Topics in Financial Asset Pricing:
Equity Premium Puzzle and HMM models for Equity Market Returns

BY

JING CAI
B.S., Peking University, Beijing, 2006
M.S., Columbia University, New York, 2008

THESIS
Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Business Administration
in the Graduate College of the
University of Illinois at Chicago, 2012

Chicago, Illinois

Defense Committee:
Stanley Sclove, Chair and Advisor
Gilbert Bassett
John Sparks
Fangfang Wang
Jing Wang, Mathematics, Statistics, and Computer Science
To my family
ACKNOWLEDGMENTS

I want to thank my advisor, my committee member and my family. Without their support, I could not have made it.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> ASSET PRICING: OVERVIEW AND RECENT DEVELOPMENTS</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Empirical Asset Returns Study</td>
<td>4</td>
</tr>
<tr>
<td>1.2 Present-Value Approach to Asset Pricing</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Investors-Preference-Dependent Asset Pricing Approach</td>
<td>8</td>
</tr>
<tr>
<td>1.4 The Equity Premium Puzzle in Financial Economics</td>
<td>11</td>
</tr>
<tr>
<td>1.4.1 The Equity Premium Puzzle</td>
<td>12</td>
</tr>
<tr>
<td>1.4.2 Hansen &amp; Jagannathan Bound</td>
<td>14</td>
</tr>
<tr>
<td>1.4.3 Epstein-Zin Investor’s Preference</td>
<td>15</td>
</tr>
<tr>
<td><strong>2</strong> LONG RUN RISKS MODEL WITH TWO SHOCKS: A POSSIBLE RESOLUTION OF THE EQUITY PREMIUM PUZZLE</td>
<td>21</td>
</tr>
<tr>
<td>2.1 Long-Run Risks Model, Overview of BY(2004)</td>
<td>21</td>
</tr>
<tr>
<td>2.2 My Work: Long-Run Model with Two Shocks</td>
<td>24</td>
</tr>
<tr>
<td>2.3 Model Construction</td>
<td>26</td>
</tr>
<tr>
<td>2.4 Data</td>
<td>29</td>
</tr>
<tr>
<td>2.5 Long-Run Model Estimation</td>
<td>30</td>
</tr>
<tr>
<td>2.6 Impulse Response Analysis</td>
<td>32</td>
</tr>
<tr>
<td>2.7 More to say about the Impulse Response Analysis</td>
<td>34</td>
</tr>
<tr>
<td>2.8 GMM Method and the Estimations of the Epstein-Zin Preference Parameters</td>
<td>36</td>
</tr>
<tr>
<td>2.8.1 GMM overview</td>
<td>36</td>
</tr>
<tr>
<td>2.8.2 The Application of GMM to My Model</td>
<td>38</td>
</tr>
<tr>
<td>2.9 How My Model Helps to Resolve the Equity Premium Puzzle</td>
<td>42</td>
</tr>
<tr>
<td>2.10 Appendix</td>
<td>44</td>
</tr>
<tr>
<td>2.10.1 The formulas to derive the long-run risks VAR</td>
<td>44</td>
</tr>
<tr>
<td>2.10.2 Summary Statistics and Figures</td>
<td>51</td>
</tr>
<tr>
<td>2.10.3 Figures for Major Variables</td>
<td>52</td>
</tr>
<tr>
<td><strong>3</strong> HIDDEN MARKOV MODELS AND ASSET PRICING</td>
<td>62</td>
</tr>
<tr>
<td>3.1 A Brief Introduction to Hidden Markov Models (HMMs)</td>
<td>62</td>
</tr>
<tr>
<td>3.2 Formulation of HMMs</td>
<td>65</td>
</tr>
<tr>
<td>3.2.1 Task 1: computing the probability of an observed series</td>
<td>66</td>
</tr>
<tr>
<td>3.2.2 Task 2: finding the best state sequence</td>
<td>67</td>
</tr>
<tr>
<td>3.2.3 Task 3: estimating parametric model</td>
<td>68</td>
</tr>
<tr>
<td>3.3 Financial Asset Pricing Using HMM</td>
<td>70</td>
</tr>
<tr>
<td>3.4 HMM Segmentation According to Mean</td>
<td>71</td>
</tr>
<tr>
<td>3.4.1 Two-state HMM</td>
<td>73</td>
</tr>
</tbody>
</table>
## TABLE OF CONTENTS (Continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.2</td>
<td>Three-state HMM</td>
</tr>
<tr>
<td>3.4.3</td>
<td>2-states HMM vs 3-states HMM</td>
</tr>
<tr>
<td>3.5</td>
<td>HMM Segmentation According to Variance</td>
</tr>
<tr>
<td>3.6</td>
<td>Posterior Probabilities from BIC and HMM Model Comparison</td>
</tr>
<tr>
<td>3.7</td>
<td>A HMM Case Study of FTSE/Xinhua China A50 Index</td>
</tr>
<tr>
<td>4</td>
<td>CONCLUSION</td>
</tr>
<tr>
<td></td>
<td>CITED LITERATURE</td>
</tr>
<tr>
<td></td>
<td>VITA</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>41</td>
</tr>
<tr>
<td>II</td>
<td>42</td>
</tr>
<tr>
<td>III</td>
<td>52</td>
</tr>
<tr>
<td>IV</td>
<td>52</td>
</tr>
<tr>
<td>V</td>
<td>53</td>
</tr>
<tr>
<td>VI</td>
<td>53</td>
</tr>
<tr>
<td>VII</td>
<td>70</td>
</tr>
<tr>
<td>VIII</td>
<td>76</td>
</tr>
<tr>
<td>IX</td>
<td>76</td>
</tr>
<tr>
<td>X</td>
<td>81</td>
</tr>
<tr>
<td>XI</td>
<td>82</td>
</tr>
<tr>
<td>XII</td>
<td>82</td>
</tr>
<tr>
<td>XIII</td>
<td>83</td>
</tr>
<tr>
<td>XIV</td>
<td>84</td>
</tr>
</tbody>
</table>
### LIST OF TABLES (Continued)

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>XV</td>
<td>86</td>
</tr>
<tr>
<td>XVI</td>
<td>88</td>
</tr>
<tr>
<td>XVII</td>
<td>92</td>
</tr>
</tbody>
</table>

- XV  STANDARD DEVIATION COMPARISON
- XVI COMPUTATION OF POSTERIOR PROBABILITIES
- XVII COMPUTATION OF BIC STATISTICS
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
</tr>
<tr>
<td>7</td>
<td>58</td>
</tr>
<tr>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>59</td>
</tr>
<tr>
<td>10</td>
<td>59</td>
</tr>
<tr>
<td>11</td>
<td>59</td>
</tr>
<tr>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>60</td>
</tr>
<tr>
<td>14</td>
<td>61</td>
</tr>
<tr>
<td>15</td>
<td>61</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>16</td>
<td>Monthly Continuous Consumption Growth Rate</td>
</tr>
<tr>
<td>17</td>
<td>Hidden Markov Model</td>
</tr>
</tbody>
</table>
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMS</td>
<td>American Mathematical Society</td>
</tr>
<tr>
<td>CTAN</td>
<td>Comprehensive \TeX Archive Network</td>
</tr>
<tr>
<td>TUG</td>
<td>\TeX Users Group</td>
</tr>
<tr>
<td>UIC</td>
<td>University of Illinois at Chicago</td>
</tr>
<tr>
<td>UICTHESI</td>
<td>Thesis formatting system for use at UIC.</td>
</tr>
</tbody>
</table>
SUMMARY

My thesis concerns asset pricing in finance. In Chapter 1, I draw an overview of the financial economics field and in specific several mainstream studies of asset pricing. The finance field can be divided into two sub-fields, asset pricing and corporate finance. In asset pricing, the empirical study of asset returns and the study of asset prices are two divisions. For asset prices, two mainstream approaches are introduced: one is the "prevent value" approach and the other is the "investors preference dependent" approach. In addition in Chapter 1, I address a prevailing interesting topic in the economics and finance field, the problem of the equity premium puzzle.

In Chapter 2, I propose a long run risk model with two shocks based on Bansal and Yaron’s well-known long-run risk model. My model combines the two divisions of the asset pricing field and consolidates the two mainstream approaches, to better capture the movements of financial markets and the US economy. My model helps resolve the equity premium puzzle. Moreover, I demonstrate that two distinct shocks, the consumption growth shock and the dividend growth shock are the driving factors of the persistent long run movements of asset prices, asset returns, dividend growth and consumption growth. The GMM(Generalized Method of Moments) method is used to estimate the unknown parameters of the model.

In Chapter 3, I apply the Hidden Markov Model to the same package of monthly data sets which are used in Chapter 2. Two kinds of HMM segmentation are applied to the SP500 monthly data, one is "mean segmentation" and the other is "variance segmentation”. In each segmentation, two- and three-state HMMs are fitted in this chapter. In mean segmentation, for
SUMMARY (Continued)

each of the 2-state HMM and 3-state HMM, three sub-cases are studied: one is with constant mean for each state’s distribution function and constant transition matrix, another is with time varying mean but constant transition matrix, and the other is with time varying mean and time varying transition matrix. In variance segmentation, 2-state HMM and 2-state HMM with constant mean and transition matrix are studied. In total, 8 HMM models are fitted. The result shows that the winner today will have much higher probability to win tomorrow and the loser today will have much higher probability to lose tomorrow. BIC statistics are used to compare these 8 models. The comparison shows that in mean segmentation, the 2-state HMM with constant transition matrix and constant mean of the state’s normal density function, is the best model; in variance segmentation, the 2-state HMM with the constant mean and transition matrix is better. Moreover, the posterior probability statistics, which are based on BIC statistics, are calculated for all eight models. The result is consistent with that of the BIC statistics. The 2-state HMM with constant transition matrix and constant mean of the density function is the best model. This suggests that the conventional motion of Bull and Bear markets may make some sense.
CHAPTER 1

ASSET PRICING: OVERVIEW AND RECENT DEVELOPMENTS

When it comes to research in finance, many studies have been conducted, by different groups of people such as economists and statisticians, from different perspectives. In the academic world of finance there are two main directions: the corporate finance field and the asset pricing field. Corporate finance focuses mainly on the business enterprises’ corporate values and concentrates on firms’ management decisions, business strategies and financial risk management. In contrast, researchers in asset pricing emphasize market performance and the movements of different asset classes, stocks, bonds, real estate, and others. They pay attention to the relationships between asset returns/prices and economic factors, the goal being to find out which factors decide or influence the asset returns/prices. Here I will focus on the asset pricing field only. Compared with corporate finance, the study of asset pricing in general contains more interdisciplinary research, and usually implements relatively rigorous mathematical derivations and uses statistical analytical methodology.

Research on asset returns exclusively and research on the asset prices represent two dimensions in the asset pricing field. At first glance, this statement seems confusing: It is true that asset returns and asset prices can be written as functions of each other, so how does it come to two dimensions and not one? To answer this question, first we need to clarify two definitions.
• The Study of asset prices analyzes asset price movements and the internal economic structures which could be incorporated into the asset returns. It aims at answering the question of what drives the asset price or in other words how to price the asset based on the economic framework.

• On the other side, when talking about the study of asset returns, academically this usually refers to studies focusing on returns’ movements and interrelationships exclusively without saying anything explicitly about asset prices. In general, a study of asset returns puts more weight on statistical analysis but not economic interpretations from the economics academic setting. Time series analysis of asset returns such as ARMA, GARCH modeling and Minimum Variance portfolio analysis belong to this area.

From the academic point of view, the study of asset prices is more fundamental than that of returns. There are several reasons for this: First, return is usually defined and expressed as a function of this period and next period’s prices; in this way price movement has its direct link to returns. Second, while return takes in the effect of one-lag time series price, it may screen out some important information in prices. Third, many economic factors should have their direct impacts on prices but not returns. Take a simple example, when consumers buy a coat, the first thing that jumps into their eyes is the price tag. The same is true when investors buy a stock. Investors won’t calculate the stock return at first. Thus, returning to the study of prices helps to bring attention back to the economic foundation and to emphasize the market factors which could have a direct effect on asset prices.
In the following section 1.1, I will briefly discuss the study of asset returns. For the rest of sections 1.2 to 1.4 and the other chapters of my thesis, the study of asset pricing is the theme. Section 1.2 and section 1.3 represents two mainstreams of asset pricing theory: the present-value model and investors’ preference-dependent model.

Figure 1. Financial Economics Academic Tree Chart
1.1 **Empirical Asset Returns Study**

I will not elaborate the discussion here. Most of time series analysis, volatility studies and portfolio analysis of the asset returns belong to this type of research. Usually asset prices are not included explicitly in this return-exclusive research. In Chapter 3, I will present my work on the uncertainty of stock returns, and there I use the methods of time series and the Hidden Markov Model.

1.2 **Present-Value Approach to Asset Pricing**

In this section, I introduce one of the two main asset pricing theories: *the present value model*. This model is also called the “discounted-cash flow” model. Unlike the returns study mentioned in the section above, which centers on statistical description and time series analysis of historical returns, the present value model views the prices of assets in a “forward-looking” way by defining the price of a stock as the present value of its expected future cash flows, discounted by a reasonable discount rate. For stocks, the future cash flow is usually its dividend payouts. The discount rate can be either constant or time-varying.

There are two key merits of this theory. First, it takes the “time value of money” into consideration. The “time value of money” means one dollar today is more valuable than one dollar tomorrow, even without mentioning whether one dollar tomorrow is guaranteed to gain, saving one dollar today to bank deposit we can at least earn a return equal to the risk-free rate. Thus the model removes the drawback which views all the future cash flows of investors at the same value level. Second, two uncertainties about the future are structured into the model. One is the uncertainty of the long-lasting future cash flows and the other is the possible
changing discount rate. In this way, the long run persistent time series movements of future dividends and of the discount rate could give a solid interpretation to the long run movement of prices thus to the long run movement of returns. This interpretation of long run return movements is also called the “predictability of stock returns at long horizon”. As in the return studies mentioned above, when using only past returns to forecast future returns, the evidence of predictability is weak. However, researchers have shown that when bringing dividends into the model, the dividend yield has significant larger forecasting power on stock returns. This is often viewed as one of the big contributions of the present-value model to the study of financial time series. Relevant papers in this field discussing stocks’ predictability and present values are Fama and French (1988), Lo and Mackinlay (1988), Hodrick (1992), Stambaugh (1999), Binsbergen and Koijen (2007).

Below I give detailed description and mathematical formulas of the model.

Relationship between prices, dividends and returns

\[ R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1 = \frac{(P_{t+1} - P_t) + D_{t+1}}{P_t} \]

Let \( r_{t+1} \equiv \log(1 + R_{t+1}) \), \( p_{t+1} \equiv \log(P_{t+1}) \), \( d_{t+1} \equiv \log(D_{t+1}) \); this notation will be used throughout the thesis.

When the expected return \( R \) is constant, by taking expectations with condition at time \( t \), the present-value relation is expressed as:

\[ R = E_t[R_{t+1}] \]
\[ P_t = E_t \left[ \frac{P_{t+1} + D_{t+1}}{1 + R} \right] \]

Where \( E_t \) denotes conditional expectation given the history of the process up to and including time \( t \).

Using the law of iterated expectation \( E_t(X) = E_t(E_{t+1}(X)) \), then \( P_t \) goes to:

\[ P_t = E_t \left[ \sum_{i=1}^{k} \left( \frac{1}{1 + R} \right)^i D_{t+i} \right] + E_t \left[ \left( \frac{1}{1 + R} \right)^k P_{t+k} \right] \]

Let \( k \to \infty \) and assume \( \lim_{k \to \infty} E_t \left[ \left( \frac{1}{1 + R} \right)^k P_{t+k} \right] = 0 \),

\[ P_t = E_t \left[ \sum_{i=1}^{\infty} \left( \frac{1}{1 + R} \right)^i D_{t+i} \right] \]

\[ \Rightarrow P_t - \frac{D_t}{R} = \left( \frac{1}{R} \right) E_t \left[ \sum_{i=0}^{\infty} \left( \frac{1}{1 + R} \right)^i \Delta D_{t+1-i} \right] \]

where \( E_t \left[ \sum_{i=0}^{\infty} \left( \frac{1}{1 + R} \right)^i \Delta D_{t+1-i} \right] \) is the discounted changes of dividends.

As in real data, stock prices and dividends are more likely to grow exponentially over time rather than linearly. Campbell and Shiller (1988 a, b) derived the log linearization approximation of the present value model with time varying expected returns. The log linear framework enables us to start with asset prices but get to the behavior of the returns and, furthermore, to link prices and returns with each other in a time series format. The calculation of log linearization is as below. Taylor expansion is used in linearizing.

\[ r_{t+1} \equiv \log(P_{t+1} + D_{t+1}) - \log(P_t) \]
\[= \log(P_{t+1} + D_{t+1}) - \log(P_{t+1}) + \log(P_{t+1}) + \log(P_t)\]

\[= \log(1 + \frac{D_{t+1}}{P_{t+1}}) + p_{t+1} - p_t\]

\[= p_{t+1} - p_t + \log(1 + \exp(d_{t+1} - p_{t+1}))\]

After Taylor expansion, this is approximately

\[r_{t+1} = \rho_0 + \rho p_{t+1} + (1 - \rho)\Delta d_{t+1} - p_t\]

where \(\rho = (1 + D/P)^{-1}\).

\[\rho_0 + \rho p_{t+1} - \rho d_{t+1} + d_{t+1} - p_t + d_t - d_t\]

\[= \rho_0 + \rho(p_{t+1} - d_{t+1}) + (d_{t+1} - d_t) + (d_t - p_t)\]

\[= \rho_0 - \rho \Delta p_{t+1} + \Delta d_{t+1} + \Delta p_t,\]

where \(\Delta p_t = d_t - p_t\).

Empirically in US data over the period 1928 to 1994, economists found that the average dividend-price ratio (dividend yield) was about 4% annually, implying that \(\rho\) should be about 0.96 in annual data or about 0.997 per year in monthly data. In Chapter 2, with a recent real dataset I estimate \(\rho\) to be 0.9679 in annual data and 0.9991 in monthly data.
The forward looking formula for $p_t$ is:

$$p_t = \frac{\rho_0}{1 - \rho} + \sum_{j=1}^{\infty} \rho^j [(1 - \rho)d_{t+1-j} - r_{t+1+j}]$$

Thus, after taking the expectation, the formula is:

$$p_t = \frac{\rho_0}{1 - \rho} + E_t \left[ \sum_{j=1}^{\infty} \rho^j [(1 - \rho)d_{t+1-j} - r_{t+1+j}] \right]$$

This formula tells us that if the stock price is high today, then investors must be expecting some combination of high future dividends and low future returns.

In addition the dividend price ratio and unexpected stock returns can be expressed as:

$$d_t - p_t = -\frac{\rho_0}{1 - \rho} + E_t \left[ \sum_{j=0}^{\infty} \rho^j [\Delta d_{t+1+j} + r_{t+1+j}] \right]$$

$$r_{t+1} - E_t[r_{t+1}] = E_{t+1} \left[ \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \right] - E_t \left[ \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \right] - \left( E_{t+1} \left[ \sum_{j=0}^{\infty} \rho^j r_{t+1+j} \right] - E_t \left[ \sum_{j=0}^{\infty} \rho^j r_{t+1+j} \right] \right)$$

The change of the unexpected stock returns today are associated with the changes in expectations of future dividends and the change of the real returns respectively.

1.3 **Investors-Preference-Dependent Asset Pricing Approach**

The Investors/consumers preference-dependent model is another mainstream asset pricing theory. The idea of this theory is to link asset prices with investors’ preferences and decisions.
It views the asset price behavior from the perspective of the asset holders. And the idea is intuitive, since fundamentally the asset prices will be determined in the buying and selling processes of the investors. In this setting, to simplify the problem, it is assumed that investors of assets are also consumers of consumption goods, which means people choose how much of their wealth to consume and how much to invest in assets in order to get future returns. This is related to the marginal propensity to consumer, MPC. At each period of time, investors decide their consumption level and portfolio choices simultaneously. Another common assumption within this approach is that the investors are homogeneous so that in this way all investors share the same preference function over different consumption levels.

The theory derived from the idea of maximizing investors expectation of future utility on consumption is as follows: Assuming there is one asset to invest in, the problem is to solve

$$\max E_t \left[ \sum_{j=0}^{\infty} \delta^j U(C_{t+j}) \right]$$

s.t. $W_{t+1} = (1 + R_{t+1})(W_t - C_t)$

where $\delta$ is the time discount factor, $C_{t+j}$ is the investor’s consumption at time $t+j$, $1 + R_{t+1}$ is the asset return and $W_t$ is time t’s investor’s total wealth. The First-Order condition of the above optimization problem gives the result:

$$U'(C_t) = \delta E_t[(1 + R_{a,t+1})U'(C_{t+1})] \implies 1 = E_t[(1 + R_{a,t+1})M_{t+1}]$$

where $M_{t+1} = U'(C_{t+1})/U'(C_t)$ is the stochastic discount factor, also called the pricing kernel.
For multiple assets, the calculation is similar:

\[
V(W_t) \equiv \max_{\{c_t, c_{t+1}, \ldots, w_t, w_{t+1}, \ldots\}} E_t \left[ \sum_{j=0}^{\infty} \beta^j U(C_{t+j}) \right]
\]

s.t. \quad W_{t+1} = R_{t+1}^w (W_t - C_t)

\[
R_t^w = w_t' R_t, \quad w_t' 1 = 1
\]

Under the assumption that the value function \( V(W_t) \) is a concave function, the procedure of solving the maximization problem is as follows.

\[
V(W_t) = \max E_t [U(C_t) + \beta E_t [V(W_{t+1})]]
\]

\[
\Rightarrow V(W_t) = \max E_t [U(C_t) + \beta E_t [V(R_{t+1}^w (W_t - C_t))]]
\]

\[
\Rightarrow V'(W_t) = U'(C_t) - \beta E_t [V'(W_{t+1}) R_{t+1}^w],
\]

where \( U(C_t) = \beta E_t [R_{t+1}^w U'(C_{t+1})] \) and \( R_{t+1}^w = 1 + R_{t,t+1} \).

The unconditional version is:

\[
1 = E[(1 + R_{it}) M_t] = E[1 + R_{it}] E[M_t] + Cov(R_{it}, M_t)
\]

\[
\Rightarrow E[1 + R_{it}] = \frac{1}{E[M_t]} \left( 1 - Cov(R_{it}, M_t) \right)
\]
With a commonly used preference function \( U(C_t) = \frac{C_t^{1-\gamma} - 1}{1-\gamma} \), the pricing kernel \( M_{t+1} \) is equal to

\[
\frac{U'(C_{t+1})}{U'(C_t)} = \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma}
\]

1.4 The Equity Premium Puzzle in Financial Economics

The Equity Premium Puzzle was brought out by Mehra & Prescott (1985). When using consumption-based asset pricing with the power utility function, researchers found that the historical high mean and volatility of stock excess log return, combined with the very low historical covariance between stock returns and consumption growth, imply a very large coefficient of relative risk aversion for consumers/investors in the US. Under this large risk aversion coefficient, within the model’s setting, the risk free rates (Treasury bond rates) should be much higher than they are/were during the past 80 years, which is called the “risk free rate puzzle” in Weil (1989). In other words, under a normal risk aversion coefficient for investors as a whole who would like to hold Treasury bonds, the mean and the volatility of equities in the US are abnormally high. For an overview of equity premium puzzle, Campbell (2002) “Consumption-based asset pricing” and Constantinides (2006) “Understanding the Equity Risk Premium Puzzle” are two papers giving great lectures on the first fifteen years of research developments on this topic.

While twenty-five years have passed, researchers are still trying to resolve this puzzle under the academic setting. One approach is to adjust/change the preference/utility function assumed for investors; bunches of papers have been written along these lines. A recent innovative and pioneering approach is to take consumers/investors concerns about long-run expected growth of
consumptions and dividends and the time varying future economic prospects into consideration. A related paper is by Bansal and Yaron (2004). This is a new start to incorporate the study of financial economics into the traditional macro economics research setting. I will elaborate Bansal and Yaron’s work in the next chapter when I open the main body of the thesis.

To give more detailed descriptions on the puzzle, I will use mathematical formulas and economic models in the following two sections 1.4.1 and 1.4.2. First, I will derive and discuss the puzzle based on the investors’ preference-dependent asset pricing model, as originally the equity premium puzzle was brought out in this exact macroeconomic setting. Next in section 1.4.2, I will use Hansen and Jagannathan (1991)’s approach to view the puzzle from another angle and in a visualized way.

One important thing need to be pointed out explicitly is, to solve the puzzle always means to find a “right”, or to say a better model, under the current academic setting, so that applying this model to real financial data, we can better explain the equity premium puzzle.

1.4.1 The Equity Premium Puzzle

As I have shown in the previous section, \( \log E_t[X] = E_t[\log X] + \frac{1}{2} \text{Var}_t[\log X] \). Taking logs on both sides of

\[
1 = E_t[(1 + R_{t,t+1})\delta(C_{t+1}/C_t)^{-\gamma}],
\]

we have,

\[
0 = E_t[r_{t+1} + \log \delta - \gamma E_t[\Delta C_{t+1}] + \frac{1}{2} \sigma_i^2 + \gamma^2 \sigma_i^2 - 2\gamma \delta_i c]
\]
and

\[ r_{f,t+1} = -\log \delta - \frac{\gamma^2 \sigma_c^2}{2} + E_t[\Delta C_{t+1}], \]

where \( \sigma_c^2 \) is consumption volatility and \( E_t[\Delta C_{t+1}] \) is expected consumption growth. Furthermore, we get the expression for expected excess log return of stocks:

\[ E_t[r_{i,t+1} - r_{f,t+1}] + \frac{\sigma_i^2}{2} = \gamma \sigma_{ic}. \]

Using 1889 to 1994 annual data set, the historical excess log stock returns is 4.2% with variance \( \sigma_i^2 = 17.7% \) and the covariance between log consumption (nondurables and services) growth and excess log stock return is 0.003 which is small. Based on the above formulas, the risk-aversion coefficient is \( \gamma = 19 \) which is much greater than the maximum acceptable value 10.

This is the equity premium puzzle. Under such a large \( \gamma = 19 \), another puzzle about risk free rate arises. As

\[ E[r_{ft}] = -\log \delta + \gamma g - \frac{\gamma^2 \delta^2}{2} \]

where \( g \) represents the continuous growth rate of consumption, and \( \delta \) is the discounted factor as previously stated.

The historical treasury bond rate has mean equal to 1.01% and the historical continuous consumption growth rate has mean 1.8% and standard deviation equal to 3.3% , as a result, a \( \gamma = 19 \) implies that the discounted factor \( \delta \) is equal to 1.15 which is > 1. A larger than one \( \delta \) contradicts with the time value of money as it means the value of tomorrow’s one dollar with
unknown uncertainty is 1.15 times larger than today’s one dollar for sure in hand. This cannot
be true. Weil (1989) calls this the risk free rate puzzle which was born with the equity premium
puzzle.

1.4.2 Hansen & Jagannathan Bound

The HJ bound gives a view to help us further address the EP puzzle. As

\[
E[M_t \tilde{R}_{it}] = 1; \ i = 1, \ldots, N
\]

\[
\tilde{R}_{it} = 1 + R_{it}
\]

For any portfolio \( \tilde{R}_p \),

\[
1 = E[M \tilde{R}_p]
\]

\[
= E[M] \tilde{R}_p + sd(M)sd(\tilde{R}_p) corr(M, \tilde{R}_p)
\]

This implies that

\[
corr(M, \tilde{R}_p) = \frac{1 - E[M] \tilde{R}_p}{sd(M)sd(\tilde{R}_p)}
\]

Therefore,

\[
\left| \frac{1 - E[M] \tilde{R}_p}{sd(M)sd(\tilde{R}_p)} \right| \leq 1.
\]
Since $E[M] > 0$,

$$\frac{sd(M)}{E[M]} \geq \left| \frac{\tilde{R}_p - (E[M])^{-1}}{sd(R_p)} \right| \geq |\text{Sharpe}(R_p)|$$

$$\frac{sd(M)}{E[M]} \geq \text{the Sharpe ratio of the tangency portfolio in Min-Var portfolio analysis}$$

$$\geq \left[ (R - (E[M])^{-1} \right)^T V^{-1} (R - (E[M])^{-1} \right]^{1/2},$$

That is

$$sd(M) \geq \left[ (R(E[M]) - 1)^T V^{-1} (R(E[M]) - 1) \right]^{1/2}$$

$$\geq \left[ a(E[M])^2 - 2bE[M] + c \right]^{1/2}$$

Figure 2 in this chapter will help us understand the Equity Risk Premium and HJ bound better:

The case with a risk free asset is,

$$sd(M) \geq [a(R_F)^{-2} - 2b(R_F)^{-1} + c]^{-1/2},$$

where $R_F = (E[M])^{-1}$.

### 1.4.3 Epstein-Zin Investor’s Preference

As stated before, one mainstream research to solve the equity premium puzzle is to change/adjust the investors’ preference functions, which are also called utility functions of consumers. People
argue that the equity premium puzzle is there because the power preference function is not sufficient or not good to explain the markets.

Epstein-Zin (1989) preference is one of the preference functions proposed by researchers. The economic logic contained in this preference is very convincing and it plays an important role in the long-run risk model which I will discuss in detail in a later chapter.

The core adjustment of the Epstein-Zin preference, compared to other preference functions is that it allows for a separation between risk aversion (RA) of investors and intertemporal elasticity of substitution (IES) of investors. RA describes the consumer’s reluctance to substitute consumption across uncertain states of the asset payoffs. It is related to the investor decision at each time point, about whether to consume less and to buy more assets with taking more risks in exchange for future returns. IES describes the consumer’s willingness to substitute consumptions over time. In summary, RA is about the risk preference across holding cash and different kind of assets, IES is about the risk preference across time. In a power preference function, RA is the inverse of IES, which is not a very reasonable assumption.

The IES, RRA and ARA formulas are as follows:

\[
IES = \frac{d\ln(C_{t=2}/C_{t=1})}{d\ln(U''(C_{t=1})/U''(C_{t=2}))}
\]

Relative risk aversion

\[
RRA = -\frac{CU''(C)}{U'(C)}
\]
Absolute risk aversion

\[ ARA = -\frac{U''(C)}{U'(C)} \]

EZ preference is defined as

\[ U_t = \left\{ (1 - \delta)C_t^{\gamma/(1-\gamma)} + \delta(E_t[U_{t+1}^{1-\gamma}])^{\theta/(1-\gamma)} \right\}^{\theta/(1-\gamma)} \]

where \( \delta > 0 \) is the subjective discount factor, \( \gamma \) is the RRA coefficient, and \( \theta = (1-\gamma)/(1-1/\psi) \) with \( \psi = \partial E[\Delta C_{t+1}/\partial r_{t,t+1}] \) being the elasticity of intertemporal substitution. When \( \theta = 1 \), then the Epstein-Zin preference collapses to the usual case of power preference with RRA equal to the inverse of EIS, because

\[ \theta = \frac{1 - \gamma}{1 - 1/\psi} = 1 \implies \gamma = \frac{1}{\psi} \]

The Epstein-Zin preference can be written with value function in the form discussed in section 1.3:

\[ U_t = \left\{ (1 - \delta)C_t^{\gamma/(1-\gamma)} + \delta(E_t[U_{t+1}^{1-\gamma}])^{\theta/(1-\gamma)} \right\}^{\theta/(1-\gamma)} \]

\[ \implies U_t^{1-\gamma} = \left\{ (1 - \delta)C_t^{\gamma/(1-\gamma)} + \delta(E_t[U_{t+1}^{1-\gamma}])^{\theta/(1-\gamma)} \right\}^{\theta/(1-\gamma)} \]

\[ \implies V_t = \frac{U_t^{1-\gamma}}{1 - \delta} = C_t^{1-\gamma} + \delta E_t \left[ U_{t+1}^{1-\gamma} \right] \]

\[ \implies V_t = C_t^{1-\gamma} + \delta E_t [V_{t+1}] \]
The investors make the optimization decision as usual:

$$\max U_t$$

s.t. \( W_{t+1} = (1 + R_{a,t+1})(W_t - C_t) \)

Taking the first derivative with respect to \( C_t \) yields

$$1 = E_t \left[ \delta^\theta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma+1-\theta} (1 + R_{M,t+1})^{-1+\theta}(1 + R_{a,t+1}) \right]$$

$$1 = E_t \left[ \delta^\theta G_{t+1}^{-\gamma+1-\theta}(1 + R_{M,t+1})^{-1+\theta}(1 + R_{a,t+1}) \right],$$

where \( G_{t+1} = C_{t+1}/C_t \). Define

$$g_{t+1} = \log(C_{t+1}/C_t)$$

$$r_{M,t+1} = \log(1 + R_{M,t+1})$$

$$r_{f,t+1} = \log(1 + R_{f,t+1})$$

$$r_{a,t+1} = \log(1 + R_{a,t+1})$$

$$m_{t+1} = \theta \log(\delta) - \frac{\theta}{\psi} g_{t+1} + (\theta - 1)r_{a,t+1}$$

$$r_{f,t+1} = -\log \delta + \frac{\theta - 1}{2} \sigma_M^2 - \frac{\theta}{2\psi^2} \sigma_c^2 + \frac{1}{\psi} E_t[\Delta C_{t+1}]$$,
where $\psi = \partial E[\Delta C_{t+1}] / \partial r_{f,t+1}$. Thus,

$$E[r_{i,t+1}] - r_{f,t+1} + \frac{\sigma_i^2}{2} = \frac{\theta}{\psi} \sigma_{ic} + (1 - \theta) \sigma_{iM}.$$

If $\theta = 1$, $RRA = 1$, then the Epstein-Zin preference is reduced to the power utility case

$$E[r_{i,t+1}] - r_{f,t+1} + \frac{\sigma_i^2}{2} = \gamma \sigma_{ic}.$$

EZ collapses to the log power utility case,

Typically, $\gamma \gg 1$ and $1/\psi \gg 1$, thus

$$\theta = \frac{1 - \gamma}{1 - 1/\psi} \approx \gamma \psi.$$
Figure 2. HJ Bound and Equity Premium Puzzle with Power Utility Function
2.1 Long-Run Risks Model, Overview of BY(2004)

Bansal & Yaron (2004), denoted BY (2004), propose a model which consolidates Epstein-Zin preferences, long run persistent risks, and time varying consumption growth rates and volatility. The model puts these four features all together in a nutshell. It is the first recognizable research combining the two dimensions of the asset pricing field, and at the same time includes the two theories in asset price studies to explain financial economics. In short, it contains the contents I mentioned in section 1.1, 1.2 and 1.3.

Two economic channels are created by BY (2004) in its long-run risks model to resolve the equity premium puzzle:

- One channel is through the long run persistent effects of consumption and dividend growth rates. Here BY combines the asset returns time series study with the present value model. This channel is constructed into the asset returns time series model and the present value asset pricing model. As discussed in section 1.2, researchers have found that dividend price ratios have significant power in forecasting future returns; in other words, time series of asset returns can be treated as containing a long run persistent effect which is driven by the dividend price ratio. BY (2004) goes one step further to make the dividend growth
rates and the consumption growth rates both contain this long-run component. The authors argue that “current shocks to expected growth alter expectations about future economic growth not for short horizons but also for the very long run”. In this way, the model is able to explain a large and long-lasting equity mean and variance thus matching the real data.

- The second channel is through the time-varying conditional mean and volatility of the consumption growth. This is a feature imbedded into the investor’s preference dependent asset pricing model by the authors. The idea is to let fluctuations in consumption and its volatility lead to time variation in equity risk premium thus to explain the high volatility of equity premium.

BY (2004) model’s setting is:

\[
E_t \left[ \delta^\theta G_{c,t+1}^{-\theta/\psi} R_{a,t+1}^{(1-\theta)} R_{M,t+1} \right] = 1
\]

Here \( G_{c,t+1} \) is the aggregate gross rate of consumption. \( R_{M,t+1} \) is observable return on the market portfolio. \( R_{a,t+1} \) is unobservable return on a claim to aggregate consumption. As a commonly used method, economists take consumption as the “dividends” of the human capital asset of the investors, since human capital, as an image asset whose “asset value” is unobservable, \( R_{a,t+1} \) is unobservable accordingly.
BY (2004) brought the observable log linearized formula into the modeling of the unobservable $R_{a,t+1}$ and stated the following two equations:

\[
\begin{align*}
\log(R_{a,t+1}) &= r_{a,t+1} = k_0 - k\bar{c}_{t+1} + \bar{c}_t + \Delta c_{t+1} \\
\log(R_{M,t+1}) &= r_{M,t+1} = \rho_0 - \rho\bar{d}_{t+1} + \bar{d}_t + \Delta d_{t+1}
\end{align*}
\]

Two cases are proposed in BY (2004). In case I, time varying volatility of consumption is not considered. In case II, this feature is added and expressed as following a GARCH time series process. However, in Bansal and Yaron’s 2006 paper, they further illustrate that these two cases lead to similar results in modeling financial markets.

- Case I

\[
\begin{align*}
x_{t+1} &= \alpha^{BY} x_t + e_{x,t+1}^x \\
\Delta c_{t+1} &= \mu_c + x_t + e_{c,t+1}^c \\
\Delta d_{t+1} &= \mu_d + \beta^{BY} x_t + e_{d,t+1}^d
\end{align*}
\]

where $e_{x,t+1}^x$, $e_{c,t+1}^c$, and $e_{d,t+1}^d$ are i.i.d. $N(0, 1)$.

- Case II

\[
\begin{align*}
x_{t+1} &= \alpha^{BY} x_t + \sigma_t e_{x,t+1}^x \\
\Delta c_{t+1} &= \mu_c' + x_t + \sigma_t e_{c,t+1}^c
\end{align*}
\]
\[
\Delta d_{t+1} = \mu'_d + \beta^{BY} x_t + \sigma_t e_{t+1}^d
\]
\[
\sigma_{t+1}^2 = \sigma^2 + \nu(\sigma_t^2 - \sigma^2) + \omega_t + 1
\]

where \( e_{t+1}^d \), \( e_{t+1}^e \), \( e_{t+1}^d \) and \( W_{t+1} \) are i.i.d. \( N(0,1) \). BY(2004) used \( \bar{dp}_t \) as the long-run persistent factor \( x_{t+1} \).

- Drawbacks of BY(2004) model

\( r_{a,t+1} \) is unobservable, so it is difficult to estimate the model’s parameters out. BY(2004) didn’t find a way to estimate the model even for case I, not to mention case II. They simply gave qualitative not quantitative results, by trying some specific numbers for key parameters such as \( \theta, \psi \).

2.2 My Work: Long-Run Model with Two Shocks

My work extends and makes adjustments to BY (2004)’s model. By clearing out and exploring the relationships between asset returns, dividend yield and dividend growth, I construct a long run uncertainty model with two shocks and estimate all the parameters using some statistical methods and particularly GMM methodology. The model helps solve the equity premium puzzle.

I give an extension to their first model. I show that without imposing the time varying uncertainty term on the model, we can get better performance just by clearing out the relationships between asset return, dividend yield and dividend growth. Case II in BY (2004) is somehow unnecessary or we can say redundant, probably because the Case I model already contains the volatility information of consumption growth rate. Thus it is unnecessary to impose
another time series structure on the volatility of consumption only. This result is consistent with what BY (2006) found, that the two cases give similar results. In addition I resolve the latent variable difficulty by finding that the unobservable rate of return on consumption claim can be written as functions of observable variables, specifically the log dividend price ratio.

In my model I demonstrate that two distinct shocks, the consumption growth shock and the dividend growth shock, contribute to all the variations of the following variables: the return on the stock market return $r_m$, the return on the unobservable consumption claim $r_a$, the rate of the consumption growth, the rate of the dividend growth, the log dividend-price ratio, the log consumption-price ratio, the change of the asset price and the Epstein-Zin preference pricing kernel. The data tells us these two shocks are uncorrelated due to the way in which the model is constructed.

By drawing impulse responses functions, I show that both of these two shocks’ effects on asset prices are long run persistent. This result reconciles the fact that the price is much more volatile than the consumption and dividend growth. A small shock in consumption growth or dividend growth has an amplified and persistent effect on the asset price, as well as on the dividend growth and the consumption growth themselves. The result is consistent with what BY(2004) argued: the current shocks to expected growth alter expectations about future economic growth not only for short horizons but also for the very long run, and as investors care about long run components, a small variation in them lead to larger the changes in asset prices.
As in BY(2004), I apply Epstein and Zin(1989)’s preference to the model which allows the separation between the intertemporal elasticity of substitution (IES) and the relative risk aversion (RRA). By using real data and applying GMM methodology, I estimate the IER is about 3.5 and the RRA is about 1.57. While comparing with BY(2004) which doesn’t really estimate out the parameters of its models, this is one of the main contributions of my work. When IER is larger than one, the substitution effect dominates the income effect of higher returns: today’s consumption relative to wealth falls when expected return rises. The IER has a magnitude larger than 1 is an important feature of the model. This IER=3.5 estimates is consistent with Bansal and Yaron(2004), Hansen and Singleton(1982), Attanasio and Weber(1989), Attanasio and Vissing-Jorgensen(2003) which all have an IER estimate well above 1. On the other hand, Hall(1988) and Campbell(1999) give a IER value well below one.

2.3 Model Construction

The two equations below are used as in BY(2004):

\[ r_{m,t+1} = -\rho \delta p_{t+1} + \delta p_t + \Delta d_{t+1} \]  
\[ r_{a,t+1} = -k \delta p_{t+1} + \delta p_t + \Delta c_{t+1} \]  

Epstein-Zin preference and its corresponding log pricing kernel are:

\[ E_t = [\delta^{\theta} \Delta C_{t+1}^{-\theta/\psi} R_{a,t+1}^{\psi(1-\theta)} R_{i,t+1}] = 1 \]
I construct the time series part of the model as:

\[ x_{t+1} = ax_t + \epsilon_{x_{t+1}} \]  \hspace{1cm} (2.4)

\[ \Delta d_{t+1} = bx_t + \epsilon_{d_{t+1}} \]  \hspace{1cm} (2.5)

\[ \Delta c_{t+1} = cx_t + \epsilon_{c_{t+1}} \]  \hspace{1cm} (2.6)

and

\[ r_{m,t+1} = A_1 x_t + \epsilon_{r_{m,t+1}} \]  \hspace{1cm} (2.7)

\[ r_{a,t+1} = A_2 x_t + \epsilon_{r_{a,t+1}} \]  \hspace{1cm} (2.8)

The last two equations are not included in Bansal and Yaron(2004). At a first look, these two extra equations are too restrictive; they assume that the observable \( r_{m,t+1} \) and the unobservable \( r_{a,t+1} \) are both linear functions of the long-run persistent factor \( x_t \). Here I will take the same \( \log p_t \) which is also used in BY(2004) as the specific form of \( x_t \). Actually, it is the only way they can be in BY(2004) model’s setting. Plug the three equations (Equation 2.5) to (Equation 2.7) into (Equation 2.1) and (Equation 2.2), then you will see that \( r_{m,t+1} \) and \( r_{a,t+1} \) should be written as a linear function of \( x_t \).

In Bansal and Yaron’s paper, the coefficient \( c \) in equation (Equation 2.8) is equal to 1; I relax this restriction. All the equations are in demeaned form just for simplification, as in the
following sections I will focus on estimating the coefficients but not the intercept. In addition, the intercepts do not influence the estimation of the coefficients and other properties of the model such as equity premium. This is the common approach in academic finance.

I assume \( \epsilon^x_t, \epsilon^r_{t+1}, \epsilon^{ra}_{t+1}, \epsilon^d_{t+1} \) and \( \epsilon^c_{t+1} \) are all following normal distribution with 0 mean, but with different standard deviations. I also assume that these six error term time series are IID through time, however, they are not IID with each other. BY(2004) assumed that \( \epsilon^x_t, \epsilon^d_{t+1} \) and \( \epsilon^c_{t+1} \) are mutually independent. I show that under the model setting, they are not. Actually, one is a linear combination of the others.

Take \( x_t = \overline{dp_t} \) as BY(2004), I rewrite the full regression equations of my model. See section 6 for detailed calculation about how to get these equations:

\[
\begin{align*}
    r_{m,t+1} &= A_1 T_2^{-1} \overline{dp_t} + \epsilon^d_{t+1} - \rho T_2 \epsilon^x_{t+1} \quad (2.9) \\
    r_{a,t+1} &= A_2 T_2^{-1} \overline{dp_t} + \epsilon^r_{t+1} - k T_3 \epsilon^x_{t+1} \quad (2.10) \\
    \Delta d_{t+1} &= b T_2^{-1} \overline{dp_t} + \epsilon^d_{t+1} \quad (2.11) \\
    \Delta c_{t+1} &= c T_2^{-1} \overline{dp_t} + \epsilon^c_{t+1} \quad (2.12) \\
    \overline{dp_{t+1}} &= \overline{dp_t} + T_2 \epsilon^x_{t+1} \quad (2.13) \\
    \overline{cp_{t+1}} &= a T_3 T_2^{-1} \overline{dp_t} + T_3 \epsilon^x_{t+1} \quad (2.14) \\
    \Delta p_{t+1} &= (1 - a + b T_2^{-1}) \overline{dp_t} + \epsilon^d_{t+1} - T_2 \epsilon^x_{t+1} \quad (2.15)
\end{align*}
\]
Also from section 6.9’s calculation, we have

\[ \epsilon_{t+1} = (T_2 - T_3)^{-1}(\epsilon_{t+1}^d - \epsilon_{t+1}^c) \]  

(2.16)

where

\[ T_2 = (1 - \rho A_1)^{-1} - b(1 - \rho a)^{-1} \]  

(2.17)

\[ T_3 = (1 - k A_2)^{-1} - c(1 - ka)^{-1} \]  

(2.18)

\[ A_1 = b + (1 - \rho a)T_2 \]  

(2.19)

\[ A_2 = c + (1 - ka)T_3 \]  

(2.20)

\[ (1 - a)T_2 + b = (1 - a)T_3 + c \]  

(2.21)

2.4 Data

I used the annual and monthly data sets of the Personal Consumption Expenditure (nondurable goods) from Federal Reserve Bank’s Economic data base. The source of the data is the same as BY(2004) and many other relative researches. I got and calculated the annually and monthly SP500’s continues rate of return, log dividend price ratio, dividend growth rate, and risk free rate from Standard & Poor’s and Robert Shiller’s public data set base. The annual data is from 1930-01-01 to 2010-01-01 containing 81 time series points. The monthly data is from 1929-02-01 to 2010-12-01 containing 623 time series points.

The basic statistics and time series plots of these variables are attached in the Appendix.
2.5 Long-Run Model Estimation

There are 7 unknown parameters in the above equations, let’s call them the long-run uncertainty parameters to distinguish them from the Epstein-Zin preference’s parameters.

These 7 unknown parameters are $A_1$, $A_2$, $a$, $b$, $c$, $k$, $\rho$. If adding $T_2$ and $T_3$ which can be written in terms of the other unknowns, there are 9 unknowns. The $r_{m,t+1}, \Delta c_{t+1}, \Delta d_{t+1}$ and $\overline{d}_{t+1}$ equations are observable and can be estimated by running regressions. Plus the (Equation 2.17) to (Equation 2.21) ’s 5 equations, there are 9 equations in total, so we can solve for the 9 unknowns. All the equations and regressions are written down in demeaned form for simplicity only. Annual and monthly data are studied respectively. Full regression results and details are attached in the Appendix.

The 3 error terms denoted in the regressions should satisfy the following relationship for which I’ve shown the detailed calculation in section 6.9:

$$\epsilon_x^{t+1} = (T_2 - T_3)^{-1}(\epsilon_d^{t+1} - \epsilon_c^{t+1})$$

(Equation 2.22)

This equation tells us the 3 error terms $\epsilon_x^{t+1}$, $\epsilon_d^{t+1}$ and $\epsilon_c^{t+1}$ (we call them ‘shocks”), are not IID with each other, one is a linear combination of the others. This feature is ignored in BY(2004).
The linear regressions results of the observable variables:

\[ r_{m,t+1} = 0.07118dp_t + (1 - \frac{\rho T_2}{T_2 - T_3}) \epsilon_{t+1}^d + (\frac{\rho T_2}{T_2 - T_3}) \epsilon_{t+1}^c, \]

\[ \Delta d_{t+1} = -0.05670dp_t + \epsilon_{t+1}^d, \]

\[ \Delta c_{t+1} = 0.00521dp_t + \epsilon_{t+1}^d, \]

\[ dp_{t+1} = 0.90104dp_t + (1 - \frac{T_2}{T_2 - T_3}) \epsilon_{t+1}^d + (\frac{T_2}{T_2 - T_3}) \epsilon_{t+1}^c. \]

Compare the coefficients of these 4 regressions with those in the (Equation 2.9), (Equation 2.11), (Equation 2.12) and (Equation 2.14), and use (Equation 2.17) to (Equation 2.21), I calculated out:

\[ a = 0.90104, b = -0.7038, c = 0.0647, k = 0.9143, \rho = 0.9679. \]

And \( T_2 = 12.4130, T_3 = 4.6473, A_1 = 0.8836, A_2 = 0.8836. \)

The two returns’ demeaned processes are \( r_{m,t+1} = 0.8836x_t + \epsilon_{t+1}^r \) and \( r_{a,t+1} = 0.8836x_t + \epsilon_{t+1}^r \). \( A_1 \) is equal to \( A_2 \). We conclude that the unobservable return on consumption claim and the observable return on market portfolio claim follow the same demeaned time series process, so that the \( r_{a,t+1} = 0.8836x_t + \epsilon_{t+1}^r \) can be seen as redundant under the models’ internal structure. Imposing this view, the asset pricing equation cannot give more information about the model.
Similar calculation for monthly data: The the linear regression results of the observable variables:

\[
\begin{align*}
    r_{m,t+1} &= 0.02878dp_t + (1 - \frac{\rho T_2}{T_2 - T_3})\epsilon_{t+1}^d + (\frac{\rho T_2}{T_2 - T_3})\epsilon_{t+1}^c, \\
    \Delta d_{t+1} &= -0.0014dp_t + \epsilon_{t+1}^c, \\
    \Delta c_{t+1} &= 0.00127dp_t + \epsilon_{t+1}^d, \\
    \overline{dp}_{t+1} &= 0.9966dp_t + (1 - \frac{T_2}{T_2 - T_3})\epsilon_{t+1}^d + (\frac{T_2}{T_2 - T_3})\epsilon_{t+1}^c,
\end{align*}
\]

Comparing the coefficients of these 4 regressions with those in (Equation 2.9), (Equation 2.11), (Equation 2.12) and (Equation 2.14), and use (Equation 2.17) to (Equation 2.21), I calculated:

\[
\begin{align*}
    a &= 0.9966, b = -0.0466, c = 0.0424, k = 0.9958, \rho = 0.9991. \text{ And } T_2 = 33.308, T_3 = 7.1122, A_1 = 0.0965, A_2 = 0.0965.
\end{align*}
\]

### 2.6 Impulse Response Analysis

In this section, I will analyze the impulse responses time series of the five key variables \( r_{m,t+1}, \Delta d_{t+1}, \Delta c_{t+1}, \overline{dp}_{t+1}, \Delta p_{t+1} \) after the 1% shock of dividend growth or consumption growth hit the market.
The result generated from the Annual data is as follows:

\[
\begin{pmatrix}
    r_{m,t+1} \\
    \Delta d_{t+1} \\
    \Delta c_{t+1} \\
    \bar{d}p_{t+1} \\
    \Delta p_{t+1}
\end{pmatrix} =
\begin{pmatrix}
    0 & 0 & 0 & 0.0712 & 0 \\
    0 & 0 & 0 & -0.0567 & 0 \\
    0 & 0 & 0 & 0.0052 & 0 \\
    0 & 0 & 0 & 0.9010 & 0 \\
    0 & 0 & 0 & 0.0423 & 0
\end{pmatrix}
\begin{pmatrix}
    r_{m,t} \\
    \Delta d_{t} \\
    \Delta c_{t} \\
    \bar{d}p_{t} \\
    \Delta p_{t}
\end{pmatrix} + 
\begin{pmatrix}
    0 \\
    -0.5475 \\
    1 \\
    1.5988 \\
    -0.5988
\end{pmatrix} \epsilon_{t+1}^{d} + 
\begin{pmatrix}
    1.5475 \\
    0 \\
    0 \\
    -1.5988 \\
    1.5988
\end{pmatrix} \epsilon_{t+1}^{c}.
\]

I set the initial variables’ values in the vector as the time 2010-01-01’s real value. It aims to show that when the dividend growth shock or the consumption growth shock happens in 2010-01-01, the other values move correspondingly in the future.

\[(r_{m,t}, \Delta d_{t}, \Delta c_{t}, \bar{d}p_{t}, \Delta p_{t}) = (0.2546, -0.2565, 0.0589, 0.6831, 0.1020)\]

The impulse response pictures tell us, 1% dividend growth shock at time t=0 will give a persistent influence to the stock rate of return (ROR) and make the cumulated change of ROR go to about 0.6 at time t=15. Moreover, the dividend growth shock has a persistent negative effect on the dividend yield but a positive effect on the price growth rate. After 1% dividend growth’s sudden increase happens, the shocks on the dividend growth rate and the consumption growth rate are enormous at t=0 but are dropping to a very low level from next period and converge to 0 afterward.
Similar results happen with a the 1 % consumption growth shock. The similarity between the dividend growth shock and the consumption growth shock implies that a gain in consumption and a gain in dividend have the similar effect on the consumers’ investment decision and therefore on the stock market’s performance.

The monthly result is as follows:

\[
\begin{pmatrix}
    r_{m,t+1} \\
    \Delta d_{t+1} \\
    \Delta c_{t+1} \\
    \Delta p_{t+1}
\end{pmatrix}
= \begin{pmatrix}
    0 & 0 & 0 & 0.0288 & 0 \\
    0 & 0 & 0 & -0.0014 & 0 \\
    0 & 0 & 0 & 0.0013 & 0 \\
    0 & 0 & 0 & 0.0020 & 0
\end{pmatrix}
\begin{pmatrix}
    r_{m,t} \\
    \Delta d_t \\
    \Delta c_t \\
    \Delta p_t
\end{pmatrix}
+ \begin{pmatrix}
    -0.2712 \\
    0 \\
    1 \\
    -0.2724
\end{pmatrix}
\epsilon_{t+1}^d + \begin{pmatrix}
    1.2712 \\
    0 \\
    1 \\
    0
\end{pmatrix}
\epsilon_{t+1}^c.
\]

I set the initial variable values in the vector as the actual values for 2010-01-01. This aims at showing when the dividend growth shock or the consumption growth shock happens in 2010-01-01, what the other values move correspondingly in future.

\[
(r_{m,t}, \Delta d_t, \Delta c_t, \Delta p_t, \Delta p_t) = (0.0514, 0.0035, 0.1462, 0.6043, 0.0144)
\]

In monthly data, the 1% consumption and dividend growth shocks’ effects on the markets have a similar trend as that in the yearly data, but have less magnitude.

2.7 More to say about the Impulse Response Analysis

I’m taking the Annual data result for example.
Based on the 2010-01-01’s initial value vector listed above and the impulse response vector, a 1% increase of the dividend shock $\epsilon_t^{d,t+1}$ will make $r_m$ drop from $r_{m,t} = 0.2546$ to $r_{m,t+1} = 0.0431$ and $\Delta p$ drop from $\Delta p_t = 0.1020$ to $\Delta p_{t+1} = 0.0229$. Here $t$ represents the year 2010, and $t+1$ represents for the year 2011.

A similar story applies to the consumption shock $\epsilon_t^{c,t+1}$. 1% increase of the consumption shock will make $r_m$ drop from $r_{m,t} = 0.2546$ to $r_{m,t+1} = 0.0641$ and $\Delta p$ drop from $\Delta p_t = 0.1020$ to $\Delta p_{t+1} = 0.0449$.

It is worth to mentioning that a 1% increase of the dividend shock $\epsilon_t^{d,t+1}$ will contribute gains to the next period’s consumption as well. It will help maintain a persistent trend the the $\Delta c$. Without the 1% increase in dividend shock, $\Delta c$ will drop from this period’s 0.0589 to next period’s 0.0003. But with this 1% increase in dividend shock, $\Delta c$ will drop from this period’s 0.0589 to next period’s 0.0103. This is a 1% difference in terms of $\Delta c$. Like $\Delta c$ is consumption growth rate, and a positive consumption growth rate will push next period consumption to an even higher level.

The above increase effect, transferred from this period’s dividend shock to next period’s consumption level, can be considered as the effect of the Marginal Propensity to Consume, (MPC). The MPC is known as the proportion of the disposable income which individuals desire to spend on consumption. In our generalized and simplified model, the next period disposable income comes from two channels: one is the dividend payment which depends on the last period investment and consumption allocations; the other is the granted wealth.
2.8 GMM Method and the Estimations of the Epstein-Zin Preference Parameters

2.8.1 GMM overview

GMM, Generalized Method of Moments, was first brought out by Hansen (1982) and was further discussed in Hansen and Singleton (1982). It applied a widely used statistical method, the method of moments, into the finance study which usually has a time series related information set. GMM is considered as a pioneering step in the Econometrics field. In theory, GMM shows that Maximum likelihood estimation is one of the special cases of GMM estimation. GMM helps to estimate parameters when we have more moment conditions than unknown parameters. GMM is a very useful and meaningful tool in the finance and economics field because these studies usually have many factors’ historical information that need to be considered, thus will have more moment conditions generated. Take the asset pricing formula I discussed in the previous section for example:

\[ p_t = E_t[m_{t+1}x_{t+1}], \]

where \( p_t \) is price, \( m_{t+1} \) is the pricing kernel as discussed before and \( x_{t+1} \) is payoffs. Since

\[ \frac{x_{t+1}}{p_t} = R_{t+1} \]

then,

\[ E_t[m_{t+1}R_{t+1} - 1] = 0, \]
which can also be written in a form of a conditional expectation:

$$E[(m_{t+1}R_{t+1} - 1) \mid I_t] = 0.$$  

Here $I_t$ represents the information set until time $t$ and should include all information that the investors can get at time $t$. The conditional expectation should be equal to zero, otherwise there is an arbitrage opportunity which when it occurs will disappear in the next second. The practical idea of GMM’s implementation in a finance study is straightforward: under a relative efficient market when all information in the economy is accessible to the investors, no matter what pricing kernel we are using, the stock price should fully reflect all the discounted future pay-outs information. Otherwise, there would be an obvious arbitrage opportunity for all the investors and then this opportunity usually wouldn’t last longer than a second in the market.

$$E[ E_t[m_{t+1}R_{t+1} - 1] ] = 0 \implies E[ m_{t+1}R_{t+1} - 1 ] = 0$$

Thus,

$$E_t [(m_{t+1}R_{t+1} - 1)z_t ] = 0$$

Here the instrumental variable set $z_t$ can be any factors in the information set $I_t$, for example, the consumption growth rate, the risk free rate, the inflation rate, so on and so forth. By introducing $z_t$ we bring more moment conditions to the unconditional equation. If the model
with certain estimated parameters is correct, all moment conditions should be approximately equal to 0. For example, when \( z_t = (1, R_t) \), the unconditional moments condition goes to

\[
E \begin{bmatrix} m_{t+1} R_{t+1} - 1 \\ (m_{t+1} R_{t+1} - 1) R_t \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\]

If the \( m_{t+1} \) is the commonly used power utility case,

\[
m_{t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} = \beta \frac{u'(C_{t+1})}{u'(c_t)},
\]

then we have one unknown parameter \( \gamma \) but two moment conditions.

### 2.8.2 The Application of GMM to My Model

Based on the theoretical foundation of GMM, I used the lag-1 asset return \( r_{m,t} \) and risk free rate \( r_{f,t} \) as instrumental variables to estimate the Epstein-Zin preference parameters. In addition, I perform the GMM statistic test to give evidence for accepting the model.

This is one of the contributions I made for the long-run uncertainty model. BY(2004), BY(2006) and other relevant research didn’t apply any econometric method to estimate their model’s parameters. They simply try several values of the risk aversion coefficient \( \gamma \) and say something about the tried-and-error results. The model implies that \( \gamma \) should be a number within a relatively fixed range, which can be obtained and accepted statistically by using econometric techniques.
The pricing kernel derived from Epstein-Zin preference is

\[ m_{t+1} = \theta \log \delta - \theta \psi^{-1} \Delta c_{t+1} + (\theta - 1) r_{a,t+1} \]

Under the no-arbitrage condition, the following equation should be satisfied, otherwise a short position on one and long position on the other will generate arbitrage profits which will thereafter disappear in a second:

\[ E_t[\exp(\theta \log \delta - \theta \psi^{-1} \Delta c_{t+1} + (\theta - 1) r_{a,t+1} + r_{f,t+1})] = 1 \]

I use two observable assets’ rate of returns in the above equation, one is the market portfolio \( r_{m,t+1} \) which is SP500 specifically and the other is Treasury bond yield \( r_{f,t+1} \).

The object is to estimate the three unknown parameters \( \delta, \theta \) and \( \psi \) so that they satisfy the following vector equation. In other words, we need to find the good parameters which can make the left hand side of the below equation equal not significantly different from zero.

\[
E \begin{bmatrix}
\exp(m_{t+1}(\delta, \theta, \psi)r_{m,t+1}) \\
\exp(m_{t+1}(\delta, \theta, \psi)r_{f,t+1}) \\
\exp(m_{t+1}(\delta, \theta, \psi)r_{m,t+1})z_t \\
\exp(m_{t+1}(\delta, \theta, \psi)r_{f,t+1})z_t \\
\end{bmatrix} - \begin{bmatrix}
1 \\
1 \\
z_t \\
z_t \\
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
z_t \\
z_t \\
\end{bmatrix}
\]

\[(2.23)\]

The \( z_t \) are instrumental variables which contains all possible useful information at time \( t \). As the usual way in GMM, I use \( \delta C_t \) and \( r_{f,t} \) as the instrumental variable \( z_t \).
Let

\[ e_t(b) = \begin{bmatrix} \exp(m_{t+1}r_{m,t+1}) - 1 \\ \exp(m_{t+1}r_{f,t+1}) - 1 \\ \exp(m_{t+1}r_{m,t+1})z_t - z_t \\ \exp(m_{t+1}r_{f,t+1})z_t - z_t \end{bmatrix}, \]

where \( b \) represents \([\delta, \theta, \psi]\). Then, let

\[ f_T = \mathbf{E}_T[e_t] \]

I run 2-stage GMM to estimate the parameters:

In the first stage, I get

\[ \hat{b}_1 = \arg\min_b f_T(b)'W_f f_T(b), \]

where \( W \) represents the identity matrix \( I \).

In the second stage, I get

\[ \hat{b}_2 = \arg\min_b f_T(b)'\hat{S}^{-1}f_T(b), \]

where the \( \hat{S} \) is the Newey-West variance covariance matrix.

Using yearly data, the optimization function inside GMM gives me the following estimates:

\( \theta = -0.7936, \delta = 0.1, \psi = 3.5 \). The \( \psi \) is the intertemporal elasticity of substitution, and the
relative risk aversion $\gamma = 1 - \theta (1 - 1/\psi) = 1.5669$. Under this estimation, the value of the quadratic error terms $f_T(b)'S^{-1}f_T(b) = 0.01756$, which is very near 0 so that we can say that these parameter estimates do satisfy our object function (14) statistically.

To justify the goodness-of-fit of the model using these estimation, I do a $\chi^2$ test:

$$TJ_T = T[f_T(\hat{b})'S^{-1}f_T(\hat{b})] \sim \chi^2 (m),$$

Where $m$ is the number of moments minus the number of parameters. The above test gives p-value=0.2307 for yearly data. The model is accepted.

When using monthly data, GMM gives me the following estimates: $\theta = -5.7017$, $\delta = 6.0267$, $\psi = 1.4028$. The $\psi$ is the intertemporal elasticity of substitution. And the relative risk aversion $\gamma = 1 - \theta (1 - 1/\psi) = 2.6372$. The p-value of the GMM test is 0.1197, which also accepts the model.

| Table I |
|---|---|---|---|
| GMM ESTIMATION FOR YEARLY AND MONTHLY DATA |
| $\theta$ | $\delta$ | $\psi$ | $\gamma$ |
| Yearly Data | -0.7936 | 0.1 | 3.5 | 1.5669 |
| Monthly Data | -5.7017 | 6.0267 | 1.4028 | 2.6372 |
TABLE II

GMM TEST STATISTICS FOR $E_{T}B = 0$

<table>
<thead>
<tr>
<th></th>
<th>error term</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yearly Data</td>
<td>0.01796</td>
<td>0.2307</td>
</tr>
<tr>
<td>Monthly Data</td>
<td>0.00389</td>
<td>0.1197</td>
</tr>
</tbody>
</table>

Comparing the annually and monthly relative risk aversion $\gamma$’s which are 1.3929 and 2.6372 respectively, we know the RRA monthly is larger than the RRA yearly. This result matches our common sense that investors are more risk averse in the short run than in the long run.

2.9 How My Model Helps to Resolve the Equity Premium Puzzle

The equity premium can be written as in the Epstein-Zin preference section. I use the proved result that the observable $r_m$ and the unobservable $r_a$ follow the same time series process, thus $\sigma_{ma} = \sigma_m^2$. To be comparable with the existing literature which uses yearly data, I got the yearly equity risk premium calculated from the model as:

$$E(r_{m,t+1}) - E(r_{f,t+1}) = \frac{\theta}{\phi}\sigma_{mc} + (1-\theta)\sigma_{ma} - \frac{1}{2}\sigma_m^2$$

$$= \frac{\theta}{\phi}Cov(r_m, \Delta c) + (1-\theta-0.5)\sigma_m^2$$

$$= 0.04528 \approx 4.5\%$$
The equity risk premium from real data is

\[ E(r_{m,t+1}) - E(r_{f,t+1}) = 0.05412 - 0.000438 = 0.053682 \approx 5.3\% \]

There is a 0.8% difference of the equity risk premium between the model and the real data. Though the result of my model is not perfect, it is much better than the previous models based on the power utility. The relative risk aversion of my model is 1.5669, when using this reasonable RRA, the power utility model’s equity risk premium is calculated as follows:

\[ E(r_{m,t+1}) - E(r_{f,t+1}) = \frac{\gamma}{\phi} \text{Cov}(r_m, \Delta c) - 0.5\sigma_m^2 \]

\[ = -0.0109 \]

The number is negative and far from matching the real data’s equity premium 5.3%. If we were using the old model, even when \( \gamma = 10 \) the risk premium is only about 2.5%, which is still far away from 5.3%. When \( \gamma = 15 \), which is a very large number, the equity risk premium can be around the number obtained by my model.
2.10 Appendix

2.10.1 The formulas to derive the long-run risks VAR

- 

\[ r_{m,t+1} = -\rho \overline{dp}_{t+1} + \overline{dp}_t + \Delta d_{t+1} \]
\[ r_{a,t+1} = -k \overline{cp}_{t+1} + \overline{cp}_t + \Delta c_{t+1} \]
\[ x_t = ax_{t-1} + \epsilon^x_t \]
\[ r_{m,t+1} = A_1 x_t + \epsilon^r_{t+1} \]
\[ r_{a,t+1} = A_2 x_t + \epsilon^{ra}_{t+1} \]
\[ \Delta d_{t+1} = bx_t + \epsilon^d_{t+1} \]
\[ \Delta c_{t+1} = cx_t + \epsilon^c_{t+1} \]

The pricing kernel is written as:

\[ m_{t+1} = \theta \log \delta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1)r_{a,t+1} \]

- Get the \( \overline{dp}_{t+1} \)'s Expression

Rewrite \( r_{m,t+1} = -\rho \overline{dp}_{t+1} + \overline{dp}_t + \Delta d_{t+1} \) in a recursive form of \( \overline{dp}_t \)

\[ \overline{dp}_t = E_t \sum_{j=1}^{\infty} \rho^{j-1}(-\Delta d_{t+j} + r_{m,t+j}) \]
Plugging $r_{m,t+1} = A_1 x_t + c^r_{t+1}$ and $\Delta d_{t+1} = b x_t + c^d_{t+1}$ into above equation, and use $x_t = a x_{t-1} + c^x_t$:

$$dp_t = (1 - \rho A_1)^{-1} x_t - b (1 - \rho a)^{-1} x_t, \quad (2.24)$$

Let $T_2 = (1 - \rho A_1)^{-1} - b (1 - \rho a)^{-1}$. Plugging (Equation 2.24) into $x_{t+1} = a x_t + c^r_{t+1}$ yields

$$dp_{t+1} = adp_t + T_2 c^x_{t+1}$$

$$\epsilon^{dp}_{t+1} = T_2 c^x_{t+1}$$

• Get the expression for $cp_{t+1}$’s

Similar to obtaining $dp_{t+1}$, we have

$$cp_t = E_t \sum_j k^{j-1} (r_{a,t+j} - \Delta c_{t+j})$$

$$= T_3 x_t = (1 - k A_2)^{-1} x_t - c (1 - k a)^{-1} x_t, \quad (2.25)$$

where $T_3 = (1 - k A_2)^{-1} - c (1 - k a)^{-1}$. Plugging (Equation 2.25) into $x_{t+1} = a x_t + c^r_{t+1}$ yields

$$cp_{t+1} = a cp_t + T_3 c^x_{t+1}$$

$$\epsilon^{cp}_{t+1} = T_3 c^x_{t+1}$$
Get the expression for $r_{m,t+1}$

Since $x_t = T_2^{-1} \overline{dp}_t = T_3^{-1} \overline{cp}_t$ and

$$r_{m,t} = A_1 x_t + \epsilon^r_{t+1} \quad (2.26)$$
$$r_{m,t+1} = -\rho \overline{dp}_{t+1} + \overline{dp}_t + \Delta d_{t+1} \quad (2.27)$$

Plugging $x_t$’s expression into (Equation 2.26) and plug $\overline{dp}_{t+1}$, $\overline{dp}_t$ and $\Delta d_{t+1}$’s expressions into (Equation 2.27) yields

$$r_{m,t+1} = A_1 T_2^{-1} \overline{dp}_t + \epsilon^r_{t+1} \quad (2.28)$$

$$r_{m,t+1} = -\rho a \overline{dp}_t - \rho T_2^{-1} \epsilon^r_{t+1} + b x_t + \epsilon^d_{t+1} \quad (2.29)$$

Compare (Equation 2.28) and (Equation 2.29), we should have

$$A_1 T_2^{-1} = 1 + b T_2^{-1} - \rho a \Rightarrow A_1 = b + (1 - \rho a) T_2$$
$$\epsilon^r_{t+1} = \epsilon^d_{t+1} - \rho T_2 \epsilon^x_{t+1}$$
• Get the expression $r_{a,t+1}$

Similarly we have

$$r_{a,t+1} = A_2 x_t + \epsilon^r_{t+1}, \quad (2.30)$$

$$r_{a,t+1} = -k \overline{c p}_{t+1} + \overline{c p}_t + \Delta c_{t+1}, \quad (2.31)$$

then

$$r_{a,t+1} = A_2 T_3^{-1} \overline{c p}_t + \epsilon^r_{t+1} \quad (2.32)$$

$$r_{a,t+1} = -k (a \overline{c p}_t + T_3 \epsilon^x_{t+1}) + \overline{c p}_t + (c x_t + \epsilon^c_{t+1})$$

$$= -ka \overline{c p}_t - kT_3 \epsilon^x_{t+1} + \overline{c p}_t + cT_3^{-1} \overline{c p}_t + \epsilon^c_{t+1}$$

$$= (1 + cT_3^{-1} - ka) \overline{d p}_t + \epsilon^c_{t+1} - kT_3 \epsilon^x_{t+1} \quad (2.33)$$

Compare (Equation 2.32) and (Equation 2.33), we should have

$$A_2 T_3^{-1} = 1 + cT_3^{-1} - ka \Rightarrow A_2 = c + (1 - ka)T_3$$

$$\epsilon^r_{t+1} = \epsilon^c_{t+1} - kT_3 \epsilon^x_{t+1}$$

• Update the expression for $\overline{c p}_{t+1}$

Since we have

$$\overline{d p}_t = T_2 x_t,$$
\[ \overline{c}p_t = T_3 x_t, \]

So,

\[ \overline{c}p_t = T_3 T_2^{-1} dp_t, \]
\[ \overline{c}p_{t+1} = a \overline{c}p_t + T_3 \epsilon^c_{t+1} = aT_3 T_2^{-1} dp_t + T_3 \epsilon^c_{t+1}. \]

Since we also have

\[ r_{a,t+1} = (1 + c T_3^{-1} - ka) \overline{c}p_t + \epsilon^a_{t+1}, \]

which is equivalent to

\[ r_{a,t+1} = (1 + c T_3^{-1} - ka) T_3 T_2^{-1} dp_t + \epsilon^a_{t+1}. \]

• Get the expression for \( \Delta c_{t+1} \) and \( \Delta d_{t+1} \)

\[ \Delta c_{t+1} = cx_t + \epsilon^c_{t+1} = c T_2^{-1} dp_t + \epsilon^c_{t+1} \]
\[ \Delta d_{t+1} = b T_2^{-1} dp_t + \epsilon^d_{t+1} \]

• Complete List of Equations
\[ r_{m,t+1} = A_1 T_2^{-1} \overline{dp}_t + \epsilon_{t+1}^d - \rho T_2 \epsilon_t^x, \]

\[ r_{a,t+1} = A_2 T_2^{-1} \overline{cp}_t + (\epsilon_{t+1}^c - k T_3 \epsilon_{t+1}^x) \]

\[ \Delta d_{t+1} = b T_2^{-1} \overline{dp}_t + \epsilon_{t+1}^d, \]

\[ \Delta c_{t+1} = c T_2^{-1} \overline{dp}_t + \epsilon_{t+1}^c, \]

\[ \overline{dp}_{t+1} = a \overline{dp}_t + T_2 \epsilon_{t+1}^x, \]

\[ \overline{cp}_{t+1} = a T_3 T_2^{-1} \overline{dp}_t + T_3 \epsilon_{t+1}^x. \]

- What about \( \Delta p_{t+1} \)?

\[ \Delta p_{t+1} = -\overline{dp}_{t+1} + \overline{dp}_t + \Delta d_{t+1} \]

\[ = -a \overline{dp}_t - b T_2 \epsilon_{t+1}^x + \overline{dp}_t + b T_2^{-1} \overline{dp}_t + \epsilon_{t+1}^d \]

\[ = (1 - a + b T_2^{-1}) \overline{dp}_t + \epsilon_{t+1}^d - b T_2 \epsilon_{t+1}^x. \]

\( \Delta p_{t+1} \) can also be written down as a function of \( \overline{cp}_t \),

\[ \Delta p_{t+1} = -\overline{cp}_{t+1} + \overline{cp}_t + \Delta c_{t+1} \]

\[ = T_3 T_2^{-1} (1 - a + c T_3^{-1}) \overline{dp}_t + \epsilon_{t+1}^c - T_3 \epsilon_{t+1}^x. \]
So,

\begin{align*}
1 - a + bT_2^{-1} &= T_2^{-1}T_3^{-1}(1 - a + cT_3^{-1}) \\
\epsilon_{t+1}^d - T_2\epsilon_{t+1}^x &= \epsilon_{t+1}^c - T_3\epsilon_{t+1}^x,
\end{align*}

implying that

\[\epsilon_{t+1}^x = (T_2 - T_3)^{-1}(\epsilon_{t+1}^d - \epsilon_{t+1}^c).\]

**• Full Model**

From section 6.2 to 6.5, we have the all constrains for \(T_2\) and \(T_3\).

\begin{align*}
T_2 &= (1 - \rho A_1)^{-1} - b(1 - \rho a)^{-1} \\
T_3 &= (1 - k A_2)^{-1} - c(1 - ka)^{-1} \\
A_1 &= b + (1 - \rho a)T_2 \\
A_2 &= c + (1 - ka)T_3 \\
1 - a + bT_2^{-1} &= T_2^{-1}(c - aT_3 + T_3)
\end{align*}

Combining the above equations yields that

\[r_{m,t+1} = A_1T_2^{-1}\tilde{d}_{p_t} + \epsilon_{t+1}^d - \rho T_2\epsilon_{t+1}^x,\]
\[ r_{a,t+1} = A_2 T_2^{-1} dp_t + \epsilon^c_{t+1} - k T_3 \epsilon^x_{t+1} , \]
\[ \Delta d_{t+1} = b T_2^{-1} dp_t + \epsilon^d_{t+1} , \]
\[ \Delta c_{t+1} = c T_2^{-1} dp_t + \epsilon^c_{t+1} , \]
\[ dp_{t+1} = a dp_t + T_2 \epsilon^r_{t+1} , \]
\[ c p_{t+1} = a T_3 T_2^{-1} dp_t + T_3 \epsilon^x_{t+1} , \]
\[ \Delta p_{t+1} = (1 - a + b T_2^{-1}) dp_t + \epsilon^d_{t+1} - T_2 \epsilon^r_{t+1} , \]

- What about \( M \)?

\[
m_{t+1} = \theta \log \delta - \theta \psi^{-1} \Delta c_{t+1} + (\theta - 1) r_{a,t+1} \]
\[
= \theta \log \delta - \theta \psi^{-1} (c T_2^{-1} dp_t + \epsilon^c_{t+1}) + (\theta - 1) (A_2 T_2^{-1} dp_t + \epsilon^c_{t+1} - k T_3 \epsilon^x_{t+1}) \\
= \theta \log \delta + (A_2 \theta - A_2 - c \theta \psi^{-1}) T_2^{-1} dp_t - \theta \psi^{-1} \epsilon^c_{t+1} + (\theta - 1) \epsilon^c_{t+1} - (\theta - 1) k T_3 \epsilon^x_{t+1} \\
= \theta \log \delta + (A_2 \theta - A_2 - c \theta \psi^{-1}) T_2^{-1} dp_t + (\theta - 1 - \theta \psi^{-1}) \epsilon^c_{t+1} - \\
\quad (\theta - 1) k T_3 (T_2 - T_3)^{-1} (\epsilon^d_{t+1} - \epsilon^c_{t+1}) \\
= \theta \log \delta + (A_2 \theta - A_2 - c \theta \psi^{-1}) T_2^{-1} dp_t + (\theta - 1 - \theta \psi^{-1} - (\theta - 1) k T_3 \nu) \epsilon^c_{t+1} + \\
\quad (\theta - 1) k T_3 \nu \epsilon^d_{t+1} ,
\]

2.10.2 Summary Statistics and Figures
### TABLE III

**SUMMARY STATISTICS OF MAJOR VARIABLES IN YEARLY DATA**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Autocorr.(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_m$</td>
<td>5.412%</td>
<td>18.728%</td>
<td>0.0229</td>
</tr>
<tr>
<td>dp</td>
<td>1.2734</td>
<td>0.4564</td>
<td>0.9010</td>
</tr>
<tr>
<td>$\Delta d$</td>
<td>0.8812%</td>
<td>10.7168%</td>
<td>0.1838</td>
</tr>
<tr>
<td>$\Delta c$</td>
<td>5.2267%</td>
<td>6.0308%</td>
<td>0.6014</td>
</tr>
</tbody>
</table>

### TABLE IV

**CORRELATION MATRIX OF THE VARIABLES IN YEARLY DATA**

<table>
<thead>
<tr>
<th></th>
<th>$r_m(r_a)$</th>
<th>dp</th>
<th>$\Delta d$</th>
<th>$\Delta c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_m(r_a)$</td>
<td>1</td>
<td>-0.2304</td>
<td>0.0848</td>
<td>0.3732</td>
</tr>
<tr>
<td>dp</td>
<td>1</td>
<td>-0.0397</td>
<td>-0.1263</td>
<td></td>
</tr>
<tr>
<td>$\Delta d$</td>
<td>1</td>
<td></td>
<td>-0.1263</td>
<td></td>
</tr>
<tr>
<td>$\Delta c$</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

#### 2.10.3 Figures for Major Variables
TABLE V
SUMMARY STATISTICS OF MAJOR VARIABLES IN MONTHLY DATA

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Autocorr.(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_m$</td>
<td>3.217%</td>
<td>3.745%</td>
<td>0.3131</td>
</tr>
<tr>
<td>dp</td>
<td>1.0568</td>
<td>0.3976</td>
<td>0.9966</td>
</tr>
<tr>
<td>$\Delta d$</td>
<td>0.086%</td>
<td>0.6097%</td>
<td>0.7190</td>
</tr>
<tr>
<td>$\Delta c$</td>
<td>0.4759%</td>
<td>0.9048%</td>
<td>0.0988</td>
</tr>
</tbody>
</table>

TABLE VI
CORRELATION MATRIX OF THE VARIABLES IN YEARLY DATA

<table>
<thead>
<tr>
<th></th>
<th>$r_m(r_a)$</th>
<th>dp</th>
<th>$\Delta d$</th>
<th>$\Delta c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_m(r_a)$</td>
<td>1</td>
<td>0.2162</td>
<td>0.0102</td>
<td>0.0811</td>
</tr>
<tr>
<td>dp</td>
<td>1</td>
<td>-0.0794</td>
<td>0.0475</td>
<td></td>
</tr>
<tr>
<td>$\Delta d$</td>
<td>1</td>
<td>-0.1256</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta c$</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 3. Annual Data Top: $\Delta d$ shock on $r_m$ and cumulative $r_m$; Middle: $\Delta d$ shock on dividend growth rate and consumption growth rate; Bottom: $\Delta d$ shock on dividend yield and price growth rate
Figure 4. Annual Data Top: $\Delta c$ shock on $r_m$, cumulated $r_m$; Middle: $\Delta c$ shock on dividend growth rate and consumption growth rate; Bottom: $\Delta c$ shock on dividend yield and price growth rate
Figure 5. Month Data Top: $\Delta d$ shock on $r_m$ and cumulative $r_m$; Middle: $\Delta d$ shock on dividend growth rate and consumption growth rate; Bottom: $\Delta d$ shock on dividend yield and price growth rate
Figure 6. Month Data Top: $\Delta c$ shock on $r_m$, cumulated $r_m$; Middle: $\Delta c$ shock on dividend growth rate and consumption growth rate; Bottom: $\Delta c$ shock on dividend yield and price growth rate
Figure 7. Yearly Continuous Rate of Return

Figure 8. Yearly Continuous Risk Free Rate
Figure 9. Yearly Dividend Yield

Figure 10. Yearly Continuous Dividend Growth Rate

Figure 11. Yearly Continuous Consumption Growth Rate
Figure 12. Monthly Continuous Rate of Return

Figure 13. Monthly Continuous Risk Free Rate
Figure 14. Monthly Dividend Yield

Figure 15. Monthly Continuous Dividend Growth Rate

Figure 16. Monthly Continuous Consumption Growth Rate
CHAPTER 3

HIDDEN MARKOV MODELS AND ASSET PRICING

This chapter presents Hidden Markov Models (HMMs) and their application in financial asset pricing, where the financial data can be considered as being governed by some underlying stochastic process that is hidden. In the literature, what pattern financial time series follows has long been an interest of the financial industry, attracting economists, statisticians and scientists and researchers from many other disciplines.

3.1 A Brief Introduction to Hidden Markov Models (HMMs)

Hidden Markov models can be defined in term of \((O_t, X_t)\). Here \(O_t\) is observable, e.g. price or ROR, \(X_t\) is an unobservable discrete "state", e.g., "Bull" or "Bear".

Markov models are widely used to model time series data, as well as other sequential data. In a Markov model, the next observation in the series depends on historical data only through the current observation; this is known as the Markov property. Consider a time series \(x_t\) with \(t\) denoting the time points; then the Markov property can be precisely described as

\[
P(X_{t+1} | X_t, X_{t-1}, \ldots, X_1) = P(X_{t+1} | X_t). \tag{3.1}
\]

If we further assume that the conditional distribution of \(X_{t+1} | X_t\) is independent of time \(t\), then

\[
P(X_{t+1} = j | X_t = i) = p_{ij}, \text{ for } i, j = 1, \ldots, N,
\]
where $1, \ldots, N$ are elements in the support of $X_t$. A transition probability matrix $A$ can be defined with entries $p_{ij}$, and

$$p_{ij} \geq 0, \quad \text{and} \quad \sum_{j=1}^{N} p_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, N.$$ 

In order to facilitate the Markov model, the probability distribution of the initial state is modeled as

$$P(X_1 = i) = \pi_i, \quad \text{for} \quad i = 1, \ldots, N,$$

where $\sum_{i=1}^{N} \pi_i = 1$. This may be given, or estimated.

In many real applications, the $X_i$’s are unobservable. For example, to model the movement of a particular stock, Markov models are often employed to model its two possible short-term states: Bull and Bear, or three possible short-term states: Bull, Contraction and Bear. However, these movements are usually hard to define due to the high variability of stock markets within each given short term. Therefore we propose to model the stock market movement as the hidden states of the HMMs, for which one needs not to know anything about what short-term movement generates the actual observable stock price $O_t$. Instead, each state of the HMM is associated with a probability distribution, which governs the distribution of the actual stock prices. Specifically, at time $t$, an observation $o_t$ is generated by a probabilistic function $f(o_t | X_t = i)$, which is associated with state $i$, and the short-term movement $X_t$ is modeled as the hidden states of a HMM, where the next state only relies on the current state and is independent of all the previous states. An unknown transition probability matrix $A$ which
quantifies the dependence between $X_t$ and $X_{t+1}$ is to be constructed. Figure below shows the example of an HMM that models the prices of certain stock.

Assume the short-term movement of a stock has two distinct states, Bull and Bear. Also assume that the stock price within each short term follows a particular probability distribution associated with the short-term movement of that period. The next short-term movement (state) is determined by current movement according to an unknown transition probability matrix $A$. 

![Figure 17. Hidden Markov Model](image-url)
The primary goal of the HMM is based on the observed stock price series $o_t$, to estimate the transition probability matrix $A$ and the most likely short-term movement series $x_t$, to predict the potential future short-term movement, and to provide a guideline for modeling future stock prices.

3.2 Formulation of HMMs

An typical HMM is composed of four elements: $(X, \Pi, A, f)$.

- The set of states $X = (1, \ldots, N)$, where the state at time $t$ is denoted as $X_t$.
- Initial state distribution $\Pi = \{\pi_i\}; i \in X$ is defined as
  \[
  \pi_i = P(X_1 = i). 
  \]

- State transition probability matrix $A = \{p_{ij}\}; i, j \in X$, where
  \[
  p_{ij} = P(X_{t+1} = j | X_t = i). 
  \]

- The probability distribution of observed series $f(o_t | X_t = i)$, which can be modeled as a normal density with mean $\mu_i$ and variance $\sigma_i^2$,
  \[
  f(o_t | X_t = i) \sim Normal(\mu_i, \sigma_i^2) 
  \]

After modeling the stock price series as an HMM, we are able to calculate the likelihood of the observation price series and the most probable underlying state sequences. Also we can
estimate the model \( f(o_t|X_t = j) \) based on the observed price series, and then use the estimated model to predict future stock price.

Given the HMM and an observed stock price series \( O = (o_1, \ldots, o_T) \), the primary purpose is to estimate the transition probability matrix \( A \), the parametric model \( f(o_t|X_t = i) \) and the most probable state sequence, so that the likelihood of the observed stock price series \( P(O|A,f,X_i) \) is maximized. In specific, it consists of three tasks. The first task is to determine which of the HMMs is most likely when the stock price series is given. The second task concerns estimating the hidden state sequence, which best explains the observed price series. The third task is to construct a prediction model based on the historical data, which can be used for future prediction.

3.2.1 Task 1: computing the probability of an observed series

Given an observation series \( O = (o_1, \ldots, o_T) \) and an HMM \((\Pi, A, f)\), we want to compute the conditional probability of the series \( P(O|\Pi, A, f) \). Since the observations are independent of each other, the probability of a state sequence \( X = (X_1, \ldots, X_T) \) generating the observation sequence can be calculated as

\[
P(O|\Pi, A, f, X) = P(O|f, X) = \prod_{t=1}^{T} P(o_t|X_t, f) = \prod_{t=1}^{T} f(o_t|X_t).
\]

The state transition probability is

\[
P(X|\Pi, A, f) = P(X|\Pi, A) = \prod_{t=1}^{T} \pi_{X_t} p_{X_1 X_2} \cdots p_{X_{T-1} X_T}.
\]
Therefore, the conditional probability $P(O|\Pi, A, f)$ is

\[
P(O|\Pi, A, f) = \sum_X P(X|\Pi, A, f)P(O|\Pi, A, f, X)
= \sum_{X_1, \ldots, X_T} \pi_{X_1} \prod_{t=1}^T p_{X_t|X_{t+1}} f(o_t|X_t).
\]

### 3.2.2 Task 2: finding the best state sequence

The second task is to find the best state sequence given an HMM and the observed stock price series. There are several possible ways to attack this task. One of them is the Viterbi algorithm, which proceeds as follows. For time $1 \leq t \leq T + 1$, define

\[
\gamma_i(t) = P(X_t = i|O, \Pi, A, f) = \frac{P(X_t = i, O|\Pi, A, f)}{P(O|\Pi, A, f)} = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^N \alpha_j(t)\beta_j(t)},
\]

Here $\alpha_i(t) = P(o_1, o_2, \ldots, o_{t-1}, X_t = i|\Pi, A, f)$, which stores the total probability of ending up in state $i$ at time $t$, given the observation sequence $o_1, o_2, \ldots, o_{t-1}$, and

\[
\beta_i(t) = P(o_{t+1}, o_{t+2}, \ldots, o_T|s_t = i, \Pi, A, f)
\]

which calculates the probability of the partial observation sequence from $t+1$ to the end, given the HMM model $(\Pi, A, f)$ and state $i$ at time $t$. The individually most likely state sequence $\tilde{X}$ is defined as

\[
\tilde{X} = \arg \max_i \gamma_i(t), \quad \text{for } 1 \leq t \leq T + 1, \quad 1 \leq i \leq N.
\]
Clearly, $\tilde{X}$ maximizes the expected number of correct states. However, directly maximizing $\gamma_i(t)$ may generate an unlikely state sequence as it does not incorporate the state transition probabilities. The Viterbi algorithm proposes to find the most likely state sequence $\tilde{X}$ by solving

$$\arg\max_{\tilde{X}} P(\tilde{X} | O, \Pi, A, f).$$

For a fixed observation sequence $O$, the above maximization problem is equivalent to

$$\arg\max_{\tilde{X}} P(\tilde{X}, O | \Pi, A, f).$$

A forward-backward algorithm (Baum, Welch, Weiss, Sooles 1969) can be constructed to solve the maximization.

### 3.2.3 Task 3: estimating parametric model

The last problem about HMMs is the parameters estimation. Given an observation sequence $O$, we want to find the model parameters $(\Pi, A, f)$ that best explains the observation sequence. The problem can be reformulated to find the parameters that solves

$$\arg\max_{\Pi,A,f} P(O | \Pi, A, f).$$

However, there is no analytic solution to maximize $P(O | \Pi, A, f)$, and thus a local maximization algorithm, called the Baum-Welch algorithm, is often employed; it has been reported to achieve the global maximum with high probability. The Baum-Welch algorithm can be thought
of as a special case of the Expectation Maximization (EM) algorithm, which iteratively improves the value of $P(O|\Pi, A, f)$. The Baum-Welch algorithm is numerically stable with the likelihood nondecreasing at each iteration, and it has linear convergence to a local optima.

The Baum-Welch algorithm works as follows. Define intermediate variables $q_t(i, j)$ as the probability of being at state $i$ at time $t$, and at state $j$ at time $t+1$, given the HMM model $(\Pi, A, f)$ and the observation sequence $O$.

$$q_t(i, j) = P(s_t = i, s_{t+1} = j|O, \Pi, A, f) = \frac{P(s_t = i, s_{t+1} = j, O|\Pi, A, f)}{P(O|\Pi, A, f)}$$

$$= \frac{\alpha_i(t)f_i(o_t)p_{ij}\beta_j(t + 1)}{\sum_{l=1}^{N}\alpha_l(t)\beta_l(t)} = \frac{\alpha_i(t)f_i(o_t)p_{ij}\beta_j(t + 1)}{\sum_{l=1}^{N}\sum_{m=1}^{N}\alpha_l(t)f_i(o_t)p_{lm}\beta_m(t + 1)}.$$

Then $\gamma_i(t)$ can be rewritten as

$$\gamma_i(t) = P(s_t = i|O, \Pi, A, f) = \sum_{j=1}^{N} P(s_t = i, s_{t+1} = j|O, \Pi, A, f) = \sum_{j=1}^{N} q_t(i, j).$$

The above equation can be expected as $\gamma_i(t)$ is the expected number of transition from state $i$ and $q_t(i, j)$ is the expected number of transitions from state $i$ to $j$. Given the above equality, we begin with an initial model $\Pi, A, f$ and calculate the likelihood of the observation sequence $O$ based on the current model and estimate the expectations of each model parameter. Then the model can be updated to maximize the values of the paths that are used. This process can be iterated until convergence to the optimal values of the model parameters.
3.3 Financial Asset Pricing Using HMM

In this chapter, I will use the same pack of data sets that I used in Chapter 2. I applied HMM in modeling the monthly S &P 500 rate of returns from 1959/02/01 to 2010/12/01. The ROR is defined as in Chapter 2:

\[
R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1 = \frac{(P_{t+1} - P_t) + D_{t+1}}{P_t}
\]

The continuous ROR is \( r_{t+1} \equiv \log(1 + R_{t+1}) \)

The monthly ROR and growth rate of consumption statistics is as follows:

<table>
<thead>
<tr>
<th>TABLE VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMARY STATISTICS OF MAJOR VARIABLES IN MONTHLY DATA</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( r_m )</td>
</tr>
<tr>
<td>( \Delta c )</td>
</tr>
</tbody>
</table>

Two segmentations are modeled. One is segmentation based on the mean, and the other is segmentation based on variance.
3.4 HMM Segmentation According to Mean

Six different HMM models are studied and compared. In some of these 6 cases which have covariation factors, I used the consumption growth rate and the risk free rate as covariates. Both of these two covariates are also used in Chapter 2’s asset pricing models to estimate the models' parameters. Consistency is maintained between Chapter 2 and Chapter 3. These two variables are serving as GMM’s instrumental variables. By fitting HMM on the hidden states, a model-based segmentation is obtained.

When adding the time-varying effects of the covariates to the probabilities of the transition matrix, a logistic regression is used. Here I take $P_{11}$ and $\mu_1$ as an example:

$$\log \left(\frac{P_{11}}{1 - P_{11}}\right) = \beta_{11,0} + \beta_{11,1} \times \Delta c + \beta_{11,2} \times r_f$$  \hspace{1cm} (3.2)

When taking the mean as a time-varying factor co-moving with the covariates, a regression model is applied. Here is the model:

For states $k=1,2,\ldots,K$,

$$\mu_k = \gamma_{0k} + \gamma_{1k} \times \Delta c_t + \gamma_{2k} \times r_{ft}$$  \hspace{1cm} (3.3)

For 2-state HMM, a time varying mean but constant transition matrix case has two means $\mu_1$ and $\mu_2$ to estimate; a time varying mean and time varying transition matrix case has two mean $\mu_1,\mu_2$ but four probabilities $P_{11}, P_{12}, P_{21}$ and $P_{22}$ to estimate.
For 3-state HMM, the former has three means $\mu_1, \mu_2$ and $\mu_3$ to estimate, the later has six probabilities $P_{11}, P_{12}, P_{13}, P_{21}, P_{22}, P_{23}, P_{31}, P_{32}, P_{33}$ to estimate.

The 6 HMM models based on mean segmentation are categorized into two forms, one is 2-state and the other is 3-state:

- 2-state HMM with constant transition matrix, normal density function for each state
- 2-state HMM with constant transition matrix, normal density function with mean depending on covariates
- 2-state HMM with changing transition matrix, normal density function with both transition matrix and normal mean depending on covariates

- 3-state HMM with constant transition matrix, normal density function for each state
- 3-state HMM with constant transition matrix, normal density function with mean depending on covariates
- 3-state HMM with changing transition matrix, normal density function with both transition matrix and normal mean depending on covariates

All 6 cases assume Normal distributions with constant standard deviation. However the mean of the distribution could be time varying and co-moving with the two covariates in different cases.

The statistical software R is used to estimate the model.
3.4.1 Two-state HMM

- constant transition matrix and constant mean

This is the simplest case, the estimates are as follows:

\[
\begin{align*}
\hat{\pi} &= (0, 1) \\
\hat{P} &= \begin{pmatrix} 0.9485 & 0.0515 \\ 0.1408 & 0.8592 \end{pmatrix} \\
\hat{\mu}_1 &= 0.03793 \\
\hat{\sigma}_1 &= 0.02611 \\
\hat{\mu}_2 &= 0.01482 \\
\hat{\sigma}_2 &= 0.05585
\end{align*}
\]

The estimation outputs show that the mean of state 1 ROR is 3.793%, the mean of state 2 ROR is 1.482%. Standard deviations of the two states are 0.0261 and 0.0558. The transition matrix tells us when today’s SP500 ROR is at state 1, there is 0.948 probability that tomorrow the SP500 will stay at state 1, while only 0.0515 probability it will switch to state 2 tomorrow. When today’s ROR is at state 2, there is 0.859 probability that tomorrow the ROR will stay at state 2, while 0.141 probability that will switch to state 1. These means the winner today has significant higher chance to win tomorrow, and the loser today also has significant higher chance to lose tomorrow.
constant transition matrix but time-varying mean

\[ \hat{\pi} = (0, 1) \]
\[ \hat{P} = \begin{pmatrix} 0.9209 & 0.0791 \\ 0.1196 & 0.8804 \end{pmatrix} \]
\[ \hat{\mu}_1 = 0.03741 \]
\[ \hat{\sigma}_1 = 0.02837 \]
\[ \hat{\mu}_2 = 0.02033 \]
\[ \hat{\sigma}_2 = 0.10588 \]

In this case, each of the two means of the assumed normal distributions of the unobservable states is following a logistic function of the two covariates which are the consumption growth rate and market risk free rate (10-year treasury bond rate). The mean of state 1 ROR is 3.741%, the mean of state 2 ROR is 2.033%. Standard deviations for the two states are 0.0283 and 0.1059. The transition matrix tells us when today’s SP500 ROR is at state 1, there is 0.921 probability that tomorrow the SP500 will stay at state 1, while only 0.0791 probability tomorrow it will switch to state 2. When today’s ROR is at state 2, there is 0.880 probability that tomorrow the ROR will stay at state 2, while 0.120
probability that will switch to state 1. The winner wins and loser loses trend maintains in this case.

- time-varying transition matrix and time-varying mean

\[
\hat{\pi} = (0, 1)
\]
\[
\hat{P} = \begin{pmatrix}
0.9210 & 0.0790 \\
0.0965 & 0.9035
\end{pmatrix}
\]
\[
\hat{\mu}_1 = 0.03960
\]
\[
\hat{\sigma}_1 = 0.02599
\]
\[
\hat{\mu}_2 = 0.01146
\]
\[
\hat{\sigma}_2 = 0.10388
\]

In this case, not only the means but also each of the four elements of the transition matrix is following a logistic regression of the same two covariates. The mean of state 1 ROR is 3.959% the mean of state 2 ROR is 1.146%. Standard deviations for the two states are 0.0259 and 0.104. The transition matrix tells us when today’s SP500 ROR is at state 1, there is 0.921 probability that tomorrow the SP500 will stay at state 1, while only 0.0790 probability tomorrow it will switch to state 2. When today’s ROR is in state 2, there is 0.904 probability that tomorrow the ROR will stay in state 2, while 0.096 probability
that will switch to state 1. As case 1 and case 2, this case has the same winner and loser pattern. Winner has much higher chance to win tomorrow and loser has much higher chance to lose tomorrow.

- Three cases comparison in 2-state HMM

**TABLE VIII**

THREE MODELS OF 2-STATE HMM MEAN AND VARIANCE COMPARISON

<table>
<thead>
<tr>
<th></th>
<th>$\mu_1$</th>
<th>$\sigma_1$</th>
<th>$\mu_2$</th>
<th>$\sigma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>model 1</td>
<td>3.793%</td>
<td>2.611%</td>
<td>1.482%</td>
<td>5.582%</td>
</tr>
<tr>
<td>model 2</td>
<td>3.741%</td>
<td>2.837%</td>
<td>2.032%</td>
<td>10.588%</td>
</tr>
<tr>
<td>model 3</td>
<td>3.960%</td>
<td>2.599%</td>
<td>1.146%</td>
<td>10.388%</td>
</tr>
</tbody>
</table>

**TABLE IX**

THREE MODELS OF 2-STATE HMM MEAN AND PROBABILITY COMPARISON

<table>
<thead>
<tr>
<th></th>
<th>Mean of state 1</th>
<th>mean of state 2</th>
<th>prob. of 1 to 1</th>
<th>prob. of 2 to 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>model 1</td>
<td>3.793%</td>
<td>1.482%</td>
<td>0.9485</td>
<td>0.8592</td>
</tr>
<tr>
<td>model 2</td>
<td>3.741%</td>
<td>2.032%</td>
<td>0.9209</td>
<td>0.8804</td>
</tr>
<tr>
<td>model 3</td>
<td>3.960%</td>
<td>1.146%</td>
<td>0.9210</td>
<td>0.9035</td>
</tr>
</tbody>
</table>
As the mean of the monthly SP500 ROR is 3.217%, for each case in 2-state HMM, the state 1’s mean is larger than the mean of the whole data set, so it is "Bull" market. And the state 2 is "Bear" market. In addition, the Bear market usually has higher volatility as indicated in the table of mean and variance comparison.

From the above comparison tables, we can see that model 3 has the highest mean at state 1 and model 2 has the highest mean at state 2. Model 1 has the highest probability in 1 to 1 case, and model 3 has the highest probability 3 to 3 case. In the 3rd table, using time varying factors makes tomorrow’s ROR has higher possibility to stay in the same state when today’s state is 2, but lower possibility to stay the same when today’s state is 1. A suddenly rise or suddenly drop in ROR out of its current state range is a rare event.

3.4.2 Three-state HMM

• constant transition matrix and constant mean The estimates are as follows:

\[
\hat{\pi} = (0, 1, 0)
\]

\[
\hat{P} = \begin{pmatrix}
0.9407 & 0.0070 & 0.0523 \\
0.2116 & 0.6156 & 0.1729 \\
0.1188 & 0.0442 & 0.8370
\end{pmatrix}
\]

\[
\hat{\mu}_1 = 0.04056
\]

\[
\hat{\sigma}_1 = 0.02603
\]

\[
\hat{\mu}_2 = 0.02967
\]
\[
\hat{\sigma}_2 = 0.08245 \\
\hat{\mu}_3 = 0.00697 \\
\hat{\sigma}_3 = 0.04055
\]

The estimation outputs show that the mean of state 1 ROR is 4.056%, the mean of state 2 ROR is 2.967%, the mean of state 3 ROR is 0.697%. Standard deviations for the three states are 0.026, 0.082 and 0.041 respectively. When today’s ROR is at state 1, there is 0.941 probability that tomorrow the ROR will stay at state 1, and the chance of switching to other two states are low. When today’s SP500 ROR is at state 2, the probability that tomorrow the SP500 ROR will stay at state 2 is 0.616, the probability that the ROR will switch to state 1 and state 3 are 0.212 and 0.173 separately. When today’s ROR is at state 3, there is 0.837 probability that tomorrow the ROR will stay at the same state, and there is 0.119 probability that it will switch to state 1. The results are consistent with the two-state HMM models that SP500 ROR has significant higher chance to stay in the same state. 1 to 1, 2 to 2 and 3 to 3 have the largest probability. 2 to 1, 2 to 3 and 3 to 1 will happen with possibilities falling into the range between 0.1 and 0.25. But 1 to 2, 1 to 3, 3 to 2 seem rarely happen based on the transition matrix.

- constant transition matrix but time-varying mean

\[
\hat{\pi} = (0, 1, 0)
\]
\[ \hat{\mu}_1 = 0.03593 \]
\[ \hat{\sigma}_1 = 0.02521 \]
\[ \hat{\mu}_2 = 0.01602 \]
\[ \hat{\sigma}_2 = 0.05200 \]
\[ \hat{\mu}_3 = 0.01328 \]
\[ \hat{\sigma}_3 = 0.06800 \]

In this case, each of the three means of the assumed Normal distributions of the unobservable states is following a logistic function of the two covariates which are the consumption growth rate and market risk free rate (10-year treasury bond rate). The mean of state 1 ROR is 3.593%, the mean of state 2 ROR is 1.602% and the mean of state 3 ROR is 1.328%. Standard deviations for the three states are 0.0252, 0.0520 and 0.0680 respectively. The transition matrix tells us when today’s SP500 ROR is at state 1, there is 0.923 probability that tomorrow the SP500 will stay at state 1, while the probabilities of switching to state 2 and state 3 are very small. When today’s ROR is at state 2, there is 0.7120 probability that tomorrow the ROR will stay at state 2, when there is 0.2524 probability that will switch to state 1. When ROR is at state 3 today, the probability of staying at state 3 is 0.8676, while the probabilities of switching to state 1 and state 2.
are very small. As in the first model of 3-state HMM, 1 to 1, 2 to 2 and 3 to 3 have the largest probability. 2 to 1 has the probability between 0.1 to 0.25. The other switching cases are not very possible to happen. The winner wins and loser loses trend maintains in this case.

• time-varying transition matrix and time-varying mean

\[
\hat{\pi} = (0, 1, 0)
\]

\[
\hat{P} = \begin{pmatrix}
0.9048 & 0.0486 & 0.0466 \\
0.0530 & 0.8964 & 0.0506 \\
0.0545 & 0.0472 & 0.8983
\end{pmatrix}
\]

\[
\hat{\mu}_1 = 0.03671
\]

\[
\hat{\sigma}_1 = 0.02986
\]

\[
\hat{\mu}_2 = 0.01315
\]

\[
\hat{\sigma}_2 = 0.12428
\]

\[
\hat{\mu}_3 = -0.01206
\]

\[
\hat{\sigma}_3 = 0.07657
\]

In this case, not only the means but also each of the four elements of the transition matrix vary, the latter according to a logistic function of the same two covariates. The mean of state 1 ROR is 3.671%, the mean of state 2 ROR is 1.315% and the mean of state 3
ROR is -1.206%. Standard deviations for the three states are 0.0299, 0.1243 and 0.0766 respectively. The transition matrix tells us when today’s SP500 ROR is at state 1, there is 0.9048 probability that tomorrow the SP500 will stay at state 1. When today’s ROR is at state 2, there is 0.8964 probability that tomorrow the ROR will stay at state 2. When ROR is at state 3 today, the probability of staying at state 3 is 0.8983. The probabilities of switching to the other states are all very small, no matter which state today’s ROR is. Winner has more chance win tomorrow and loser has more chance lose tomorrow. In addition, state switching is a small-probability event.

- Three cases comparison in 3-state HMM

<table>
<thead>
<tr>
<th></th>
<th>Mean of state 1</th>
<th>Mean of state 2</th>
<th>Mean of state 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>model 1</td>
<td>4.056%</td>
<td>2.967%</td>
<td>0.697%</td>
</tr>
<tr>
<td>model 2</td>
<td>3.593%</td>
<td>1.602%</td>
<td>1.328%</td>
</tr>
<tr>
<td>model 3</td>
<td>3.671%</td>
<td>1.315%</td>
<td>-1.206%</td>
</tr>
</tbody>
</table>

State 1’s mean is above the average of the monthly SP500 3.217%, while state 2 and state 3’s are both below the average. We can call state 1 ”Bull”, state 2 ”Contraction” and the state 3 ”Bear”. State 2 has the highest volatility and the state 1 has the lowest.
TABLE XI

VARIANCE OF THE THREE CASES IN 3-STATE HMM

<table>
<thead>
<tr>
<th></th>
<th>Variance of state 1</th>
<th>Variance of state 2</th>
<th>Variance of state 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>model 1</td>
<td>2.603%</td>
<td>8.245%</td>
<td>4.055%</td>
</tr>
<tr>
<td>model 2</td>
<td>2.521%</td>
<td>5.200%</td>
<td>6.800%</td>
</tr>
<tr>
<td>model 3</td>
<td>2.986%</td>
<td>12.428%</td>
<td>7.657%</td>
</tr>
</tbody>
</table>

TABLE XII

PROBABILITY OF MAINTAINING THE SAME STATE IN 3-STATE HMM

<table>
<thead>
<tr>
<th></th>
<th>1 to 1</th>
<th>2 to 2</th>
<th>3 to 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>model 1</td>
<td>0.9407</td>
<td>0.6156</td>
<td>0.8370</td>
</tr>
<tr>
<td>model 2</td>
<td>0.9225</td>
<td>0.7120</td>
<td>0.8676</td>
</tr>
<tr>
<td>model 3</td>
<td>0.9048</td>
<td>0.8964</td>
<td>0.8983</td>
</tr>
</tbody>
</table>

From the above comparison tables, we can see that with regard to state 2 and state 3, model 2 has lower mean than model 1 and model 3 has lower mean than model 2. Model 1 has the highest probability in 1 to 1 case, model 3 has highest probability in 2 to 2 and 3 to 3 cases. In the 3rd table, using time varying factors makes tomorrow’s ROR has higher possibility to stay in the same state as today. A sudden rise or sudden drop in ROR is a rare event.
TABLE XIII

OVERVIEW OF THE TRANSITION MATRIX IN THE 3-STATE HMM

<table>
<thead>
<tr>
<th></th>
<th>elements prob. &gt; 0.7</th>
<th>elements 0.25 &gt; prob. &gt; 0.1</th>
<th>elements prob. &lt; 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>model 1</td>
<td>1 to 1 2 to 2 3 to 3</td>
<td>2 to 1 2 to 3 3 to 1</td>
<td>all others</td>
</tr>
<tr>
<td>model 2</td>
<td>1 to 1 2 to 2 3 to 3</td>
<td>2 to 1</td>
<td>all others</td>
</tr>
<tr>
<td>model 3</td>
<td>1 to 1 2 to 2 3 to 3</td>
<td>none</td>
<td>all others</td>
</tr>
</tbody>
</table>

3.4.3 **2-states HMM vs 3-states HMM**

The Bayesian Information Criterion (BIC) method is used to compare these two HMM models. The formula for BIC is, for alternative models indexed by k=1,2,...,K,

\[
BIC_k = -2 \times LL_k + \log(n) \times m_k
\]

where

\[
L_k = \text{maximized likelihood for Model } k,
\]

\[
LL_k = \text{natural log of } L_k,
\]

\[
m_k = \text{number of independent parameters in Model } k,
\]

and \( n \) is the sample size. Note that \( m_k = 7, 11, 15, 14, 20, 32 \) for HMM1-HMM6, respectively.

Here are the values of BIC for the six HMM models:
The smaller the BIC statistic, the better the fitted model is. The above summary of the BIC statistics tells us the most sophisticated case with time varying mean and time varying transition matrix (model 3) has the largest BIC, in both the 2-state HMM and the 3-state HMM. The constant mean and transition matrix case (model 1) has the smallest BIC value. For each case, the 3-state HMM has bigger BIC than the 2-state HMM. Thus the 2-state HMM model with constant mean and constant transition matrix is the best model.

3.5 HMM Segmentation According to Variance

The above section, I used HMM to model the SP500 monthly data by letting the ”hidden” states vary in the mean. In this section, I will let the ”hidden” states differ with respect to the variance and keep the mean fixed at the average of the data set. Only two models are studied here, one with two states and the other with three states. Both are with constant mean and transition matrix.

- 2-states HMM based on variance segmentation
This is the simplest case, the estimates are as follows:

\begin{align*}
\hat{\pi} &= (0, 1) \\
\hat{P} &= \begin{pmatrix}
0.9598 & 0.0402 \\
0.1092 & 0.8908 \\
\end{pmatrix} \\
\hat{\mu} &= 0.0347 \\
\hat{\sigma}_1 &= 0.0260 \\
\hat{\sigma}_2 &= 0.0583
\end{align*}

In 2-state HMM, state 1’s volatility is 2.60% and state 2’s volatility is 5.83%. The mean is fixed to the average of the SP500 which is 3.45%. The state 1 is superior to the state 2, as it has smaller variance than state 2 does when gaining the same rate of return. The transition matrix tells us there is much higher probability of staying in the same state than moving to the other.

- 3-states HMM based on variance segmentation

\begin{align*}
\hat{\pi} &= (0, 1, 0) \\
\hat{P} &= \begin{pmatrix}
0.9278 & 0.0588 & 0.0135 \\
0.0514 & 0.8552 & 0.0934 \\
0.0769 & 0.1310 & 0.7922 \\
\end{pmatrix} \\
\hat{\mu} &= 0.0345
\end{align*}
In 2-state HMM, state 1’s volatility is 2.17%, state 2’s volatility is 3.34% and state 3’s volatility is 6.31%. The mean is fixed to the average of the SP500 which is 3.45% as well. The state 1 is superior to the state 2 and the state 2 is superior to the state 3. The state 1 has the smallest variance and the state 3 has the largest variance when gaining the same rate of return.

- 2-state and 3-state HMM segmented by variance comparison

The BIC statistics for 2-state and 3-state HMM are -2384.506 and -2345.102. The 2-state HMM segmented according to variance is better than the 3-state one. The standard deviation of each state is as below:

<table>
<thead>
<tr>
<th>TABLE XV</th>
<th>STANDARD DEVIATION COMPARISON</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>state 1</td>
</tr>
<tr>
<td>2-state HMM</td>
<td>2.60%</td>
</tr>
<tr>
<td>3-state HMM</td>
<td>2.17%</td>
</tr>
</tbody>
</table>
3.6 Posterior Probabilities from BIC and HMM Model Comparison

Posterior Probabilities based on BIC values provide another angle to compare the goodness
of the models.

Suppose that there are $K$ alternative models, indexed by $k = 1, \ldots, K$. Let $L_k$ denote the
likelihood for model $k$. Then

$$BIC_k = -2 \times \log L_k + \log(n) \times m_k.$$ 

Let $p_k$ denote the prior probability of model $k$, and then the posterior probability of model $k$ is

$$\Pr(\text{model } k \mid \text{data}) = \text{Const} \cdot p_k L_k,$$

$$-2 \log \Pr(\text{model } k \mid \text{data}) = -2 \log \text{Const} - 2 \log p_k - 2 \log L_k.$$

Also, since $-2 \log \Pr(\text{model } k \mid \text{data}) \approx \text{Const} \cdot BIC_k$,

$$\Pr(\text{model } k \mid \text{data}) \approx \text{Const} \cdot \exp(-BIC_k/2), \quad k = 1, \ldots, K.$$

In calculating $\Pr(\text{model } k \mid \text{data})$, we first calculate $\exp(-BIC_k/2)$ for each $k$, and then find
Const. by normalizing the sum of $\Pr(\text{model } k \mid \text{data})$ to 1. The details are presented in Table
below.

The result is consistent with the BIC statistics result. The 2-state HMM segmented according
to mean with constant transition matrix and mean of the density function is the best.
### TABLE XVI

COMPUTATION OF POSTERIOR PROBABILITIES

<table>
<thead>
<tr>
<th>Model k, State i</th>
<th>( BIC_k )</th>
<th>post. prob. of Model k</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>segmentation according to mean</td>
<td></td>
</tr>
<tr>
<td>1,2</td>
<td>-2390.993</td>
<td>0.962</td>
</tr>
<tr>
<td>2,2</td>
<td>-2300.694</td>
<td>0.000</td>
</tr>
<tr>
<td>3,2</td>
<td>-2292.286</td>
<td>0.000</td>
</tr>
<tr>
<td>4,3</td>
<td>-2354.055</td>
<td>0.000</td>
</tr>
<tr>
<td>5,3</td>
<td>-2295.766</td>
<td>0.000</td>
</tr>
<tr>
<td>6,3</td>
<td>-2154.691</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>segmentation according to variance</td>
<td></td>
</tr>
<tr>
<td>7,2</td>
<td>-2384.506</td>
<td>0.038</td>
</tr>
<tr>
<td>8,3</td>
<td>-2345.102</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### 3.7 A HMM Case Study of FTSE/Xinhua China A50 Index

When China is becoming more and more important to the world economy, a HMM case study on China’s equity index will be interesting. Like SP500 which is considered a bellwether for the American economy, FTSE/Xinhua China A50 Index is considered a bellwether for the China economy. In the above several sections, I have already used HMM to fit the SP500 index.

The data I will use on this study is the daily data of FTSE/Xinhua China A50 index form 16-Nov-04 to 28-Sep-11. Two kinds of segmentations are applied, one is segmentation based on mean and the other is segmentation based on variance. 2-state and 3-state models are used for each of the segmentations.

- 2-states HMM based on mean segmentation
This is the simplest case, the estimates are as follows:

\[
\hat{\pi} = (0, 1)
\]

\[
\hat{P} = \begin{pmatrix}
0.8209 & 0.1791 \\
0.3978 & 0.6022
\end{pmatrix}
\]

\[
\hat{\mu}_1 = -0.0014
\]

\[
\hat{\sigma}_1 = 0.0118
\]

\[
\hat{\mu}_2 = 0.0012
\]

\[
\hat{\sigma}_2 = 0.0318
\]

State 1 is Bear state with -0.0014 as its mean and state 2 is Bull state with 0.0012 as its mean. The probability transition matrix tells us when the China equity market is in Bear state today, it will have around 0.82 probability to be in Bear state tomorrow; while when the Chine equity market is in Bull state today, it will have only around 0.60 probability to stay in Bull state tomorrow. A Bull market today has 40% possibility to move to Bear market tomorrow.

- 3-states HMM based on mean segmentation

\[
\hat{\pi} = (0, 1, 0)
\]
\[ \hat{P} = \begin{pmatrix} 0.8396 & 0.1296 & 0.0307 \\ 0.0336 & 0.9456 & 0.0208 \\ 0.0353 & 0.2977 & 0.6670 \end{pmatrix} \]

\[ \hat{\mu}_1 = 0.0019 \]

\[ \hat{\sigma} = 0.0310 \]

\[ \hat{\mu}_2 = -0.0017 \]

\[ \hat{\sigma} = 0.0127 \]

\[ \hat{\mu}_3 = 1.0609 \]

\[ \hat{\sigma} = 0.1147 \]

In this 3-state HMM, state 2 is the Bear state with mean -0.0017 and state 3 is Bull state with mean 1.0609. Similar story applies to 3-state HMM as to 2-state HMM. When China market is in Bear state today, it will have around 0.95 probability to stay in Bear state; when China market is in Bull state, it will have 0.30 probability to switch to Bear state tomorrow.

- 2-states HMM based on variance segmentation

This is the simplest case, the estimates are as follows:

\[ \hat{\pi} = (0, 1) \]
\[ \hat{P} = \begin{pmatrix} 0.7179 & 0.2821 \\ 0.4145 & 0.5855 \end{pmatrix} \]

\[ \hat{\mu} = 0.0347 \]

\[ \hat{\sigma}_1 = 0.0260 \]

\[ \hat{\sigma}_2 = 0.0583 \]

In this variance segmentation HMM, mean is fixed. State 2 has higher volatility than state 1 but has the same mean. The probability of staying in state 1 is 0.7179 and the probability of staying in state 2 is 0.5855. Switching to the other states is much more probable in the 2-state HMM based on variance segmentation.

• 3-states HMM based on variance segmentation

\[ \hat{\pi} = (0, 1, 0) \]

\[ \hat{P} = \begin{pmatrix} 0.9863 & 0.0133 & 0.0004 \\ 0.0192 & 0.9803 & 0.0005 \\ 0.1514 & 0.8461 & 0.0025 \end{pmatrix} \]

\[ \hat{\mu} = -0.0005 \]

\[ \hat{\sigma}_1 = 0.0125 \]

\[ \hat{\sigma}_2 = 0.0275 \]

\[ \hat{\sigma}_3 = 0.0096 \]
In the 3-state variance segmentation HMM, state 2 has the highest volatility with 98% probability of staying in the same state tomorrow. State 3 has the lowest volatility with a near 0 probability to stay in the same state tomorrow. This model tells us state 3 is a very unstable state: when the market hits this state, we can expect tomorrow it will almost certainly move back to the other two higher volatility states.

- BIC Comparison

<table>
<thead>
<tr>
<th>State</th>
<th>$BIC_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>segmentation according to mean</td>
<td></td>
</tr>
<tr>
<td>2 state</td>
<td>-8243.69</td>
</tr>
<tr>
<td>3 state</td>
<td>-8173.29</td>
</tr>
<tr>
<td>segmentation according to variance</td>
<td></td>
</tr>
<tr>
<td>2 state</td>
<td>-8227.84</td>
</tr>
<tr>
<td>3 state</td>
<td>-8303.97</td>
</tr>
</tbody>
</table>

According to the BIC criterion, the best model is 3-state HMM based on variance segmentation. This result is different from HMM on US SP500 index which gave us 2-state HMM based on mean segmentation as the best model.

The HMM case study on China A50 index gives us some interesting points:
• Compared to the SP500 index, the China A50 index seems easier to switch from Bull market today to Bear market tomorrow.

• The China A50 index is much harder to stay in a state with lowest volatility.

• While "the simpler the model is, the better it is" happens on SP500, the reverse situation happens with the China A50.
CHAPTER 4

CONCLUSION

My thesis, on asset pricing topics, contains three chapters. Chapter 1 is an overview and presents recent development of the financial economics field. The difference between asset returns study and asset prices study is discussed and the ‘present value’ approach and ‘investor preference’ approach are introduced. The ‘present value’ approach brings the time value of money over long period of time into consideration to price assets. And the ‘investor preference’ approach also called consumption growth approach is targeting in linking the investors’ preferences and decisions with the asset prices. Furthermore, the topics on equity premium puzzle and Epstein-Zin investor preference function are introduced. The puzzle is that when using a power utility preference function to price the stock log return under the consumption based asset pricing setting, the very low historical covariance between stock return and consumption growth will imply an unreasonable large coefficient of relative risk aversion for investors in US and thus will imply a much higher risk free rates than the true risk free rates in the markets.

In Chapter 2, I construct a long run risk model deriving from Bansal and Yaron(2004), in which Epstein-Zin preference is incorporated. I show that two shocks, the consumption growth shock and the dividend growth shock have an amplified and persistent effect on the asset price and asset return as well as on the dividend growth rate and consumption growth rate. By applying GMM methodology and using yearly and monthly SP500 stock data and the personal consumption expenditure data from the Fed’s database, I estimate all the parameters, and get...
the intertemporal elasticity of substitution to be about 3.5 and the relative risk aversion (RRA) to be about 1.57. The substitution effect dominates the income effect of higher returns which means today’s consumption level relative to total wealth falls when the expected rate of asset return rise. At last, based on my model’s calculation the equity premium is 4.5% which is quite near the true equity premium 5.4%. Comparing with the original power utility model which gives a negative equity premium when the RRA is 1.57 and a premium equal to 2.5% when the RRA is 10, my model’s improvement is significant.

Another important part of my thesis concerns the Hidden Markov Model. I proposed modeling the observable stock price movements in terms of the hidden states of the HMMs and each of the states is associated with a probability distribution which governs the distribution of the real stock prices. Eight models are fitted here.

Six are based on mean segmentation and they are: 2-state HMM with constant transition matrix, and constant normal density function for each state; 2-state HMM with constant transition matrix, normal density function with mean depending on covariates; 2-state HMM with changing transition matrix, normal density function with both transition matrix and normal mean depending on the covariates. 3-state HMM with constant transition matrix, and constant normal density function for each state; 3-state HMM with constant transition matrix, normal density function with mean depending on covariates; 3-state HMM with changing transition matrix, normal density function with both transition matrix and normal mean depending on the covariates. The two covariates are the consumption growth rate and the risk free rate which are used to estimate the model in chapter 2 as well. The empirical results show that no matter
whether it is a 2-state HMM or a 3-state HMM, buying winners and selling losers is a right strategy. Because the estimated transition probability matrix reveals much higher probabilities of staying at the same state as today for tomorrow but very low probabilities of switching to other states. The BIC statistics are calculated; and indicate that the 2-state HMM with constant transition matrix and constant mean of the state's density function is the best model.

Two models are based on segmentation according to variances alone and they are: 2-state and 3-state with constant transition matrix and mean. The results show that with the same mean, the state 1 in 2-state HMM and the state 1 in 3-state HMM have the lowest volatility, so the state 1 can be called the min-variance state. BIC statistics pick the 2-state HMM as the best segmentation HMM according to variance.

In addition, the posterior probabilities based on BIC statistics, are calculated for all of the eight models. The result is consistent with that of the BIC statistics.

At last, I applied HMM on a case study of the China A50 index which is considered as bellwether for the China economy. Some interesting results different from those on the SP500 have been found.
CITED LITERATURE


VITA

Ph.D. in Business Administration, University of Illinois, Chicago 2009 - 2012

- Advisor: Dr. Stanley Sclove

M.S. in Statistics, Columbia University, New York City, 2006 - 2008

B.S. in Economics, Peking University, Beijing 2002 - 2006