0.0011	-0.0008	0.012	0.0125	0.5032	0.5033
					0.4968	0.4967
KO	0.0007529	0.0007585	0.0075	0.0075	0.5001	0.5001
					0.4999	0.4999
BA	-0.0002	-0.0018	0.0191	0.0152	0.4753	0.4732
					0.5247	0.5268
Stocks in S&P500						
GE	-0.0009352	-0.0009009	0.0069	0.0073	0.5365	0.5355
					0.4635	0.4645
MSFT	-0.0018	-0.0043	0.0083	0.0269	0.4496	0.4752
					0.5504	0.5248
XOM	-0.0005	0.0019	0.0146	0.0131	0.4959	0.4997
					0.5041	0.5003
PFE	0.0002	-0.0021	0.0139	0.0085	0.2216	0.2582
					0.7784	0.7418
C	-0.0006061	-0.0006063	0.0114	0.0113	0.4955	0.4955
					0.5045	0.5045
Stocks in S&P400						
WPO	0.0033	-0.0022	0.0158	0.0079	0.4139	0.4069
					0.5861	0.5931
NYB	0.0011	-0.0005	0.0137	0.0078	0.3221	0.3146
					0.6779	0.6854
TSN	-0.002	0.0033	0.0119	0.0319	0.7139	0.4690
					0.2861	0.5310
VLO	-0.0005117	0.0002455	0.0212	0.0213	0.4654	0.4560
					0.5346	0.5440
LLL	-0.0013	-0.001	0.0102	0.0265	0.6881	0.6222
					0.3119	0.3778
Stocks in S&P600						
NVR	-0.0115	-0.0075	0.0308	0.0159	0.5883	0.6008
					0.4117	0.3992
URBN	-0.0088	-0.0049	0.0178	0.0243	0.5207	0.5823
					0.4793	0.4177
MRX	-0.0026	-0.0081	0.0113	0.0324	0.5621	0.5790

† Appendix A. Estimates of Drift, Volatility, Transition Matrix (2) Model: HMM

					0.4379	0.4210
IDXX	-0.0055	-0.0002	0.0128	0.0083	0.4612	0.2218
					0.5388	0.7782
ROP	-0.0037	-0.0012	0.0248	0.0097	0.4286	0.4083
					0.5714	0.5917
Stocks in NASDAQ100						
QCOM	-0.0041	-0.0076	0.0149	0.028	0.5064	0.4577
					0.4936	0.5423
INTC	-0.0018	-0.0015	0.0137	0.0216	0.5718	0.5635
					0.4282	0.4365
CSCO	-0.0024	-0.0035	0.012	0.0214	0.5848	0.5693
					0.4152	0.4307
EBAY	-0.0033	-0.007	0.014	0.0368	0.6321	0.5643
					0.3679	0.4357
DELL	-0.0023	-0.0047	0.0101	0.0208	0.5320	0.5003
					0.4680	0.4997
ETF, Tech						
MTK	-0.0034	-0.0031	0.012	0.0118	0.5057	0.5059
					0.4943	0.4941
XLK	-0.0028	-0.0026	0.0106	0.0104	0.5049	0.5052
					0.4951	0.4948
IYW	-0.0032	-0.003	0.0121	0.0119	0.5051	0.5053
					0.4949	0.4947
ETF, Large-cap						
IVV	-0.0011	-0.0007	0.0088	0.0084	0.5168	0.5171
					0.4832	0.4829
VV	0.0014	0.0009	0.0076	0.0052	0.4266	0.4292
					0.5734	0.5708
ELV	-0.0008255	-0.0003611	0.0092	0.0069	0.4819	0.4750
					0.5181	0.5250
ETF, Middle-cap						
IJH	-0.0031	-0.0001	0.0129	0.0103	0.5561	0.5511
					0.4439	0.4489
IWR	-0.0016	-0.0007	0.011	0.0106	0.5045	0.5015
					0.4955	0.4985
VO	0.0002	0.0011	0.0096	0.0068	0.5418	0.5141
					0.4582	0.4859
ETF, Small-cap						
VB	-0.0004	0.0013	0.0112	0.0099	0.5139	0.5091
					0.4861	0.4909
IWC	-0.0003025	0.0006296	0.0111	0.011	0.5036	0.5028
					0.4964	0.4972
DSV	-0.0029	0.0007	0.0116	0.0099	0.5358	0.4887
					0.4642	0.5113
Mutual Funds						
VFINX	-0.0011	-0.0008	0.0084	0.0083	0.51149	0.5121
					0.48851	0.4879
AGTHX	-0.0016	-0.0009	0.0087	0.0087	0.50595	0.50475
					0.49405	0.49525
AIVSX	-0.0007	-0.0003	0.0069	0.0069	0.51088	0.51016
					0.48912	0.48984
AWSHX	-0.0006	-0.0004	0.0074	0.0073	0.5222	0.52009
					0.4778	0.47991
FCNTX	-0.0010	-0.0014	0.0102	0.0102	0.47697	0.46803
					0.52303	0.53197

† Appendix A. Estimates of Drift, Volatility, Transition Matrix (3) Model: HMM

HMM	alpha	relative alpha	beta	S. E. for regression	relative S.E. for regression	D-W
Data: adjusted price: 3/12/2001---10/3/2006, data length = 1938						
Indices, corresponding ETFs						
^DJI	1120.2000	0.1005	0.8996	90.7144	0.0081	1.7699
DIA	12.3891	0.1126	0.8875	0.9059	0.0082	1.7235
^RUA	47.4651	0.0643	0.9356	6.8564	0.0093	1.7818
IWV	5.6249	0.0767	0.9234	0.6528	0.0089	1.7782
^IXIC	81.3016	0.0373	0.9629	24.3120	0.0112	1.7350
ONEQ	8.5784	0.1039	0.8962	0.9739	0.0118	1.7610
^NDX	62.3106	0.0392	0.9609	18.7758	0.0118	1.8321
QQQQ	1.5101	0.0387	0.9615	0.4581	0.0117	1.7874
^OEX	51.3247	0.0883	0.9118	4.5312	0.0078	1.8363
OEF	5.5014	0.0945	0.9056	0.4447	0.0076	1.8402
^GSPC	96.8597	0.0762	0.9239	10.5351	0.0083	1.7897
SPY	11.1868	0.0886	0.9116	1.0798	0.0086	1.7804
^RUI	48.5067	0.0702	0.9302	5.8032	0.0084	1.7622
IWB	5.8763	0.0856	0.9145	0.5968	0.0087	1.7905
^MID	35.9552	0.0472	0.9534	8.5745	0.0113	1.6092
MDY	7.8859	0.0569	0.9433	1.6021	0.0116	1.6525
^SML	17.2268	0.0461	0.9537	4.9565	0.0133	1.6683
IJR	4.1207	0.0666	0.9337	0.8000	0.0129	1.7084
^RUT	40.8194	0.0567	0.9436	10.0405	0.0139	1.5902
IWM	4.5731	0.0642	0.9363	1.0423	0.0146	1.6908
^NYA	477.4242	0.0584	0.9416	80.5900	0.0099	1.6987
^HSI	1060.6000	0.0650	0.9349	160.3885	0.0098	1.8891
^N225	1400.6000	0.0918	0.9081	210.5372	0.0138	1.9442
^FTSE	1367.7000	0.2332	0.7670	42.5662	0.0073	2.0863

† Appendix B. Prediction Performance (1) Model: HMM

Gov Bonds						
^TNX	0.8439	0.1653	0.8347	0.0383	0.0075	2.1038
^TYX	0.8492	0.1643	0.8358	0.0381	0.0074	2.0085
Bond Funds						
INBNX	0.8570	0.1884	0.8116	0.0094	0.0021	1.8945
AFTEX	0.8510	0.0711	0.9289	0.1287	0.0108	1.9444
PTHYX	0.7951	0.0631	0.9370	0.0119	0.0009	2.1950
ELFTX	0.7863	0.0703	0.9297	0.0167	0.0015	2.0154
MDXBX	0.8734	0.0857	0.9142	0.0151	0.0015	1.9702
Stocks in DJI						
MMM	8.0556	0.0999	0.9001	1.1656	0.0145	2.1643
JNJ	2.6273	0.0443	0.9556	0.3976	0.0067	1.9630
WMT	7.8045	0.1667	0.8334	0.5356	0.0114	1.7717
KO	4.1096	0.0968	0.9035	0.3004	0.0071	1.8809
BA	12.9808	0.1577	0.8424	1.3503	0.0164	1.7374
Stocks in S&P500						
GE	4.2936	0.1290	0.8711	0.2266	0.0068	2.2466
MSFT	-0.7063	-0.0310	1.0322	0.5085	0.0223	2.1383
XOM	5.4137	0.0896	0.9106	0.8200	0.0136	2.0890
PFE	1.5870	0.0686	0.9316	0.2335	0.0101	1.6592
C	14.0730	0.2960	0.7041	0.4869	0.0102	1.9529
Stocks in S&P400						
WPO	110.9407	0.1433	0.8571	8.9835	0.0116	1.7870
NYB	3.4763	0.2189	0.7812	0.1539	0.0097	2.0443
TSN	1.9725	0.1306	0.8692	0.3251	0.0215	1.7898
VLO	5.2865	0.0855	0.9149	1.2826	0.0207	1.8359
LLL	3.8091	0.0496	0.9509	1.3093	0.0170	1.9824
Stocks in S&P600						
NVR	14.4651	0.0247	0.9772	12.2259	0.0209	2.3116
URBN	0.0865	0.0047	0.9964	0.3865	0.0210	2.2012
MRX	0.3109	0.0112	0.9904	0.6544	0.0236	1.8190

† Appendix B. Prediction Performance (2) Model: HMM

IDXX	0.3488	0.0045	0.9963	0.7659	0.0099	1.6610
ROP	2.9197	0.0634	0.9369	0.8326	0.0181	2.0071
Stocks in NASDAQ 100						
QCOM	1.1622	0.0264	0.9739	0.9805	0.0223	1.7194
INTC	1.0860	0.0597	0.9408	0.3172	0.0174	2.0516
CSCO	2.2606	0.1128	0.8872	0.3255	0.0162	1.9988
EBAY	1.9060	0.0622	0.9383	0.7823	0.0255	1.7731
DELL	3.1291	0.1278	0.8721	0.3792	0.0155	1.9549
ETF, Tech						
MTK	1.0769	0.0216	0.9786	0.5931	0.0119	1.7069
XLK	0.3599	0.0176	0.9826	0.2164	0.0106	1.8059
IYW	0.9460	0.0197	0.9806	0.5805	0.0121	1.8454
ETF, Large-cap						
IVV	11.3092	0.0890	0.9111	1.0790	0.0085	1.7426
VV	3.4679	0.0611	0.9389	0.3535	0.0062	1.9442
ELV	6.7927	0.0937	0.9063	0.5793	0.0080	1.7316
ETF, Middle-cap						
IJH	4.5919	0.0603	0.9400	0.8945	0.0117	1.6429
IWR	4.9994	0.0553	0.9450	0.9733	0.0108	1.8088
VO	7.7205	0.1178	0.8821	0.5333	0.0081	1.8228
ETF, Small-cap						
VB	9.2734	0.1515	0.8484	0.6208	0.0101	1.8832
IWC	7.0269	0.1347	0.8653	0.5622	0.0108	1.7518
DSV	5.3191	0.0828	0.9175	0.6910	0.0108	1.6529
Mutual Funds						
VFINX	9.6930	0.0837	0.9164	0.9590	0.0083	1.7844
AGTHX	1.3391	0.0441	0.9559	0.2644	0.0087	1.5347
AIVSX	2.6823	0.0885	0.9116	0.2083	0.0069	1.6366
AWSHX	3.6646	0.1185	0.8815	0.2244	0.0073	1.7192
FCNTX	3.4812	0.0576	0.9425	0.6138	0.0102	1.5971

† Appendix B. Prediction Performance (3) Model: HMM

HMM	t_alpha H0: alpha=0	t_beta H0: beta=0	t_beta H0: beta=1 beta=1	P_alpha H0: alpha=0	P_ beta H0: beta=0 beta=0	P_beta H0: beta=1
Data: adjusted price: 3/12/2001---10/3/2006, data length = 1938						
Indices, corresponding ETFs						
^DJI	1.89546	16.9745	-1.8934	0.063389	2.6356E-23	0.063665
DIA	1.99576	15.7297	-1.9932	0.051014	8.1618E-22	0.051306
^RUA	1.19516	17.399	-1.1978	0.23725	8.5049E-24	0.23622
IWV	1.4786	17.8069	-1.4768	0.14506	2.9199E-24	0.14553
^IXIC	0.946197	24.4674	-0.94278	0.34826	6.67E-31	0.34999
ONEQ	1.41215	12.1822	-1.4107	0.16364	4.0265E-17	0.16406
^NDX	1.01528	24.8931	-1.0118	0.3145	2.8695E-31	0.31614
QQQQ	1.01061	25.1208	-1.0068	0.31671	1.8238E-31	0.31852
^OEX	1.5765	16.2779	-1.575	0.12075	1.7612E-22	0.1211
OEF	1.66026	15.905	-1.6588	0.10266	4.9802E-22	0.10295
^GSPC	1.4516	17.601	-1.4503	0.1524	4.9976E-24	0.15277
SPY	1.58098	16.2712	-1.5787	0.11972	1.7943E-22	0.12024
^RUI	1.35648	17.9848	-1.3493	0.18059	1.8422E-24	0.18289
IWB	1.54379	16.5034	-1.5423	0.12848	9.467E-23	0.12885
^MID	1.13841	22.9971	-1.1252	0.25998	1.44E-29	0.26546
MDY	1.3185	21.8814	-1.3142	0.1929	1.6352E-28	0.19432
^SML	1.00255	20.7732	-1.0084	0.32055	2.0194E-27	0.31778
IJR	1.47756	20.7362	-1.4716	0.14534	2.2006E-27	0.14693
^RUT	1.28734	21.4427	-1.2822	0.20347	4.3689E-28	0.20524
IWM	1.37396	20.0494	-1.3649	0.17513	1.1046E-26	0.17793
^NYA	1.29312	20.8432	-1.2919	0.20148	1.7176E-27	0.20191
^HSI	1.79136	25.764	-1.7954	0.07884	5.1631E-32	0.078186
^N225	1.61802	16.0201	-1.6211	0.11148	3.6072E-22	0.11083
^FTSE	2.6573	8.7393	-2.6552	0.010338	6.4407E-12	0.010395

† Appendix B. Prediction Performance (4) Model: HMM

Gov Bonds						
^TNX	2.2332	11.2774	-2.2333	0.029701	8.1857E-16	0.029697
^TYX	2.20028	11.1959	-2.1992	0.032085	1.079E-15	0.032166
Bond Funds						
INBNX	2.47841	10.6737	-2.4784	0.016354	6.4667E-15	0.016353
AFTEX	1.31522	17.192	-1.3161	0.19399	1.4727E-23	0.1937
PTHYX	1.34526	19.9875	-1.3449	0.18416	1.2802E-26	0.18429
ELFTX	1.28854	17.0514	-1.2886	0.20305	2.1442E-23	0.20302
MDXBX	1.53204	16.3389	-1.5329	0.13135	1.4882E-22	0.13114
Stocks in DJI						
MMM	2.6919	24.2835	-2.6942	0.0094398	9.827E-31	0.0093821
JNJ	0.937564	20.2479	-0.94026	0.35264	6.8989E-27	0.35127
WMT	3.0864	15.4406	-3.086	0.0031939	1.8579E-21	0.0031974
KO	1.36341	12.7239	-1.3583	0.17841	6.9768E-18	0.18003
BA	2.34447	12.5295	-2.3434	0.022762	1.3031E-17	0.022823
Stocks in S&P500						
GE	2.48836	16.8031	-2.4872	0.01595	4.1857E-23	0.015998
MSFT	-0.375436	12.4944	0.38937	0.70881	1.4592E-17	0.69854
XOM	1.73143	17.601	-1.7273	0.089084	4.9983E-24	0.089826
PFE	1.68097	22.845	-1.6771	0.098546	1.9939E-29	0.099311
C	3.6868	8.7711	-3.6861	0.00052795	5.731E-12	0.00052924
Stocks in S&P400						
WPO	2.12063	12.6894	-2.1156	0.038562	7.7919E-18	0.039003
NYB	2.6009	9.2818	-2.5994	0.011971	8.9367E-13	0.012017
TSN	2.00678	13.3676	-2.0113	0.04979	9.125E-19	0.049301
VLO	1.52361	16.329	-1.5192	0.13344	1.5295E-22	0.13455
LLL	1.13783	21.8292	-1.1267	0.26021	1.8367E-28	0.26483
Stocks in S&P600						
NVR	1.52996	61.424	-1.436	0.13186	1.187E-51	0.15677
URBN	0.203457	43.8927	-0.1606	0.83954	6.3213E-44	0.87301
MRX	0.482552	42.9694	-0.41749	0.63136	1.9309E-43	0.67798

† Appendix B. Prediction Performance (5) Model: HMM

IDXX	0.0994003	21.7896	-0.081229	0.92119	2.006E-28	0.93556
ROP	1.15593	17.1114	-1.1515	0.2528	1.8264E-23	0.2546
Stocks in NASDAQ						
QCOM	0.947595	35.1494	-0.94199	0.34756	6.857E-39	0.3504
INTC	1.07358	16.9268	-1.0647	0.28778	2.9969E-23	0.29177
CSCO	2.30201	18.1257	-2.3035	0.025218	1.2819E-24	0.025131
EBAY	1.36875	20.7129	-1.3613	0.17674	2.3224E-27	0.17908
DELL	2.62966	17.962	-2.6334	0.011111	1.954E-24	0.011004
ETF, Tech						
MTK	0.666582	30.2366	-0.66258	0.50788	1.5802E-35	0.51042
XLK	0.466443	26.0645	-0.46232	0.64277	2.8894E-32	0.64571
IYW	0.556198	27.763	-0.55025	0.58037	1.206E-33	0.58442
ETF, Large-cap						
IVV	1.59638	16.3459	-1.594	0.11624	1.4595E-22	0.11676
VV	1.59152	24.4721	-1.5925	0.11733	6.6959E-31	0.1171
ELV	1.60374	15.5068	-1.603	0.1146	1.5379E-21	0.11477
ETF, Middle-cap						
IJH	1.37594	21.4658	-1.3709	0.17452	4.1467E-28	0.17608
IWR	1.03635	17.7156	-1.0318	0.30466	3.7038E-24	0.30679
VO	2.09619	15.7	-2.0978	0.040766	8.8786E-22	0.040615
ETF, Small-cap						
VB	2.30832	12.9329	-2.3111	0.024839	3.5837E-18	0.024673
IWC	2.04825	13.1588	-2.0482	0.045411	1.7547E-18	0.045411
DSV	1.6228	17.9878	-1.6172	0.11046	1.8277E-24	0.11166
Mutual Funds						
VFINX	1.52787	16.737	-1.5266	0.13238	5.0068E-23	0.13271
AGTHX	1.05174	22.7877	-1.0517	0.2976	2.25E-29	0.29764
AIVSX	1.56829	16.1619	-1.5677	0.12266	2.43E-22	0.1228
AWSHX	1.85907	13.8266	-1.8583	0.06847	2.21E-19	0.068587
FCNTX	1.2887	21.0937	-1.2875	0.203	9.66E-28	0.2034

† Appendix B. Prediction Performance (6) Model: HMM

GARCH(1,1)	Predicted Next Day Volatility	alpha	relative alpha	beta	S. E. for regression	relative S. E. for regression	D-W
Data: adjusted price: 3/12/2001---10/3/2006, data length = 1938							
Indices, corresponding ETFs							
^DJI	0.005335	1136	0.1019	0.89949	90.78	0.008143	1.7704
DIA	0.00515	12.533	0.11394	0.88741	0.90652	0.008241	1.7226
^RUA	0.00579	48.706	0.065965	0.93521	6.8618	0.009293	1.777
IWV	0.005541	5.7742	0.078731	0.92292	0.65332	0.008908	1.776
^IXIC	0.008732	87.716	0.040203	0.96274	24.367	0.011168	1.7404
ONEQ	0.008834	8.7411	0.10588	0.89509	0.97413	0.0118	1.7652
^NDX	0.009903	67.066	0.042219	0.96099	18.821	0.011848	1.8378
QQQQ	0.009757	1.6288	0.041755	0.96149	0.45921	0.011772	1.7933
^OEX	0.005108	52.33	0.09005	0.91115	4.5334	0.007801	1.8358
OEF	0.005012	5.5915	0.096092	0.90504	0.44488	0.007645	1.8384
^GSPC	0.005372	99.281	0.078117	0.92331	10.543	0.008295	1.7882
SPY	0.005181	11.42	0.090435	0.91099	1.0805	0.008557	1.7774
^RUI	0.005425	49.649	0.071829	0.92967	5.8068	0.008401	1.7636
IWB	0.005304	6.0181	0.087632	0.91389	0.59725	0.008697	1.7892
^MID	0.00751	37.254	0.04893	0.95359	8.5884	0.01128	1.6088
MDY	0.008259	8.1078	0.05848	0.94388	1.605	0.011576	1.6469
^SML	0.009517	17.919	0.047913	0.95427	4.9657	0.013277	1.6626
IJR	0.009717	4.2181	0.068156	0.93444	0.80157	0.012952	1.7034
^RUT	0.010071	42.071	0.058438	0.94431	10.061	0.013974	1.5868
IWM	0.010269	4.6808	0.06573	0.93693	1.0443	0.014664	1.6857
^NYA	0.005379	488.61	0.059814	0.9419	80.692	0.009878	1.6953
^HSI	0.008009	1064.7	0.065297	0.93345	160.16	0.009822	1.8835
^N225	0.010519	1404.8	0.092037	0.90715	210.36	0.013782	1.945
^FTSE	0.007055	1368.5	0.23336	0.7671	42.576	0.00726	2.0872

† Appendix C. Prediction Performance (1) Model: GARCH(1,1)

Gov Bonds							
^TNX	0.007513	0.84508	0.16554	0.83477	0.038352	0.007513	2.1034
^TYX	0.008051	0.85185	0.16479	0.8356	0.038102	0.007371	2.0085
Bond Funds							
INBNX	0.002344	0.85634	0.1883	0.81176	0.009375	0.002062	1.8936
AFTEX	0.001633	0.83655	0.069859	0.93025	0.017521	0.001463	1.9361
PTHYX	0.001172	0.78598	0.06234	0.93776	0.011891	0.000943	2.2002
ELFTX	0.00177	0.77704	0.069432	0.93066	0.016653	0.001488	2.0135
MDXBX	0.001608	0.85975	0.084388	0.91578	0.015094	0.001482	1.9631
Stocks in DJI							
MMM	0.01076	8.1952	0.10165	0.90193	1.1723	0.014541	2.1625
JNJ	0.005396	2.6219	0.044167	0.9556	0.39774	0.0067	1.963
WMT	0.011161	7.7789	0.16611	0.83484	0.53651	0.011457	1.7718
KO	0.005762	4.1579	0.097964	0.902	0.30013	0.007071	1.8795
BA	0.014595	13.063	0.15872	0.84375	1.3534	0.016443	1.7387
Stocks in S&P500							
GE	0.007282	4.3289	0.13007	0.87124	0.22677	0.006814	2.2422
MSFT	0.012354	-0.52359	-0.02301	1.0269	0.5104	0.022426	2.1506
XOM	0.013677	5.4134	0.08963	0.91061	0.81994	0.013576	2.0891
PFE	0.010214	1.6203	0.070024	0.93149	0.23361	0.010096	1.6495
C	0.007381	14.051	0.29555	0.70557	0.48784	0.010261	1.942
Stocks in S&P400							
WPO	0.012602	110.89	0.1432	0.85729	8.9835	0.011602	1.787
NYB	0.010551	3.4934	0.22001	0.78103	0.15407	0.009703	2.0364
TSN	0.019915	1.9844	0.13141	0.86982	0.32611	0.021595	1.7864
VLO	0.026261	5.322	0.086043	0.91579	1.2843	0.020765	1.8332
LLL	0.011255	4.0316	0.052533	0.951	1.3118	0.017094	1.9736
Stocks in S&P600							
NVR	0.026957	15.847	0.027072	0.98566	12.37	0.021132	2.312
URBN	0.029183	0.14895	0.008011	1.0022	0.39091	0.021024	2.1972
MRX	0.017269	0.33987	0.012273	0.99399	0.65829	0.023771	1.8155

† Appendix C. Prediction Performance (2) Model: GARCH(1,1)

IDXX	0.013804	0.64488	0.008406	0.9951	0.77708	0.010129	1.6779
ROP	0.015833	3.0433	0.066038	0.9373	0.83488	0.018116	2.0247
Stocks in NASDAQ 100							
QCOM	0.022383	1.2507	0.028418	0.9784	0.98711	0.022429	1.7248
INTC	0.017371	1.1202	0.061601	0.94071	0.31755	0.017463	2.0512
CSCO	0.019498	2.2999	0.11476	0.88869	0.32634	0.016283	1.9998
EBAY	0.027195	2.005	0.065413	0.94104	0.78644	0.025657	1.7755
DELL	0.017766	3.2011	0.13074	0.87275	0.37987	0.015515	1.9516
ETF, Tech							
MTK	0.011073	1.2227	0.024523	0.97918	0.59517	0.011937	1.7157
XLK	0.009139	0.43575	0.021307	0.9818	0.2171	0.010616	1.8224
IYW	0.01013	1.1032	0.022932	0.98065	0.58251	0.012109	1.859
ETF, Large-cap							
IVV	0.005221	11.535	0.090776	0.91064	1.0797	0.008497	1.7395
VV	0.005993	3.4956	0.061562	0.9378	0.3532	0.00622	1.9367
ELV	0.005637	6.8778	0.094922	0.90611	0.57967	0.008	1.7295
ETF, Middle-cap							
IJH	0.007768	4.7174	0.061938	0.94047	0.89618	0.011766	1.6372
IWR	0.006511	5.2374	0.057933	0.9442	0.97465	0.010781	1.7984
VO	0.006477	7.7147	0.11771	0.88238	0.53337	0.008138	1.8244
ETF, Small-cap							
VB	0.008733	9.2704	0.15141	0.84857	0.62092	0.010141	1.884
IWC	0.008993	7.0394	0.13496	0.86543	0.56236	0.010782	1.7521
DSV	0.008246	5.4202	0.084382	0.91765	0.69279	0.010785	1.6487
Mutual Funds							
VFINX	0.005347	9.9172	0.085609	0.91581	0.95969	0.008284	1.783
AGTHX	0.005459	1.392	0.045879	0.9559	0.26481	0.008728	1.5335
AIVSX	0.004202	2.7144	0.089526	0.91143	0.20841	0.006874	1.6357
AWSHX	0.004985	3.7057	0.11987	0.88114	0.22446	0.007261	1.7178
FCNTX	0.006062	3.5638	0.058973	0.94287	0.61471	0.010172	1.5926

† Appendix C. Prediction Performance (3) Model: GARCH(1,1)

GARCH(1,1)	t for alpha H0: alpha=0	t for beta H0: beta=0	t for beta H0: beta=1	p for alpha H0: alpha=0	p for beta H0: beta=0	p for beta H0: beta=1
Data: adjusted price: 3/12/2001---10/3/2006, data length = 1938						
Indices, corresponding ETFs						
^DJI	1.9208	16.959	-1.895	0.060046	2.75E-23	0.063446
DIA	2.0175	15.716	-1.994	0.04863	8.47E-22	0.051215
^RUA	1.2254	17.378	-1.2039	0.22573	8.98E-24	0.23387
IWV	1.5165	17.782	-1.4851	0.13522	3.12E-24	0.14332
^IXIC	1.0185	24.408	-0.94454	0.31296	7.62E-31	0.3491
ONEQ	1.4386	12.164	-1.4257	0.15604	4.28E-17	0.15971
^NDX	1.0901	24.834	-1.0081	0.2805	3.23E-31	0.31791
QQQQ	1.0874	25.06	-1.0037	0.2817	2.06E-31	0.32001
^OEX	1.6066	16.259	-1.5855	0.11397	1.86E-22	0.1187
OEF	1.6867	15.888	-1.6671	0.097435	5.22E-22	0.10128
^GSPC	1.4868	17.578	-1.46	0.14287	5.31E-24	0.15008
SPY	1.6129	16.251	-1.5878	0.1126	1.9E-22	0.11817
^RUI	1.3875	17.963	-1.3589	0.17098	1.95E-24	0.17982
IWB	1.5799	16.48	-1.5528	0.11997	1.01E-22	0.12631
^MID	1.1776	22.965	-1.1178	0.24412	1.54E-29	0.26861
MDY	1.3532	21.854	-1.2994	0.18164	1.74E-28	0.19932
^SML	1.0409	20.747	-0.99431	0.30255	2.15E-27	0.32451
IJR	1.5095	20.711	-1.453	0.137	2.33E-27	0.15201
^RUT	1.3242	21.417	-1.2631	0.19103	4.63E-28	0.21197
IWM	1.4037	20.026	-1.348	0.16614	1.17E-26	0.18329
^NYA	1.3217	20.823	-1.2844	0.19183	1.8E-27	0.20448
^HSI	1.801	25.762	-1.8367	0.077292	5.18E-32	0.071756
^N225	1.6242	16.017	-1.6394	0.11016	3.64E-22	0.10693
^FTSE	2.6582	8.7387	-2.6531	0.010314	6.45E-12	0.010452

† Appendix C. Prediction Performance (4) Model: GARCH(1,1)

Gov Bonds						
^TNX	2.236	11.276	-2.2319	0.029509	8.21E-16	0.029791
^TYX	2.2071	11.193	-2.2022	0.031579	1.09E-15	0.031943
Bond Funds						
INBNX	2.4762	10.675	-2.4754	0.016445	6.45E-15	0.016478
AFTEX	1.292	17.204	-1.2899	0.20186	1.42E-23	0.20258
PTHYX	1.3354	20.088	-1.3332	0.18735	1.01E-26	0.18805
ELFTX	1.2743	17.081	-1.2726	0.20801	1.98E-23	0.2086
MDXBX	1.5052	16.334	-1.5022	0.13811	1.51E-22	0.13886
Stocks in DJI						
MMM	2.7229	24.193	-2.6306	0.008697	1.18E-30	0.011085
JNJ	0.93524	20.239	-0.94025	0.35383	7.05E-27	0.35128
WMT	3.0711	15.441	-3.0546	0.003337	1.86E-21	0.003496
KO	1.3805	12.712	-1.3812	0.17311	7.24E-18	0.17292
BA	2.354	12.521	-2.3186	0.02224	1.34E-17	0.024233
Stocks in S&P500						
GE	2.5067	16.793	-2.4818	0.015227	4.3E-23	0.016216
MSFT	-0.27725	12.384	0.32426	0.78264	2.09E-17	0.74699
XOM	1.7314	17.601	-1.7279	0.089089	4.99E-24	0.089725
PFE	1.7151	22.827	-1.6789	0.09207	2.07E-29	0.098944
C	3.674	8.7722	-3.6606	0.00055	5.71E-12	0.000573
Stocks in S&P400						
WPO	2.1196	12.692	-2.1129	0.038654	7.72E-18	0.039251
NYB	2.6115	9.2715	-2.5993	0.011649	9.27E-13	0.01202
TSN	2.0125	13.334	-1.9956	0.049169	1.01E-18	0.051029
VLO	1.5317	16.323	-1.5009	0.13143	1.56E-22	0.13922
LLL	1.202	21.789	-1.1227	0.23463	2.01E-28	0.26655
Stocks in S&P600						
NVR	1.6567	61.237	-0.89107	0.10339	1.40E-51	0.37685
URBN	0.34632	43.65	0.097401	0.73045	8.46E-44	0.92277
MRX	0.52444	42.873	-0.2592	0.60212	2.17E-43	0.79647

† Appendix C. Prediction Performance (5) Model: GARCH(1,1)

IDXX	0.18111	21.45	-0.10558	0.85696	4.30E-28	0.91631
ROP	1.2016	17.071	-1.1419	0.23478	2.04E-23	0.25854
Stocks in NASDAQ 100						
QCOM	1.0129	35.075	-0.77437	0.3156	7.65E-39	0.44209
INTC	1.1059	16.904	-1.0653	0.27365	3.19E-23	0.29147
CSCO	2.3361	18.109	-2.2683	0.02323	1.34E-24	0.027336
EBAY	1.4322	20.662	-1.2946	0.15784	2.61E-27	0.20095
DELL	2.685	17.94	-2.6159	0.009612	2.07E-24	0.011516
ETF, Tech						
MTK	0.75416	30.15	-0.64098	0.45403	1.83E-35	0.52425
XLK	0.56312	25.966	-0.48125	0.57568	3.49E-32	0.63228
IYW	0.64642	27.672	-0.54594	0.52074	1.42E-33	0.58736
ETF, Large-cap						
IVV	1.6271	16.326	-1.6021	0.10954	1.54E-22	0.11497
VV	1.6058	24.467	-1.6227	0.11416	6.76E-31	0.11047
ELV	1.6228	15.494	-1.6054	0.11045	1.59E-21	0.11423
ETF, Middle-cap						
IJH	1.411	21.438	-1.3569	0.16398	4.41E-28	0.18046
IWR	1.0843	17.678	-1.0448	0.28307	4.09E-24	0.30077
VO	2.0942	15.701	-2.0929	0.040954	8.85E-22	0.041068
ETF, Small-cap						
VB	2.3072	12.933	-2.3081	0.024905	3.58E-18	0.024853
IWC	2.0514	13.158	-2.0459	0.04509	1.76E-18	0.045651
DSV	1.6495	17.945	-1.6105	0.10486	2.04E-24	0.11313
Mutual Funds						
VFINX	1.5621	16.715	-1.5365	0.1241	5.32E-23	0.13025
AGTHX	1.0918	22.756	-1.0499	0.27978	2.41E-29	0.29846
AIVSX	1.5861	16.15	-1.5695	0.11855	2.51E-22	0.12238
AWSHX	1.879	13.814	-1.8634	0.065644	2.30E-19	0.067846
FCNTX	1.3174	21.072	-1.2768	0.19327	1.01E-27	0.20714

† Appendix C. Prediction Performance (6) Model: GARCH(1,1)

	| relative alpha - 0 | HMM-GARCH(1,1)	| beta - 1 | HMM-GARCH(1,1)	| relative S.E. - 0 | HMM-GARCH(1,1)	| DW - 2 | HMM-GARCH(1,1)
Indices and ETFs				
^DJI	-0.0014	-0.00011	-0.000004	0.0005
DIA	-0.00134	-9E-05	-0.000008	-0.0009
^RUA	-0.001665	-0.00039	-0.000005	-0.0048
IWV	-0.002031	-0.00048	-0.000007	-0.0022
^IXIC	-0.002903	-0.00016	-0.000014	0.0054
ONEQ	-0.00198	-0.00111	-0.000004	0.0042
^NDX	-0.003019	9E-05	-0.000036	0.0057
QQQQ	-0.003055	-1E-05	-0.000032	0.0059
^OEX	-0.00175	-0.00065	-0.000006	-0.0005
OEF	-0.001592	-0.00056	-0.000007	-0.0018
^GSPC	-0.001917	-0.00059	-0.000007	-0.0015
SPY	-0.001835	-0.00061	-0.000005	-0.003
^RUI	-0.001629	-0.00053	-0.000002	0.0014
IWB	-0.002032	-0.00061	-0.000003	-0.0013
^MID	-0.00173	0.00019	-0.000024	-0.0004
MDY	-0.00158	0.00058	-0.000016	-0.0056
^SML	-0.001813	0.00057	-0.000013	-0.0057
IJR	-0.001556	0.00074	-0.000022	-0.005
^RUT	-0.001738	0.00071	-0.000027	-0.0034
IWM	-0.00153	0.00063	-0.000032	-0.0051
^NYA	-0.001414	0.0003	-0.000020	-0.0034
^HSI	-0.000297	-0.00145	0.000007	-0.0056
^N225	-0.000237	-0.00095	0.000017	0.0008
^FTSE	-0.00016	1E-04	-0.000003	-0.0009
Gov Bonds				
^TNX	-0.00024	7E-05	-0.000011	0.0004
^TYX	-0.00049	-0.0002	0.000001	0
Bond Funds				
INBNX	0.0001	0.00016	0.000005	-0.0009
AFTEX	0.001241	0.00135	0.009290	-0.0083
PTHYX	0.00076	0.00076	0.000001	-0.0052
ELFTX	0.000868	0.00096	0.000005	0.0019
MDXBX	0.001312	0.00158	0.000000	-0.0071
Stocks in DJI				
MMM	-0.00175	0.00183	-0.000086	0.0018
JNJ	0.000133	0	0.000004	0
WMT	0.00059	0.00144	-0.000017	1E-04
KO	-0.001164	-0.0015	0.000004	-0.0014
BA	-0.00102	0.00135	-0.000039	0.0013
Stocks in SP500				
GE	-0.00107	0.00014	-0.000006	0.0044
MSFT	0.007994	0.0053	-0.000108	-0.0123
XOM	-3E-05	1E-05	-0.000005	-0.0001

† Appendix D. Model Comparison (1) HMM and GARCH(1,1)

PFE	-0.001424	-0.00011	-0.000003	-0.0097
C	0.00045	0.00147	-0.000020	-0.0109
Stocks in SP400				
WPO	0.0001	0.00019	0.000002	0
NYB	-0.00111	-0.00017	-0.000012	0.0079
TSN	-0.00081	0.00062	-0.000070	-0.0034
VLO	-0.000543	0.00089	-0.000021	-0.0027
LLL	-0.002933	1E-04	-0.000045	-0.0088
Stocks in SP600				
NVR	-0.002372	0.00846	-0.000256	-0.0004
URBN	-0.003311	0.0014	-0.000023	0.004
MRX	-0.001073	0.00359	-0.000197	-0.0035
IDXX	-0.0039058	-0.0012	-0.000248	0.0169
ROP	-0.002638	0.0004	-0.000036	-0.0176
Stocks in NDX				
QCOM	-0.002018	0.0045	-0.000156	0.0054
INTC	-0.001901	-9E-05	-0.000026	0.0004
CSCO	-0.00196	0.00149	-0.000041	0.001
EBAY	-0.003213	0.00274	-0.000128	0.0024
DELL	-0.00294	0.00065	-0.000028	-0.0033
ETF, Tech				
MTK	-0.002923	0.00058	-0.000041	0.0088
XLK	-0.003707	-0.0008	-0.000034	0.0165
IYW	-0.003232	5E-05	-0.000020	0.0136
ETF, Large-cap				
IVV	-0.001776	-0.00046	-0.000005	-0.0031
VV	-0.000462	-0.0011	0.000008	-0.0075
ELV	-0.001222	-0.00019	-0.000009	-0.0021
ETF, Middle-cap				
IJH	-0.001638	0.00047	-0.000020	-0.0057
IWR	-0.002633	-0.0008	-0.000015	-0.0104
VO	9E-05	0.00028	-0.000001	0.0016
ETF, Small-cap				
VB	9E-05	0.00017	0.000001	0.0008
IWC	-0.00026	0.00013	-0.000005	0.0003
DSV	-0.001582	0.00015	-0.000029	-0.0042
Mutual Funds				
VFINX	-0.001935	-0.0006	-0.000006	-0.0014
AGTHX	-0.001744	1E-05	-0.000012	-0.0012
AIVSX	-0.001058	-0.00015	-0.000004	-0.0009
AWSHX	-0.00133	-0.00039	-0.000004	-0.0014
FCNTX	-0.001367	0.0004	-0.000015	-0.0045

† Appendix D. Model Comparison (2) HMM and GARCH(1,1)

	HMM t for alpha H0: alpha=0	GARCH t for alpha H0: alpha=0	HMM t for beta H0: beta=0	GARCH t for beta H0: beta=0	HMM t for beta H0: beta=1	GARCH t for beta H0: beta=1
Indices and ETFs						
^DJI	1.8955	1.9208	16.9745	16.959	-1.8934	-1.895
DIA	1.9958	2.0175	15.7297	15.716	-1.9932	-1.994
^RUA	1.1952	1.2254	17.399	17.378	-1.1978	-1.2039
IWV	1.4786	1.5165	17.8069	17.782	-1.4768	-1.4851
^IXIC	0.9462	1.0185	24.4674	24.408	-0.94278	-0.94454
ONEQ	1.4122	1.4386	12.1822	12.164	-1.4107	-1.4257
^NDX	1.0153	1.0901	24.8931	24.834	-1.0118	-1.0081
QQQQ	1.0106	1.0874	25.1208	25.06	-1.0068	-1.0037
^OEX	1.5765	1.6066	16.2779	16.259	-1.575	-1.5855
OEF	1.6603	1.6867	15.905	15.888	-1.6588	-1.6671
^GSPC	1.4516	1.4868	17.601	17.578	-1.4503	-1.46
SPY	1.5810	1.6129	16.2712	16.251	-1.5787	-1.5878
^RUI	1.3565	1.3875	17.9848	17.963	-1.3493	-1.3589
IWB	1.5438	1.5799	16.5034	16.48	-1.5423	-1.5528
^MID	1.1384	1.1776	22.9971	22.965	-1.1252	-1.1178
MDY	1.3185	1.3532	21.8814	21.854	-1.3142	-1.2994
^SML	1.0026	1.0409	20.7732	20.747	-1.0084	-0.99431
IJR	1.4776	1.5095	20.7362	20.711	-1.4716	-1.453
^RUT	1.2873	1.3242	21.4427	21.417	-1.2822	-1.2631
IWM	1.3740	1.4037	20.0494	20.026	-1.3649	-1.348
^NYA	1.2931	1.3217	20.8432	20.823	-1.2919	-1.2844
^HSI	1.7914	1.8010	25.764	25.762	-1.7954	-1.8367
^N225	1.6180	1.6242	16.0201	16.017	-1.6211	-1.6394
^FTSE	2.6573	2.6582	8.7393	8.7387	-2.6552	-2.6531

† Appendix D. Model Comparison (3) HMM and GARCH(1,1)

Gov Bonds						
^TNX	2.2332	2.2360	11.2774	11.276	-2.2333	-2.2319
^TYX	2.2003	2.2071	11.1959	11.193	-2.1992	-2.2022
Bond Funds						
INBNX	2.4784	2.4762	10.6737	10.675	-2.4784	-2.4754
AFTEX	1.3152	1.2920	17.192	17.204	-1.3161	-1.2899
PTHYX	1.3453	1.3354	19.9875	20.088	-1.3449	-1.3332
ELFTX	1.2885	1.2743	17.0514	17.081	-1.2886	-1.2726
MDXBX	1.5320	1.5052	16.3389	16.334	-1.5329	-1.5022
Stocks in DJI						
MMM	2.6919	2.7229	24.2835	24.193	-2.6942	-2.6306
JNJ	0.9376	0.9352	20.2479	20.239	-0.94026	-0.94025
WMT	3.0864	3.0711	15.4406	15.441	-3.086	-3.0546
KO	1.3634	1.3805	12.7239	12.712	-1.3583	-1.3812
BA	2.3445	2.3540	12.5295	12.521	-2.3434	-2.3186
Stocks in SP500						
GE	2.4884	2.5067	16.8031	16.793	-2.4872	-2.4818
MSFT	-0.3754	-0.2773	12.4944	12.384	0.38937	0.32426
XOM	1.7314	1.7314	17.601	17.601	-1.7273	-1.7279
PFE	1.6810	1.7151	22.845	22.827	-1.6771	-1.6789
C	3.6868	3.6740	8.7711	8.7722	-3.6861	-3.6606
Stocks in SP400						
WPO	2.1206	2.1196	12.6894	12.692	-2.1156	-2.1129
NYB	2.6009	2.6115	9.2818	9.2715	-2.5994	-2.5993
TSN	2.0068	2.0125	13.3676	13.334	-2.0113	-1.9956
VLO	1.5236	1.5317	16.329	16.323	-1.5192	-1.5009
LLL	1.1378	1.2020	21.8292	21.789	-1.1267	-1.1227
Stocks in SP600						
NVR	1.5300	1.6567	61.424	61.237	-1.436	-0.89107
URBN	0.2035	0.3463	43.8927	43.65	-0.1606	0.097401
MRX	0.4826	0.5244	42.9694	42.873	-0.41749	-0.2592

† Appendix D. Model Comparison (4) HMM and GARCH(1,1)

IDXX	0.0994	0.1811	21.7896	21.45	-0.081229	-0.10558
ROP	1.1559	1.2016	17.1114	17.071	-1.1515	-1.1419
Stocks in NDX						
QCOM	0.9476	1.0129	35.1494	35.075	-0.94199	-0.77437
INTC	1.0736	1.1059	16.9268	16.904	-1.0647	-1.0653
CSCO	2.3020	2.3361	18.1257	18.109	-2.3035	-2.2683
EBAY	1.3688	1.4322	20.7129	20.662	-1.3613	-1.2946
DELL	2.6297	2.6850	17.962	17.94	-2.6334	-2.6159
ETF, Tech						
MTK	0.6666	0.7542	30.2366	30.15	-0.66258	-0.64098
XLK	0.4664	0.5631	26.0645	25.966	-0.46232	-0.48125
IYW	0.5562	0.6464	27.763	27.672	-0.55025	-0.54594
ETF, Large-cap						
IVV	1.5964	1.6271	16.3459	16.326	-1.594	-1.6021
VV	1.5915	1.6058	24.4721	24.467	-1.5925	-1.6227
ELV	1.6037	1.6228	15.5068	15.494	-1.603	-1.6054
ETF, Middle-cap						
IJH	1.3759	1.4110	21.4658	21.438	-1.3709	-1.3569
IWR	1.0364	1.0843	17.7156	17.678	-1.0318	-1.0448
VO	2.0962	2.0942	15.7	15.701	-2.0978	-2.0929
ETF, Small-cap						
VB	2.3083	2.3072	12.9329	12.933	-2.3111	-2.3081
IWC	2.0483	2.0514	13.1588	13.158	-2.0482	-2.0459
DSV	1.6228	1.6495	17.9878	17.945	-1.6172	-1.6105
Mutual Funds						
VFINX	1.5279	1.5621	16.737	16.715	-1.5266	-1.5365
AGTHX	1.0517	1.0918	22.7877	22.756	-1.0517	-1.0499
AIVSX	1.5683	1.5861	16.1619	16.15	-1.5677	-1.5695
AWSHX	1.8591	1.8790	13.8266	13.814	-1.8583	-1.8634
FCNTX	1.2887	1.3174	21.0937	21.072	-1.2875	-1.2768

† Appendix D. Model Comparison (5) HMM and GARCH(1,1)

	HMM p for alpha H0: alpha=0	GARCH p for alpha H0: alpha=0	HMM p for beta H0: beta=0	GARCH p for beta H0: beta=0	HMM p for beta H0: beta=1	GARCH p for beta H0: beta=1
Indices and ETFs						
^DJI	0.063389	0.060046	2.6356E-23	2.7457E-23	0.063665	0.063446
DIA	0.051014	0.04863	8.1618E-22	8.4742E-22	0.051306	0.051215
^RUA	0.23725	0.22573	8.5049E-24	8.9817E-24	0.23622	0.23387
IWV	0.14506	0.13522	2.9199E-24	3.1161E-24	0.14553	0.14332
^IXIC	0.34826	0.31296	6.67E-31	7.6207E-31	0.34999	0.3491
ONEQ	0.16364	0.15604	4.0265E-17	4.2757E-17	0.16406	0.15971
^NDX	0.3145	0.2805	2.8695E-31	3.2291E-31	0.31614	0.31791
QQQQ	0.31671	0.2817	1.8238E-31	2.0577E-31	0.31852	0.32001
^OEX	0.12075	0.11397	1.7612E-22	1.8567E-22	0.1211	0.1187
OEF	0.10266	0.097435	4.9802E-22	5.2165E-22	0.10295	0.10128
^GSPC	0.1524	0.14287	4.9976E-24	5.312E-24	0.15277	0.15008
SPY	0.11972	0.1126	1.7943E-22	1.8999E-22	0.12024	0.11817
^RUI	0.18059	0.17098	1.8422E-24	1.9496E-24	0.18289	0.17982
IWB	0.12848	0.11997	9.467E-23	1.0097E-22	0.12885	0.12631
^MID	0.25998	0.24412	1.44E-29	1.5404E-29	0.26546	0.26861
MDY	0.1929	0.18164	1.6352E-28	1.7374E-28	0.19432	0.19932
^SML	0.32055	0.30255	2.0194E-27	2.1457E-27	0.31778	0.32451
IJR	0.14534	0.137	2.2006E-27	2.3337E-27	0.14693	0.15201
^RUT	0.20347	0.19103	4.3689E-28	4.6349E-28	0.20524	0.21197
IWM	0.17513	0.16614	1.1046E-26	1.1674E-26	0.17793	0.18329
^NYA	0.20148	0.19183	1.7176E-27	1.8016E-27	0.20191	0.20448
^HSI	0.07884	0.077292	5.1631E-32	5.1779E-32	0.078186	0.071756
^N225	0.11148	0.11016	3.6072E-22	3.6415E-22	0.11083	0.10693
^FTSE	0.010338	0.010314	6.4407E-12	6.4545E-12	0.010395	0.010452

† Appendix D. Model Comparison (6) HMM and GARCH(1,1)

Gov Bonds						
^TNX	0.029701	0.029509	8.1857E-16	8.2141E-16	0.029697	0.029791
^TYX	0.032085	0.031579	1.079E-15	1.0909E-15	0.032166	0.031943
Bond Funds						
INBNX	0.016354	0.016445	6.4667E-15	6.445E-15	0.016353	0.016478
AFTEX	0.19399	0.20186	1.4727E-23	1.4249E-23	0.1937	0.20258
PTHYX	0.18416	0.18735	1.2802E-26	1.0076E-26	0.18429	0.18805
ELFTX	0.20305	0.20801	2.1442E-23	1.9817E-23	0.20302	0.2086
MDXBX	0.13135	0.13811	1.4882E-22	1.5086E-22	0.13114	0.13886
Stocks in DJI						
MMM	0.0094398	0.0086969	9.827E-31	1.183E-30	0.0093821	0.011085
JNJ	0.35264	0.35383	6.8989E-27	7.0509E-27	0.35127	0.35128
WMT	0.0031939	0.0033367	1.8579E-21	1.8582E-21	0.0031974	0.0034964
KO	0.17841	0.17311	6.9768E-18	7.2424E-18	0.18003	0.17292
BA	0.022762	0.02224	1.3031E-17	1.3409E-17	0.022823	0.024233
Stocks in SP500						
GE	0.01595	0.015227	4.1857E-23	4.3031E-23	0.015998	0.016216
MSFT	0.70881	0.78264	1.4592E-17	2.0891E-17	0.69854	0.74699
XOM	0.089084	0.089089	4.9983E-24	4.99E-24	0.089826	0.089725
PFE	0.098546	0.09207	1.9939E-29	2.07E-29	0.099311	0.098944
C	0.00052795	0.00054962	5.731E-12	5.71E-12	0.00052924	0.00057306
Stocks in SP400						
WPO	0.038562	0.038654	7.7919E-18	7.72E-18	0.039003	0.039251
NYB	0.011971	0.011649	8.9367E-13	9.27E-13	0.012017	0.01202
TSN	0.04979	0.049169	9.125E-19	1.01E-18	0.049301	0.051029
VLO	0.13344	0.13143	1.5295E-22	1.56E-22	0.13455	0.13922
LLL	0.26021	0.23463	1.8367E-28	2.01E-28	0.26483	0.26655
Stocks in SP600						
NVR	0.13186	0.10339	1.187E-51	1.40E-51	0.15677	0.37685
URBN	0.83954	0.73045	6.3213E-44	8.46E-44	0.87301	0.92277
MRX	0.63136	0.60212	1.9309E-43	2.17E-43	0.67798	0.79647

† Appendix D. Model Comparison (7) HMM and GARCH(1,1)

IDXX	0.92119	0.85696	2.006E-28	4.30E-28	0.93556	0.91631
ROP	0.2528	0.23478	1.8264E-23	2.04E-23	0.2546	0.25854
Stocks in NDX						
QCOM	0.34756	0.3156	6.857E-39	7.65E-39	0.3504	0.44209
INTC	0.28778	0.27365	2.9969E-23	3.19E-23	0.29177	0.29147
CSCO	0.025218	0.02323	1.2819E-24	1.34E-24	0.025131	0.027336
EBAY	0.17674	0.15784	2.3224E-27	2.61E-27	0.17908	0.20095
DELL	0.011111	0.0096119	1.954E-24	2.07E-24	0.011004	0.011516
ETF, Tech						
MTK	0.50788	0.45403	1.5802E-35	1.83E-35	0.51042	0.52425
XLK	0.64277	0.57568	2.8894E-32	3.49E-32	0.64571	0.63228
IYW	0.58037	0.52074	1.206E-33	1.42E-33	0.58442	0.58736
ETF, Large-cap						
IVV	0.11624	0.10954	1.4595E-22	1.54E-22	0.11676	0.11497
VV	0.11733	0.11416	6.6959E-31	6.76E-31	0.1171	0.11047
ELV	0.1146	0.11045	1.5379E-21	1.59E-21	0.11477	0.11423
ETF, Middle-cap						
IJH	0.17452	0.16398	4.1467E-28	4.41E-28	0.17608	0.18046
IWR	0.30466	0.28307	3.7038E-24	4.09E-24	0.30679	0.30077
VO	0.040766	0.040954	8.8786E-22	8.85E-22	0.040615	0.041068
ETF, Small-cap						
VB	0.024839	0.024905	3.5837E-18	3.58E-18	0.024673	0.024853
IWC	0.045411	0.04509	1.7547E-18	1.76E-18	0.045411	0.045651
DSV	0.11046	0.10486	1.8277E-24	2.04E-24	0.11166	0.11313
Mutual Funds						
VFINX	0.13238	0.1241	5.0068E-23	5.32E-23	0.13271	0.13025
AGTHX	0.2976	0.27978	2.25E-29	2.41E-29	0.29764	0.29846
AIVSX	0.12266	0.11855	2.43E-22	2.51E-22	0.1228	0.12238
AWSHX	0.06847	0.065644	2.21E-19	2.30E-19	0.068587	0.067846
FCNTX	0.203	0.19327	9.66E-28	1.01E-27	0.2034	0.20714

† Appendix D. Model Comparison (8) HMM and GARCH(1,1)

Bibliography

[1] N. Amenc, F. Goltz, and V. L. Sourd, Assessing the Quality of Stock Market Indices, *EDHEC Publication* (2006)

[2] R.D. Arnott, J. Hsu, and P. Moore, Fundamental Indexation, *Financial Analysts Journal* **60(2)** (2005) 83-99.

[3] L.E.Baum and T.Petrie, Statistical Inference for Probabilistic Functions of Finite State Markov Chains, *Ann. Math. Statist* **37** (1966) 1554–1563.

[4] L.E.Baum, T.Petrie, G.Soules, and N.Weiss, A Maximization Technique Occuring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *Ann. Math. Statist* **41** (1970) 164–171.

[5] F.Black and M.Scholes, The Pricing of Options and Corporate Liabilities, *J. Pol. Econ.* **81** (1973) 637–659.

[6] F.Black, The Pricing of Commodity Contracts, *Journal of Financial Economics* **3** (1976) 167–179.

[7] N.Bollen, Valuing Options in Regime-Switching Models, *Journal of Derivatives* **6** (1998) 38–49

[8] T.Bollerslev, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31** (1986) 307–327

[9] D.A.Coast, G.G.Cano, and S.A.Briller, Use of Hidden Markov Models for Electrocardiographic Signal Analysis *Journal of Electrocardiol.* **23 Suppl.** (1990) 184–191

[10] J.C.Cox, S.A. Ross and M. Rubinstein, Option Pricing: A Simplified Approach, *Journal of Financial Economics* **7** (1979) 229-263.

[11] J.C.Cox and M. Rubinstein, Options Markets, *Prentice Hall* (1985)

[12] N.J. Cutland,P.E.Kopp and W.Willinger, Stock Price Returns and the Joseph Effect, A Fractional Version of the Black-Sholes Model, Seminar on Stochastic Analysis, Random Fields and Applications, *Progress in Probability, Ascona* **36** (1993) *Birkhauser* (1995), 327–351.

[13] A. Dembo and O.Zeitouni, Parameter Estimation of Partially Observed Continuous Time Stochastic Processes via the EM Algorithm, *Stochastic Processes and their Applications* **23** (1986) 91–113.

[14] J.L. Doob, Regularity Properties of Certain Families of Chance Variables, *Transactions of the American Mathematical Society* **47** (1940) 455–486.

[15] J.L. Doob, Stochastic Processes *Wiley New York* (1953).

[16] R.J. Elliott, Stochastic Calculus and Applications, *Springer Verlag New York* (1995).

[17] R.J. Elliott, L.Aggoun and J.B.Moore, Hidden Markov Models: Estimation and Control; Vol.18 of Applications of Mathematics, *Springer Verlag New York* (1982).

[18] R.J.Elliott,W.C.Hunter and B.M.Jamieson, Drift and Volatility Estimation in Discrete Time, *Journal of Economic Dynamics and Control* **22**(1998), 209-218.

[19] R.J. Elliott, W.P.Malcolm and A.H.Tsoi, Robust Parameter Estimation for Asset Price Models with Markov Modulated Volatilities, *Journal of Economic Dynamics and Control* **27(8)** (2003) 1391–1409.

[20] R.J. Elliott and R.W.Rishel, Estimating the Implicit Interest Rate of a Risky Asset, *Stochastic Processes and their Applications* **49** (1982) 199–206.

[21] R.J. Elliott and J.van der Hoek, An Application of Hidden Markov Models to Asset Allocation Problems, *Finance and Stochastics* **13(2)** (1997) 229–238.

[22] R.J. Elliott, Exact Adaptive Filters for Markov Chains Observed in Gaussian Noise, *Automatica* **30(9)** (1994) 1399–1408.

[23] R.J. Elliott, New Finite Dimensional Filters and Smoothers for Noisily Observed Markov Chains, *IEEE Transactions on Information Theory* **39(1)** (1993) 265–271.

[24] R.J. Elliott and H.Yang, Forward and Backward Equations for an Adjoint Process, *Festschrift for G.Kallianpur, Springer Verlag* (1992) 61–70.

[25] R. Engle, Autorregressive Conditional Heteroskedasticity with estimates of United Kingdom Inflation, *Econometrica* **50** (1982) 987-1008 61–70.

[26] Y. Ephraim and N.Merhav, Hidden Markov Processes, *IEEE Transactions on Information Theory* **48(6)** (2002) 1518–1569.

[27] A. Etheridge, A Course in Financial Calculus *Cambridge University Press* (2002).

[28] E.F. Fama and M.R. Gibbons, A Comparison of Inflation Forecasts, *Journal of Monetary Economics* **13** (1984) 327–348.

[29] R.Fernholz, R. Garvy, and J. Hannon, Diversity-Weighted Indexing, *Journal of Portfolio Management***24(2)** (1998) 74-82

[30] G. D. Forney, The Viterbi algorithm, *Proceedings of the IEEE* **61(3)** (1973) 268-278

[31] R. Frey and W.Runggaldier, A Nonlinear Filtering Approach to Volatility Estimation with a View Towards High Frequency Data, *International Journal of Theoretial and Applied Finance* **4** (2001) 199–210.

[32] D. Gamerman, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, *Chapman & Hall CRC* (1997).

[33] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice, *Chapman & Hall CRC* (1996).

[34] L. Glosten, R. Jagannathan, and D. Runkle, Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* **48** (1993) 1779-1801.

[35] J.D. Hamilton, Rational Expectations Econometric Analysis of Changes in Regimes: An Investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control* **12** (1988) 385–423

[36] J.D. Hamilton, A New Approach to the Economic Analysis of Nonstationary Time Series, *Econometrica* **57** (1989) 357–384

[37] R.A.Haugen and N.L. Baker, The Efficient Market Inefficiency of Capitalization-Weighted Stock Portfolios, *Journal of Portfolio Management* (1991)

[38] J. Hsu and C.Jason, Cap-Weighted Portfolios are Sub-optimal Portfolios, *Journal of Investment Management* **4(3)** (2006) 1-10

[39] Y. Hu and B.Oksendal, Fractional White Noise Calculus and Applications to Finance, *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **6(1)** (2003) 1–32.

[40] J.C.Hull, Options, Futures,and Other Derivatives Securities, *Prentice Hall* (1993).

[41] R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering* **82** (1960) 35–45

[42] A.N. Kolmogorov, Grundbegriffe der wahrsch, *Ergodic Mathematics* (1933).

[43] G.C. Lamoureux and W.D. Lastrapes, Heteroskedasticity in Stock Return Data: Volumn versus GARCH Effects, *preprint.*(2005).

[44] S. Luo and A.H.Tsoi, Optimal Stopping via Partial Observation, *Journal of Finance* **45** (1990) 221-229.

[45] S. Luo and A.H.Tsoi, American Option Pricing under Weak Markov Environment, *Working Paper.*(2005).

[46] S. Luo and A.H.Tsoi, Weak Markov Renewal Process, *Working Paper.*(2005).

[47] B.B. Mandelbrot and J.W.van Ness, Fractional Brownian Motions, Fractional Noises and Applications, *SIAM Rev.* **10** (1968) 422–437.

[48] I. Manno, Introduction to the Monte Carlo Method, *Akademiai Kiado* (1999).

[49] R.C.Merton, Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science* **4** (1973) 141–183.

[50] R.C.Merton, Option Pricing when Underlying Stock Returns Are Discontinuous, *Journal of Financial Economics***3** (1976) 125–144.

[51] R.C.Merton, On the Pricing of Contingent Claims and the Modigliani-Miller Theorem, *Journal of Financial Economics* **5** (1977) 241–249.

[52] N. Metropolis and S. Ulam, The Monte Carlo Method, *Journal of American Staistics Associasion* **44** (1949) 335-341.

[53] N. Metropolis, The Beginning of the Monte Carlo Method, *Los Alamos Science Special Issue* **15** (1987).

[54] G. A. Mikhailov, Parametric Estimates by the Monte Carlo Method, *Utrecht, Netherlands: VSP* (1999).

[55] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE* **77(2)** (1989) 257-286

[56] I. M. Sobol, A Primer for the Monte Carlo Method, *CRC; Subsequent edition* (1994).

[57] A.H. Tsoi, H.Yang and S.N.Yeung, European Option Pricing when the Risk-free Interest Rate Follows a Jump Process, *Stochastic Models* **16(1)** (2000) 143–166.

[58] J. Ville, Étude Critique de la Notion de Collectif, *Gauthier-Villars, Paris* (1939).

[59] G.Welch and G.Bishop, An Introduction to the Kalman Filter, *UNC-Chapel Hill, TR:* **95-041** (2006)

[60] Lloyd R. Welch, Hidden Markov Models and the Baum-Welch Algorithm, *IEEE Information Theory Society Newsletter* **53(4)** (2003)

[61] P. Wilmott, J. Dewynne, and S. Howison, Option Pricing, Mathematical Models and Computation,*Oxford Financial Press* (1995).

VITA

Pei Yin was born on July 23, 1978 in Wuhan, China. After attending public schools in Wuhan, she received the following degrees: B.A. in Economics and B.S. in Mathematics from Wuhan University, China (2000); M.A. in Economics and M.S. in Applied Mathematics from the University of Missouri-Columbia (2003). Then she started her Ph.D. study in the Mathematics Department at the University of Missouri-Columbia in Aug 2003. She plans to graduate in July 2007.