ESSAYS ON VOLATILITY RISK, ASSET RETURNS
AND CONSUMPTION-BASED ASSET PRICING

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By
Young Il Kim, M.A.

The Ohio State University
2008

Dissertation Committee:
J. Huston McCulloch, Advisor
Paul Evans
Pok-sang Lam

Approved by

_________________________ Advisor
Economics Graduate Program
© Copyright by

Young Il Kim

2008
ABSTRACT

My dissertation studies and addresses two main issues regarding asset returns: an econometric modeling framework for asset returns in two essays and puzzling features of the standard consumption-based pricing model in two other essays. The first two essays provide a new empirical guidance for modeling a skewed and thick-tailed error distribution along with GARCH effects for financial data with application to U.S. stock returns based on the theoretical derivation for the GARCH-skew-t model. The third essay addresses asset return puzzles such as high equity premium and low riskfree rate associated with the standard consumption-based asset pricing model. The fourth essay investigates a possible link between stock market volatility and macroeconomic risk within a general equilibrium asset pricing framework.

The first essay, “The Skew-Student Distribution with Application to U.S. Stock Market Returns and the Equity Premium,” provides a new theoretical derivation for the skew-Student (skew-t) distribution and applies it to an 80-year sample of U.S. monthly excess stock market returns. This essay derives the skew-t as a normal variance-mean mixture distribution \( N(\mu + \beta \tilde{V}, \tilde{V}) \) with the latent variance \( \tilde{V} \) drawn from the inverse-chi-square distribution. The skew-t is same as the Student t if \( \beta = 0 \). The skew-t can capture the substantially skewed fat-tails that characterize U.S. stock data. It is found that a time-
dependent skew-t model represents the excess returns very accurately when GARCH effects are incorporated. Based on the estimated skew-t distribution, the equity premium is computed. This essay shows that skewness and kurtosis can have significant effect on the equity premium and that with sufficiently negatively skewed distribution of the excess returns, a finite equity premium can be assured, contrary to the case of the Student t in which an infinite equity premium arises.

The second essay, “Realized Volatility and Modeling Stock Returns as A Mixture of Normal Random Variables: The GARCH-Skew-t Model,” relates the skew-t distribution to Realized Volatility ($RV$) measures, constructed from the summation of higher-frequency squared (demeaned) returns. This essay provides a new empirical guidance for modeling a skewed and thick-tailed error distribution along with GARCH effects for financial data. Based on an 80-year sample of U.S. daily stock market returns, it is found that the distribution of monthly $RV$ conditional on past returns is approximately the inverse-chi-square. It is also found that monthly market returns, conditional on $RV$ and past returns are normally distributed with $RV$ in both mean and variance $N(\mu + \beta RV, RV)$. These empirical findings serve as the building blocks for the distributional assumption underlying the GARCH-skew-t model, which is derived in the first essay. Thus, it provides a new empirical support for the GARCH-skew-t modeling of equity returns. It is shown that the implied GARCH-skew-t model adequately describes the pattern of U.S. stock market returns. Moreover, the GARCH-skew-t accurately represents the three important stylized facts of stock market returns: volatility clustering, fat-tails and negative skewness.
The third essay, “Solving Asset Return Puzzles By Force of Parameter Uncertainty: A Mixture-of-distributions Approach With the NIG Distribution” provides a potential solution to asset return puzzles such as high equity premium and low riskfree rate for the standard consumption-based asset pricing model. This essay shows that a negatively skewed and thick-tailed error distribution underlying a perceived consumption endowment process can explain a 6-8% equity premium and 1-2% riskfree rate even with the degree of risk aversion below 10 in the CRRA utility function. This essay considers parameter uncertainty underlying the Data Generating Process (DGP) as a main source of negatively skewed and leptokurtic distribution. Thus, a consumer investor with uncertainty about the DGP may perceive the consumption growth as drawn from a negatively skewed and fat-tailed probability distribution so that he/she may require high risk premium for holding equity while accepting a low riskfree rate.

The last essay, “Stock Market Volatility And Macroeconomic Risk: Asset Pricing Under Two Factor Production Economy,” investigates a possible link between stock market and macroeconomic volatility. We have observed in the mid-1980s that U.S. economy experienced a great moderation in macroeconomic volatility in variables such as output growth and consumption. But during the same period, we did not experience a similar moderation in stock market volatility. The stock market volatility in the face of macroeconomic moderation is a conundrum for the standard consumption-based asset pricing model. This essay explains this conundrum by decomposing the source of consumption into payouts from equity and non-equity assets each of which serves as a factor of production. Thus, stock price and dividend are distinguished from total wealth and consumption respectively. The model can predict the high volatility of stock return,
smooth consumption growth and their weak correlation at the same time. It is shown that the great moderation of macroeconomic risk must have originated from declining risk of the relatively large non-equity assets while the relatively small equity-based income has been persistently volatile during the moderation era. Furthermore, the model explicitly shows that the consumption risk of equity-holding is positively associated with the stock share of total wealth.
To my parents
ACKNOWLEDGMENTS

I would like to thank my advisor, J. Huston McCulloch, for his teaching, advice, support and enthusiasm which have greatly contributed to my intellectual growth as a professional economist. He has guided me throughout my PhD years ever since the first year statistics course in the graduate program. As a student, I have learned importance of data, technical skills to deal with them, and how to think about or approach important economic issues, especially those of this dissertation. Chapter 2 and 5 of this dissertation are fruitful outcomes of intensive intellectual interaction with Dr. McCulloch.

I am also indebted to Paul Evans for his teaching, advice and intellectual support. I learned greatly not only from his lectures but also from his comments and conversations during seminars in the department of economics. I was quite amused by listening to his reaction to and his conversation with seminar speakers; in this respect, Dr. Evans provided me with great motivation for attending macro seminars. My appreciation also goes to Pok-sang Lam who first introduced and discussed important issues of consumption-based asset pricing models in the first year graduate course which two chapters of my dissertation deal with. My thanks also go to Lucia Dunn for her guidance to be a better teacher. I also wish to thank Hajime Miyazaki for his thoughtful guidance since I began my PhD study at the Ohio State University. I would also like to express my
sincere appreciation to Andrew P. Blake for his warm and intellectual encouragement during my PhD internship in Bank of England.

I appreciate financial supports from the Department of Economics at the Ohio State University in the form of the Graduate Teaching Associateship during my PhD study. My research also benefited from the JMCB travel grant for presentation in several conferences. I also thank the seminar invitation from the Department of Economics at Bowling Green State University. My dissertation also benefited from Bank of England in the form of PhD internship. During this period, I initiated two research projects both of which form two chapters of this dissertation: chapter 3 and 4.

My special thanks go to my mother who has shown me endless support and love throughout my entire life. I would also like to thank my father who takes care of me from the heaven even though I have no way to return even a part of what I have received from him. I also thank my brother and sister-in-law for their support and warm encouragement. Any words of expression may not be sufficient enough to describe their love and support for me.
VITA

July 25, 1975…………………………... Born – Kimcheon, Kyungbuk, Korea

2003……………………………………… B.S. Physics, Korea University, Korea

2003……………………………………… B.A. Economics, Korea University, Korea


2003-present…………………………... Graduate Teaching Associate,
The Ohio State University, U.S.A.

FIELDS OF STUDY

Major Field: Economics

Studies in:
   Money and Macroeconomics
   Econometrics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vii</td>
</tr>
<tr>
<td>Vita</td>
<td>ix</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xiii</td>
</tr>
<tr>
<td>List of Graphs</td>
<td>xv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xvi</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. THE SKEW-STUDENT DISTRIBUTION WITH APPLICATION TO U.S. STOck MARKET RETURNS AND THE EQUITY PREMIUM</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2.2 The skew-Student distribution and the moment generating function</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Data and estimation of the skew-Student distribution</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1 Data</td>
<td>13</td>
</tr>
<tr>
<td>2.3.2 Estimation</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Discussion of the goodness of fit</td>
<td>16</td>
</tr>
<tr>
<td>2.5 The equity premium</td>
<td>19</td>
</tr>
<tr>
<td>2.6 Dynamic skew-Student distribution</td>
<td>22</td>
</tr>
<tr>
<td>2.6.1 Extension to a dynamic model</td>
<td>22</td>
</tr>
<tr>
<td>2.6.2 GARCH models with alternative distributional assumptions</td>
<td>25</td>
</tr>
<tr>
<td>2.6.2.1 GARCH-t model as a conditional normal variance mixture</td>
<td>26</td>
</tr>
<tr>
<td>2.6.2.2 GARCH-N model</td>
<td>27</td>
</tr>
<tr>
<td>2.6.3 Estimation</td>
<td>27</td>
</tr>
<tr>
<td>2.6.4 Conditional variance</td>
<td>29</td>
</tr>
<tr>
<td>2.6.5 Distribution of GARCH-standardized residuals</td>
<td>29</td>
</tr>
<tr>
<td>2.6.6 Analysis of residuals for alternative distributions</td>
<td>29</td>
</tr>
<tr>
<td>2.6.7 Conditional equity premium</td>
<td>30</td>
</tr>
<tr>
<td>2.7 Conclusion</td>
<td>31</td>
</tr>
</tbody>
</table>
5.4.1 Model implication ................................................................. 138
5.4.2 Empirical evidence ................................................................. 139
5.5 Stock market capitalization and the systematic risk of equity-holding .... 141
5.6 Conclusion .................................................................................. 142

6. CONCLUDING REMARKS ................................................................. 153

Appendices:

A. DERIVATIONS FOR CHAPTER 2 ......................................................... 157

A.1 Derivation of the skew-Student density function ................................. 157
A.2 Derivations for the mean, variance, skewness and kurtosis of the skew-
Student distribution ........................................................................... 158
A.3 Derivation of the MGF of the skew t distribution ................................. 160
A.4 Derivation of the CDF and numerical approximation for the skew Student
distribution ....................................................................................... 161
A.5 The MGF, CGF, moments, and cumulants in general ............................. 162

B. INVERTED CHI-SQUARE DISTRIBUTION AND THE CDF .............. 163

C. NOTES FOR CHAPTER 4 ................................................................. 165

C.1 Cumulant Generating Function (CGF) and cumulants ........................... 165
C.2 The MGF (CGF) for Gaussian distribution .......................................... 166

D. DERIVATIONS FOR CHAPTER 5 ......................................................... 167

D.1 Price-dividend ratio under iid shocks to each pay-out (stock and non-
stock) ............................................................................................... 167
D.2 Computing $Var_{t}(\log C_{t+1})$, $Var_{t}(\log \varepsilon_{x,t+1})$ and $Cov_{t}(\log C_{t+1}, \log \varepsilon_{x,t+1})$... 168

Bibliography ......................................................................................... 171
LIST OF TABLES

Table                  Page

2.1 Parameter estimates and the annualized mean of excess log returns for alternative
distributional models……………………………………………………………………….33

2.2 Mean, standard deviation, skewness and kurtosis of monthly excess log returns
for alternative estimated distributional models against those of the empirical
distribution………………………………………………………………………………34

2.3 Equity premium under alternative distributions……………………………………34

2.4 Cumulant-based decomposition of the equity premium for Gaussian and the skew
t distribution…………………………………………………………………………34

2.5 Estimation of GARCH models with alternative distributional assumptions:
Gaussian, symmetric t and skew-t………………………………………………………35

2.6 Mean, standard deviation, skewness and kurtosis of each estimated GARCH-
standardized error distribution in comparison with summary statistics of
residuals………………………………………………………………………………36

2.7 Annualized average equity premium under alternative distributions based on the
long sample period (1926:1-2003:12)………………………………………………36

3.1 Estimation and statistical properties of alternative models……………………79

3.2 Ljung-Box portmanteau statistics for the realized volatility $RV_i$ and the scale-
adjusted realized volatility $RV_i / h_i^2$ with p-values in parentheses………………80

3.3 Parameter estimates of Gaussian, symmetric t, and skew t distribution with
standard errors in parentheses……………………………………………………80

3.4 Mean, standard deviation, skewness and kurtosis of each estimated distributional
model for monthly returns in comparison with those of the sample………………80
3.5 Estimation and statistical properties of alternative models..............................81
3.6 Ljung-Box portmanteau statistics for $R_V$ and the scale-adjusted $R_V$ with p-values in parentheses.................................................................83
3.7 Estimation of GARCH models with alternative distributional assumptions: Gaussian, symmetric t and skew-t.................................................................83
3.8 Parameter estimates of the GARCH (1,1) model with Gaussian and Student t error distribution respectively..........................................................84
3.9 Parameter estimates of Gaussian, Student t and the skew-Student distribution for GLS standardized residuals $(r_t - \mu_{GLS})/h_t$ with the log-likelihood and model selection criteria of the AIC and BIC..........................................................84
3.10 Mean, standard deviation, skewness and kurtosis of GLS residuals for estimated alternative distributional models against the sample summary statistics..........85
4.1 Price dividend ratio, expected return, riskfree rate and equity premium under the NIG and Gaussian distribution..........................................................116
4.2 Asset-pricing features corresponding to statistical properties of parameter uncertainty under the CRRA utility function........................................116
4.3 Asset-pricing features corresponding to statistical properties of parameter uncertainty under the EZ utility function........................................117
5.1 Variance and covariance of productivity shocks corresponding to the historical variance and covariance of stock returns and consumption growth for different degrees of substitutability $\sigma$ between factors of production at $K = 0.03$ ........144
5.2 Change in the standard deviation of consumption growth and stock returns in response to 25% decline of productivity shocks to the non-stock factor (asset)..........................................................144
<table>
<thead>
<tr>
<th>Graph</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Standard deviation of consumption growth $\text{sig}(g)$, standard deviation of stock returns $\text{sig}(\log R)$, their correlation $\text{corr}(g, \log R)$ and the covariance $\text{cov}(g, \log R)$ (with the benchmark case of $K = 0.03$ indicated by the vertical line) for $\rho = 0.8$.</td>
</tr>
<tr>
<td>5.2</td>
<td>Standard deviation of consumption growth $\text{sig}(g)$, standard deviation of stock returns $\text{sig}(\log R)$, their correlation $\text{corr}(g, \log R)$ and the covariance $\text{cov}(g, \log R)$ (with the benchmark case of $K = 0.03$ indicated by the vertical line) for $\rho = -2.5$.</td>
</tr>
<tr>
<td>5.3</td>
<td>The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 1$, $\sigma_\beta = 0.15$ and $\sigma_y = 0.01$.</td>
</tr>
<tr>
<td>5.4</td>
<td>The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 0.5$, $\sigma_\beta = 0.15$ and $\sigma_y = 0.01$.</td>
</tr>
<tr>
<td>5.5</td>
<td>The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 0.8$, $\sigma_\beta = 0.12$, $\sigma_y = 0.012$ and $\sigma_{\beta y} = 0.12 \times 0.012 \times 0.05$.</td>
</tr>
<tr>
<td>5.6</td>
<td>The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 0.8$, $\sigma_\beta = 0.12$, $\sigma_y = 0.012$ and $\sigma_{\beta y} = 0.12 \times 0.012 \times 0.2$.</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Normal QQ-plot of excess log returns over the postwar period.</td>
<td>37</td>
</tr>
<tr>
<td>2.3</td>
<td>Estimated density functions for the alternative distributional models.</td>
<td>38</td>
</tr>
<tr>
<td>2.4</td>
<td>Estimated density functions for the alternative distributional models.</td>
<td>38</td>
</tr>
<tr>
<td>2.5</td>
<td>Empirical and estimated CDF’s for alternative distributions in logarithmic (base 10) scale over ordered sample data.</td>
<td>39</td>
</tr>
<tr>
<td>2.6</td>
<td>Empirical and estimated (1-CDF)'s for alternative distributions in logarithmic (base 10) scale over ordered sample data.</td>
<td>39</td>
</tr>
<tr>
<td>2.7</td>
<td>PP-plot of Gaussian distribution for excess log returns.</td>
<td>40</td>
</tr>
<tr>
<td>2.8</td>
<td>PP-plot of the symmetric t distribution for excess log returns.</td>
<td>40</td>
</tr>
<tr>
<td>2.9</td>
<td>PP-plot of the skew t distribution for excess log returns.</td>
<td>41</td>
</tr>
<tr>
<td>2.10</td>
<td>Absolute demeaned log excess returns and the GARCH standard deviations.</td>
<td>41</td>
</tr>
<tr>
<td>2.11</td>
<td>QQ-plot of GARCH standardized residuals over the entire sample period.</td>
<td>42</td>
</tr>
<tr>
<td>2.12</td>
<td>Empirical and estimated CDF’s for the alternative distributions in logarithmic (base 10) scale over the sample quantiles.</td>
<td>42</td>
</tr>
<tr>
<td>2.13</td>
<td>Empirical and estimated (1-CDF)'s for the alternative distributions in logarithmic (base 10) scale over the sample quantiles.</td>
<td>43</td>
</tr>
<tr>
<td>2.14</td>
<td>PP-plots for Gaussian and Student t distributions.</td>
<td>43</td>
</tr>
</tbody>
</table>
2.15 PP-plots for the skew-Student distribution………………………………………….44
2.16 The conditional equity premium computed under the dynamic skew-t model and
the GARCH(1,1) with Gaussian error………………………………………………..44
3.1 Monthly value-weighted NYSE returns and the realized standard deviations
\(\sqrt{RV_i}\) from Jan. 1926 to Dec. 2005…………………………………………..86
3.3 Gaussian PP-plots of normalized returns from each model……………………..87
3.4 Histogram and the estimated inverted chi-square density function are shown on
the left panel while the inverted-chi-square PP-plot is on the right panel for the
monthly realized volatilities from Jan. 1966 to Dec. 2005…………………………87
3.5 PP-plots of estimated Gaussian, Student t and the skew-t distribution…………...88
3.6 Tail probability distribution of returns with CDF and 1-CDF in logarithmic scale
(base 10)…………………………………………………………………………..88
3.7 QQ-plot of monthly U.S. stock market returns from Jan. 1926 to Dec. 2005……89
3.8 Gaussian PP-plots of normalized returns for each alternative model……………89
3.9 Histogram and the inverted chi-square PP-plot for the scale-adjusted realized
volatilities \(RV_i / h_i^2\) from Jan. 1926 to Dec. 2005………………………………..90
3.10 Histogram and the inverted chi-square PP-plot for the scale-adjusted realized
volatilities \(RV_i / h_i^2\) from Jan. 1966 to Dec. 2005………………………………90
3.11 Histogram and the inverted chi-square PP-plot for the scale-adjusted realized
volatilities \(RV_i / h_i^2\) from Jan. 1976 to Dec. 2005………………………………90
3.12 QQ-plot of GARCH(1,1)-standardized residuals…………………………………….91
3.13 PP-plots of Gaussian, Student t and the skew-Student distribution for
GARCH(1,1)-standardized residuals……………………………………………….91
3.14 Absolute demeaned monthly returns \(|r_t - \bar{r}|\) and GARCH(1,1) standard
deviations (std. dev.) with Gaussian and Student t error distribution……………92
3.15 QQ-plot of the GLS-standardized residuals \( (r_i - \mu_{GLS})/h_i \) from Jan. 1926 to Dec. 2005………………………………………………………………………………………………93

3.16 PP-plots of Gaussian, Student t and the skew-t distribution for the GLS normalized residuals \( (r_i - \mu_{GLS})/h_i \)……………………………………………..93

3.17 Probability distribution of tails for GLS residuals with CDF and 1-CDF in logarithmic scale (base 10)……………………………………………………………94

4.1 Probability density function for log consumption growth rate and the stochastic parameter………………………………………………………………………..118

4.2 Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -3.49 and 35.23 respectively while the mean and variance for the IG are 0.001 and 0.0022 each……………...119

4.3 Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -1.0 and 10.7 respectively while the mean and variance for the IG are 0.0012 and 0.0018 each……………...120

4.4 Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -0.01 and 5.45 respectively while the mean and variance for the IG are 0.0013 and 0.0012 each……………...121

4.5 Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -0.001 and 3.01 respectively while the mean and variance for the IG are 0.0013 and 0.0001 each……………...122

5.1 Growth rate of quarterly per capita real personal consumption expenditure for nondurable goods and services…………………………………………………148

5.2 Quarterly New York Stock Exchange real stock returns………………………………………148

5.3 Log ratio of quarterly stock wealth to nondurable consumption expenditures…149

5.4 Quarterly real stock return and quarterly growth rate of per capita real personal consumption expenditure for nondurable goods and services……………………149

5.5 Scatter-plot of quarterly real stock return and growth rate of per capita real personal consumption expenditure for nondurable goods and services………..150

5.6 Quarterly per capita real net corporate dividend growth at an annual rate……….150

5.7 Quarterly per capita real corporate profits after tax at an annual rate……………..151
5.8 Quarterly real growth rate of unit labor cost of non-farm business at an annual rate.

5.9 Quarterly HP-filtered average hourly earnings growth at an annual rate.

5.10 Quarterly per capita real compensation of employees at an annual rate.
CHAPTER 1

INTRODUCTION

My dissertation addresses two main issues regarding asset returns. Two chapters of this dissertation develops an econometric modeling framework for asset returns while two other chapters provide potential solutions to puzzling features of the standard consumption-based asset pricing model. Chapter 2 and 3 analyze U.S. equity market returns and provide an empirical guidance for modeling a skewed and fat-tailed error distribution along with GARCH effects for the stock data. Chapter 4 addresses asset return puzzles such as high equity premium and low riskfree rate based on a negatively skewed and fat-tailed non-Gaussian error distribution that may result from parameter uncertainty underlying the data generating process. Chapter 5 then investigates volatility of U.S. equity market in relation to macroeconomic risk within a general equilibrium asset pricing framework.

Chapter 2 provides a new theoretical derivation for the skew-Student (skew-t) distribution and applies it to an 80-year sample of U.S. monthly excess stock market returns. As a further application, it computes the equity premium based on the estimated skew-t distribution so that skewness and kurtosis of the excess returns are taken into account. The skew-t is derived as a normal variance-mean mixture distribution
$N(\mu + \beta \tilde{V}, \tilde{V})$ with the latent variance $\tilde{V}$ drawn from the inverted-chi-square distribution.

The skew-t turns out to be the same – but with different parameterization – as that derived as a special limiting case of the Generalized Hyperbolic (GH) distribution, which Aas & Haff (2006) call the GH skew Student’s t distribution. The skew-t results in the Student t if $\beta = 0$. The skew-t can capture the substantially skewed fat-tails that characterize U.S. stock data. It is found that a time-dependent skew-t model represents the excess returns very accurately when GARCH effects are incorporated. Based on the estimated skew-t distribution, I compute the equity premium and find that taking into account the substantially negative skewness of excess returns results in a finite equity premium that is close to the one under Gaussian assumption in contrast to the infinite equity premium under the symmetric Student t assumption.

In chapter 3, I provide a new empirical guidance for modeling a skewed and fat-tailed error distribution along with GARCH effects for equity returns based on empirical findings on Realized Volatility ($RV$), constructed from the summation of higher-frequency squared (demeaned) returns. Based on an 80-year sample of U.S. daily stock market returns, I find that the distribution of monthly $RV$ conditional on past returns is approximately the inverted-chi-square. I also find that monthly market returns, conditional on $RV$ and past returns are normally distributed with $RV$ in both mean and variance $N(\mu + \beta RV, RV)$. These empirical findings serve as building blocks for the distributional assumption underlying the GARCH-skew-t model, which is derived as a mixture distribution of normal and inverted-chi-square in chapter 2. Thus, I provide a new empirical support for the GARCH-skew-t modeling of equity returns. I show that the implied GARCH-skew-t model adequately describes the pattern of U.S. stock market
returns. Moreover, the GARCH-skew-t accurately represents the three important stylized facts of stock market returns: volatility clustering, fat-tails and negative skewness.

In chapter 4, I provide a potential solution to asset return puzzles as been noted in macro-finance literature since the seminal work of Mehra & Prescott (1985). It is well known that the standard consumption-based asset pricing model (Lucas 1978) leads to asset return puzzles such as high equity premium (6-8%) and low riskfree rate (1-2%) that are hard to explain with a reasonable degree of risk aversion in the CRRA utility function, which are commonly employed in macroeconomic literature. Hence, these puzzles raise critical questions for the common macroeconomic modeling exercise. I address these puzzles by showing that a negatively skewed and thick-tailed error distribution, arising mainly due to parameter uncertainty underlying the data generating process, can explain the historically high equity premium and low riskfree rate; thus, providing a potential solution to those puzzles. Mehra & Prescott (1985) find that about 0.35% is the maximum risk premium the standard C-CAPM can explain with a risk-free rate between 0 and 4% when the mean and standard deviation of consumption growth is assumed 1.8% and 3.6% respectively. I also take 1.8% mean and 3.6% standard deviation for consumption growth and then explain 6-7% equity risk premium, 0.5-2% risk-free rate and 5-7% dividend yield with the risk aversion coefficient below 10 solely relying on negative skewness and leptokurtosis that might result from the stochastic nature of the uncertain (variance) parameter underlying the endowment process.

Chapter 5 investigates a possible link between stock market and macroeconomic volatility. We have observed in the mid-1980s that U.S. economy experienced a great moderation in macroeconomic volatility in variables such as output and consumption
growth. But during the same period, we did not experience a similar moderation in stock market volatility. The stock market volatility in the face of macroeconomic moderation is a conundrum for the standard consumption-based asset pricing model. We explain this conundrum by decomposing the source of consumption into payouts from equity and non-equity assets each of which serves as a factor of production. Thus, we distinguish stock price and dividend from total wealth and consumption. The model can predict the high volatility of stock returns, smooth consumption growth and their weak correlation at the same time. It is shown that the great moderation of macroeconomic risk must have originated from declining risk of the relatively large non-equity assets while the relatively small equity-based income has been persistently volatile during the moderation era. Furthermore, the model shows that the consumption risk of equity-holding is positively associated with stock share of total wealth.

Chapter 6 concludes this dissertation.
CHAPTER 2

THE SKEW-STUDENT DISTRIBUTION WITH APPLICATION TO U.S. STOCK MARKET RETURNS AND THE EQUITY PREMIUM

2.1 Introduction

The skew-Student distribution, which is derived as a normal variance-mean mixture with the inverted chi-square as a mixing distribution is employed to study U.S. stock returns. The skew-t distribution derived in this paper turns out to be the same - but with different parameterization - as that derived as a special limiting case of the Generalized Hyperbolic (GH) distribution (Barndorff-Nielsen 1977), which Aas & Haff (2006) call the GH skew Student’s t-distribution. The skew-Student distribution is so flexible that it can represent substantially skewed fat-tails that characterize the empirical distribution of U.S. stock data. The skew-t model also contains the symmetric t distribution as a special case in which the skewness parameter is set to zero. As more skewness is imposed, the skew-t distribution can represent various combinations of asymmetric tail behaviors.

The empirical distribution of log excess U.S. stock market returns over T-bill rates is substantially skewed to the left with a heavy left tail and a Gaussian-like right tail - this is confirmed by the normal QQ-plot of log returns.\footnote{See Singleton (2006) for an overview of U.S. stock return distributions.} That is, the left tail of the empirical distribution may be characterized by a power function while the right tail by an
exponential function. We find that the left tail of the empirical distribution is even heavier than can be represented by the fitted symmetric t distribution while its right tail is close to that of the Gaussian. Such substantially negative skeweness and leptokurtic property of log excess returns are well represented by the skew-Student distribution.

The skew-t model is estimated with log excess U.S. stock returns. To examine the goodness of fit, we analyze the estimated skew-t Cumulative Distribution Function (CDF) in comparison with the empirical distribution of log returns. We find that the estimated skew-t CDF quite well represents the empirical CDF of log returns. The PP-plot also shows that log excess returns are well represented by the skew-t distribution. Based on the goodness of fit, the skew-t outperforms the symmetric t, which again outperforms Gaussian. The advantage of the skew-t over the symmetric t comes from its ability to represent quite asymmetric tail behaviors while the advantage of the symmetric t over Gaussian comes from its ability to represent leptokurtic tails.

The equity risk premium is then computed as the log expectation of excess arithmetic stock returns over T-bill rates over the postwar period. The equity premium is simply the log of the Moment Generating Function (MGF) of log excess returns or Cumulant Generating Function (CGF) evaluated at $t = 1$. The MGF of the Student t does not exist because the expectation of the exponential function goes to infinity due to its heavy-tail density determined by a power function, implying the equity premium under the symmetric-t goes to infinity as noted in Geweke (2001) and Weitzman (2007). The MGF of the skew-Student may also explode unless skewness is sufficiently negative. However, the substantially negative skewness of excess log returns results in a finite equity premium for the postwar period of $0.00627 / \text{mo.} = 7.52 \% / \text{yr}$. This is close to the
equity premium of 0.00628/mo. = 7.53 %/yr. under Gaussian distribution over the same sample period. This result must be interesting considering that Gaussian distribution very poorly represents the empirical distribution of excess log returns, even poorer than the symmetric Student t in terms of the goodness of fit.

To understand why the equity premium under the skew-t distribution is so close to that of Gaussian, we decompose the equity premium into different orders of cumulants. Decomposition of the equity premium under the skew-t shows that cumulants higher than the 2nd order decline very rapidly while the first and second order cumulants are very close to those of Gaussian respectively. In other words, the equity premium under the skew-t distribution is mainly determined by the first two lower order cumulants; hence, it is very close to that of Gaussian.

Financial time series often show skewed and leptokurtic distribution conditionally as well as unconditionally – i.e. with and without ARCH or GARCH effects. The conditional distribution of U.S. stock market returns also shows substantially negative skewness and leptokurtic tails. This statistical property of the conditional distribution motivates the extension of the skew-Student distribution to incorporate the time-dependent feature of second moments, or volatility clustering.

Volatility clustering is easily incorporated into the skew-t model by combining observation-driven modeling approach such as GARCH-type specifications. Barndorff-Nielsen (1997) points out that a variance-mean mixture distribution such as the Generalized Hyperbolic (GH) distribution can be extended to represent volatility-clustering by incorporating observation-driven modeling approach. Andersson (2001) and Jensen & Lunde (2001) apply GARCH-type specifications to the Normal Inverse
Gaussian (NIG) distribution which is another sub-class of the GH distribution as suggested by Barndorff-Nielsen (1997). The present paper extends the skew-t model by incorporating observation-driven modeling approach into the underlying stochastic variance but adopts a simple GARCH (1,1) specification rather than more complicated ones because the GARCH (1,1) accounts well for volatility clustering of log excess returns and also allows easy comparison with GARCH(1,1) models with such familiar distributional assumptions as Gaussian and the symmetric Student t.

It is shown that the dynamic skew-t model with GARCH effects added fits U.S. log excess stock returns quite well whereas the same GARCH-type specification with both Gaussian and Student-t fits the sample data poorly. This empirical result reflects the significance of skewness and kurtosis inherent in U.S. stock market returns remaining even after adjusting for GARCH effects, or volatility clustering.

The equity premium is then computed under the common GARCH(1,1) specification but with different distributional errors to examine the impact of skewness and kurtosis on the equity premium. The cumulant-based analysis of the time-varying equity premium shows that the equity premium under the dynamic skew-t is also very close to that of GARCH(1,1) with Gaussian errors because higher-order cumulants above the second order decline very rapidly while the conditional mean and variance are set equal to those of GARCH(1,1) with Gaussian errors by construction. This empirical result must reflect the Gaussian-like right tail of the conditional distribution of log excess returns, or the GARCH-standardized residuals.

The rest of this paper is organized in the following order. The skew-Student distribution and its moment generating function are derived and discussed in section 2.2.
Log excess U.S. stock market returns are analyzed, and the skew t model is then estimated with the excess returns in section 2.3. Section 2.4 discusses its goodness of fit in comparison with other familiar distributions such as Gaussian and the symmetric Student t. Based on the estimated model, the equity premium is computed in section 2.5. Section 2.6 extends the skew-t model to a dynamic one based on the normal inverted-chi-square mixture setting so that it can incorporate volatility clustering as well as skewed and leptokurtic distribution. Conclusion follows in section 2.7.

2.2 The skew-Student distribution and the moment generating function

It is well known that a variance mixture distribution of normal random variables with the inverted chi-square as the mixing distribution leads to the symmetric Student t. That is, a constant-mean normal random variable conditional on its latent variance whose reciprocal is drawn from the chi-square distribution gives rise to Student t as a mixture distribution. However, if the mean of the conditionally Gaussian random variable moves in proportion to its stochastic latent variance whose reciprocal is drawn from the chi-square, as shown in (2.1), then the joint mixture distribution turns out to be the skew-Student. In other words, the skew-Student distribution is derived as a normal variance-mean mixture with the inverted chi-square as a mixing distribution.²

\[
\tilde{r} \mid \tilde{V} \sim N(\mu + \beta \tilde{V}, \tilde{V}) \quad \text{where} \quad \frac{1}{\tilde{V}} \sim \frac{1}{\nu^2} \chi^2
\]

That is, the random variable, \( \tilde{r} \) is Gaussian conditional on its latent variance, \( \tilde{V} \) whose reciprocal is drawn from the chi-square distribution. In addition, the mean of the

² If the latent variance is drawn from the Inverse Gaussian (IG) distribution as another example of the normal variance-mean mixture distribution, then the joint mixture distribution results in the Normal Inverse Gaussian (NIG) distribution, which may be used to model asymmetric semi-heavy tails – see Barndorff-Nielsen (1995, 1997).
conditionally Gaussian distribution moves in proportion to its variance. The variance-mean mixture representation in terms of probability density functions is shown below by defining \( \tilde{\theta} = 1/\tilde{V} \), called the precision.

\[
\tilde{r} | \tilde{\theta} \sim f_{\tilde{r} | \tilde{\theta}}(r | \tilde{\theta}, \beta, \mu) = \frac{\tilde{\theta}^{1/2}}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \tilde{\theta} \left( r - \mu \frac{\beta}{\tilde{\theta}} \right)^2 \right\}
\]

where \( \tilde{\theta} \sim f_\tilde{\theta}(\theta | \nu, h) = \frac{h^\nu (\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \theta^{\nu-2} \exp \left( -\frac{1}{2} h^2 \nu \theta \right) \)

The marginal density function of \( \tilde{r} \) is then derived as a mixture of the two distributions as

\[
f_r(r | \nu, \beta, h, \mu) = \int_0^\infty f_{\tilde{r} | \tilde{\theta}}(r | \beta, \mu, \tilde{\theta}) f_\tilde{\theta}(\theta | \nu, h) d\tilde{\theta}.
\] (2.2)

Substituting for each density function and further simplifying,

\[
f_r(r | \nu, \beta, h, \mu) = \frac{h^\nu (\nu/2)^{\nu/2}}{\sqrt{2\pi \Gamma(\nu/2)}} \exp(\beta r^\nu) \int_0^\infty \theta^{(\nu+1)/2-1} \exp \left( -\frac{1}{2} \theta(\nu h^2 + \nu \beta^2 / \theta) \right) d\theta
\]

If \( \beta = 0 \), the density function simply turns out to be Student t.\(^3\) On the other hand, if \( \beta \neq 0 \), the density function turns out to be the skewed Student t - the derivation of the density function is shown in detail in Appendix A.1. The result for the density function is shown below.

\[
f_r(r | \nu, \beta = 0, h, \mu) = \frac{1}{h} t_r \left( \frac{r - \mu}{h} \right)
\]

where \( t_r \) is the Student t distribution with degree of freedom \( \nu \).

\[
f_r(r | \mu, \nu, \beta \neq 0, h) = \frac{h^\nu (\nu/2)^{\nu/2}}{\sqrt{\pi \Gamma(\nu/2)}} \frac{K_{(\nu+1)/2}((\beta \sqrt{(\nu+1)/2}) \sqrt{(r - \mu)^2 + h^2 \nu})}{2^{(\nu+1)/2 - 1} ((r - \mu)^2 + h^2 \nu)^{(\nu+1)/4}} \exp(\beta (r - \mu))
\] (2.3)

\(^3\) The integration can be easily computed by utilizing the Gamma function.
where \( K_\nu(\cdot) \) is the modified Bessel function of the second type of order \( \nu \), and \( \Gamma(\cdot) \) is the Gamma function.\(^4\)

As shown above, the mixture of normal random variables turns out to be the symmetric t distribution if the conditionally Gaussian distribution has a constant mean independent of its variance (\( \beta = 0 \)), whereas it becomes the skew t distribution if the mean of the conditionally Gaussian random variable moves in proportion to its variance (\( \beta \neq 0 \)). That is, the coefficient parameter \( \beta \) determines the degree of skewness imposed on the Student t distribution. As a result of skewness imposed on Student t, one tail may become thicker than that of the symmetric t while the other one becomes thinner – i.e. one heavy and the other semi-heavy or Gaussian-like tail. This is how the skew-Student distribution represents substantial skewness as well as heavy tails.

The mean, variance, skewness and kurtosis of the skew-t distribution are derived based on the normal inverted-chi-square mixture specification (2.1) in the present paper. Note that the mean is computed as a sum of the location parameter \( \mu \) and \( \beta \) times the expected value of the stochastic variance \( E[\tilde{V}] \). Hence, the mean of the skew-t distributed random variable could be much different from the location parameter \( \mu \) depending on the degree of skewness reflected in the skewness parameter \( \beta \).

\[
E[\tilde{r}] = \mu + \beta E[\tilde{V}] = \mu + \frac{\beta h^2 \nu}{\nu - 2} \tag{2.4}
\]

The variance is also derived in a similar way as

\[
Var[\tilde{r}] = \frac{h^2 \nu}{(\nu - 2)} + \frac{2 \beta^2 h^4 \nu^2}{(\nu - 2)^2 (\nu - 4)} \tag{2.5}
\]

\(^4\) See Abramowitz and Stegun (1972).
The skewness and kurtosis follow subsequently.

\[
Skw[\tilde{\tau}] = \frac{2\beta h \nu^{1/2} (\nu - 4)^{1/2}}{\{2\beta^2 h^2 \nu + (\nu - 2)(\nu - 4)\}^{3/2}} \{3(\nu - 2) + 8\beta^2 h^2 \nu \over (\nu - 6)\} \tag{2.6}
\]

\[
Kts[\tilde{\tau}] = \frac{3(\nu - 2)^3 (\nu - 4)}{\{2\beta^2 h^2 \nu + (\nu - 2)(\nu - 4)\}^2} \{1 + 4\beta^2 h^2 \nu (\nu + 2) \over (\nu - 2)^2 (\nu - 6) + 4\beta^4 h^4 \nu^2 (\nu + 10) \over (\nu - 2)^3 (\nu - 6)(\nu - 8)\} \tag{2.7}
\]

Derivations are shown in Appendix A.2 utilizing the mixture representation. Note that the mean, variance, skewness and kurtosis of the symmetric Student t distribution are obtained simply as special cases of the skew-t as skewness is turned off by \( \beta = 0 \).

The Moment Generating Function (MGF) of the symmetric t distribution does not exist because the tail density declining with a power function is dominated by the exponential integrand in computing the expectation; hence, the MGF explodes. In contrast, the MGF of the skew t distribution may exist provided skewness, as imposed by \( \beta \) is sufficiently large. The MGF of the skew t distribution is derived in Appendix A.3 and the result is shown below.

The MGF exists provided \( t \geq 0 \) with \( \beta < -t / 2 \) or \( t \leq 0 \) with \( \beta > -t / 2 \).

\[
M_{\tilde{\tau}}(t | \nu, \beta, h, \nu) = E[e^{\nu \tilde{\tau}}] = 2e^{\nu h \nu^{1/2} (\nu / 4)^{1/4} (-t(t + 2\beta))^{1/4}} \over \Gamma(\nu / 2) \} K_{\nu / 2}(h \sqrt{-\nu t(t + 2\beta)}) \tag{2.8}
\]

It is noteworthy to point out that the skew-Student distribution (2.3) is in fact the same - but with different parameterization - as that derived as a special limiting case of the Generalized Hyperbolic (GH) distribution, which Aas & Haff (2006) call the GH skew Student’s t distribution. In contrast, the present paper derives the skew-Student distribution as a variance-mean mixture of normal random variables under the familiar chi-square distribution without any reference to the GH distribution; hence, this
derivation is simpler as well as more conceptual. The parameterization of this derivation for the skew t distribution is used throughout the present paper. The dynamic skew-t model, which will be discussed in section 2.6 is also established as a mixture distribution of a conditionally normal and inverted chi-square distribution.

2.3 Data and estimation of the skew-Student distribution

2.3.1 Data

The value-weighted NYSE monthly index returns, including distributions, and Fama 1-month Treasury bill rates from the CRSP are used to construct the log excess stock market returns over Treasury bill rates. The excess log returns during the postwar period from Jan. 1947 to Dec. 2003 (684 observations) are displayed in Figure 2.1. The prewar period, especially the 1930’s is excluded from the sample because the level of volatility during the prewar period is much higher than the postwar period. Since the unconditional distribution of the excess returns is considered in this section, it might be inappropriate to assume that stock returns of the prewar period are drawn from the same distribution as that of the postwar period.5

To analyze the empirical distribution of excess log returns over the postwar period, especially the asymmetric tail behavior, the normal QQ-plot is drawn in Figure 2.2. As shown in the QQ-plot, the left-tail of the sample data is much thicker than that can be represented by Gaussian distribution while the right-tail is roughly close to that of Gaussian. It is also clear that excess log returns are substantially skewed to the left – i.e. worse events tend to occur more frequently than better events do. Even excluding such extreme events as stock market crashes in October 1987 and August 1998 and the market

---

5 Kim, Morley & Nelson (2005) among others report a structural break in the level of volatility for various econometric specifications at around the WWII.
boom in October 1974, the overall feature of the empirical distribution of excess log returns does not change much. As is shown in Figure 2.5, the left tail is even heavier than can be represented by the fitted Student t distribution. That is, the left-tail of excess log returns may be characterized by a power function while the right-tail can be approximated by an exponential function. This distributional feature of excess log returns is the motivation for the use of the skew-Student distribution.\(^6\)

### 2.3.2 Estimation

Parameters of the skew t distribution are estimated by maximum likelihood. The skew t density function for excess log returns \( f_i(r_i | \mu, h, \nu, \beta) \) is as shown in (2.3).

\[
(\hat{\mu}, \hat{h}, \hat{\nu}, \hat{\beta}) = \arg \max_{\mu, h, \nu, \beta} \sum_{i=1}^{n} \log f_i(r_i | \mu, h, \nu, \beta)
\]

Estimated parameters of the skew t distribution and the annualized mean of log excess returns over the postwar period are shown in Table 2.1 in comparison with parameter estimates of the familiar Gaussian and Student t distribution. Gaussian and Student t distributions might be considered benchmark distributions whose tails are determined by exponential and power functions respectively, but with symmetry. In addition, the distributions are related to one another based on variance uncertainty and variance-mean mixture representation. That is, the difference between Gaussian and Student t comes from variance uncertainty characterized by the inverted chi-square distribution. If the mean of conditionally Gaussian distribution moves in proportion to its latent stochastic variance drawn from the inverted chi-square distribution, then the

\(^6\) Alternative distributions such as the NIG (Barndorff-Nielsen 1995, 1997) and other skew Student’s t distributions – Fernandez & Steel (1998), Jones & Faddy (2003) and Azzalini & Capitanio (2003) for example - may be inferior in fitting the excess log returns because the NIG is characterized by semi-heavy tails while other skew Student’s t distributions are by power functions. Tail behaviors of those alternative distributions are summarized in Aas & Haff (2006).
marginal distribution turns out to be the skew-Student as discussed above. If there is no skewness ($\beta = 0$), then the skew t model turns out to be the symmetric t.

While all parameter estimates are significant, the skew-t outperforms the symmetric t which again outperforms Gaussian according to the log-likelihood and model selection criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The annualized sample mean of log excess returns is 6.52%/yr. - the sample mean is the same as the mean of estimated Gaussian distribution. In contrast, the annualized Student t mean is much bigger, 8.38%/yr. That is, there is a non-trivial difference in the mean between the estimated Gaussian – characterized by exponential tails - and Student t distribution – characterized by power function tails. This empirical result indicates that estimation of mean is very sensitive to distributional assumptions. Statistical descriptions based on these two distributions might be invalidated given that excess log returns are characterized by a significantly heavy left-tail and a Gaussian-like right-tail as discussed above. Nonetheless, the skew t model whose mean is 6.52%/yr. – which fits the empirical distribution much better – appears to favor the Gaussian over Student t distribution based on estimated means. The mean of the skew t distribution is much smaller than that of the symmetric t distribution due to the substantially negative skewness inherent in U.S. stock market returns. This is because the mean of the skew t distribution is determined not only by the location parameter $\mu$ but also by the expectation of the stochastic variance as shown in (2.4).

Table 2.2 shows the mean, standard deviation, skewness and kurtosis of each fitted distributional model in comparison with those of the empirical distribution. The Gaussian and Student t are unable to represent skewness at all due to their symmetry by
construction. The Gaussian is further disadvantaged by its fixed kurtosis at 3 while Student t is able to capture a higher degree of kurtosis. In contrast, the skew t distribution quite well represents both skewness and kurtosis in addition to mean and variance. Thus, the skew-Student distribution may be employed to represent the distribution of financial data that are characterized by substantial skewness as well as excess kurtosis as of stock market returns we are considering in this paper.

The skew t density function with estimated parameters is drawn in Figure 2.3 in comparison with estimated Gaussian and Student t density functions. Because tail behaviors are of interest, densities are rescaled in logarithm (base 10) in Figure 2.4. As displayed in Figure 2.4, the fitted skew t distribution has a heavier left tail than that of both Student t and Gaussian while its right tail lies in between Student t and Gaussian. That is, the estimated skew-Student distribution is characterized by a heavy left-tail and a semi-heavy right-tail. We can also see that the skew t density goes below Gaussian over the locally positive range of excess log returns until it finally cuts through Gaussian. This feature of the estimated skew t well represents the empirical distribution of returns that lies below Gaussian over the local range of 5–10% log returns as shown in the QQ-plot of Figure 2.2.

2.4 Discussion of the goodness of fit

To investigate the goodness of fit, the estimated skew t model should be analyzed in relation to the sample data. One simple way is to examine how well the CDF (or 1-CDF) of the estimated skew t distribution represents the empirical CDF (1-CDF) as in Figure
2.5 and 2.6. This simple graphical test allows easy assessment of the goodness of fit especially for tail behaviors. A Probability-Probability (PP) plot may also be employed to see how likely the sample data would be drawn from a specific distributional model. If sample data were actually drawn from a certain distribution, estimated probability distribution, or estimated CDF of sample quantiles should be approximated by uniform distribution – the former is the empirical probability and the latter is the theoretical probability. Hence, the plot of empirical probability vs. theoretical probability would be well aligned with the 45 degree line if sample data were generated by the distribution. The PP-plot is useful to examine the overall fit of a distributional model as shown in Figure 2.7, 2.8, and 2.9. In addition to these graphical analyses of the goodness of fit, the Neyman Smooth Test may be employed to provide the metric for the goodness of fit for alternative distributional models. For all these purposes, the CDF of each distributional model is required. The CDF of the estimated skew-Student is derived and numerically approximated to analyze the fit - see Appendix A.4 for detail.

To test the goodness of fit especially for tail behaviors, the CDF (or 1-CDF) of estimated distributions of the skew t, symmetric t and Gaussian are drawn and are compared with the empirical CDF (1-CDF) over sample quantiles in Figure 2.5 and 2.6. As clearly shown in Figure 2.5, the left tail of the empirical distribution is quite well represented by the skew t model, whereas it is systematically underestimated by both the Student t and Gaussian distribution. This implies that worse events in U.S. equity markets tend to occur more frequently than the estimated symmetric t distribution can represent.

---

7 The CDF \( \Pr[\tilde{r} \leq r] \) in logarithmic scale is used to compare left-tail distributions while 1-CDF \( \Pr[\tilde{r} > r] \) is used for right-tail distributions. Note that empirical CDF (1-CDF) is a step function as shown in Figure 2.5 and 2.6.

Figure 2.6 shows that the right tail of the empirical distribution is much better represented by the skew t model than by the symmetric t and Gaussian distribution although it is locally over- or underestimated over some range of log excess returns. The symmetric t distribution systematically overstates the probability distribution of the right tail. This implies that better events in U.S. equity markets tend to occur less often than the symmetric t distribution predicts. Gaussian distribution also much overstates right-tail probabilities roughly over 5-10 % range of excess log returns though it is quite well aligned with the empirical distribution over the range of 10-15% excess log returns.\textsuperscript{9}

A PP-plot of each distributional model can be examined to assess how well the sample data are explained by each specified distribution. The PP-plot is constructed by plotting cumulative probabilities corresponding to ordered sample data via the CDF of the assumed distribution – empirical probabilities - against the uniform distribution – theoretical probabilities. If the assumed distribution is the true distribution that generated the sample data, then the pairs of theoretical and empirical probabilities should lie along the 45° line.

Comparing across PP-plots of each assumed distribution in Figure 2.7, 2.8, and 2.9, we find that the skew t distribution best fits excess log returns during the postwar period. In addition, it appears based on PP-plots in Figure 2.7 and 2.8 that Gaussian distribution very poorly fits the sample data, even poorer than the symmetric t distribution. Taking all these empirical results into account, the skew-Student may be employed to represent the distribution of excess log returns. The success of the skew t at

\textsuperscript{9} The QQ-plot of Figure 2.2 also shows that Gaussian distribution overstates the probability distribution of excess log returns over the range of about 5-10% log returns. Note that much more data points are concentrated on the range of 5-10 % log returns compared with about dozen points over 10-15%.
representing the empirical distribution of excess returns motivates the use of the skew t to compute the equity risk premium in the next section.

2.5 The equity premium

The equity risk premium is computed as the log expectation of excess arithmetic stock returns over T-bill rates – i.e. the logarithm of the expected ratio of gross returns of stocks to T-bill rates.\textsuperscript{10} To compute the expectation for the equity premium, the probability distribution of excess log returns should be available or assumed. Hence, distributional assumptions for log excess returns may play a significant role to explain the equity premium.\textsuperscript{11} This is the motivation for the use of the skew-Student distribution which quite accurately represents the empirical distribution of excess log returns over the postwar period as discussed above.

If log excess returns were drawn from Gaussian, the equity premium would be

\[ EP = \log(E[e^r]) = E[\bar{r}] + Var[\bar{r}]/2 = 7.53\% \text{ / yr.} \]

as shown in Table 2.3. In contrast, the equity premium under the symmetric t distribution employed by Shephard (1994) and Weitzman (2006) goes to infinity because its heavy tail declining in a power function are dominated by the exponentially increasing integrand in computation of the expectation, just as the MGF of the Student t explodes – also noted by Geweke (2001).

The expectation of arithmetic returns is simply the MGF of the distribution of log returns evaluated at \( t = 1 \), where \( t \) is the coefficient of the power. Utilizing the MGF of

\textsuperscript{10} See Campbell, Lo & MacKinlay (1997) for example.

\textsuperscript{11} In fact, distributional assumptions might be one approach to address asset pricing issues such as equity premium puzzles that have been investigated since the seminal work of Mehra & Prescott (1985).
the skew-Student distribution (2.8), the expectation of excess arithmetic stock returns is computed as below. For $t = 1$ with $\beta < -1/2$,

$$E[e^r] = \frac{2e^{\mu}h^{\nu/2}(\nu/4)^{\nu/4}(-1+2\beta))^{\nu/4}}{\Gamma(\nu/2)}K_{\nu/2}(h\sqrt{-\nu(1+2\beta)})$$ (2.9)

The equity premium under the skew-Student is then the logarithm of the expectation

$$EP = \log(E[e^r])$$

Utilizing the Cumulant Generating Function (CGF) $K(t) = \log\{M(t)\}$, which is a logarithm of the MGF $M(t)$, the equity premium is also given by $K(1)$.

$$EP = K(1)$$

$$= \mu + \frac{\nu}{4}\log\nu - \frac{\nu-2}{2}\log 2 + \frac{\nu}{2}\log h + \frac{\nu}{4}\log\{-1+2\beta\}$$

$$+ \log\{K_{\nu/2}(h\sqrt{-\nu(1+2\beta)})\} - \log\{\Gamma(\nu/2)\}$$ (2.10)

Note that the equity premium exists only if the skewness is sufficiently negative, i.e. $\beta < -1/2$. If skewness is not negative enough, $\beta \geq -1/2$, which includes the symmetric t distribution, then, the MGF explodes – i.e. the equity premium goes to infinity. However, the substantially negative skewness present in U.S. excess stock market returns as shown in Figure 2.2, 2.5 and 2.6 can prevent such explosion; thus, leading to a finite expectation of excess arithmetic stock returns. The equity premium under the skew t is then computed from (2.10) as 7.52 %/yr., which is quite close to that of Gaussian as shown in Table 2.3. This result is interesting considering that the Gaussian distribution fits log excess returns very poorly in contrast to the skew t, which quite well represents excess log returns – note that the Gaussian distribution is in fact even poorer than the symmetric t distribution based on the goodness of fit.

---

12 Details on the CGF and cumulants relevant to the discussion of this section are given in Appendix A.5.
13 The estimate of $\beta$ is -13.94 as shown in Table 2.1.
To understand why the equity premium under the skew-t distribution is close to that of Gaussian, the CGF may be expanded as the summation of cumulants as the following.

\[ K(t) = \sum_{n=1}^{\infty} k_n t^n/n! = k_1 t + \frac{1}{2} k_2 t^2 + \frac{1}{3!} k_3 t^3 + \frac{1}{4!} k_4 t^4 + \ldots \]

where \( k_n \) is the \( n \)th cumulant, which is the \( n \)th derivative of the CGF evaluated at zero – i.e. \( K^{(n)}(0) \). The equity premium is then expressed in terms of cumulants:

\[ EP = K(1) = k_1 + \frac{1}{2} k_2 + \frac{1}{6} k_3 + \frac{1}{24} k_4 + \ldots \]

Note that cumulants are related to central moments and thus to mean, variance, skewness, and kurtosis. That is,

\[
\begin{align*}
    k_1 &= E[\bar{r}] \\
    k_2 &= Var[\bar{r}] \\
    k_3 &= (k_2^{3/2}) Skw[\bar{r}] \\
    k_4 &= (k_2)^2 (Kts[\bar{r}] - 3)
\end{align*}
\]

The first four cumulants of the skew t and Gaussian distribution are computed based on Table 2.2 respectively. Hence, we can identify the contribution of each order of cumulants for the equity premium and thus explain why the equity premia under the skew-t and Gaussian are close to each other. Table 2.4 shows the first four components of the CGF decomposition. For Gaussian distribution, the sum of the first two terms corresponding to the first and second cumulants explains the equity premium because all higher-order terms are zero. For the skew-t, the sum of the first two terms is quite close to that of Gaussian while higher-order terms decline very rapidly. Thus, the equity premium under the skew t is very close to that of Gaussian.
2.6 Dynamic skew-Student distribution

2.6.1 Extension to a dynamic model

Financial time series often show both skewed and leptokurtic property conditionally as well as unconditionally. We have discussed above that the skew-Student distribution quite well represents both skewness and kurtosis of log excess returns – outperforming both Gaussian and the symmetric t distribution. The fair amount of evidence that the conditional distribution of stock returns is also skewed as well as leptokurtic motivates the extension of the skew-Student distribution to model the conditional distribution by incorporating the time-dependent structure – especially of volatility clustering, or GARCH effects.

Volatility clustering can be incorporated into the skew-t model by combining observation-driven modeling approach such as GARCH-type specifications. Barndorff-Nielsen (1997) points out that a normal variance-mean mixture distribution such as the GH distribution can be extended to represent volatility-clustering by combining it with an observation-driven modeling approach. Andersson (2001) and Jensen & Lunde (2001) apply GARCH-type specifications to the NIG distribution which is another sub-class of the GH distribution as suggested by Barndorff-Nielsen (1997). The present paper also incorporates observation-driven modeling approach but adopts a simple GARCH (1,1) specification rather than investigating more sophisticated optimal ones because the GARCH (1,1) well represents the volatility clustering of log U.S. excess stock returns and also allows easy comparison with GARCH(1,1) with such familiar distributional assumptions as Gaussian and the symmetric Student t.
For the *skew t* distribution \( f(r \mid \mu, h, \nu, \beta) \), the location-scale family of the distribution is given by 
\[
\frac{1}{h} f\left(\frac{r - \mu}{h} \mid 0, 1, \nu, \bar{\beta}\right)
\]
with \( \bar{\beta} = h\beta \) which we denote as 
\( \tilde{r} \sim \text{Skew}(\mu, h, \nu, \bar{\beta}) \). That is \( (r - \mu)/h \sim \text{Skew}(0, 1, \nu, \bar{\beta}) \), where \( \nu \) and \( \bar{\beta} \) are invariant parameters of the location-scale family. The temporal dependence of the second moments or volatility clustering can be incorporated into the model via the scale parameter \( h \) as may be specified as a function of past observations just like GARCH-type specifications – this is how the *observation-driven* modeling approach is combined with the *parameter-driven* stochastic variance.

The one-period prediction distribution for the *skew-t* is then given by 
\[
\tilde{r}_t \mid I_{t-1} \sim \text{Skew}(\mu, h(I_{t-1}; \varphi), \nu, \bar{\beta}) \text{ for } t = 1, 2, \ldots, T
\]
Note that this is a conditional mixture distribution of the conditionally normal and the conditionally inverted chi-square distribution as follows.

\[
r_t \mid (\tilde{V}_t, I_{t-1}) \sim N(\mu + \beta\tilde{V}_t, \tilde{V}_t') \text{ for } t = 1, 2, \ldots, T
\]
\[
\tilde{V}_t^{-1} \mid I_{t-1} \sim \frac{1}{h(I_{t-1}; \varphi)^2} \chi\nu^2 \text{ for } t = 1, 2, \ldots, T
\]

where \( I_{t-1} = \{r_{t-1}, r_{t-2}, \ldots, r_1\} \) is the information set up to time \( t - 1 \) and \( h_t = h(I_{t-1}; \varphi) \) is the scale parameter through which the temporal dependence of the second moments is captured. Once the functional form of \( h(\cdot) \) is determined - for instance, a GARCH-type specification may be adopted -, the conditional mean and variance are given by

\[
E[r_t \mid I_{t-1}] = \mu + \frac{\beta\nu}{\nu - 2} h_t^2 \\
\text{Var}[r_t \mid I_{t-1}] = \frac{\nu}{(\nu - 2)} h_t^2 + \frac{2\beta^2\nu^2}{(\nu - 2)^2(\nu - 4)} h_t^4
\]
The dynamic skew-t model is cast into the common GARCH form for easy interpretation and comparison with GARCH models with different error distributions.

\[ r_t = \mu + h_t \varepsilon_t \]
\[ \varepsilon_t \sim \text{Skewt}(\mu^*, h^*, \nu, \beta) \]

where \( \mu^* = -\frac{\beta \sqrt{\nu(\nu - 4)}}{\sqrt{2\nu\beta^2 + (\nu - 2)(\nu - 4)}} \) and \( h^{*2} = \frac{(\nu - 2)^2 \mu^{*2}}{\beta^2 \nu^2} \) with \( \beta = \beta h^* \) (2.11)

That is,

\[ \frac{r_t - \mu}{h_t} \mid I_{t-1} \sim \text{Skewt}(\mu^*, h^*, \nu, \beta) \] (2.12)

Note that the restriction allows \( E[r_t \mid I_{t-1}] = \mu \) and \( Var[r_t \mid I_{t-1}] = h_t^2 \) because \( E[\varepsilon_t \mid I_{t-1}] = 0 \) and \( Var[\varepsilon_t \mid I_{t-1}] = 1 \) under the restriction.

The conditional distribution is then given by

\[ r_t \mid I_{t-1} \sim \text{Skewt}(\mu + \mu^* h_t, h^* h_t, \nu, \beta / h_t) \]

The distribution can also be represented as a mixture of conditionally normal and inverted chi-square distributions as below.

\[ r_t \mid (\tilde{V}_t, I_{t-1}) \sim N(\mu + \mu^* h_t, \frac{\beta}{h_t}, \tilde{V}_t) \]
\[ \tilde{V}_t^{-1} \mid I_{t-1} \sim \frac{1}{(h^* h_t)^2} \chi^2 \nu \] (2.13)

The conditional mean and variance of the latent stochastic variance \( \tilde{V}_t \) are given by

\[ E[\tilde{V}_t \mid I_{t-1}] = \frac{\nu h^{*2}}{(\nu - 2)} h_t^2 \]
\[ Var[\tilde{V}_t \mid I_{t-1}] = \frac{2\nu^2 h^{*4}}{(\nu - 2)^2 (\nu - 4)} h_t^4 \] (2.14)
Thus, the conditional stochastic variance $\widetilde{V}_t$ is time-varying with the inverted chi-square distribution as the scale parameter $h_t$ changes over time. This is how the dynamic skew-t model is distinguished from the purely observation-driven modeling approach such as GARCH-type models.

The function $h(I_{t-1}; \varphi)$ may be specified as GARCH (1,1) $h_t^2 = c + bh_{t-1}^2 + a\varepsilon_{t-1}^2$ because it quite well represents volatility clustering of log excess returns as will be shown in Figure 2.10. A more sophisticated specification may also be considered but it would go beyond the scope of this paper.

The log-likelihood function is then given by

$$\log L(r_0, r_1, ..., r_T; \nu, \beta, \mu, c, b, a) = \sum_{t=1}^{T} \log f_{r_t}^{LS}(r_t | I_{t-1}; \nu, \beta, \mu, c, b, a)$$

Note that the dynamic skew-t model allows a closed-form analytic expression for the likelihood function in contrast to such stochastic volatility models as Taylor (1986); hence, it is more tractable.

2.6.2 GARCH models with alternative distributional assumptions

As popular benchmark conditional distributions of stock returns, GARCH models are often used with the auxiliary Gaussian error. GARCH(1,1) with Gaussian errors is often applied in literature to describe especially stock market returns. Given the leptokurtic property of conditional distribution of returns, the Student t may be considered as a natural alternative (Bollerslev 1987). This section considers these two conditional distributions as alternative counterparts of the dynamic skew-t model. Moreover, the GARCH-t model is derived as a variance mixture of normal random variables as a special case of the dynamic skew-t model.
2.6.2.1 *GARCH*-t model as a conditional normal variance mixture distribution

The auxiliary error distribution behind the GARCH model is assumed to be the Student t.

\[ r_t = \mu + h_t \epsilon_t \]

\[ h_t^2 = c + bh_{t-1}^2 + ae_{t-1}^2 \]

\[ \epsilon_t \sim iid \frac{1}{h_t} t_{\nu}\left(\frac{\epsilon_t}{h_t}\right) \]

where \( t_\nu(\cdot) \) is the standard Student t distribution. The distribution of \( \epsilon_t \) is scaled in order to have a unit variance, \( \text{Var}[\epsilon_t] = 1 \). The constraint is satisfied if \( h^* = \sqrt{\nu - 2}/\nu \) because \( \text{Var}[\epsilon_t] = h^{*2} \nu / (\nu - 2) \).

Hence,

\[ \frac{(r_t - \mu)/h_t}{\sqrt{(\nu - 2)/\nu}} | I_{t-1} \sim t_\nu \]

(2.15)

Note that the GARCH-t model is a special case of the *dynamic skew-t* with skewness turned off by \( \beta = 0 \). From (2.11), \( \mu^* = 0 \) and \( h^* = \sqrt{(\nu - 2)/\nu} \) if \( \beta = 0 \).

Hence, (2.12) leads to \( (r_t - \mu)/h_t | I_{t-1} \sim \text{Skewt}(0, \sqrt{(\nu - 2)/\nu}, \nu, 0) \). By standardization,

\[ \frac{(r_t - \mu)/h_t}{\sqrt{(\nu - 2)/\nu}} | I_{t-1} \sim \text{Skewt}(0, 1, \nu, 0) \]

where \( \text{Skewt}(0, 1, \nu, 0) \) is the standard Student t distribution \( t_\nu \). Thus, we see that (2.15) is derived directly from the *dynamic skew-t* model by shutting off the skewness by \( \beta = 0 \).

We can also establish the GARCH-t as a variance mixture distribution of normal random variables as (2.13) for the *dynamic skew-t*. Just by imposing \( \beta = 0 \) on (2.13), the mixture-of-distributions representation of the GARCH-t is derived as below.
\[
\begin{align*}
r_t | (\tilde{V}_t, I_{t-1}) &\sim N(\mu, \tilde{V}_t) \\
\tilde{V}_t^{-1} | I_{t-1} &\sim \frac{1}{(\nu - 2)h_t^2} \chi^2_{\nu}
\end{align*}
\]  

(2.16)

The mean and variance of the stochastic variance \( \tilde{V}_t \) can also be derived straightforwardly from (2.14) by substituting in \( h^* = \sqrt{(\nu - 2)/\nu} \) if \( \beta = 0 \).

\[
\begin{align*}
E[\tilde{V}_t | I_{t-1}] &= h_t^2 \\
Var[\tilde{V}_t | I_{t-1}] &= \frac{2}{(\nu - 4)} h_t^4
\end{align*}
\]

2.6.2.2 GARCH-N model

The error term of the GARCH model is assumed to be drawn from the standard normal distribution as below.

\[
r_t = \mu + h_t \epsilon_t
\]

\[
h_t^2 = c + bh_{t-1}^2 + ae_{t-1}^2
\]

\[
\epsilon_t \sim iidN(0,1)
\]

That is, \( r_t | I_{t-1} \sim N(\mu, h_t^2) \) or \( \frac{r_t - \mu}{h_t} | I_{t-1} \sim N(0,1) \)  

(2.17)

Note that the GARCH-N model is a special limiting case of the GARCH-t with \( \nu \to \infty \). As \( \nu \to \infty \), (2.15) leads to (2.17) because the Student t distribution \( t_\nu \) approaches the standard normal \( N(0,1) \).

2.6.3 Estimation

We use monthly log excess returns of a longer sample period (1926:1-2003:12) to estimate the dynamic skew-t model because the model allows changes in the level of volatility so that it can represent the high volatility of the prewar period as displayed in Figure 2.10.
Estimation of the dynamic skew-t model is computationally demanding due to combination of GARCH specifications and the skew-t distribution. Hence, a Quasi-MLE (QMLE) is employed to estimate the GARCH parameters $c, b, a$, which are asymptotically consistent under certain regularity conditions, following Bollerslev & Wooldridge (1992) and Glosten, Jagannathan & Runkle (1989) among others. At the second stage, the constrained skew-t and Student-t model discussed above are then fit to the GARCH-standardized residuals by MLE. Table 2.5 shows parameter estimates with standard errors in parentheses.

Mean, standard deviation, skewness, and kurtosis of the GARCH-standardized residuals are shown in Table 2.6. The distribution of residuals is very negatively skewed and leptokurtic. That is, the conditional distribution of log excess returns is characterized by substantial skewness as well as leptokurtosis. Hence, both Gaussian and the symmetric t distribution are limited by their symmetry while the symmetric t distribution is able to capture some degree of the excess kurtosis. In contrast, the skew t distribution well represents the skewed and leptokurtic property of sample residuals. These empirical results are also well reflected in the model selection criteria of AIC and more conservative BIC in Table 2.5b. Based on both model selection criteria, the skew-t outperforms the symmetric t which again outperforms Gaussian.

---

14 The location parameter is estimated simply as a sample average given that estimation of the location parameter is sensitive to distributional assumption. For instance, the Student t results in much bigger estimate of the location parameter than under Gaussian assumption by a few percentage points at an annual rate. Hence, employing the QMLE even for the location parameter seems to give advantage to the location parameter estimated under Gaussian assumption even though the Student t fits the data much better.
2.6.4 Conditional variance

Conditional variance is obtained from the GARCH(1,1) model based on the long sample period (1926:1-2003:12). Figure 2.10 shows GARCH(1,1) standard deviations on top of the absolute values of demeaned excess log returns. The conditional standard deviations quite well represent the volatility of U.S. excess stock market returns. It is shown that the level of volatility over the postwar period is much stabilized and lower than the prewar period.

2.6.5 Distribution of GARCH-standardized residuals

The QQ-plot of GARCH-standardized residuals $\varepsilon_i = (r_i - \mu)/h_i$ is drawn based on the long sample period (1926:1-2003:12) in Figure 2.11. As illustrated by the QQ-plot, the left-tail of the residual distribution is much thicker than can be represented by Gaussian while the right tail is close to that of Gaussian. Hence, this empirical result confirms that log excess stock market returns are negatively skewed and leptokurtic even conditionally. That is, log excess returns are negatively skewed and leptokurtic even after adjusting for the GARCH effects.

2.6.6 Analysis of residuals for alternative distributions

Graphical test of each distributional model are employed as before. Recall our discussion of the goodness of fit in section 2.4. To analyze the goodness of tail-fit, the CDF (or 1-CDF) of each estimated distributional model is compared with the empirical CDF (1-CDF) of log excess returns in Figure 2.12 and 2.13. Both figures show that GARCH-standardized residuals are quite well represented by the skew-t distribution. Note that both Gaussian and Student t systematically under-estimate the probability distribution of
the left tail while they tend to over-estimate probability mass in the right tail of the residuals.

To assess the overall fit of each distributional model for GARCH-standardized residuals, PP-plots of each distribution are analyzed in Figure 2.14 and 2.15. Recall the discussion on PP-plots in section 2.4. Figure 2.14 shows PP-plots of Gaussian and the Student t distribution for residuals. The large discrepancy between the PP-plots and the 45-degree line suggest that it is hard to accept that the conditional errors are drawn from Gaussian or the Student t distribution. In contrast, Figure 2.15 shows a nice alignment of the PP-plot of the skew-t with the 45-degree line. This empirical result supports that the conditional distribution of excess log returns is very well represented by the skew-Student distribution.

2.6.7 Conditional equity premium

Provided conditional distributions of log excess returns – the dynamic skew-t, GARCH-t and GARCH-N models for example –, we can compute the conditional equity premium based on each estimated conditional distribution. Recall that the symmetric t distribution leads to infinite equity premium because the MGF explodes due to the heavy-tail density as discussed in section 2.5. Hence, we compute the conditional equity premium under the dynamic skew-t and the GARCH-N model. The estimated equity premia under these two conditional distributions are shown in Figure 2.16. As shown in the figure, they are quite close to each other – even hard to distinguish one from the other. Note that both the dynamic skew-t and GARCH-N model share the same mean $\mu$ and conditional standard deviation $\sigma$, by construction. That is, the sum of the first two cumulant components of the equity premia is the same for both the dynamic skew-t and GARCH-N model while
higher order cumulant terms are allowed to differ. Hence, we can see that the trivial difference between the two conditional equity premia comes from the weak contribution of higher order cumulants to the equity premium. In other words, only rapidly declining high-order cumulant components beyond the second order explain the trivial difference between the conditional equity premium of the GARCH-N and the dynamic skew-t. Recall the discussion on the cumulant-based decomposition of the equity premium in section 2.5.

Table 2.7 shows the average of the time-varying equity premium. As implied by Figure 16, the average equity premia are quite close to each other under the two distributions.

2.7 Conclusion

The skew-Student distribution is derived as a normal variance-mean mixture with the inverted chi-square as a mixing distribution in this paper. The distribution turns out to be the same - but with different parameterization - as that derived as a special limiting case of the Generalized Hyperbolic (GH) distribution, which Aas & Haff (2006) call the GH skew Student’s t distribution. The moments, cumulants and MGF are also derived based on the mixture-of-distributions setting and their properties have been discussed. The skew-t is then extended to a dynamic model to represent volatility clustering as well as the skewed and leptokurtic property of financial data. The dynamic skew-t model is derived by incorporating observation-driven modeling approach such as GARCH-type specifications so that the stochastic latent variance is updated upon arrival of new observations. Hence, the mixture-of-distributions setting is preserved over time while the
scale parameter of the conditionally inverted-chi-squared variance changes as the information set is updated.

The skew-Student distribution is then applied to U.S. excess stock market returns and is shown much better than other familiar distributions such as Gaussian and Student t based on the goodness of fit both conditionally and unconditionally – i.e. with and without GARCH effects. The equity premium is then computed based on estimated distributional models. The MGF of the Student t explodes due to the heavy-tail density; hence, leading to an infinite equity premium as noted by Weitzman (2006). The skew-t may also give an infinite equity premium unless the skewness is sufficiently negative. However, the substantially negative skewness inherent in U.S. stock market returns results in a finite equity premium which is close to that of Gaussian both conditionally and unconditionally. The cumulant-based decomposition of the equity premium under Gaussian and the skew-t distribution implies that the impact of skewness and kurtosis on the equity premium is relatively trivial compared with the first two central moments for both conditional and unconditional distributional models.
<table>
<thead>
<tr>
<th></th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>0.00543</td>
<td>0.00699</td>
<td>0.0264</td>
</tr>
<tr>
<td></td>
<td>(0.00157)</td>
<td>(0.00151)</td>
<td>(0.0075)</td>
</tr>
<tr>
<td>( h )</td>
<td>0.0410</td>
<td>0.0351</td>
<td>0.0343</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0016)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>( \nu )</td>
<td>( \infty )</td>
<td>7.70</td>
<td>9.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.07)</td>
<td>(2.65)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0</td>
<td></td>
<td>-13.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5.74)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>1213.7</td>
<td>1230</td>
<td>1236</td>
</tr>
<tr>
<td>AIC</td>
<td>-2423.4</td>
<td>-2454</td>
<td>-2464</td>
</tr>
<tr>
<td>BIC</td>
<td>-2414.3</td>
<td>-2440.4</td>
<td>-2445.9</td>
</tr>
<tr>
<td>( E[\tilde{r}] )</td>
<td>6.52%/yr.</td>
<td>8.38%/yr.</td>
<td>6.52%/yr.</td>
</tr>
</tbody>
</table>

Table 2.1: Parameter estimates and the annualized mean of excess log returns for alternative distributional models. In parentheses are standard errors. The log-likelihood and model selection criteria such as AIC and more conservative BIC are also provided.
<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[\bar{r}]$</td>
<td>0.00543</td>
<td>0.00543</td>
<td>0.00699</td>
<td>0.00543</td>
</tr>
<tr>
<td>$Std[\bar{r}]$</td>
<td>0.0411</td>
<td>0.0410</td>
<td>0.0407</td>
<td>0.0410</td>
</tr>
<tr>
<td>$Skw[\bar{r}]$</td>
<td>-0.649</td>
<td>0</td>
<td>0</td>
<td>-0.688</td>
</tr>
<tr>
<td>$Kts[\bar{r}]$</td>
<td>5.62</td>
<td>3</td>
<td>4.62</td>
<td>6.40</td>
</tr>
</tbody>
</table>

Table 2.2: Mean, standard deviation, skewness and kurtosis of monthly excess log returns for alternative estimated distributional models against those of the empirical distribution (Sample summary statistics are computed with bias correction.)

<table>
<thead>
<tr>
<th></th>
<th>Equity premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>0.00628/mo. = 7.53 %/yr</td>
</tr>
<tr>
<td>Symmetric t</td>
<td>$\infty$ (infinite)</td>
</tr>
<tr>
<td>Skew t</td>
<td>0.00627/mo. = 7.52 %/yr</td>
</tr>
</tbody>
</table>

Table 2.3: Equity premium under alternative distributions

<table>
<thead>
<tr>
<th></th>
<th>$k_1$</th>
<th>$k_2/2$</th>
<th>$k_3/6$</th>
<th>$k_4/24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>0.00543</td>
<td>0.000843</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Skew t</td>
<td>0.00543</td>
<td>0.000841</td>
<td>$-7.91\times10^{-6}$</td>
<td>$4.01\times10^{-7}$</td>
</tr>
</tbody>
</table>

Table 2.4: Cumulant-based decomposition of the equity premium for Gaussian and the skew t distribution
<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$c$</th>
<th>$b$</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00501</td>
<td>$6.56 \times 10^{-5}$</td>
<td>0.8671</td>
<td>0.1129</td>
</tr>
<tr>
<td>(0.00176)</td>
<td>(1.27 $\times 10^{-9}$)</td>
<td>(0.0012)</td>
<td>(0.0009)</td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu^*$</td>
<td>0</td>
<td>0</td>
<td>0.8105</td>
</tr>
<tr>
<td>$\nu^*$</td>
<td>1</td>
<td>0.8526</td>
<td>0.8107</td>
</tr>
<tr>
<td>$\nu$</td>
<td>$\infty$</td>
<td>7.327</td>
<td>10.795</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0</td>
<td>(1.428)</td>
<td>(3.321)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-1326.7</td>
<td>-1304.2</td>
<td>-1288.3</td>
</tr>
<tr>
<td>AIC</td>
<td>2657.4</td>
<td>2614.4</td>
<td>2584.6</td>
</tr>
<tr>
<td>BIC</td>
<td>2667.1</td>
<td>2629</td>
<td>2604</td>
</tr>
</tbody>
</table>

(b)

Table 2.5: Estimation of GARCH models with alternative distributional assumptions: Gaussian, symmetric t and skew-t

a) Parameter estimates of the generic GARCH model with standard errors in parentheses over the long sample period (1926:1-2003:12)

b) Parameter estimates of each distribution for GARCH-standardized residuals
Table 2.6: Mean, standard deviation, skewness and kurtosis of each estimated GARCH-standardized error distribution in comparison with summary statistics of residuals

<table>
<thead>
<tr>
<th></th>
<th>Sample residuals</th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[\tilde{r}]$</td>
<td>0.00371</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\text{Std}[\tilde{r}]$</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\text{Skw}[\tilde{r}]$</td>
<td>-0.815</td>
<td>0</td>
<td>0</td>
<td>-0.839</td>
</tr>
<tr>
<td>$\text{Kts}[\tilde{r}]$</td>
<td>5.37</td>
<td>3</td>
<td>4.80</td>
<td>6.21</td>
</tr>
</tbody>
</table>

Table 2.7: Annualized average equity premium under alternative distributions based on the long sample period (1926:1-2003:12)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Equity premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH-N</td>
<td>7.803%/yr.</td>
</tr>
<tr>
<td>GARCH-t</td>
<td>$\infty$ (infinite)</td>
</tr>
<tr>
<td>Dynamic skew-t</td>
<td>7.768 %/yr.</td>
</tr>
</tbody>
</table>
Figure 2.1: Excess log monthly stock market returns over T-bill rates during the postwar period from Jan. 1947 – Dec. 2003 (684 observations)

Figure 2.2: Normal QQ-plot of excess log returns over the postwar period
Figure 2.3: Estimated density functions for the alternative distributional models

Figure 2.4: Estimated density functions for the alternative distributional models
- Densities are rescaled in logarithm (base 10) to highlight tail behaviors
Figure 2.5: Empirical and estimated CDF’s for alternative distributions in logarithmic (base 10) scale over ordered sample data – the logarithmic scale highlights the left tail of each distribution

Figure 2.6: Empirical and estimated (1-CDF)’s for alternative distributions in logarithmic (base 10) scale over ordered sample data – the logarithmic scale highlights the right tail of each distribution
Figure 2.7: PP-plot of Gaussian distribution for excess log returns

Figure 2.8: PP-plot of the symmetric t distribution for excess log returns
Figure 2.9: PP-plot of the skew t distribution for excess log returns

Figure 2.10: Absolute demeaned log excess returns and the GARCH standard deviations
Figure 2.11: QQ-plot of GARCH standardized residuals over the entire sample period (1926:1-2003:12)

Figure 2.12: Empirical and estimated CDF’s for the alternative distributions in logarithmic (base 10) scale over the sample quantiles – the logarithmic scale highlights the left tail of each distribution
Figure 2.13: Empirical and estimated (1-CDF)’s for the alternative distributions in logarithmic (base 10) scale over the sample quantiles – the logarithmic scale highlights the right tail of each distribution

Figure 2.14: PP-plots for Gaussian and Student t distributions (1926:1-2003:12)
Figure 2.15: PP-plots for the skew-Student distribution (1926:1-2003:12)

Figure 2.16: The conditional equity premium computed under the dynamic skew-t model and the GARCH(1,1) with Gaussian error (1926:1-2003:12)
CHAPTER 3

REALIZED VOLATILITY AND MODELING STOCK RETURNS AS A MIXTURE OF NORMAL RANDOM VARIABLES: THE GARCH-SKEW-T MODEL

3.1 Introduction

The distribution of equity returns is characterized by three important stylized facts - volatility clustering, thick tails and negative skewness. Volatility clustering has been of great interest and has resulted in the ARCH class of models since the seminal work of Engle (1982) and further development by McCulloch (1985) and Bollerslev (1986). It is well documented that GARCH models represent volatility clustering effectively. However, it is also known that GARCH models with conditionally Gaussian errors are unable to account for the asymmetric fat-tailed conditional distribution of equity returns. This limitation of GARCH models with Gaussian errors has triggered alternative distributional assumptions for the conditional error distribution. McCulloch (1985) uses the symmetric Stable distribution while Bollerslev (1987) proposes Student t distribution and Nelson (1991) employs the General Error Distribution (GED) to capture the non-normality of conditional distribution. The Normal Inverse Gaussian (NIG) distribution is proposed by Barndorff-Nielsen (1995, 1997) and further investigated by Andersson (2001) and Jensen & Lunde (2001). More recently, the skew-Student (skew-t) distribution is discussed by Aas & Haff (2006) and further developed by Kim & McCulloch (2007). As pointed out by Forsberg & Bollerslev (2002), a large part of the literature in this line has explored
alternative error distributions for the GARCH models mainly for empirical and pragmatic purposes in nature. Rather than exploring different classes of distributions for the auxiliary error distribution behind the traditional GARCH modeling, the present paper derives the non-Gaussian distribution of equity returns based on empirical findings on Realized Volatility \((RV)\) measures, constructed from the summation of higher-frequency squared (demeaned) returns. We find that the distribution of stock returns conditional on RV with RV in both mean and variance is (approximately) normal while the distribution of RV is (nearly) inverted-chi-square. This paper shows that these empirical findings serve as the building blocks for the GARCH-skew-t model; thus, providing a new empirical justification for the GARCH-skew-t modeling of equity returns.

As originally proposed by Clark (1973), the Mixture-of-Distributions Hypothesis (MDH) postulates that the distribution of returns conditional on the rate of information arrival to the market is normal. Recent studies on the realized volatility \((RV)\) and stock and foreign exchange rate returns report that the distribution of returns standardized by the realized standard deviations \(\sqrt{RV}\) is approximately normal – see Andersen, Diebold, Bollerslev & Labys (2000, 2001), Ebens (1999), Areal & Taylor (2002) and Forsberg & Bollerslev (2002) among others. Hence, the realized volatility may be interpreted as an ideal measure of the latent information process. In the present paper, we find empirical evidence for the normality of monthly U.S. stock market returns conditional on \(RV\), constructed from the summation of daily squared (demeaned) returns, but with \(RV\) in both mean and variance – i.e. \(N(\mu + \beta RV, RV)\). It is noteworthy to point out that the coefficient \(\beta\) of the \(RV\)-in-mean is significantly negative so that it explains the strong association between the negative sequence of
location-adjusted returns and increases in $RV$, as reflected in the negatively skewed distribution of returns.

The realized volatility measures have been discussed and used since they were employed by French, Schwert & Stambaugh (1987), Schwert (1989) and recently by Campbell, Lettau, Malkiel, & Xu (2001) among others. As high frequency data became available more recently, realized volatility is discussed even at a daily frequency. In the present paper, we find that the distribution of monthly realized volatility is approximately the inverted-chi-square, or the scale-inverse-chi-square. It is also found that the simple GARCH(1,1) variance accounts for most of the autocorrelation of $RV$ so that $RV$ scaled by the GARCH(1,1) variance is well approximated by the inverted chi-square density function. That is, the distribution of $RV$ conditional on lagged squared returns is nearly the inverted chi-square. In short, the empirical distribution of $RV$ is close to the scale-inverse-chi-square both conditionally and unconditionally – i.e. with and without GARCH effects.

These empirical findings suggest that the marginal distribution of returns should be well approximated by the joint mixture distribution of normal and inverted chi-square. If returns are normal with $RV$ in both mean and variance, $N(\mu + \beta RV, RV)$, while $RV$ is drawn from the inverted chi-square, then the marginal distribution of returns turns out to be the skew-Student as derived by Kim & McCulloch (2007b). Thus, the building blocks underlying the skew-Student distribution are provided from empirical studies of realized volatility. Moreover, we find that this normal inverted-chi-square mixture distribution holds even conditional on past returns; implying that the conditional distribution of returns should also be well approximated by the conditional skew-t, or the GARCH-skew-t model as derived as a mixture distribution in Kim & McCulloch (2007b). This inference on both
conditional and unconditional distribution of returns based on the MDH and the
distribution of $RV$ is actually confirmed by monthly U.S. stock market returns.

Motivated by the empirical finding of the building blocks behind the skew-
Student, we estimate the skew-Student with a 40-year sample of monthly U.S. stock
market returns from Jan. 1966 to Dec. 2005 to examine if the empirical distribution of
returns is actually well approximated by the unconditional skew-Student model. We
find that the unconditional distribution of U.S. stock market returns is actually well
represented by the skew-t model as predicted based on the empirical confirmation of
the two building blocks. For the conditional distribution of returns, we use an 80-year
sample of monthly returns from Jan. 1926 to Dec. 2005 to estimate the conditional
skew-t, or the GARCH-skew-t model. It is shown that the GARCH-skew-t model
represents the three important stylized facts – volatility clustering, fat tails and
negative skewness - of U.S. stock market returns very accurately. It is noteworthy to
point out that the conditional distribution of monthly returns is substantially skewed to
the left in the sense that the left tail is very thick while the right tail is Gaussian-like.
The skew-t model is so flexible that it accounts for the substantially skewed
distribution of returns.¹

Since Clark (1973) modeled returns as a normal distribution but conditional on
latent (iid lognormal) stochastic variance $N(\mu, \tilde{V})$, Stochastic Volatility ($SV$) models
have been often employed to account for fat-tailed distribution of returns. This (latent)
stochastic variance ($SV$) model is further developed by Taylor (1982, 1986) by
incorporating volatility clustering with autoregressive $SV$ process. However, the joint
mixture distribution of normal and lognormal is known only in an integration form
without a closed-form density function. Moreover, the variance mixture distribution

¹ See Aas & Haff (2006) for discussion on the tail property of the skew-t distribution.
of normal random variables, \( N(\mu, \tilde{V}) \) is characterized by symmetry so that it is unable to capture substantial skewness as observed in stock market returns unless the two distributions are correlated.

The significantly skewed fat-tailed distribution of stock market returns motivates modeling returns as a normal variance-mean mixture distribution, \( N(\mu + \beta \tilde{V}, \tilde{V}) \), where the stochastic variance \( \tilde{V} \) in the mean captures the skewness. Building on the conditional normality along with the distribution of the \( SV \), the marginal distribution of returns is derived as a joint mixture distribution. Barndorff-Nielsen (1995, 1997) draws the \( SV \) from the Inverse Gaussian (IG) distribution and labels the consequent mixture distribution the Normal Inverse Gaussian (NIG). The NIG distribution is analytically tractable and is able to represent asymmetric tail behaviors but is limited by semi-heavy tail densities. Given the substantial skewness in addition to the excess-kurtosis of stock market returns, the skew-Student distribution is motivated as an alternative (Aas & Haff, 2006). Kim & McCulloch (2007b) derive the skew-Student as the normal variance-mean mixture distribution with the inverted chi-square as the mixing distribution. This mixture-of-distributions setting serves as the theoretical basis in the present paper because they provide the building blocks underlying the skew-t model.

In addition to asymmetric fat-tails, equity returns are characterized by serially correlated variances, known as volatility clustering, for which ARCH/GARCH class of models have been successful. Such volatility clustering is incorporated into the normal variance-mean mixture setting underlying the skew-Student model by combining observation-driven modeling approach such as GARCH-type specifications - see Kim & McCulloch (2007b) for details. Hence, the latent stochastic
variance is updated upon arrival of new observations of returns so that the conditional joint mixture distribution evolves over time with GARCH effects. This conditional skew-t model is easily cast into the common form of GARCH models with the skew-t error distribution; hence, resulting in the GARCH-skew-t model.

The success of GARCH models at capturing volatility clustering is well documented while its inability of explaining other facts such as skewness and thick tails is also recognized. To overcome this limitation of GARCH models with the common Gaussian errors, alternative error distributions have been explored mainly for practical purposes. Rather than searching for different classes of error distributions, the present paper asks what kind of non-Gaussian distributional model for equity returns may arise as a joint mixture distribution based on empirical findings on the distribution of returns conditional on $RV$ and the distribution of $RV$ itself; hence, providing the auxiliary error distribution underlying the traditional GARCH modeling. This direction of approach toward the distributional assumption for GARCH models was previously adopted by Forsberg & Bollerslev (2002), who investigate the distribution of the foreign exchange rate returns with the GARCH-NIG model but with symmetric error distribution. Due to the substantial skewness – one heavy and the other Gaussian-like tail – of monthly U.S stock market returns, the skew-t is supported against the NIG. The theoretical building blocks behind the skew-t distribution based on which this paper builds are provided by Kim & McCulloch (2007b) that decompose the skew-t into normal and the inverted chi-square distribution. With empirical evidence for the building blocks, the distributional assumption underlying the GARCH-skew-t model arises accordingly.

The rest of this paper is organized in the following order. Section 3.2 discusses the MDH and the mixture-of-distributions setting underlying the skew-Student
distribution both conditionally and unconditionally. Data and the realized volatility are explained in Section 3.3. Section 3.4 investigates the distribution of returns conditional on $RV$ and the distribution of $RV$ itself to derive their implication for the unconditional distribution of stock market returns. Section 3.5 extends the analysis from section 3.4 to the time-dependent distribution for their implication for the conditional distribution of returns. Section 3.6 concludes.

3.2 The MDH and the econometric modeling framework

The Mixture-of-Distributions-Hypothesis (MDH) postulates that returns conditional on the rate of information arrival to the market are normally distributed. Since Clark (1973) proposed iid-lognormal stochastic variance, numerous stochastic volatility models have been employed to capture the thick-tailed distribution of asset returns – see also Tauchen & Pitts (1983) and Taylor (1986). Recent studies on the realized volatility ($RV$) constructed from the summation of higher-frequency squared (demeaned) returns show that the distribution of returns standardized by $RV$ is approximately normal; thus, $RV$ may be interpreted as an ideal measure of the latent information process – see Andersen et al. (2000) among others.

Motivated by the MDH and previous studies on $RV$, the present paper investigates the distribution of monthly U.S. stock market returns conditional on $RV$. It is found that the distribution of stock market returns conditional on $RV$ is Gaussian while returns are negatively correlated with the contemporaneous $RV$ - the negative association between returns and $RV$ reflects the substantially negative skewness of the distribution of returns. Consequently, returns are specified as the normal variance-mean mixture distribution with $RV$ as the mixing random variable $N(\mu + \beta RV, RV)$.

---

2 The negative association between returns and volatility as explained by the leverage effect was first discussed by Black (1976).
which is actually supported by U.S. stock market data; hence, confirming the basic
tenet of the MDH. It is also found that the distribution of $RV$ is well approximated by
the inverted chi-square distribution both conditionally and unconditionally. These
empirical findings suggest that the marginal distribution of returns should be the
skew-Student, which is derived as a mixture distribution of normal and inverted-chi-
square in Kim & McCulloch (2007b). The following shows the modeling framework
underlying the skew-Student distribution.

The Skew-Student distribution is derived as a *normal variance-mean mixture*
distribution with the inverted chi-square as the mixing distribution.\(^3\)

\[
 r_i | \tilde{V}_i \sim N(\mu + \beta \tilde{V}_i, \tilde{V}_i) \tag{3.1}
\]

where

\[
\tilde{V}_i \sim \frac{1}{h^2 V} \chi^2_v \tag{3.2}
\]

That is, the distribution of a random variable conditional on the (latent) stochastic
variance is normal while the reciprocal of the latent variance is drawn from a chi-
square distribution. This mixture-of-distributions specification is of interest for the
present study. The probability density function of the stochastic variance is easily
derived from the chi-square density function – see Appendix B for details. Hence,
(3.2) is converted to (3.3).

\[
\tilde{V}_i \sim f_{\tilde{V}}(V_i | h, \nu) = \frac{h^\nu (\nu / 2)^{\nu/2}}{\Gamma(\nu / 2)} \frac{1}{V_i^{\nu/2+1}} \exp\left(-\frac{h^2 \nu}{2V_i}\right) \tag{3.3}
\]

The marginal distribution of a random variable $r_i$ is then derived as an expectation of
the normal density function with respect to the stochastic variance.

\[
r_i \sim \int f_r(r_i | \mu, h, \nu, \beta) = \int f_{\tilde{V}}(r_i | \mu, \beta, V_i) f_{\tilde{V}}(V_i | h, \nu) dV \tag{3.4}
\]

\(^3\) Alternatively, the skew-Student is derived as a special limiting case of the Generalized Hyperbolic
(GH) distribution; hence, labeled as the GH skew-Student’s $t$ distribution as pointed out in Aas & Haff
For $\beta \neq 0$, the skew-Student density function is given by

$$
 f_{\tilde{r}}(r_t \mid \mu, h, \nu, \beta) = \frac{h^\nu \nu^{\nu/2} |\beta|^{(\nu+1)/2} K_{(\nu+1)/2}(\nu \sqrt{(r_t - \mu)^2 + h^2 \nu})}{\sqrt{\pi} \Gamma(\nu/2) 2^{(\nu+1)/2-1} ((r_t - \mu)^2 + h^2 \nu)^{(\nu+3)/4}} \exp(\beta (r_t - \mu)) \tag{3.5}
$$

where $K_{\nu}(\cdot)$ is the modified Bessel function of the second type of order $\nu$, and $\Gamma(\cdot)$ is the Gamma function. For $\beta = 0$, the skew-t density function turns out to be the symmetric Student $t$.

$$
 f_{r}(r_t \mid \mu, h, \nu) = \frac{1}{h} t_\nu \left( \frac{r_t - \mu}{h} \right) \tag{3.6}
$$

where $t_\nu(\cdot)$ is the Student $t$ distribution with the degree of freedom $\nu$. Hence, Student $t$ is a special case of the skew-Student distribution with skewness shut off by $\beta = 0$.

The present paper reports empirical evidence for the normality of equity returns conditional on $RV$ with $RV$ in both mean and variance. This empirical finding suggests that (3.1) holds even with $RV$ in place of $\bar{V}$.

$$
 r_i \mid RV_t \sim N(\mu + \beta RV_t, RV_i) \tag{3.7}
$$

We then consider the mixing distribution $f_{RV_t}(RV_t \mid \phi)$ from which $RV$ is drawn, where $\phi$ is a vector of parameters. This mixing distribution along with the normality of returns $r_i$ conditional on $RV$ gives rise to the non-Gaussian distribution of returns according to the mixing rule (3.4).

$$
 r_t \sim f_{\tilde{r}}(r_t \mid \mu, \beta, \phi) = \int_0^\infty f_{RV_t}(RV_t \mid \phi) f_{RV_t}(r_t \mid \mu, \beta, RV_t) dRV \tag{3.8}
$$

We find that the distribution of $RV$ is well approximated by the inverted chi-square. Thus, the marginal distribution of returns computed by (3.8) should be the skew-Student (3.5) as discussed above. This inference on the marginal distribution of returns is actually confirmed empirically in this paper.
Volatility clustering is also an important feature of equity returns in addition to negative skewness and excess kurtosis. Kim & McCulloch (2007b) extends the skew-Student model to incorporate the time-dependent structure of the second moments based on observation-driven modeling approach such as GARCH-type specifications as recommended by Barndorff-Nielsen (1997). The dynamic skew-Student distribution is derived as a mixture of conditionally Gaussian and conditionally scale-inverse-chi-square distribution. If returns conditional on latent variance and past returns \( I_{t-1} \) are normal while the distribution of the latent variance is the inverted chi-square conditional on past returns \( I_{t-1} \), then the marginal distribution of returns conditional on \( I_{t-1} \) turns out to be the skew-Student. The conditional mixture-of-distributions framework for return process is specified as

\[
 r_t \mid \bar{\nu}_t, I_{t-1} \sim N(\mu + \beta \bar{\nu}_t, \bar{\nu}_t) \tag{3.9}
\]

where

\[
 \bar{\nu}_t^{-1} \mid I_{t-1} \sim \frac{1}{h(I_{t-1}; \phi)\nu} \chi^2,
\]

and \( h(I_{t-1}; \phi) \) captures volatility clustering based on observation-driven modeling approach such as GARCH type specifications.\(^4\) While the optimal functional form of the scale parameter \( h(I_{t-1}; \phi) \) could be investigated, the present paper adopts the simple GARCH(1,1) specification \( h_t^2 = c + bh_{t-1}^2 + a \varepsilon_{t-1}^2 \), in which case \( \phi = \{c, b, a\} \).

Hence, the time-varying scale parameter is determined by the long-run level of variance, the most recent variance and observation of recent returns. As the time-varying parameter of the distribution for the latent variance \( \bar{\nu}_t \) is updated, the marginal distribution of returns conditional on past returns \( I_{t-1} \) evolves accordingly.

\(^4\) The time-dependent parameter \( h(I_{t-1}; \phi) \) serves as the scale parameter for the distribution of returns while it controls the location of the stochastic variance. It is same as the conditional standard deviation of returns in the GARCH framework with some restrictions.
with time-varying volatility. The time-dependent distribution of returns conditional on past returns is then derived according to the mixing rule in the same manner as above.

\[
\begin{align*}
  r_t \mid I_{t-1} & \sim f_{\tilde{r}}(r_t \mid \mu, \beta, \nu, \phi, I_{t-1}) \\
  &= \int_{0}^{\infty} f_{\tilde{r}V} (r_t \mid \mu, \beta, \nu, I_{t-1}) f_{V} (V_t \mid \nu, \phi, I_{t-1}) dV 
\end{align*}
\]

The conditional skew-t distribution then accounts for the three important stylized facts of equity market returns – (substantial) skewness, excess kurtosis and volatility clustering. The dynamic skew-t model is then cast into the common form of GARCH models with the skew-Student error distribution for easy interpretation and practical purposes. Hence, the dynamic skew-t model is labeled as the GARCH-skew-t model in the present paper.

As reported below, returns conditional on \( RV \) and lagged squared returns are approximately normal with \( RV \) in both mean and variance. Thus, \( (3.9) \) holds even with \( RV_t \) replacing the stochastic variance \( \tilde{V}_t \).

\[
r_t \mid RV_t, I_{t-1} \sim N(\mu + \beta RV_t, RV_t)
\]

The realized volatility shows a significant autocorrelation as predicted from the significant GARCH effects. It is found that most of volatility clustering effects are captured by the simple GARCH(1,1) variance. It is shown that \( RV \) scaled by the GARCH(1,1) variance is well approximated by the inverted chi-square. Hence, the distribution of the realized volatility conditional on lagged squared returns is nearly the inverted chi-square with the time-varying parameter \( h(I_{t-1}; \phi) \).

\[
RV_t^{-1} \mid I_{t-1} \sim \frac{1}{h(I_{t-1}; \phi)^2} \chi^2_v
\]

These empirical findings are then combined to derive the joint mixture distribution of returns conditional on past returns according to the mixing rule \( (3.11) \).
\[ r_t | I_{t-1} \sim f_\gamma (r_t | \mu, \beta, \nu, \phi, I_{t-1}) \]
\[ = \int_0^\infty f_\gamma (r_t | \mu, \beta, RV_t, I_{t-1}) f_{RV|I_{t-1}} (RV_t | \nu, \phi, I_{t-1}) dRV \]

As explained above, this conditional distribution should be the dynamic skew-t or the GARCH-skew-t model.

We now turn to empirical investigation of the MDH and the distribution of the realized volatility for monthly U.S. stock market returns both conditionally and unconditionally – i.e. with and without GARCH effects. We will then examine how well the unconditional distribution of returns is explained by the implied skew-Student distribution. For the unconditional distribution of returns, we analyze a 40-year sample of daily U.S. stock market returns from Jan. 1966 to Dec. 2005, which shows relatively weak volatility clustering because the unconditional skew-Student model ignores time-dependent feature. For the conditional distribution, we study an 80-year sample of daily U.S. stock market returns from Jan. 1926 to Dec. 2005, as characterized by a very significant volatility clustering because the GARCH-skew-t model is able to capture the time-dependent property of returns, or volatility clustering.

3.3 Data and construction of RV

The value-weighted NYSE daily index returns, excluding dividends, from CRSP\(^5\) are used to construct monthly returns and its corresponding RV. Daily market returns (on day \(i\) in month \(t\)) \(r_i\) are computed as a log-difference of two adjacent end-of-day index levels \((\log p_{i,t} - \log p_{i-1,t})\). The monthly RV is measured by the summation of squared demeaned daily returns as in Schwert (1989) and Campbell et al. (2001).

\[ RV_t = \sum_{i=1}^{N_t} (r_{it} - \bar{r})^2 \]  

\(^5\) The Center for Research in Security Prices
where there are $N_t$ daily returns $r_{it}$ in month $t$, and the average of daily returns in month $t$ is denoted by $r_t$. Monthly market returns from Jan. 1926 to Dec. 2005 are shown in Figure 3.1. The monthly realized standard deviations $\sqrt{RV_t}$ are shown on the second panel of Figure 3.1.

In this paper, we investigate both unconditional and conditional distribution of equity market returns. To study the unconditional distribution, we choose a sub-sample period from January 1966 to December 2005 so that the unusually volatile period is excluded from the sample – the average realized standard deviation from Jan. 1926 to Dec. 1950 is 0.0517. It also excludes 1950s and early 1960s to separate out the relatively low volatility period - the average realized standard deviation from Jan. 1951 to Dec. 1965 is 0.0245 in comparison with 0.0354 of the following period from Jan. 1966 to Dec. 2005. However, for the analysis of the conditional distribution, we employ the whole sample period because the dynamic skew-t, or the GARCH-skew-t model accounts for volatility clustering very effectively. In short, we employ the recent 40-year sub-sample period to analyze the unconditional distribution of equity returns while using the entire sample period for the conditional distribution.

3.4 Unconditional distribution
3.4.1 Distribution of stock market returns conditional on RV

- Aggregate stock market returns from Jan. 1966 to Dec. 2005

The basic tenet of the MDH is that the distribution of returns conditional on latent information arrival to the market is Gaussian (Clark 1973). In this section, we consider the Realized Volatility ($RV$) as a measure of such unobservable information flow and investigate the distribution of returns conditional on $RV$, as motivated by previous studies on $RV$ and returns as discussed above – see Anderson et al. (2000)
for example. Because we are interested in the unconditional distribution of returns in this section, we focus on the recent 40-year sub-sample of market returns, as characterized by relatively weak volatility clustering.

To understand statistical properties of monthly U.S. stock market returns from Jan. 1966 to Dec. 2005, we analyze the normal QQ-plot of returns in Figure 3.2. As shown in the plot, the left tail of the empirical distribution is much thicker than can be represented by the normal distribution while the right tail is Gaussian-like; implying that the distribution of returns is negatively skewed. The mean, standard deviation, skewness, and kurtosis of the sample are 0.00551, 0.0430, -0.631, and 5.84 respectively, consistent with the QQ-plot. The non-normality of returns is also confirmed by the Jarque-Bera test statistic of 190.4 (p = 0.000); hence, stock returns are significantly different from Gaussian mainly due to the negatively skewed and leptokurtic property. This empirical evidence provides a motivation for departure from a normal distribution \( N(\mu, h^2) \) with constant mean \( \mu \) and variance \( h^2 \) to explore alternatives.

Andersen et al. (2000, 2001) and Ebens (1999) report that daily individual stock and foreign exchange rate returns divided by the realized standard deviations are approximately Gaussian. Because the location (or mean) of the return distribution may not be simply ignored in standardization, returns may need to be adjusted by the location and then divided by the realized standard deviations. Areal & Taylor (2002) divides mean-adjusted daily returns from FTSE-100 futures contracts by the realized standard deviations \( (r_t - \bar{r})/\sqrt{RV_t} \), where \( \bar{r} \) is the sample mean. However, it would be more efficient to use a weighted least-square estimation of the mean \( \hat{\mu}_{WLS} \) rather than the simple average to account for the heteroskedasticity of the realized volatility.
Taylor (2005) provides a brief survey of previous works on the distribution of returns conditional on $RV$. Previous works in the survey have found that (mean-adjusted) returns divided by the realized standard deviation is nearly Gaussian. The present paper also investigate if the distribution of mean-adjusted returns standardized by $RV$ $(r_i - \hat{\mu}_{wls})/\sqrt{RV_i}$ is normal. If the $RV$-standardized returns are Gaussian, $r_i \sim N(\mu, RV_i)$, then the return can also be specified as

$$r_i = \mu + \sqrt{RV_i}Z_i,$$

(3.13)

where $Z_i$ is the iid standard normal random variable, $Z_i \sim N(0,1)$. To take the heteroskedasticity of $RV$ into account, we consider the following transformed regression equation.

$$r_i / \sqrt{RV_i} = \mu / \sqrt{RV_i} + Z_i$$

The weighted least square estimate of $\mu$ is 0.0108 ($s.e. = 0.00152$). The mean, standard deviation, skewness and kurtosis of the residuals are -0.0788, 1.20, 0.103, 2.89 respectively. Note that the kurtosis is very close to that of Gaussian, reduced greatly from the kurtosis of raw returns 5.84. Much of the skewness of raw returns -0.631 also disappears. The Jarque-Bera test statistic is 1.12 ($p = 0.572$) as the excess kurtosis and much of the skewness disappear; hence, the normality of mean-adjusted $\sqrt{RV}$-standardized returns can not be rejected even with 50% significance level. The normal QQ-plot of the residual is much closer to the straight line in contrast to the QQ-plot of raw returns in Figure 3.2 though not reported here.

Given the apparent asymmetry, or the negative skewness of the empirical distribution of stock market returns as shown in Figure 3.1 and 3.2, we investigate if the standardized returns are correlated with the realized standard deviations because
the negative skewness arises as volatility is more strongly associated with a negative sequence of location-adjusted returns than positive ones. The regression of the residual from (3.13) \((r_t - \mu)/\sqrt{RV_t}\) on the realized standard deviation \(\sqrt{RV_t}\) shows a significantly negative coefficient \(-7.58\ (s.e. = 3.00)\) as shown in Table 3.1.3; hence, reflecting the negatively skewed distribution of returns. The significantly negative correlation implies that the model (3.13) may be misspecified, leading to (downward) biased estimation. Thus, this empirical finding motivates for the \(RV\)-in-mean term as a conditional mean of returns – similar to the ARCH-M model (Engle et al. 1987) in the specification.\(^6\)

Thus, we investigate the normal variance-mean mixture specification \(N(\mu + \beta RV, RV)\) with \(RV\) as the mixing random variance (3.7) to model the skewed distribution of equity returns. Motivated by the Mixture-of-Distributions Hypothesis (MDH) implied by (3.7), we study if stock returns are well explained by the variance-mean mixture distribution of normal random variables \(N(\mu + \beta RV, RV)\). That is

\[
r_t = \mu + \beta RV_t + \sqrt{RV_t} Z_t
\]

where \(Z_t\) is the iid standard normal random variable. To test the normality of stock returns conditional on \(RV\), we regress the \(\sqrt{RV}\)-scaled returns \(r_t / \sqrt{RV_t}\) on both \(1/\sqrt{RV_t}\) and \(\sqrt{RV_t}\), and then analyze the distribution of residuals.

\[
r_t / \sqrt{RV_t} = \mu / \sqrt{RV_t} + \beta \sqrt{RV_t} + Z_t
\]

Estimates of parameters \(\mu\) and \(\beta\) in (3.15) are 0.0158 (s.e. = 0.0021) and -6.52 (s.e. = 1.91) respectively. Note that estimates of all parameters including the

\(^6\) Note however that the distribution of returns conditional on past squared returns based on the ARCH-M model is not normal while returns conditional on \(RV\) are normally distributed. The distribution of the GARCH-standardized returns is negatively skewed and leptokurtic in contrast to the approximate normality of returns conditional on \(RV\).
coefficient of the $RV$-in-mean are statistically significant. The regression of the normalized returns $(r_t - \mu - \beta RV_t) / \sqrt{RV_t}$ on the realized standard deviation $\sqrt{RV_t}$ shows a statistically insignificant coefficient 1.75 (s.e. = 2.99); implying that the $RV$-in-mean term effectively captures the negative correlation between returns and $RV$; hence, it effectively accounts for the negative skewness of returns.\footnote{Note that the negative skewness can be viewed as a strong association between a negative sequence of location-adjusted returns and increases in variance.}

The mean, standard deviation, skewness, and kurtosis of the normalized returns are -0.0163, 1.19, -0.0193, 2.88 respectively. Both skewness and kurtosis are very close to those of Gaussian distribution. Note that the skewness -0.0193 is almost zero with the variance-mean mixture specification. The mean -0.0163 is also very close to zero compared with the alternative variance mixture specification. Moreover, the Jarque-Bera test statistic is further reduced to 0.365 ($p = 0.833$); hence, the normality of the normalized returns can not be rejected even with 80% significance level. The QQ-plot also shows that the distribution of residuals is quite well approximated by Gaussian distribution though not reported here. The normal PP-plot shown in Figure 3.3 also confirms the normality of the normalized returns. All these empirical results support the normal variance-mean mixture distribution $N(\mu + \beta RV_t, RV_t)$ against the alternative specification (3.13) for monthly U.S. stock market returns.

In summary, we have examined three alternative models for the distribution of equity market returns conditional on the realized volatility. Note that the normal variance mixture model $r_t = \mu + \sqrt{RV_t} Z_t$ corresponds to (3.7) with the $RV$-in-mean shut-off by $\beta = 0$. (3.7) can also be written as $r_t = \mu + \beta RV_t + RV_t^{1/2} Z_t$ (3.14) with
$Z$, drawn from the iid standard normal distribution $N(0,1)$. A normal distributional model with a constant mean and variance is also analyzed as a benchmark; hence, the comparison with other models sheds light on the importance of $RV$ at capturing excess kurtosis and skewness. It is noteworthy to point out that the $RV$-in-mean contributes to accounting for skewness while the $RV$-in-variance accounts for excess kurtosis.

Model 1: $r_i = \mu + hZ_i$

Model 2 ($\beta = 0$): $r_i = \mu + \sqrt{RV_i}Z_i$

Model 3: $r_i = \mu + \beta RV_i + \sqrt{RV_i}Z_i$

Empirical results of each model are summarized in Table 3.1. We see a clear pattern such that Model 3 outperforms Model 2 which again outperforms Model 1; thus, supporting the normal variance-mean mixture distribution $N(\mu + \beta RV_i, RV_i)$ with $RV$ as the mixing random variance for monthly U.S. stock market returns. If the realized volatilities are well approximated by the inverted chi-square distribution, we may then predict based on the mixing rule (3.8) that the marginal distribution of returns should be the skew-Student. The distribution of $RV$ and the marginal distribution of returns are discussed subsequently in following sections.

For model 2, we find a significantly negative correlation between the realized standard deviations $\sqrt{RV_i}$ and the standardized returns $(r_i - \mu)/\sqrt{RV_i}$ as reported in Table 3.1c. In contrast, the correlation between the realized standard deviations $\sqrt{RV_i}$ and the normalized returns $(r_i - \mu - \beta RV_i)/\sqrt{RV_i}$ for model 3 is statistically insignificant. Note that the coefficient of $\sqrt{RV_i}$ for model 2 is -7.58, that is close to the estimate of skewness parameter $\beta$ -6.52 from Table 3.1a. This would be explained if model 3 were correctly specified while model 2 were misspecified.
Hence, the empirical evidence supports normality of returns conditional on the realized volatility based on model 3. Thus, equity returns can be modeled as the *variance-mean mixture* distribution of normal random variables with $RV$ as the mixing random variance.

Table 3.1d shows the Ljung-Box statistics $Q(s)$ and $Q^2(s)$ to test the serial correlation of residuals and the squared residuals respectively for each model, where $s$ is the number of lags up to which the autocorrelation coefficients are tested. Note that the Ljung-Box test statistic is based on the chi-square distribution. The null hypothesis is that none of the serial correlation coefficients up to lag $s$ are different from zero. For the normalized returns, we do not find significant evidence for own serial correlation for all models as consistent with the no-arbitrage condition. Moreover, we do not find significant evidence for autocorrelation of the squared normalized returns for all models according to the portmanteau test statistics. Thus, the independence assumption of the standard normal random variable $Z$, for model 3 is supported for the 40-year sub-sample period.

**3.4.2 Unconditional distribution of the realized volatility**

We have discussed in section 3.4.1 that the distribution of equity returns conditional on $RV$ is Gaussian based on the variance-mean mixture specification; hence, confirming (3.7). Via the mixing rule of (3.8), the marginal distribution of returns is related to the distribution of the realized volatility. Because we are particularly interested in the empirical evidence for the building blocks from which the skew-Student distribution is derived, we investigate if the realized volatility is well represented by the inverted chi-square distribution (3.3). If the distribution of $RV$ is
actually well approximated by the scale-inverse-chi-square, then the marginal
distribution of returns should be well represented by the skew-Student.

The inverted chi-square distribution in (3.3) is fit to the sample of monthly
realized volatilities from Jan. 1966 to Dec. 2005 as displayed as a sub-sample in
Figure 3.1. Estimates of parameters $h$ and $\nu$ in (3.3) are 0.0278 ($s.e. = 0.000479$) and
3.58 ($s.e. = 0.218$) respectively. To analyze how the realized volatilities are
distributed and how well the scale-inverse-chi-square distribution represents them, we
compare the histogram of $RV$ with the estimated density function in Figure 3.4. It is
shown from the histogram that the empirical distribution of $RV$ from Jan. 1966 to Dec.
2005 (480 observations) is very much right-skewed – i.e. the average (mean) of $RV$
sample is bigger than the most frequently occurring $RV$ (mode). The positive
skewness of $RV$ implies that equity markets are often hit by very volatile shocks while
quite a large number of market volatilities are bound locally in a certain range as
shown in the histogram. The estimated probability density function also looks very
close to the histogram of the realized volatilities.

To examine how well the scale-inverse-chi-square distribution represents the
empirical distribution of the realized volatilities, we also analyze the PP-plot of the
scale-inverse-chi-square for the Realized Volatilities ($RV$) on the right panel of Figure
3.4. If $RV$ were actually drawn from the inverted-chi-square distribution, then the
estimated CDF of sample quantiles, called empirical probability, should be very close
to the uniform probability distribution, called theoretical probability. As shown in
Figure 3.4, the plot of theoretical probability vs. empirical probability (PP) is overall
well aligned with the 45-degree line although it deviates locally over some range.
Considering some measurement errors as proxies of volatility, time-dependent
structure of variance (volatility clustering) and the long sample period of 40 years, the
unconditional scale-inverse-chi-square distribution appears to approximate the empirical distribution of $RV$ very closely. Thus, it might be argued that the overall empirical distribution of the realized volatility is well explained by the inverted chi-square distribution.

Table 3.2 reports that the realized volatility shows a moderate autocorrelation. However, if it is scaled by the simple GARCH(1,1) variance $h_t^2$ for monthly returns, most of the serial correlation disappears. This is because the moderate autocorrelation of $RV$ is effectively captured by the simple GARCH variance. In spite of the weak GARCH effects, we will not take the weak volatility clustering into account in this section because we focus on the unconditional distribution of returns ignoring the stark autocorrelation of $RV$. Volatility clustering is taken into account in the next section when we analyze the conditional distribution of returns for the entire sample period which includes even the very volatile period of returns around the Great Depression era.

### 3.4.3 Modeling stock returns with the skew-Student distribution

The above empirical results show that the distribution of monthly stock market returns conditional on $RV$ is Gaussian based on the normal variance-mean mixture specification with $RV$ as the mixing random variance $N(\mu + \beta RV, RV)$. Given the normality of returns conditional on $RV$, the distribution of stock returns is then related to the distribution of $RV$ via the mixing rule (3.8). If the distribution of $RV$ is the inverted chi-square, the marginal distribution of returns should be approximately the skew-Student as derived by Kim & McCulloch (2007b), who model the skew-Student as the normal inverted-chi-square mixture distribution both conditionally and unconditionally. In this section, we investigate if the implied skew-t distribution of
returns is actually confirmed by the unconditional empirical distribution of the stock market returns.

Motivated by the empirical finding of the building blocks underlying the skew-t distribution, we fit the skew-Student distribution (3.5) to monthly NYSE index returns, excluding dividends, from January 1966 to December 2005 as displayed in Figure 3.1. Table 3.3 reports parameter estimates of the skew-Student distribution in comparison with those of Gaussian and Student t. Note that each distribution is related to one another with corresponding parameter restrictions. With skewness shut off by \( \beta = 0 \), the skew-Student turns out to be the symmetric Student t (3.6), which is derived as the *normal variance mixture* with the inverted chi-square as the mixing distribution. As \( \nu \to \infty \), the latent variance becomes constant; thus, the Student t approaches Gaussian.

The log-likelihood of the skew-t is the largest among the three distributions; hence, the skew-t is preferred against other distributions. Based on the Akaike Information Criterion (AIC), the common model selection criterion, the skew-t outperforms Student t which again outperforms Gaussian. However, the more conservative Bayesian Information Criterion (BIC) appears to favor Student t against the skew-t due to the penalty on one more parameter. It is noteworthy to point out that the mean of the estimated Student t 0.00697 is much bigger than that of the estimated Gaussian 0.00551 by about 1.75% at an annual rate though they are within one standard-error distance; hence, indicating that estimated means are sensitive to distributional assumptions. Meanwhile, the mean of the skew-t is almost the same as that of Gaussian. Table 3.4 shows the mean, standard deviation, skewness and kurtosis of each estimated distribution compared with those of the sample. Not surprisingly, Gaussian is disadvantaged due to its inability of explaining both
skewness and excess kurtosis. The skew-t outperforms Student t mainly due to its ability to account for the asymmetry as well as thick tails. We see from Table 3.4 that the skew-t distribution accounts for the empirical moments of the sample quite well.

To examine the overall goodness of fit of each distributional model for equity returns, PP-plots of each estimated distributions are also analyzed. Based on parameter estimates in Table 3.3, we obtain the estimated CDF of each distribution – see Kim & McCulloch (2007b) for the numerical methods to compute the CDF. Given the CDF of each distribution, we can draw PP-plots of each distribution for the monthly returns as discussed above. Figure 3.5 shows PP-plots of estimated Gaussian, symmetric t, and the skew-t distribution. The PP-plot of the Gaussian distribution deviates from the 45-degree line by a large degree while the symmetric t and skew-t are well aligned with the 45-degree line. Hence, we may reject normality in favor of Student t and Skew t for monthly market returns. However, it is hard to tell which PP-plot is closer to the 45-degree line between Student t and the skew-t, as predicted from the close values of the AIC and more conservative BIC between Student t and the skew t as reported in Table 3.3.

We may also directly compare tail distributions of each estimated distributional model with those of the empirical CDF to investigate the goodness-of-fit with focus on tail behaviors. Figure 3.6 shows that the skew-Student represents the empirical tail distribution of returns accurately, much better than the other two distributions. The skew-t distribution accounts for probability mass in both tails very accurately.

As pointed out above, stock market returns are negatively skewed and leptokurtic. The advantage of Student t over Gaussian comes from its ability to capture the leptokurtic property of returns. The superiority of the skew-Student over the other two distributions is clearly demonstrated in Figure 3.6. The skew-t distribution, which includes both skewness and excess kurtosis, provides a better fit to the empirical distribution of returns than the Gaussian distribution.
the symmetric Student t and Gaussian comes from its ability to represent both skewness and kurtosis. Thus, the skew-t represents the distribution of stock returns pretty accurately, outperforming other alternative distributions as shown in Table 3.4 and Figure 3.6. Thus, it is confirmed that the implied skew-t distribution based on its underlying building blocks represents the empirical distribution of stock market returns accurately.

3.5 Conditional distribution

3.5.1 Distribution of stock market returns conditional on $RV$ and past returns

- Aggregate stock market returns from Jan. 1926 to Dec. 2005

For the entire sample period from Jan. 1926 to Dec. 2005 for which the conditional distribution of returns is investigated, we analyze the QQ-plot of stock market returns as shown in Figure 3.7. Both tails are much thicker than Gaussian distribution but with asymmetry - the left tail is much heavier than the right tail. The mean, standard deviation, skewness and kurtosis of returns are 0.00469, 0.0537, -0.517, 10.5 respectively, consistent with the QQ-plot.

Motivated by the MDH, we investigate if the distribution of returns conditional on the realized volatility is normal. We estimate the three different models from above – i.e. model 1, model 2 and model 3 in section 3.4a. The parameter estimates are reported in Table 3.5a with standard errors in parentheses. All parameter estimates including the $RV$-in-mean coefficient are statistically significant. The negative contemporaneous correlation between returns and $RV$ reflects the significantly asymmetric (negatively skewed) distribution of returns for the sample period as discussed above. Hence, model 2 might be mis-specified with the significant $RV$-in-mean omitted. We then analyze the normalized returns of each model. As reported in Table 3.5b, the residuals from model 1 with constant mean and variance
are negatively skewed (-0.517) and leptokurtic (10.46) as shown in the QQ-plot of raw returns in Figure 3.7. The Jarque-Bera test statistic is 2257.6; hence, easily rejecting normality based on the chi-square distribution. However, such non-normality disappears dramatically when returns are normalized by $RV$ for both model 2 and 3. Note that the normality can not be rejected even with 10% and 60 % significance level respectively for model 2 and for model 3 according to the Jarque-Bera test statistic. Gaussian PP-plots are also shown for normalized returns of each model in Figure 3.8; suggesting that the distribution of returns turns out to be nearly Gaussian when they are conditioned on $RV$. We also see from the PP-plot that normalized returns based on model 3 are better approximated by Gaussian distribution than those based on model 2 though the difference does not appear big.

The normalized returns from model 2 are negatively correlated with the realized standard deviation $\sqrt{RV}$ as reported in Table 3.5c – the regression of the residuals $(r_t - \mu)/\sqrt{RV_t}$ from model 2 on $\sqrt{RV_t}$ shows a significantly negative coefficient $-6.42$ $(s.e. = 1.49)$. Meanwhile, the normalized returns from model 3 does not show any significant correlation with the realized standard deviation $\sqrt{RV}$ as reported in Table 3.5.3 as the $RV$-in-mean effectively captures the negative association between returns and $RV$. Hence, model 3 is able to fit the negatively skewed sample distribution more accurately as shown in Table 3.5b and Figure 3.8. Taking all these empirical findings into account, model 3 is preferred against the alternatives. We may then interpret $RV$ as an ideal measure of the latent information process as suggested by the MDH so that the distribution of returns conditional on $RV$ is Gaussian. Moreover, the $RV$-in-mean in model 3 accounts for the negative skewness as represented by the significantly negative coefficient.
Table 3.5d shows the Ljung-Box statistics $Q(s)$ and $Q^2(s)$ to test the autocorrelation of residuals and the squared residuals respectively for each model, where $s$ is the number of lags up to which the autocorrelation coefficients are tested. Recall that the null hypothesis is that none of the serial correlation coefficients up to lag $s$ are different from zero. For the normalized returns, we do not find significant evidence for own serial correlation for both model 2 and 3 as consistent with the no-arbitrage condition. For the squared normalized returns, Model 2 shows a weak first-order autocorrelation. However, such a weak correlation of squared residuals is not observed for model 3 as the portmanteau statistic $Q^2(1)$ drops by more than half. The well-known GARCH effects of monthly returns are eliminated by the normalization of returns based on model 3, \( \frac{(r_t - \mu - \beta RV_t)}{\sqrt{RV_t}} \). Thus, the normality of returns conditional on $RV$ holds even when the information set $I_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$ is augmented as the conditioning information. That is,

$$r_t \mid RV_t, I_{t-1} \sim N(\mu + \beta RV_t, RV_t)$$

3.5.2 Conditional distribution of the realized volatility

The empirical findings in section 3.5a suggest that the distribution of stock market returns conditional on $RV$ and past returns $I_{t-1}$ is normal based on the variance-mean mixture specification with $RV$ as the mixing random variance. The conditional distribution of returns $(r_t \mid I_{t-1})$ is then related to the conditional distribution of $RV$ ($RV_t \mid I_{t-1}$) via this empirical finding according to the mixing rule (3.11). Since we are interested in empirical evidence for the distributional assumption underlying the GARCH-skew-t model, we analyze the distribution of the realized volatility based on the inverted chi-square distribution.
\[ RV^{-1}_{t-1} | I_{t-1} \sim \frac{1}{(\bar{h} h(I_{t-1}; \phi))^2} \chi^2_{\nu} \]  

(3.16)

Equivalently,

\[ RV_t | I_{t-1} \sim f_{RV}(RV_t | h_t, \bar{h}_t, \nu, I_{t-1}) = \frac{(\bar{h} h_t)^{\nu} (\nu / 2)^{\nu/2}}{\Gamma(\nu / 2)} \frac{1}{RV_t^{\nu / 2 + 1}} \exp\left(-\frac{(\bar{h} h_t)^2 \nu}{2RV_t}\right) \]

Hence, scaling the realized volatility \( RV_t \) by the time-dependent (observation-driven) scale parameter \( h_t^2 \) leads to the unconditional scale-inverse-chi-square density function.

\[ RV_t / h_t^2 \sim f_{RV}(RV_t | \bar{h}, \nu) = \frac{\bar{h}^{\nu} (\nu / 2)^{\nu/2}}{\Gamma(\nu / 2)} \frac{1}{RV_t^{\nu / 2 + 1}} \exp\left(-\frac{\bar{h}^2 \nu}{2RV_t}\right) \]  

(3.17)

As shown in Table 3.6, the raw realized volatilities show strong autocorrelation or significant GARCH effects. Such temporal dependence breaks down dramatically when \( RV \) is scaled by the simple GARCH(1,1) variance of returns.\(^8\) Note that the Q(1), Q(5) and Q(10) Ljung-Box test statistics drop by large amounts; implying that the simple GARCH(1,1) variance accounts for most of volatility clustering represented by the autocorrelation of the realized volatility. Thus, we are led to estimate the distribution of \( RV \) conditional on lagged squared returns with the inverted chi-square density function (3.17).

We fit the inverted chi-square distribution to the scaled realized volatilities \( RV_t / h_t^2 \) and then analyze the PP-plot for the goodness of fit. Estimates of parameters \( \bar{h} \) and \( \nu \) are 0.627 (s.e. = 0.00722) and 4.02 (s.e. = 0.175) respectively for the sample period from Jan. 1926 to Dec. 2005. Figure 3.9 shows the histogram of the scale-adjusted \( RV \) \( (RV_t / h_t^2) \) and the estimated scale-inverse-chi-square density

\(^8\) The GARCH(1,1) variance \( h_t^2 = c + bh_{t-1}^2 + a\epsilon_{t-1}^2 \) is computed for the stock market returns.
function. It appears that the inverted-chi-square represents the conditional distribution of $RV$ very well. Figure 3.9 also provides the PP-plot of the inverted chi-square distribution for the scale-adjusted $RV$ to examine the goodness of fit. According to the PP-plot, the conditional distribution of the realized volatility is well approximated by the conditional scale-inverse-chi-square distribution (3.16). Note however that slightly more probability mass is accumulated over a range of the second quarter of sample quantiles than can be represented by the scale-inverse-chi-square distribution. To investigate if such empirical probability accumulation is particular to the sample period or general phenomena for the realized volatility, we also analyze the distribution of the realized volatility for different sample periods. Figure 3.10 shows the histogram and estimated density function along with the PP-plot over the recent 40 years from Jan. 1966 – Dec. 2005 (480 observations). Figure 3.11 also shows them over the recent 30 years from Jan. 1976 – Dec. 2005 (360 observations). Comparing those in Figures 3.9, 3.10 and 3.11, we find that the locally over-accumulated probability mass is particular to the sample period of interest. Moreover, we see that the conditional distribution of the realized volatility over recent decades is very well represented by the conditional scale-inverse-chi-square distribution. Hence, these empirical findings support the inverted-chi-square as an approximation for the conditional distribution of the realized volatility.

3.5.3 Modeling stock market returns with the GARCH-skew-t model

The empirical findings from section 3.5.1 and 3.5.2 serve as the building blocks for the distributional assumption underlying the GARCH-skew-t model. These empirical findings suggest according to the mixing rule (3.11) that the distribution of returns conditional on past returns should be well approximated by the conditional skew-t, or the GARCH-skew-t model. In this section, we examine if the implied GARCH-skew-t
model is actually confirmed by the conditional distribution of monthly U.S. stock market returns.

The skew-Student distribution is extended to incorporate the time-dependent property of second moments (volatility clustering), which has been of great interest in financial econometrics. Volatility clustering is embedded in the skew-t by adopting observation-driven modeling approach such as GARCH-type specifications. The inverted-chi-square distribution of the latent variance evolves via the time-varying parameter $h(I_{t-1}; \phi)$ in (3.10), where $I_{t-1}$ is an information set at time $t-1$ and $\phi$ is a vector of parameters. As the scale-parameter is updated in response to new observation of returns, the stochastic variance evolves so that it accounts for the volatility clustering while the normal variance-mean mixture setting is preserved over time. The scale-parameter $h(I_{t-1}; \phi)$ is specified as GARCH(1,1) $h_t^2 = c + bh_{t-1}^2 + a\varepsilon_{t-1}^2$ in the present study rather than the optimal specification is investigated because it captures volatility clustering effectively and is commonly adopted for equity returns.

The mixing rule (3.11) with (3.9) and (3.10) leads to the time-varying skew-t distribution which is cast into the common form of GARCH models for easy analytical interpretation – see Kim & McCulloch (2007b) for details.

$$r_t = \mu + h_t \varepsilon_t,$$
$$\varepsilon_t \sim Skewt(\mu^*, h^*, \nu, \bar{\beta})$$
$$\text{with } E[\varepsilon_t] = 0 \text{ and } Var[\varepsilon_t] = 1$$

(3.18)

where $\mu^*$ and $h^*$ are set so that $\varepsilon_t$ has zero mean and unit variance. The specification (3.18) allows $h_t$ to be interpreted as the conditional standard deviation of returns $r_t$. Note that $\bar{\beta}(= \beta h^*)$ and $\nu$ are invariant parameters which make $Skewt(\mu^*, h^*, \nu, \bar{\beta})$ the location-scale family – i.e. $E(\varepsilon_t - \mu^*)/h^* \sim Skewt(0,1,\nu, \bar{\beta})$. 

73
This specification (3.18) allows easy comparison with GARCH models with different
distributional assumptions for $\varepsilon$, such as Gaussian and Student t. (3.18) with zero
skewness ($\beta = 0$) corresponds to the GARCH model with Student t error, which is
again derived as the time-dependent normal variance mixture distribution – i.e. (3.9)
and (3.10) with skewness turned off by $\beta = 0$. (3.18) with Gaussian error is simply
the case in which the degree of freedom $\nu$ approaches the infinity $\infty$ with zero
skewness.

Estimation of parameters in (3.18) might be computationally challenging due
to the combination of the GARCH specification and the skew-t distribution. One
solution to this problem is to employ Quasi-MLE (QMLE) following Bollerslev &
Wooldridge (1992) and Glosten, Jaganathan & Runkle (1989) among others assuming
normality even for non-Gaussian errors since it is known to provide asymptotically
consistent estimators under certain conditions. Once the location parameter $\mu$ and
GARCH parameters $\phi = (c, b, a)$ are estimated by the QMLE, the distribution of
standardized residuals $(r_t - \mu)/h_t$ is analyzed with distributional assumptions such as
Gaussian, Student-t and the skew-t with zero mean and unit variance restriction.
Parameter estimates are reported in Table 3.7a and 3.7b. The log-likelihood increases
from Gaussian to symmetric t and to the skew-t. Based on model selection criteria of
both AIC and more conservative BIC, skew t outperforms symmetric t which again
outperforms Gaussian.

To analyze the empirical distribution of GARCH(1,1)-standardized residuals,
the QQ-plot is drawn in Figure 3.12. The distribution of residuals is characterized by a
very thick left tail and Gaussian-like right tail, consistent with skewness and kurtosis
of -0.792 and 5.27 respectively. It is noteworthy to point out that the GARCH-
standardized returns are far different from normal while the distribution of returns conditional on $RV$ is nearly normal. We then analyze PP-plots of the GARCH standardized residuals based on such probability assumptions as Gaussian, Student t and the skew-Student. According to PP-plots of Gaussian and Student t in Figure 3.13, Student t is better aligned with the 45-degree line; indicating that Student t explains the conditional distribution of returns better than Gaussian does. If they are compared with the PP-plot of the skew-t on the right panel of Figure 3.13, we see that the skew-t outperforms both Student t and Gaussian in terms of the goodness of fit, consistent with the results from the model selection criteria of AIC and BIC in Table 3.7b. All these empirical results suggest that the skew-Student represents stock returns very well even after adjusting for the time-varying variances.

We find that estimates of the location parameter $\mu$ is very sensitive to distributional assumptions for both unconditional and conditional models – see estimates in Table 3.3 for unconditional models and Table 3.8 for conditional models. Table 3.8 reports maximum likelihood estimation of GARCH(1,1) models with both Gaussian and Student t errors. Note that estimated location parameter $\mu$ is much bigger for the Student t (0.00854/mo. = 10.3%/yr.) than for Gaussian (0.00646/mo. = 7.75%/yr.) – the difference amounts to 2.5% at an annual rate. This empirical result indicates the quantitatively important effect of distributional assumptions on parameter estimates, especially the location parameter $\mu$ for the present example. Meanwhile, based on model selection criteria of AIC and more conservative BIC, GARCH(1,1)-Student t outperforms GARCH(1,1)-normal in its overall goodness of fit mainly due to the fat-tailed property of returns remaining even after adjusting for GARCH effects as clearly shown in the QQ-plot of Figure 3.12.
Taking the above empirical results into account, adopting the QMLE may give the GARCH(1,1)-Student-t model a great disadvantage, which fits sample data much better than GARCH(1,1)-normal does. Nevertheless, the conditional standard deviation $h_t$ is not much different across different distributional assumptions as shown in Figure 3.14. Thus, we are motivated to adopt the Generalized Least Square (GLS) – i.e. Weighted Least Square (WLS) – estimation with $h_t$ as the heteroskedastic standard deviations. Note that the GLS estimators have the minimum variance among the class of all linear unbiased estimators even with non-Gaussian errors (Hamilton, 1994). Below is the transformed regression equation for the GLS estimation.

$$r_t / h_t = \mu / h_t + \varepsilon_t$$

(3.19)

with $E[\varepsilon_t] = 0$ and $Var[\varepsilon_t] = 1$

The location parameter $\mu$ is then estimated simply by the least squares, which gives estimate $\mu_{GLS}$ of 0.00536 (s.e. = 0.00135) if $h_t$ from GARCH(1,1)-normal is employed while it gives $\mu_{GLS}$ of 0.00534 (s.e. = 0.00138) if $h_t$ from GARCH(1,1)-Student t is employed.\textsuperscript{9} We then analyze the standardized residuals $(r_t - \mu_{GLS}) / h_t$ with such distributional assumptions as Gaussian, Student t and the Skew-Student.

The mean, standard deviation, skewness and kurtosis of residuals $(r_t - \mu_{GLS}) / h_t$ are -0.00580, 0.999, -0.791 and 5.27 respectively. Note that the standardized residuals are very negatively skewed as well as leptokurtic. Hence, the standard normal distribution is limited by its inability to capture skewness and kurtosis. Student t is also disadvantaged given the substantial skewness. The QQ-plot

\textsuperscript{9} We take $h_t$ computed under Gaussian error distribution in the remainder of this paper because GARCH(1,1) standard deviations are not much different across different distributional assumptions.
of residuals is also analyzed to examine both skewness and kurtosis in Figure 3.15. The left-tail of the residual distribution is much thicker than can be represented by the Gaussian while the right tail is close to the Gaussian. The substantial skewness – i.e. one tail is heavy while the other is Gaussian-like - of the distribution of the standardized residuals is well represented by the Skew-Student. This is the motivation for the use of the skew-t distribution along with the GARCH modeling.

In Table 3.9, we report parameter estimates of Student t and the skew-Student distribution with zero mean and unit variance restriction for the GLS standardized residuals \((r_j - \mu_{GLS})/\sigma_j\). According to the model selection criteria of both AIC and BIC, the skew-t outperforms Student t which again outperforms Gaussian. This result is as predicted from the analysis of the QQ-plot in Figure 3.15 in which returns are substantially skewed and leptokurtic. For the graphical test of the goodness of fit, we analyze PP-plots of each distributional model for the GLS-standardized residuals in Figure 3.16. The PP-plot of the skew-t is located almost along the 45-degree line while both Gaussian and Student t are somewhat away from the straight line, consistent with the results from the model selection criteria of AIC and BIC reported in Table 3.9. Thus, we confirm that the distribution of equity market returns is well represented by the skew-Student even after adjusting for GARCH effects.

Table 3.10 shows the mean, standard deviation, skewness and kurtosis of each estimated density function in comparison with those of the residuals. The skew-t captures both skewness and kurtosis of GARCH residuals effectively while Student t is limited in its ability of representing the asymmetric distribution of returns. Nevertheless, the Student t fits the data better than Gaussian. Because we are also interested in how well each distributional model fits tails of the residuals, we compare the CDF (or 1-CDF) of each estimated distribution with the empirical CDF (1-CDF).
As shown in Figure 3.17, both tails are quite accurately represented by the skew-t model while they are systematically under- or over-estimated by Student t and Gaussian. Taking all these empirical findings into account, we see that the implied GARCH-skew-t model is strongly confirmed by the stock market returns. Moreover, the GARCH-skew-t represents the three important stylized facts of U.S. stock market returns – volatility clustering, fat-tails and negative skewness - very accurately.

3.6 Conclusion
We find that the distribution of monthly U.S. stock market returns conditional on the realized volatility is normal based on the variance-mean mixture specification with $RV$ as the mixing random variance $N(\mu + \beta RV, RV)$; hence, confirming the basic tenet of the mixture-of-distribution hypothesis originally proposed by Clark (1973). We also find that the distribution of the realized volatility is well represented by the inverted-chi-square both conditionally and unconditionally – i.e. with and without GARCH effects. These empirical findings serve as the building blocks underlying the skew-Student distribution. Hence, the empirical finding of the building blocks suggest according to the mixing rule that the distribution of returns should be well approximated by the skew-Student as derived as a normal inverted-chi-square mixture distribution in Kim & McCulloch (2007b). We find that the implied skew-Student distribution is actually well supported by the U.S. stock data both conditionally and unconditionally. Thus, this paper provides a new empirical justification for the distributional assumption underlying the GARCH-skew-t model. Moreover, we find a direct analytical connection between the realized volatility and the latent stochastic variance in the normal variance-mean mixture modeling of U.S. stock returns.
### Table 3.1: Estimation and statistical properties of alternative models

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>0.00551</td>
<td>0.0108</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td>(0.00196)</td>
<td>(0.00152)</td>
<td>(0.00211)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>-</td>
<td>0</td>
<td>-6.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.91)</td>
</tr>
<tr>
<td>( h )</td>
<td>0.0430</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.00146)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Parameter estimates of each model with standard errors in parentheses

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0</td>
<td>-0.0788</td>
<td>-0.0163</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1</td>
<td>1.20</td>
<td>1.19</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.631</td>
<td>0.103</td>
<td>-0.0193</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.84</td>
<td>2.89</td>
<td>2.88</td>
</tr>
<tr>
<td>JB</td>
<td>190.4</td>
<td>1.12</td>
<td>0.365</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.572)</td>
<td>(0.833)</td>
</tr>
</tbody>
</table>

(b) Summary statistics of normalized returns and Jarque-Bera test statistics

<table>
<thead>
<tr>
<th></th>
<th>Model 2 ( (r_i - \mu) / \sqrt{RV_i} )</th>
<th>Model 3 ( (r_i - \mu \beta RV_i) / \sqrt{RV_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.189</td>
<td>-0.0782</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>( \sqrt{RV_i} )</td>
<td>-7.58</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>(3.00)</td>
<td>(2.99)</td>
</tr>
</tbody>
</table>

(c) Regression of normalized returns on the realized standard deviations for model 2 and 3 with standard errors in parentheses

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q(1) (p-value)</td>
<td>0.763</td>
<td>0.513</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>(0.383)</td>
<td>(0.474)</td>
<td>(0.315)</td>
</tr>
<tr>
<td>Q(10) (p-value)</td>
<td>11.0</td>
<td>11.2</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>(0.354)</td>
<td>(0.343)</td>
<td>(0.193)</td>
</tr>
<tr>
<td>Q²(1) (p-value)</td>
<td>3.61</td>
<td>2.57</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>(0.0576)</td>
<td>(0.109)</td>
<td>(0.0795)</td>
</tr>
<tr>
<td>Q²(10) (p-value)</td>
<td>13.8</td>
<td>40.4</td>
<td>43.6</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.0000145)</td>
<td>(0.0000384)</td>
</tr>
</tbody>
</table>

(d) Ljung-Box portmanteau statistics for normalized returns and squared normalized returns with p-values in parentheses
<table>
<thead>
<tr>
<th></th>
<th>$RV_i$</th>
<th>$RV_i / h_i^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q(1)</td>
<td>14.4</td>
<td>2.78</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000151)</td>
<td>(0.0957)</td>
</tr>
<tr>
<td>Q(5)</td>
<td>34.5</td>
<td>5.27</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.0000191)</td>
<td>(0.384)</td>
</tr>
<tr>
<td>Q(10)</td>
<td>41.0</td>
<td>6.67</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000114)</td>
<td>(0.757)</td>
</tr>
</tbody>
</table>

Table 3.2: Ljung-Box portmanteau statistics for the realized volatility $RV_i$ and the scale-adjusted realized volatility $RV_i / h_i^2$ with p-values in parentheses

<table>
<thead>
<tr>
<th></th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.00551</td>
<td>0.00697</td>
<td>0.0202</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.00196)</td>
<td>(0.00185)</td>
<td>(0.00732)</td>
</tr>
<tr>
<td>$h$</td>
<td>0.0430</td>
<td>0.0427</td>
<td>0.0356</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.00146)</td>
<td>(0.00196)</td>
<td>(0.00188)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>$\infty$</td>
<td>5.22</td>
<td>7.60</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0</td>
<td>(1.93)</td>
<td>(2.41)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>829.0</td>
<td>843.3</td>
<td>845.9</td>
</tr>
<tr>
<td>AIC</td>
<td>-1653.9</td>
<td>-1680.7</td>
<td>-1683.8</td>
</tr>
<tr>
<td>BIC</td>
<td>-1645.6</td>
<td>-1668.2</td>
<td>-1667.1</td>
</tr>
</tbody>
</table>

Table 3.3: Parameter estimates of Gaussian, symmetric t, and skew t distribution with standard errors in parentheses. The log-likelihood and model selection criteria of AIC and BIC are provided for each distributional model.

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[\tilde{r}]$</td>
<td>0.00551</td>
<td>0.00551</td>
<td>0.00697</td>
<td>0.00553</td>
</tr>
<tr>
<td>$Std[\tilde{r}]$</td>
<td>0.0430</td>
<td>0.0430</td>
<td>0.0427</td>
<td>0.0429</td>
</tr>
<tr>
<td>$Skw[\tilde{r}]$</td>
<td>-0.634</td>
<td>0</td>
<td>0</td>
<td>-0.646</td>
</tr>
<tr>
<td>$Kts[\tilde{r}]$</td>
<td>5.84</td>
<td>3</td>
<td>5.22</td>
<td>5.01</td>
</tr>
</tbody>
</table>

Table 3.4: Mean, standard deviation, skewness and kurtosis of each estimated distributional model for monthly returns in comparison with those of the sample
Table 3.5: Estimation and statistical properties of alternative models: model 1, 2 and 3

a) Parameter estimates with standard errors in parentheses for each alternative model

b) Summary statistics of residuals and Jarque-Bera test statistics for each alternative model

c) Regression of residuals on the realized standard deviations for model 2 and 3 with standard errors in parentheses

d) Ljung-Box portmanteau statistics for normalized returns and squared normalized returns with p-values in parentheses
### (a)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.00469</td>
<td>0.0140</td>
<td>0.0183</td>
</tr>
<tr>
<td></td>
<td>(0.00173)</td>
<td>(0.00105)</td>
<td>(0.00124)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-</td>
<td>0</td>
<td>-6.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.02)</td>
</tr>
<tr>
<td>$h$</td>
<td>0.0537</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.00129)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### (b)

<table>
<thead>
<tr>
<th></th>
<th>Model 1 $(r_i - \mu)/h$</th>
<th>Model 2 $(r_i - \mu)/\sqrt{RV_t}$</th>
<th>Model 3 $(r_i - \mu - \beta RV_r)/\sqrt{RV_t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0</td>
<td>-0.125</td>
<td>-0.0327</td>
</tr>
<tr>
<td>Standard</td>
<td>1</td>
<td>1.23</td>
<td>1.21</td>
</tr>
<tr>
<td>deviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.517</td>
<td>0.137</td>
<td>0.0428</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>10.5</td>
<td>2.83</td>
<td>2.88</td>
</tr>
<tr>
<td>JB</td>
<td>2257.6</td>
<td>4.29</td>
<td>0.974</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.117)</td>
<td>(0.615)</td>
</tr>
</tbody>
</table>

### (c)

<table>
<thead>
<tr>
<th></th>
<th>Model 2 $(r_i - \mu)/\sqrt{RV_t}$</th>
<th>Model 3 $(r_i - \mu - \beta RV_r)/\sqrt{RV_t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.122</td>
<td>-0.102</td>
</tr>
<tr>
<td></td>
<td>(0.694)</td>
<td>(0.0690)</td>
</tr>
<tr>
<td>$\sqrt{RV_t}$</td>
<td>-6.42</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>(1.49)</td>
<td>(1.48)</td>
</tr>
</tbody>
</table>

### (d)

<table>
<thead>
<tr>
<th>Test</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q(1)</td>
<td>8.25</td>
<td>0.143</td>
<td>0.155</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.00408)</td>
<td>(0.705)</td>
<td>(0.694)</td>
</tr>
<tr>
<td>Q(10)</td>
<td>33.3</td>
<td>16.0</td>
<td>23.1</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000240)</td>
<td>(0.0994)</td>
<td>(0.0105)</td>
</tr>
<tr>
<td>$Q^2$ (1)</td>
<td>49.9</td>
<td>5.24</td>
<td>2.23</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.0221)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>$Q^2$ (10)</td>
<td>431.8</td>
<td>30.3</td>
<td>25.4</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000761)</td>
<td>(0.00457)</td>
</tr>
<tr>
<td></td>
<td>( RV_t )</td>
<td>( RV_t / h_t^2 )</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>Q(1)</td>
<td>268.4</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000837)</td>
<td></td>
</tr>
<tr>
<td>Q(5)</td>
<td>713.8</td>
<td>13.4</td>
<td></td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.0197)</td>
<td></td>
</tr>
<tr>
<td>Q(10)</td>
<td>1231.4</td>
<td>15.8</td>
<td></td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.106)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.6: Ljung-Box portmanteau statistics for \( RV \) and the scale-adjusted \( RV \) with p-values in parentheses

<table>
<thead>
<tr>
<th></th>
<th>( \mu )</th>
<th>( c )</th>
<th>( b )</th>
<th>( a )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00646</td>
<td>5.99 \times 10^{-5}</td>
<td>0.120</td>
<td>0.862</td>
</tr>
<tr>
<td></td>
<td>(2.00 \times 10^{-6})</td>
<td>(8.47 \times 10^{-8})</td>
<td>(0.00199)</td>
<td>(0.000662)</td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th></th>
<th>Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaussian</td>
</tr>
<tr>
<td>( \mu^* )</td>
<td>0</td>
</tr>
<tr>
<td>( h^* )</td>
<td>1</td>
</tr>
<tr>
<td>( \nu )</td>
<td>\infty</td>
</tr>
<tr>
<td>( \bar{\beta} )</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-1360.8</td>
</tr>
<tr>
<td></td>
<td>2725.6</td>
</tr>
<tr>
<td></td>
<td>2735.4</td>
</tr>
</tbody>
</table>

(b)

Table 3.7: Estimation of GARCH models with alternative distributional assumptions: Gaussian, symmetric t and skew-t
a) QMLE estimates for GARCH(1,1) model for returns from Jan. 1926 to Dec. 2005 (960 observations)
b) Parameter estimates of Student t and the skew-Student for GARCH(1,1) standardized residuals \( (r_t - \mu) / h_t \)
Table 3.8: Parameter estimates of the GARCH (1,1) model with Gaussian and Student t error distribution respectively. In parentheses are standard errors of estimates. The log-likelihood and model selection criteria of the AIC and BIC are also provided for each model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GARCH(1,1)-normal</th>
<th>GARCH(1,1)-Student t</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>0.00646 ( (2.00 \times 10^{-6}) )</td>
<td>0.00854 ( (1.77 \times 10^{-6}) )</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>5.99 ( \times 10^{-5} )</td>
<td>0.000109</td>
</tr>
<tr>
<td>( c )</td>
<td>0.12033 ( (8.47 \times 10^{-9}) )</td>
<td>0.119 ( (1.88 \times 10^{-9}) )</td>
</tr>
<tr>
<td>( b )</td>
<td>0.862 ( (0.000199) )</td>
<td>0.839 ( (0.000246) )</td>
</tr>
<tr>
<td>( a )</td>
<td>0.000662 ( (0.000283) )</td>
<td>0.000703 ( (0.000317) )</td>
</tr>
<tr>
<td>( \nu )</td>
<td>( \infty )</td>
<td>( (1.39) )</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>1596.6</td>
<td>1622</td>
</tr>
<tr>
<td>AIC</td>
<td>-3185.1</td>
<td>-3234</td>
</tr>
<tr>
<td>BIC</td>
<td>-3165.7</td>
<td>-3209.6</td>
</tr>
</tbody>
</table>

Table 3.9: Parameter estimates of Gaussian, Student t and the skew-Student distribution for GLS standardized residuals \( (r_t - \mu_{GLS}) / h_t \) with the log-likelihood and model selection criteria of the AIC and BIC.

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu^* )</td>
<td>0</td>
<td>0</td>
<td>0.779</td>
</tr>
<tr>
<td>( h^* )</td>
<td>1</td>
<td>0.852</td>
<td>0.814</td>
</tr>
<tr>
<td>( \nu )</td>
<td>( \infty )</td>
<td>7.29</td>
<td>10.6</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0</td>
<td>(1.41)</td>
<td>(3.21)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-1360.5</td>
<td>-1337.9</td>
<td>-1322.6</td>
</tr>
<tr>
<td>AIC</td>
<td>2725</td>
<td>2681.8</td>
<td>2653.2</td>
</tr>
<tr>
<td>BIC</td>
<td>2734.7</td>
<td>2696.4</td>
<td>2672.7</td>
</tr>
</tbody>
</table>
Table 3.10: Mean, standard deviation, skewness and kurtosis of GLS residuals for estimated alternative distributional models against the sample summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Residuals</th>
<th>Gaussian</th>
<th>Symmetric t</th>
<th>Skew t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[\tilde{r}]$</td>
<td>-0.00580</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$Std[\tilde{r}]$</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$Skw[\tilde{r}]$</td>
<td>-0.791</td>
<td>0</td>
<td>0</td>
<td>-0.825</td>
</tr>
<tr>
<td>$Kts[\tilde{r}]$</td>
<td>5.27</td>
<td>3</td>
<td>4.82</td>
<td>6.21</td>
</tr>
</tbody>
</table>
Figure 3.1: Monthly value-weighted NYSE returns and the realized standard deviations $\sqrt{\bar{R}_t}$ from Jan. 1926 to Dec. 2005 (960 observations)

Figure 3.2: QQ-plot of monthly U.S. stock market returns from Jan. 1966 to Dec. 2005 (480 observations)
Figure 3.3: Gaussian PP-plots of normalized returns from each model

Figure 3.4: Histogram and the estimated inverted chi-square density function are shown on the left panel while the inverted-chi-square PP-plot is on the right panel for the monthly realized volatilities from Jan. 1966 to Dec. 2005 (480 observations)
Figure 3.5: PP-plots of estimated Gaussian, Student t and the skew-t distribution

Figure 3.6: Tail probability distribution of returns with CDF and 1-CDF in logarithmic scale (base 10)
Figure 3.7: QQ-plot of monthly U.S. stock market returns from Jan. 1926 to Dec. 2005 (960 observations)

Figure 3.8: Gaussian PP-plots of normalized returns for each alternative model
Figure 3.9: Histogram and the inverted chi-square PP-plot for the scale-adjusted realized volatilities $RV_t / h_t^2$ from Jan. 1926 to Dec. 2005 (960 observations)

Figure 3.10: Histogram and the inverted chi-square PP-plot for the scale-adjusted realized volatilities $RV_t / h_t^2$ from Jan. 1966 to Dec. 2005 (480 observations)

Figure 3.11: Histogram and the inverted chi-square PP-plot for the scale-adjusted realized volatilities $RV_t / h_t^2$ from Jan. 1976 to Dec. 2005 (360 observations)
Figure 3.12: QQ-plot of GARCH(1,1)-standardized residuals

Figure 3.13: PP-plots of Gaussian, Student t and the skew-Student distribution for GARCH(1,1)-standardized residuals
Figure 3.14: Absolute demeaned monthly returns $|r_i - \bar{r}|$ and GARCH(1,1) standard deviations (std. dev.) with Gaussian and Student t error distribution.
Figure 3.15: QQ-plot of the GLS-standardized residuals \((r_t - \mu_{GLS})/h_t\) from Jan. 1926 to Dec. 2005 (960 observations)

Figure 3.16: PP-plots of Gaussian, Student t and the skew-t distribution for the GLS normalized residuals \((r_t - \mu_{GLS})/h_t\)
Figure 3.17: Probability distribution of tails for GLS residuals with CDF and 1-CDF in logarithmic scale (base 10)
4.1 Introduction

It is well known that the standard consumption-based asset pricing model, or the C-CAPM (Lucas 1978, Mehra & Prescott 1985) suffers from asset return puzzles such as high equity premium (6-8%) and low riskfree rate (1-2%) that are hard to explain with a reasonable degree of risk aversion and the CRRA utility function which are commonly employed in macroeconomic literature. Hence, these puzzles raise critical questions for the common macroeconomic modeling exercise. This paper addresses these puzzles by showing that a negatively skewed and fat-tailed error distribution, arising mainly due to parameter uncertainty, can explain the historically high equity premium and low riskfree rate so that it provides a potential solution to those puzzles. Mehra & Prescott (1985) find that about 0.35% is the maximum risk premium the standard C-CAPM can produce with a risk-free rate between 0 and 4% when the mean and standard deviation of consumption growth are assumed 1.8% and 3.6% respectively. The present paper also takes 1.8% mean and 3.6% standard deviation for consumption growth and then generates 6-7% equity risk premium, 0.5-2% risk-free rate and 5-7% dividend yield with the risk aversion coefficient between 7 and 10 solely relying on negative skewness and excess kurtosis that might arise due to stochastic nature of the uncertain parameter underlying the endowment process.
In macroeconomic literature, the Gaussian distribution is often assumed to represent the error distribution underlying consumption or real GDP growth. A normal distribution may actually well explain the historically observed consumption growth rate especially for the low frequency data beyond monthly over the postwar period. However, agents – especially investors in financial markets - may not see future events in the same way as they do over the past. There might be some degree of asymmetry between the future and the past from investors’ perspective. For example, investors may perceive future dividend (endowment) process as drawn from a negatively skewed and fat-tailed non-Gaussian distribution so that they may put higher probability on worse outcomes than better ones compared with the Gaussian case. As a potential source of such a non-Gaussian distribution, this paper considers parameter uncertainty. Agents may have doubts or uncertainty about parameters underlying the Data Generating Process (DGP), especially of dividend or consumption growth when they look at future events which ultimately determine their (consumption) welfare. In this case, an investor’s perceived distribution of the DGP may lead to non-Gaussian as he/she would compute the probability as a mixture distribution. In this sense, the DGP perceived by agents could be different from the DGP estimated by classical econometricians relying on only past events. This is the motivation to impose stochastic feature on parameters of the Gaussian distribution underlying the endowment process.

In the model economy of this paper, the consumer-investor draws the endowment growth rate from a Gaussian distribution but conditional on a stochastic parameter. The random parameter captures the uncertainty with which the agent views

---

1 In contrast, Barro (2006) shows that real GDP and consumption growth are quite negatively skewed and thick-tailed when a longer time period is considered so that rare disaster events such as the Great Depression or wartime events may make the distribution non-Gaussian.
the DGP for the future dividend (consumption) prospect. Because the endowment process is Gaussian conditional on a random parameter, the investor may derive or compute the subsequent distribution for the dividend (consumption) growth as a mixture of Gaussian and the probability distribution from which the random parameter is drawn. This mixture of Gaussian and the other distribution leads to a skewed and thick-tailed non-Gaussian error distribution for this DGP. In particular, we consider the Normal Inverse Gaussian (NIG) distribution as a mixture-of-distributions framework (Barndorff-Nielsen 1995, 1997). A Normal random variable whose variance parameter is again drawn from the Inverse Gaussian distribution results in the NIG distribution which can represent a probability density with skewness and semi-heavy tail-thickness. The NIG distribution converges to the normal distribution if the variance parameter is constant while it leads to a skewed and leptokurtic probability density as the variance parameter becomes more uncertain. The NIG distribution gives rise to a bounded expected utility due to its semi-heavy tail density in contrast to concerns of Geweke (2001) and Weitzman (2007) that point out that leptokurtic distributions may result in explosive Moment Generating Function (MGF) and explosive expected utility when the CRRA utility function is employed; thus, suggesting a use of priors to avoid the unboundedness.

By employing the NIG distribution in a mixture-of-distributions framework, we can analyze quantitatively how parameter uncertainty may lead to a skewed and leptokurtic non-Gaussian error distribution underlying the endowment process. Furthermore, the NIG distribution leads to a finite MGF and expectation of the CRRA utility function due to its semi-heavy tail property so that it allows a tractable and quantitative analysis of how higher-order risks such as skewness and kurtosis contribute to important asset pricing features such as equity premium, riskfree rate,
dividend yield and etc. Parameter uncertainty may also raise the perceived volatility
of the dividend process; however, we focus on its impact on skewness and kurtosis
only by restricting the variance at the historically observed level. This paper shows
that negative skewness and high kurtosis alone, resulting from parameter uncertainty,
can explain most of the high equity risk premium (6-8%) as well as low riskfree rate
(1-2%) that the Gaussian distribution fails to explain with the CRRA utility function
with the degree of risk aversion below 10.

Rietz (1987) and Barro (2006) more recently show that the same outcome
presented in the present paper can also be produced by introducing a rare disaster
event into the model. That is, high equity premium and low risk-free rate could be
explained if agents believed that the DGP of consumption includes a disaster event
such as the Great Depression or wartime events even with a small probability. The
rare disaster event changes higher-order moments of expected utility so that risk
premium and risk-free rate may change subsequently so that it explains the past
performance of asset returns. The present work is similar to theirs in a sense that
parameter uncertainty due to agents’ uncertainty of parameter estimates or the DGP
when they look over the future can change higher-order moments of the expected
utility. Nevertheless, the present work is different from their approach in a sense that
this paper employs parameter uncertainty as the main source for the perceived
negatively skewed and leptokurtic distribution, whereas the Rietz-Barro model
employs a rare disaster event as a main source in order to change the higher-order
moments (or risk). In addition, the Rietz-Barro model is based on a regime switching
model to capture a rare disaster event, whereas this paper employs the NIG
distribution as a tractable non-Gaussian distribution to explain how parameter
uncertainty leads to higher-order risk such as negative skewness and excess kurtosis
that affect asset pricing features. In this aspect, this work may be close to Weitzman (2007) that also relies on parameter uncertainty but had to introduce priors to overcome the unboundedness of the MGF and expected utility as suggested by Geweke (2001) because the symmetric Student’s t distribution arises for the perceived distribution in their Bayesian framework. In contrast, the leptokurtic distribution in the present work leads to a finite MGF and expected utility because the NIG distribution has semi-heavy tails. Furthermore, the uncertainty in both mean and variance of the Gaussian error distribution underlying the DGP in this paper leads to a more flexible non-Gaussian distribution with or without asymmetry while maintaining tractability.

In response to the asset return puzzles, there have been many approaches to solve them; however, they may be categorized into two main groups – changing either the Stochastic Discount Factor (SDF) $M$ or the underlying probability distribution for the mathematical expectation $E[\cdot]$ in the canonical form of asset pricing equation $P_t = E_t[M_{t+1}\,X_{t+1}]$ where $X$ is the payoff and $P$ is asset price. For example, Campbell & Cochrane (1999) proposes the habit formation model that makes the SDF different from that of the CRRA utility function so that the risk aversion increases to between 80 and 120; thus, solving the puzzles. The other group of approaches is based on the idea that asset pricing features are influenced by the underlying probability distribution for fundamentals such as consumption (dividend) process. This aspect has been explored by Rietz (1988), Barsky & Delong (1993), Cecchetti, Lam & Mark (2000), Barro (2006) and Weitzman (2007) among many others. Along with the second line of approach, the present work investigates how such features as equity premium, risk-free rate, dividend yield and etc. are affected by parameter uncertainty which leads to a negatively skewed and thick-tailed distribution but while maintaining
the same mean and variance as those of historically observed consumption growth rates.

The remainder of this paper is organized as follows. Asset pricing features of the Lucas one-tree model with iid shocks are discussed in relation to the moment generating function in section 4.2. The NIG distribution as derived as a mixture distribution of Normal and Inverse Gaussian is discussed in section 4.3. Section 4.4 shows how parameter uncertainty as a source of negatively skewed and leptokurtic distribution can explain asset return puzzles building on discussions on section 4.2 and 4.3. Asset-pricing features under the non-Gaussian distribution are contrasted with those of the Gaussian case in section 4.5. The relevant simulation results follow in section 4.6. Section 4.7 concludes.

4.2 Consumption-based asset pricing in terms of the Moment Generating Function (MGF)

The canonical form of asset pricing equations is given by \( P_t = E_t[M_{t+1}X_{t+1}] \), where \( M \) is the Stochastic Discount Factor (SDF) and \( X \) is the underlying pay-off – see Cochrane (2005) for more detail. If the asset is equity, the equation is then given by \( P_t = E_t[M_{t+1}(P_{t+1} + D_{t+1})] \), where \( D \) is dividends. For the consumption-based asset pricing model in Lucas (1978) and Mehra & Prescott (1985), \( M \) is given by a ratio of marginal utilities – i.e. Inter-temporal Marginal Rate of Substitution (IMRS) – and a power function of gross consumption growth if the preference is given by the CRRA utility function. Thus, the equity price is given by \( P_t = E_t[M_{t+1}(P_{t+1} + C_{t+1})] = \beta E_t[(C_{t+1}/C_t)^\gamma (P_{t+1} + C_{t+1})] \). This equation can be recursively substituted and be reformulated as below with the transversality condition (TVC) imposed to prevent bubbles from occurring.
\[
P_t = \beta E_t[(C_{t+1}/C_t)^\gamma (P_{t+1} + C_{t+1})]
= E_t[\sum_{j=1}^{\infty} \beta^j (C_{t+j}/C_t)^\gamma C_{t+j}]
= C_t \sum_{j=1}^{\infty} \beta^j E_t[(C_{t+j}/C_t)^{-\gamma}]
= C_t \sum_{j=1}^{\infty} \exp(-\rho j) E_t[\exp((1-\gamma)g_{t+j, t})]
\]

where \( g_{t+j, t} = \log(C_{t+j}/C_t) \) is log consumption growth rate over a period from time \( t \) to \( t+j \) and \( \beta = \exp(-\rho) \) is a time discount factor with a time discount rate \( \rho \). Note that \( E_t[\exp((1-\gamma)g_{t+j, t})] \) is the MGF \( m(\theta) \) with \( \theta = 1-\gamma \). Assuming independent and identical distribution for consumption growth process for simplicity, \( E_t[\exp((1-\gamma)g_{t+j, t})] = (E_t[\exp((1-\gamma)g_{t+j, t})])^j \). Substituting this equality into the above equation, we can derive the price-dividend ratio as below: 

\[
\frac{P_t}{C_t} = \frac{\exp(-\rho)E_t[\exp((1-\gamma)g_{t+j, t})]}{1 - \exp(-\rho)E_t[\exp((1-\gamma)g_{t+j, t})]}
\]

The dividend yield in logarithm is then given by 

\[
\log(1 + \frac{C_t}{P_t}) = \rho - \log E_t[\exp((1-\gamma)g_{t+j, t})]
= \rho - \log(m(1-\gamma))
\]

Thus, the log dividend yield is expressed as a function of the MGF \( m(1-\gamma) \). Note that the logarithm of MGF \( m(\theta) \) is a Cumulant Generating Function (CGF) \( k(\theta) \) – i.e. \( k(\theta) = \log \{m(\theta)\} \). Utilizing the CGF expression, the dividend yield in logarithm can be simplified further as

\[
\log(1 + \frac{C_t}{P_t}) = \rho - k(1-\gamma)
\]

---

2 Note that the dividend yield, or dividend-price ratio is the same as the consumption-wealth ratio because dividend equals consumption in the Lucas one-tree model economy.
The total rate of gross stock return is computed as \( R_{t+1} = (P_{t+1} + C_{t+1}) / P_t \). Note that the dividend yield is given as a constant for the Lucas 1-tree economy with iid assumption for the endowment process. Utilizing this, the gross stock return is

\[
R_{t+1} = \frac{P_{t+1}}{P_t} (1 + \frac{C_{t+1}}{P_{t+1}}) = \frac{C_{t+1}}{C_t} (1 + \frac{C_{t+1}}{P_{t+1}}) = \frac{\exp(g_{t,t+1}) \exp(\rho)}{m(1-\gamma)}
\]

where \( g_{t,t+1} = \log(C_{t+1} / C_t) \) is log consumption growth from time \( t \) to \( t+1 \). Taking logarithm on both sides, we derive the one-period log stock return from \( t \) to \( t+1 \) as

\[
r_{t+1} = \log(R_{t+1}) = g_{t,t+1} + \rho - \log\{m(1-\gamma)\} = g_{t,t+1} + \rho - k(1-\gamma).
\]

This equation reveals the volatility mismatch puzzle because stock returns are as volatile as consumption growth rate according to the equation. One leeway to avoid this puzzling feature of the Lucas one-tree model might be to impose a levered consumption as in Campbell (1986) – i.e. \( D = C^\lambda \) where \( \lambda > 1 \) will generate higher volatility for stock returns. Another approach to avoid this discrepancy of volatility is to decompose the source of total income into stock market traded asset and non-traded one so that stock dividends are only a small part of consumption (Kim & McCulloch 2007a).

The one-period expected stock return is then computed and is expressed in terms of MGF or CGF as below.

\[
\log E_t[R_{t+1}] = \log\{m(1)\} + \rho - \log\{m(1-\gamma)\}
\]

\[
= \rho + k(1) - k(1-\gamma)
\]

The gross risk-free rate is simply the inverse of the expected SDF. The log risk-free rate can then be computed and expressed in MGF or CGF terms as below.
\[ r_t^f = \log R_t^f = -\log \{E_t[M_{t+1}]\} \]
\[ = -\log \{\beta E_t[(C_{t+1} / C_t)^{-\gamma}]\} \]
\[ = \rho - \log E_t[\exp(-\gamma g_{t+1})] \]
\[ = \rho - \log m(-\gamma) \]
\[ = \rho - k(-\gamma) \]

Given the log expected gross stock return and risk-free rate, the equity risk premium is computed as a gap between the two. The equity premium can then be expressed in MGF or CGF as below.

\[ ep_t = \log E_t[R_{t+1}] - r_t^f \]
\[ = \log \{m(1)\} + \log \{m(-\gamma)\} - \log \{m(1 - \gamma)\} \]
\[ = k(1) + k(-\gamma) - k(1 - \gamma) \]

In short, the main asset-pricing features of the Lucas 1-tree model economy with iid shocks to the log consumption growth have been discussed and computed above in terms of MGF and CGF. Note that the asset pricing features are all expressed as a linear combination of \( \rho \), \( k(1) \), \( k(-\gamma) \) and \( k(1 - \gamma) \) - these are the essential elements of the Lucas 1-tree asset-pricing model with iid shocks for the endowment process. In following sections, we will study how the MGF \( m(1) \), \( m(-\gamma) \) and \( m(1 - \gamma) \) or CGF \( k(1) \), \( k(-\gamma) \) and \( k(1 - \gamma) \) are affected by distributional assumptions. In particular, we investigate the source of non-Gaussian distribution and its impact on asset pricing features via the MGF or CGF.

**4.3 Parameter uncertainty, the MGF and the NIG distribution**

Since asset pricing features such as equity risk premium, risk-free rate, and etc. are determined by the MGF, it is worthwhile to analyze how the MGF is affected by the underlying probability distribution of payoffs, or dividends for stocks. In this and next section, we investigate how parameter uncertainty as a possible source of negatively skewed and leptokurtic distribution can affect the MGF so that it can explain the
puzzling features of consumption-based asset pricing model as noted by Mehra & Prescott (1985).

Suppose that the (consumption) endowment process of the economy is characterized by a normal distribution: \((g - \mu) \sim iidN(0,V)\) or \(g = \mu + V^{1/2} \varepsilon\) with \(\varepsilon \sim iidN(0,1)\), where \(g\) is the log growth rate of consumption (dividend) and \(V\) is the variance. If the consumer investor doubts or is uncertain about the variance parameter \(V\) so that \(V\) is perceived as a stochastic variable by the agent, then the distribution of consumption growth \(g\) is computed as a mixture of normal and the probability distribution of \(V\). This mixture of normal random variables may lead to a thick-tailed distribution and the mixture distribution is computed as an expectation of the normal density function with respect to the random variance \((g - \mu) \sim \int_{\varepsilon} f_{g|\varepsilon}(g | V)g(V)dV\), where \(f_{g|\varepsilon}(g | V)\) is the normal density function conditional on variance \(V\) and \(g(V)\) is the probability density from which the variance parameter \(V\) is drawn.

It is well known that if variance \(V\) were drawn from the inverse chi-square distribution, the unconditional distribution of consumption growth \(g\) would then be (computed and perceived by the agent as) the Student’s t distribution. In this case, the MGF explodes so that the equity premium goes to infinity while the riskfree rate goes to minus infinity as noted by Weitzman (2007); thus, leading to anti-puzzling asset pricing features. As also pointed out by Geweke (2001), the expectation of the CRRA utility function with the Student’s t distribution does not exist. This explosive nature of the MGF and expected CRRA utility due to the heavy-tail distribution requires some restriction on the tail probability by imposing priors in their Bayesian framework.
The main reason for the Student t distribution to generate such an infinite MGF - hence leading to an infinite equity premium and minus infinite risk free rate - is its heavy tail which is characterized by a power function ($\sim x^\alpha$). In order to avoid such explosiveness of MGF and expected utility, we may consider a distribution with semi-heavy tails – i.e. a product of exponential function ($\sim e^x$) and power function ($\sim x^\alpha$) - so that the tail thickness is intermediate in between power function and exponential function. This paper considers the Normal Inverse Gaussian (NIG) distribution that is characterized by semi-heavy tails and is derived as a mixture distribution of Normal and Inverse Gaussian.

In addition to thick tails, investors may believe that worse outcomes would occur more frequently than better ones so that the perceived distribution of payoffs (consumption growth) may be negatively skewed. Skewness may be interpreted as a stronger association between the growth rate of payoff and its volatility in one direction than the other. To represent skewness based on this interpretation, we may specify the distribution of log consumption growth rate as $(g - \mu) \sim iidN(\beta\tilde{V}, \tilde{V})$. For $\beta < 0$, a decline of growth rate is more strongly associated with an increase in the variance so that the distribution is characterized by negative skewness. As will be discussed in the following, if the random variance $\tilde{V}$ is drawn from the Inverse Gaussian (IG), the mixture distribution leads to the NIG distribution, which has a semi-heavy tail distribution so that the expectation of the CRRA utility function is finite in contrast to concerns of Weitzman (2007) and Geweke (2001).

The NIG distribution was introduced into finance by Barndorff-Nielsen (1995) and its extension to a dynamic modeling (by incorporating temporal dependence of random variables) is discussed by Barndorff-Nielsen (1997) and Jensen & Lunde.
The NIG distribution can be derived as a mixture of Normal (N) and Inverse Gaussian (IG) distribution – it can also be derived as a special limiting case of the Generalized Hyperbolic (GH) distribution. The NIG distribution is tractable as well as useful in a sense that its probability density is characterized by semi-heavy tails and can represent skewness, which is important for financial data.

Suppose \( \tilde{g} | \tilde{V} \sim iid N(\mu + \beta \tilde{V}, \tilde{V}) \) where \( \tilde{V} \sim IG(\delta, \sqrt{\alpha^2 - \beta^2}) \)

where \( IG(V | \delta, \sqrt{\alpha^2 - \beta^2}) = \frac{\delta V^{-3/2}}{\sqrt{2\pi}} \exp\{\delta \sqrt{\alpha^2 - \beta^2} - \frac{1}{2} (\delta^2 V^{-1} + (\alpha^2 - \beta^2) V)\} \)

The unconditional distribution is then derived as a mixture of the two distributions – i.e. Normal and Inverse Gaussian – and is denoted by \( \tilde{g} \sim NIG(\alpha, \beta, \mu, \delta) \). Defining \( \alpha = \delta \alpha \), \( \beta = \delta \beta \) and \( \rho = \beta / \alpha \) for its location-scale family, its MGF \( m(\theta) \) is then given by \( m(\theta; \alpha, \beta, \mu, \delta) = \exp[\alpha \sqrt{1 - \rho^2} - \sqrt{1 - (\rho + (\delta / \alpha) \theta)^2}] + \theta \mu \). The CGF is then given by \( k(\theta; \alpha, \beta, \mu, \delta) = \alpha \sqrt{1 - \rho^2} - \sqrt{1 - (\rho + (\delta / \alpha) \theta)^2} + \theta \mu \). The mean, variance, skewness and kurtosis of the NIG distribution are given by

\[
E[y] = \kappa_1 = \mu + \frac{\bar{\rho} \delta}{\sqrt{1 - \rho^2}} \quad Var[y] = \kappa_2 = \frac{\delta^2}{\alpha (\sqrt{1 - \rho^2})^3} \\
Skw[y] = \frac{\kappa_3}{(\kappa_2)^{3/2}} = \frac{3 \bar{\rho}}{\alpha^{3/2} (1 - \rho^2)^{1/4}} \quad Kts[y] = 3 + \frac{\kappa_4}{(\kappa_2)^2} = 3 + \frac{1 + 4 \rho^2}{\alpha \sqrt{1 - \rho^2}}
\]

where \( \kappa_1, \kappa_2, \kappa_3 \) and \( \kappa_4 \) are the first four cumulants of the NIG distribution.\(^3\)

The CGF with the relevant restrictions for the asset-pricing application discussed above are given below.

\(^3\) The 3rd and 4th cumulants are \( \kappa_3 = 3 \delta \rho \frac{\bar{\rho}}{\alpha (1 - \rho^2)^{3/2}} \) and \( \kappa_4 = 3 \delta^4 \frac{4 \bar{\rho}^2 + 1}{\alpha (1 - \rho^2)^{7/2}} \) respectively.
\[
\begin{align*}
  k(1) &= \alpha \{1 - \rho^2 - \sqrt{1 - (\rho + (\delta / \alpha))^2}\} + \mu \\
  k(-\gamma) &= \alpha \{1 - \rho^2 - \sqrt{1 - (\rho - (\delta / \alpha)\gamma)^2}\} - \gamma \mu \\
  k(1 - \gamma) &= \alpha \{1 - \rho^2 - \sqrt{1 - (\rho + (\delta / \alpha)(1-\gamma))^2}\} + (1 - \gamma) \mu
\end{align*}
\]

For a symmetric case, \( \beta = 0 \) is set because \( \beta \) is the parameter which captures asymmetry of the distribution. In other words, \( \beta < 0 \) leads to a negatively skewed distribution while \( \beta > 0 \) leads to a positively skewed distribution. Substituting \( \beta = 0 \) into the CGF, \( k(\theta; \alpha, 0, \mu, \delta) = \alpha \{1 - \sqrt{1 - \delta^2 \theta^2 / \alpha^2}\} + \theta \mu \). Since \( \theta = 1, -1, 1 - \gamma \) are of particular interest for the (Lucas 1-tree) asset-pricing model with iid shocks, the CGF with these restrictions are derived as below.

\[
\begin{align*}
  k(1) &= \alpha \{1 - \sqrt{1 - \delta^2 / \alpha^2}\} + \mu \\
  k(-\gamma) &= \alpha \{1 - \sqrt{1 - \delta^2 \gamma^2 / \alpha^2}\} - \gamma \mu \\
  k(1 - \gamma) &= \alpha \{1 - \sqrt{1 - \delta^2 (1-\gamma)^2 / \alpha^2}\} + (1 - \gamma) \mu
\end{align*}
\]

The mean, variance, skewness and kurtosis for the symmetric NIG are given by

\[
\begin{align*}
  E[g] &= \kappa_1 = \mu \\
  Var[g] &= \kappa_2 = \frac{\delta^2}{\alpha} = \frac{\delta}{\alpha} \\
  Skw[g] &= \frac{\kappa_3}{(\kappa_2)^{3/2}} = 0 \\
  Kts[g] &= 3 + 3(1 + \frac{1}{\alpha}) = 3(1 + \frac{1}{\delta \alpha})
\end{align*}
\]

Note that the NIG distribution converges to the normal distribution as \( \alpha \to \infty \). That is, adjusting \( \alpha \) controls tail thickness up to the normal distribution; thus, the impact of kurtosis on asset-pricing features can be analyzed with the NIG distribution. For example, if \( \mu = 0 \) and \( \delta = \alpha \), then the NIG has zero mean and unit variance but with \( Kts = 3(1 + 1 / \alpha^2) \). Thus, we can generate whatever kurtosis beyond that of the standard normal distribution (\( Kts = 3 \)) while keeping the same mean and variance as
that of the standard normal. Thus, the impact of high-order moment, or risk such as kurtosis on asset pricing features can be studied while the first two or three moments are kept constant at those of the standard normal distribution.

4.4 Asset-pricing features with the NIG distribution

In the above, we discussed how asset pricing features such as equity premium, risk free rate, and etc. are expressed in terms of the MGF (or CGF) and how the MGF and higher order moments are affected by underlying distribution that may arise when market participants are uncertain about the DGP in terms of parameters. In particular, we pay attention to the NIG distribution for the perceived endowment process, which is derived - or computed by the consumer investor with uncertainty about the data generating process - as a mixture of Normal and Inverse Gaussian distribution. Incorporating parameter uncertainty and its consequent (perceived) NIG distribution into the asset pricing framework, we can now analyze how asset pricing features are affected, especially in comparison with those of the Gaussian case without any parameter uncertainty.

We investigate below how asset-pricing features are determined when the log consumption growth rate is perceived by the consumer-investor as Gaussian but conditional on the uncertain (stochastic) variance parameter that is drawn from the IG distribution: \( \tilde{g} \mid \tilde{V} \sim iidN(\mu + \beta \tilde{V}, \tilde{V}) \) where \( \tilde{V} \sim IG(\delta, \sqrt{\alpha^2 - \beta^2}) \). Note that this mixture-of-distributions is able to represent a negatively skewed and semi-heavy tails so that it may capture more frequent extreme events with negative skewness in contrast to the Gaussian case. Important asset pricing features such as dividend yield, expected stock return, riskfree rate and the risk premium are then computed and expressed by its underlying parameters as below.

108
\[
\log(1 + \frac{C_t}{P_t}) = \rho - \sqrt{\alpha^2 - \beta^2} + \sqrt{\alpha^2 - (\beta + \delta (1 - \gamma))^2} - (1 - \gamma)\mu
\]

\[
\log E_t[R_{t+1}] = \rho + \gamma\mu - \sqrt{\alpha^2 - (\beta + \delta)^2} + \sqrt{\alpha^2 - (\beta + \delta (1 - \gamma))^2}
\]

\[
r_t' = \rho + \gamma\mu - \sqrt{\alpha^2 - \beta^2} + \sqrt{\alpha^2 - (\beta - \delta \gamma)^2}
\]

\[
e_p = \sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + \delta)^2} - \sqrt{\alpha^2 - (\beta - \delta \gamma)^2} + \sqrt{\alpha^2 - (\beta + \delta (1 - \gamma))^2}
\]

For the symmetric case (\( \beta = 0 \)), the main features of C-CAPM are simplified further.

\[
\log(1 + \frac{C_t}{P_t}) = \rho - \bar{\alpha} + \sqrt{\bar{\alpha}^2 - \delta^2 (1 - \gamma)^2} - (1 - \gamma)\mu
\]

\[
\log E_t[R_{t+1}] = \rho - \sqrt{\bar{\alpha}^2 - \delta^2} + \sqrt{\bar{\alpha}^2 - \delta^2 (1 - \gamma)^2} + \gamma\mu
\]

\[
r_t' = \rho + \gamma\mu + \sqrt{\bar{\alpha}^2 - \delta^2 \gamma^2} - \bar{\alpha}
\]

\[
e_p = \bar{\alpha} - \sqrt{\bar{\alpha}^2 - \delta^2} - \sqrt{\bar{\alpha}^2 - \delta^2 \gamma^2} + \sqrt{\bar{\alpha}^2 - \delta^2 (1 - \gamma)^2}
\]

The Elasticity of Inter-temporal Substitution (EIS) is determined as an exact reciprocal of the Relative Risk Aversion (RRA) coefficient for the CRRA utility function. This may lead to too low or high risk-free rate so that it might be hard to match observed data. This aspect may motivate the use of the Epstein-Zin (EZ) preferences to break the unnecessarily strong link between the EIS and the RRA – Epstein & Zin (1990, 1991); thus, additional flexibility may be attained in matching data. The CGF expression of asset-pricing features with the EZ utility function is derived in Martin (2007) and the result is summarized below. We can then compute and compare features of the asset pricing model under the NIG with those of the Gaussian distribution. The CGF expression of asset-pricing features under the EZ preference is given below (assuming iid shocks to log consumption growth) with \( \phi = (1 - \gamma)/(1 - 1/\psi) \) where \( \psi \) is the EIS. Substituting the relevant CGF of the NIG
distribution for the asset-pricing equations under the EZ preference, we can then compute them in terms of the underlying parameters.

\[
\log(1 + \frac{C_t}{P_t}) = \rho - k(1 - \gamma) / \phi
\]

\[
\log E_t[R_{t+1}] = \rho + k(1) - k(1 - \gamma) - k(1 - \gamma)(1/\phi - 1)
\]

\[
r_t^f = \rho - k(-\gamma) - k(1 - \gamma)(1/\phi - 1)
\]

\[
ep_t = k(1) + k(-\gamma) - k(1 - \gamma)
\]

### 4.5 Asset-pricing features with Gaussian distribution

As a benchmark case of the consumption-based asset pricing model, we consider the simple Gaussian distribution for the endowment process for comparison with that of Non-Gaussian case. If log consumption growth rates were drawn from the iid normal distribution, the relevant CGF \(k(1), k(-\gamma), k(1 - \gamma)\) that determine the main asset-pricing features of the Lucas 1-tree model are then computed as below.

\[
k(1) = E[g] + \text{Var}(g) / 2
\]

\[
k(-\gamma) = -\gamma E[g] + \gamma^2 \text{Var}(g) / 2
\]

\[
k(1 - \gamma) = (1 - \gamma) E[g] + (1 - \gamma)^2 \text{Var}(g) / 2
\]

Putting these CGF expressions in the asset pricing framework, we can derive the following asset pricing features for the Lucas 1-tree economy.

\[
\log(1 + \frac{C_t}{P_t}) = \rho - (1 - \gamma)E_t[g_{t,t+1}] - (1 - \gamma)^2 \text{Var}_t(g_{t,t+1}) / 2
\]

\[
\log E_t[R_{t+1}] = \rho + \gamma E_t[g_{t,t+1}] + (1 - (1 - \gamma)^2) \text{Var}_t(g_{t,t+1}) / 2
\]

\[
r_t^f = \rho + \gamma E_t[g_{t,t+1}] - \gamma^2 \text{Var}_t(g_{t,t+1}) / 2
\]

\[
ep_t = \gamma \text{Var}_t(g_{t,t+1})
\]
4.6 Simulation

To point out the significant impact of non-Gaussian distribution on asset pricing features, the resulting outcome of the Lucas asset-pricing model with the NIG distribution is shown in Table 4.1 in contrast with those of the Gaussian case for the following parameter values: the CRRA coefficient $\gamma = 9.2$, the subjective time discount rate $\rho = 0.02$, the mean growth rate $E[g] = 0.02$, variance $Var[g] = 0.05^2$, skewness $Skw[g] = -0.50001$ and kurtosis $Kts[g] = 15.001$, where $g$ is the log consumption growth rate. In this example, the first two moments – the mean and variance - are assumed the same for the NIG and Gaussian distribution in order to consider only impacts of higher order moments – skewness and kurtosis – on the asset pricing features so that the resulting difference between the two distributional assumptions comes from higher order moments perceived by the agent with uncertainty about the DGP. The mean $E[\tilde{V}]$ and standard deviation $SD[\tilde{V}]$ of the uncertain (stochastic) variance parameter underlying the conditionally Gaussian distribution are 0.0025 and 0.0049 respectively.

This example shows how higher-order moments such as skewness and kurtosis contribute to resolution of the riskfree rate puzzle which arises under the Gaussian distributional assumption for the perceived endowment process. In this example, we can see that the riskfree rate drops from 9.82% to 1.08% as extreme events are perceived by the consumer-investor to occur with negative skewness much more frequently than that of the Gaussian case. In contrast with the constant parameter, the stochastic aspect of the variance parameter can generate negative skewness -0.50001 and high kurtosis 15.001 so that the risk averse investor tends to demand more of riskfree asset (bond); hence, the bond price goes up while the return from the riskless
bond drops dramatically. We can also see that the equity risk premium also goes up as the risk-averse investor perceives the growth of payout (dividend) from stocks to be characterized by negative and more frequent extreme events. From the summation of the riskfree rate and equity premium, the expected stock return is determined as shown in Table 4.1.

The pdf of the perceived NIG and Gaussian distribution employed for this example to generate the above features are illustrated graphically in Figure 4.1.1. Note that they have the same mean and variance but with different skewness and kurtosis – also note that the degree of skewness and kurtosis are commonly observed in financial data although it may be difficult to observe such a distribution for (low frequency beyond monthly) aggregate consumption or GDP growth over the postwar era unless it is assumed that future events may be perceived differently than the past ones so that the perceived DGP is characterized as a negatively skewed and thick-tailed ones. Figure 4.1.2 shows the distribution of the random (or uncertain) parameter that generates the skewed and leptokurtic NIG distribution in Figure 4.1.1. Note that parameter uncertainty is the main source that makes the NIG distribution depart from the Gaussian distribution; otherwise, the NIG would converge to the Gaussian distribution if uncertainty of the parameter disappeared.

Taking the mean and variance of consumption growth rate in Mehra & Prescott (1985) and Rietz(1987) as given – i.e. 1.8% mean and 3.6% standard deviation, the impact of skewness and kurtosis on asset-pricing features such as equity premium, risk-free rate and price-dividend ratio are analyzed over some range of risk aversion and time discount rate coefficient in Table 4.2 and 4.3 for the CRRA and the EZ

---

4 In contrast, Barro (2006) finds empirical evidence that the distribution of GDP (consumption) growth is negatively skewed and leptokurtic considering rare disaster events such as deep recessions, economic crisis, war and etc. by taking a long sample period.
utility function respectively. They also show the corresponding uncertainty (stochastic nature) of the variance parameter that generates the skewness and kurtosis. It is also shown that such parameter uncertainty is able to generate equity risk premium, risk-free rate and price-dividend ratio close to those observed in the real economy with the degree of risk aversion between 7 and 10.

As the consumer-investor becomes uncertain about parameter values underlying the endowment process, he/she may perceive worse events may occur more frequently than better ones. This perception of investors makes risk-free bonds much more attractive than such a risky asset as equity. Thus, they would ask for higher risk premium for holding equity asset but a low rate for riskfree bonds as they become uncertain about the DGP of consumption growth. Therefore, the size of the equity premium may be explained by the degree of uncertainty with which agents look at future consumption process. It also appears that there is a trade-off between the degree of risk aversion and parameter uncertainty. In other words, the more risk-averse agents become, the less uncertainty about the economy’s DGP would be required to generate the same level of risk premium. (It may be worthwhile to investigate this aspect further and a quantitative representation of such a trade-off.)

Table 4.2 shows that a consumer investor with a CRRA utility function with a relatively low risk aversion coefficient (7-10) asks for 6-7% risk premium for holding equity while demanding 0.5-2% return for holding a riskfree bond. It is also shown that the time discount factor should be adjusted accordingly to get the risk-free rate bound in between 0.5 and 2%. This is because the Elasticity of Inter-temporal Substitution (EIS) is determined precisely as a reciprocal of the relative risk aversion coefficient for the CRRA utility function. In other words, as risk aversion increases, the EIS declines; thus, it may be hard to match data. To prevent risk-free rate from
declining too much below 0%, the subjective time discount factor should increase. That is, as agents become more impatient, they would try to save less and consume more so that the risk-free rate increases. This is how adjustment of the time discount factor makes the risk-free rate bound in between 0.5-2%. In order to break up such a strong link between the EIS and the RRA, the Epstein-Zin (EZ) utility function may be considered. Table 4.3 shows the results when the EZ utility function is used in place of the CRRA utility function. The results are quite similar to those under the CRRA utility function except with more flexibility especially for the time discount factor $\beta$.

Figures 4.2 to 4.5 illustrate graphically that uncertainty of the (variance) parameter works as an important source for non-Gaussian distributional features such as negative skewness and high kurtosis. We can see from the figures that the negative skewness and leptokurtosis tend to disappear as uncertainty of the variance parameter dies out; thus, the NIG converges to the Gaussian case. Following Figures from 4.2 to 4.5, uncertainty of the DGP in terms of the stochastic (variance) parameter diminishes so that the corresponding subjective pdf for consumption growth becomes less negatively skewed and less leptokurtic so that the NIG distribution approaches the Gaussian. Moreover, the asset pricing features in Table 4.2 and 4.3 corresponding to the skewness and kurtosis in Figures 4.2 to 4.5 imply that the required degree of risk aversion for the historic values of equity premium, riskfree rate and price-dividend ratio increases following Figures from 4.2 to 4.5. In short, declining uncertainty about the DGP appears to require higher risk aversion so that it indicates a trade-off between the degree of uncertainty and the degree of risk aversion.
4.7 Conclusion

This paper shows that a non-Gaussian distribution for the perceived endowment process in the standard consumption-based asset pricing model can provide a possible solution to asset return puzzles such as high equity premium and low risk free rate as pointed out in macro-finance literature. The consumer-investor may perceive the endowment process as non-Gaussian as he/she is uncertain about the (variance) parameter values underlying the DGP of consumption growth. In this case, the investor may compute or derive the resulting distribution for the consumption growth as a mixture of Normal and the probability distribution from which the stochastic parameter is drawn. For tractability as well as simplicity, this paper assumes that the investor draws the uncertain (variance) parameter from the Inverse Gaussian (IG) distribution so that he/she perceives the resulting distribution for the endowment process as the negatively skewed and thick-tailed NIG distribution. We find that parameter uncertainty can lead to negatively skewed and fat-tailed distribution for the consumption growth so that most of high equity premium and low riskfree rate can be explained even with the degree of risk-aversion below 10; thus, providing a possible solution to asset return puzzles as noted in literature.
### Table 4.1: Price dividend ratio, expected return, riskfree rate and equity premium under the NIG and Gaussian distribution

<table>
<thead>
<tr>
<th></th>
<th>NIG</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(1+P/D)</td>
<td>15.76</td>
<td>10.01</td>
</tr>
<tr>
<td>Expected return</td>
<td>8.470 %</td>
<td>12.12 %</td>
</tr>
<tr>
<td>Riskfree rate</td>
<td>1.081 %</td>
<td>9.820 %</td>
</tr>
<tr>
<td>Equity premium</td>
<td>7.388 %</td>
<td>2.300 %</td>
</tr>
</tbody>
</table>

### Table 4.2: Asset-pricing features corresponding to statistical properties of parameter uncertainty under the CRRA utility function

<table>
<thead>
<tr>
<th>EP</th>
<th>$r^f$</th>
<th>P/D</th>
<th>Skw</th>
<th>Kts</th>
<th>RRA</th>
<th>$\beta$</th>
<th>E[V]</th>
<th>SD[V]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4105</td>
<td>3.2343</td>
<td>14.74</td>
<td>-4.07</td>
<td>45.54</td>
<td>7.00</td>
<td>0.999</td>
<td>0.0009</td>
<td>0.0025</td>
</tr>
<tr>
<td>6.0559</td>
<td>2.6764</td>
<td>14.56</td>
<td>-3.85</td>
<td>40.33</td>
<td>7.50</td>
<td>0.999</td>
<td>0.0009</td>
<td>0.0023</td>
</tr>
<tr>
<td>6.6363</td>
<td>2.2690</td>
<td>14.35</td>
<td>-3.49</td>
<td>35.23</td>
<td>8.00</td>
<td>0.999</td>
<td>0.0010</td>
<td>0.0022</td>
</tr>
<tr>
<td>6.8370</td>
<td>2.2259</td>
<td>13.89</td>
<td>-2.90</td>
<td>30.10</td>
<td>8.50</td>
<td>0.999</td>
<td>0.0011</td>
<td>0.0025</td>
</tr>
<tr>
<td>7.5570</td>
<td>1.4508</td>
<td>14.00</td>
<td>-2.85</td>
<td>27.52</td>
<td>9.00</td>
<td>0.998</td>
<td>0.0010</td>
<td>0.0022</td>
</tr>
<tr>
<td>8.0075</td>
<td>1.0361</td>
<td>13.93</td>
<td>-2.51</td>
<td>24.47</td>
<td>9.50</td>
<td>0.998</td>
<td>0.0011</td>
<td>0.0023</td>
</tr>
<tr>
<td>7.3996</td>
<td>0.9474</td>
<td>15.42</td>
<td>-2.90</td>
<td>24.00</td>
<td>9.95</td>
<td>0.998</td>
<td>0.0009</td>
<td>0.0017</td>
</tr>
<tr>
<td>7.0552</td>
<td>1.0773</td>
<td>15.95</td>
<td>-2.90</td>
<td>22.95</td>
<td>10.25</td>
<td>0.998</td>
<td>0.0009</td>
<td>0.0015</td>
</tr>
<tr>
<td>7.7280</td>
<td>1.2278</td>
<td>14.10</td>
<td>-1.72</td>
<td>18.30</td>
<td>11.00</td>
<td>0.998</td>
<td>0.0012</td>
<td>0.0023</td>
</tr>
<tr>
<td>7.8410</td>
<td>1.1181</td>
<td>14.09</td>
<td>-1.67</td>
<td>15.80</td>
<td>12.00</td>
<td>0.990</td>
<td>0.0012</td>
<td>0.0020</td>
</tr>
<tr>
<td>8.3779</td>
<td>1.0725</td>
<td>13.18</td>
<td>-1.10</td>
<td>13.90</td>
<td>13.00</td>
<td>0.985</td>
<td>0.0012</td>
<td>0.0022</td>
</tr>
<tr>
<td>7.8407</td>
<td>1.2238</td>
<td>13.89</td>
<td>-1.05</td>
<td>12.20</td>
<td>14.00</td>
<td>0.980</td>
<td>0.0012</td>
<td>0.0020</td>
</tr>
<tr>
<td>7.0336</td>
<td>1.2664</td>
<td>15.54</td>
<td>-1.00</td>
<td>10.70</td>
<td>15.00</td>
<td>0.980</td>
<td>0.0012</td>
<td>0.0018</td>
</tr>
<tr>
<td>6.9730</td>
<td>1.2289</td>
<td>15.78</td>
<td>-0.95</td>
<td>9.60</td>
<td>16.00</td>
<td>0.970</td>
<td>0.0012</td>
<td>0.0016</td>
</tr>
<tr>
<td>6.7247</td>
<td>1.2798</td>
<td>16.29</td>
<td>-0.95</td>
<td>8.60</td>
<td>17.00</td>
<td>0.960</td>
<td>0.0012</td>
<td>0.0015</td>
</tr>
<tr>
<td>6.3587</td>
<td>1.1233</td>
<td>17.80</td>
<td>-0.80</td>
<td>7.80</td>
<td>18.00</td>
<td>0.960</td>
<td>0.0012</td>
<td>0.0014</td>
</tr>
<tr>
<td>6.0432</td>
<td>0.8075</td>
<td>20.05</td>
<td>-0.70</td>
<td>7.10</td>
<td>19.00</td>
<td>0.960</td>
<td>0.0012</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0201</td>
<td>0.6721</td>
<td>20.71</td>
<td>-0.40</td>
<td>6.95</td>
<td>20.00</td>
<td>0.960</td>
<td>0.0013</td>
<td>0.0014</td>
</tr>
<tr>
<td>6.0816</td>
<td>0.8577</td>
<td>19.71</td>
<td>-0.23</td>
<td>6.80</td>
<td>21.00</td>
<td>0.950</td>
<td>0.0013</td>
<td>0.0014</td>
</tr>
<tr>
<td>6.1393</td>
<td>1.2322</td>
<td>18.16</td>
<td>-0.23</td>
<td>6.35</td>
<td>22.00</td>
<td>0.930</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.1998</td>
<td>0.7817</td>
<td>19.99</td>
<td>-0.15</td>
<td>6.10</td>
<td>23.00</td>
<td>0.925</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>5.9421</td>
<td>0.6934</td>
<td>20.96</td>
<td>-0.06</td>
<td>5.90</td>
<td>24.00</td>
<td>0.920</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0012</td>
<td>1.0854</td>
<td>19.15</td>
<td>-0.03</td>
<td>5.67</td>
<td>25.00</td>
<td>0.900</td>
<td>0.0013</td>
<td>0.0012</td>
</tr>
<tr>
<td>6.0623</td>
<td>0.8304</td>
<td>19.89</td>
<td>-0.01</td>
<td>5.45</td>
<td>26.00</td>
<td>0.885</td>
<td>0.0013</td>
<td>0.0012</td>
</tr>
<tr>
<td>6.0613</td>
<td>0.6336</td>
<td>20.70</td>
<td>-0.005</td>
<td>5.20</td>
<td>27.00</td>
<td>0.870</td>
<td>0.0013</td>
<td>0.0011</td>
</tr>
<tr>
<td>6.1311</td>
<td>0.5617</td>
<td>20.71</td>
<td>-0.001</td>
<td>5.00</td>
<td>28.00</td>
<td>0.850</td>
<td>0.0013</td>
<td>0.0011</td>
</tr>
<tr>
<td>6.0207</td>
<td>1.1882</td>
<td>18.71</td>
<td>-0.001</td>
<td>4.75</td>
<td>29.00</td>
<td>0.830</td>
<td>0.0013</td>
<td>0.0010</td>
</tr>
<tr>
<td>6.0195</td>
<td>0.7807</td>
<td>20.26</td>
<td>-0.001</td>
<td>4.56</td>
<td>30.00</td>
<td>0.815</td>
<td>0.0013</td>
<td>0.0009</td>
</tr>
<tr>
<td>EP</td>
<td>$\beta$</td>
<td>P/D</td>
<td>Skw</td>
<td>Kts</td>
<td>RRA</td>
<td>$\beta$</td>
<td>EIS</td>
<td>E[V]</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>---------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>5.4105</td>
<td>1.0083</td>
<td>21.94</td>
<td>-4.07</td>
<td>45.54</td>
<td>7.00</td>
<td>0.96</td>
<td>0.6667</td>
<td>0.0009</td>
</tr>
<tr>
<td>6.0559</td>
<td>1.0560</td>
<td>19.05</td>
<td>-3.85</td>
<td>40.33</td>
<td>7.50</td>
<td>0.96</td>
<td>0.4545</td>
<td>0.0009</td>
</tr>
<tr>
<td>6.6363</td>
<td>1.0907</td>
<td>17.05</td>
<td>-3.49</td>
<td>35.23</td>
<td>8.00</td>
<td>0.96</td>
<td>0.3448</td>
<td>0.0010</td>
</tr>
<tr>
<td>6.8370</td>
<td>1.0138</td>
<td>16.70</td>
<td>-2.90</td>
<td>30.10</td>
<td>8.50</td>
<td>0.96</td>
<td>0.3226</td>
<td>0.0011</td>
</tr>
<tr>
<td>7.5570</td>
<td>1.0837</td>
<td>14.75</td>
<td>-2.85</td>
<td>27.52</td>
<td>9.00</td>
<td>0.96</td>
<td>0.2381</td>
<td>0.0010</td>
</tr>
<tr>
<td>8.0075</td>
<td>1.0584</td>
<td>13.88</td>
<td>-2.51</td>
<td>24.47</td>
<td>9.50</td>
<td>0.96</td>
<td>0.2041</td>
<td>0.0011</td>
</tr>
<tr>
<td>7.3996</td>
<td>1.0260</td>
<td>15.24</td>
<td>-2.90</td>
<td>24.00</td>
<td>9.95</td>
<td>0.96</td>
<td>0.2151</td>
<td>0.0009</td>
</tr>
<tr>
<td>7.0552</td>
<td>1.0057</td>
<td>16.13</td>
<td>-2.90</td>
<td>22.95</td>
<td>10.25</td>
<td>0.96</td>
<td>0.2299</td>
<td>0.0009</td>
</tr>
<tr>
<td>7.7280</td>
<td>0.9614</td>
<td>14.65</td>
<td>-1.72</td>
<td>18.30</td>
<td>11.00</td>
<td>0.96</td>
<td>0.1961</td>
<td>0.0012</td>
</tr>
<tr>
<td>7.8410</td>
<td>0.9041</td>
<td>14.53</td>
<td>-1.67</td>
<td>15.80</td>
<td>12.00</td>
<td>0.96</td>
<td>0.1613</td>
<td>0.0012</td>
</tr>
<tr>
<td>8.3779</td>
<td>0.9858</td>
<td>13.33</td>
<td>-1.10</td>
<td>13.90</td>
<td>13.00</td>
<td>0.96</td>
<td>0.1266</td>
<td>0.0012</td>
</tr>
<tr>
<td>7.8407</td>
<td>1.0236</td>
<td>14.29</td>
<td>-1.05</td>
<td>12.20</td>
<td>14.00</td>
<td>0.96</td>
<td>0.1176</td>
<td>0.0012</td>
</tr>
<tr>
<td>7.0336</td>
<td>1.0168</td>
<td>16.16</td>
<td>-1.00</td>
<td>10.70</td>
<td>15.00</td>
<td>0.96</td>
<td>0.1266</td>
<td>0.0012</td>
</tr>
<tr>
<td>6.9730</td>
<td>0.9605</td>
<td>16.48</td>
<td>-0.95</td>
<td>9.60</td>
<td>16.00</td>
<td>0.96</td>
<td>0.0971</td>
<td>0.0012</td>
</tr>
<tr>
<td>7.6247</td>
<td>0.9587</td>
<td>17.18</td>
<td>-0.95</td>
<td>8.60</td>
<td>17.00</td>
<td>0.96</td>
<td>0.0685</td>
<td>0.0012</td>
</tr>
<tr>
<td>6.3587</td>
<td>1.0091</td>
<td>18.17</td>
<td>-0.80</td>
<td>7.80</td>
<td>18.00</td>
<td>0.96</td>
<td>0.0595</td>
<td>0.0012</td>
</tr>
<tr>
<td>6.0432</td>
<td>0.9718</td>
<td>19.41</td>
<td>-0.70</td>
<td>7.10</td>
<td>19.00</td>
<td>0.96</td>
<td>0.0455</td>
<td>0.0012</td>
</tr>
<tr>
<td>6.0201</td>
<td>1.0207</td>
<td>19.32</td>
<td>-0.40</td>
<td>6.95</td>
<td>20.00</td>
<td>0.96</td>
<td>0.0357</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0816</td>
<td>1.0147</td>
<td>19.11</td>
<td>-0.23</td>
<td>6.80</td>
<td>21.00</td>
<td>0.96</td>
<td>0.0300</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.1393</td>
<td>1.2182</td>
<td>18.21</td>
<td>-0.23</td>
<td>6.35</td>
<td>22.00</td>
<td>0.945</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0845</td>
<td>1.2689</td>
<td>18.22</td>
<td>-0.15</td>
<td>6.10</td>
<td>23.00</td>
<td>0.945</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>5.9421</td>
<td>0.9086</td>
<td>20.06</td>
<td>-0.06</td>
<td>5.90</td>
<td>24.00</td>
<td>0.95</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0012</td>
<td>0.8437</td>
<td>20.08</td>
<td>-0.03</td>
<td>5.67</td>
<td>25.00</td>
<td>0.95</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0623</td>
<td>1.2766</td>
<td>18.27</td>
<td>-0.01</td>
<td>5.45</td>
<td>26.00</td>
<td>0.945</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0613</td>
<td>1.2721</td>
<td>18.29</td>
<td>-0.005</td>
<td>5.20</td>
<td>27.00</td>
<td>0.945</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.1311</td>
<td>1.1961</td>
<td>18.31</td>
<td>-0.001</td>
<td>5.00</td>
<td>28.00</td>
<td>0.945</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0207</td>
<td>1.3025</td>
<td>18.32</td>
<td>-0.001</td>
<td>4.75</td>
<td>29.00</td>
<td>0.945</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
<tr>
<td>6.0195</td>
<td>1.2985</td>
<td>18.34</td>
<td>-0.001</td>
<td>4.56</td>
<td>30.00</td>
<td>0.945</td>
<td>0.9091</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

Table 4.3: Asset-pricing features corresponding to statistical properties of parameter uncertainty under the EZ utility function.
Figure 4.1: Probability density function for log consumption growth rate and the stochastic parameter
a) Probability density function of the Gaussian and NIG distribution for the log consumption growth rate
b) Probability density function of the Inverse Gaussian distribution from which the stochastic variance parameter is drawn
Figure 4.2: Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -3.49 and 35.23 respectively while the mean and variance for the IG are 0.001 and 0.0022 each.

a) Probability density function of the NIG distribution for consumption growth with skewness -3.49 and kurtosis 35.23 in comparison with the Gaussian distribution

b) Probability density function of the Inverse Gaussian distribution from which the stochastic variance parameter for the NIG in Figure 4.2.1 is drawn (the mean of the stochastic variance parameter is 0.001 and its standard deviation is 0.0022)
Figure 4.3: Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -1.0 and 10.7 respectively while the mean and variance for the IG are 0.0012 and 0.0018 each. 

a) Probability density function of the NIG distribution for consumption growth with skewness -1.0 and kurtosis 10.7 in comparison with the Gaussian distribution 

b) Probability density function of the Inverse Gaussian distribution from which the stochastic variance parameter for the NIG in Figure 4.3.1 is drawn (the mean of the stochastic variance parameter is 0.0012 and its standard deviation is 0.0018)
Figure 4.4: Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -0.01 and 5.45 respectively while the mean and variance for the IG are 0.0013 and 0.0012 each.

a) Probability density function of the NIG distribution for consumption growth with skewness -0.01 and kurtosis 5.45 in comparison with the Gaussian distribution

b) Probability density function of the Inverse Gaussian distribution from which the stochastic variance parameter for the NIG in Figure 4.4.1 is drawn (the mean of the stochastic variance parameter is 0.0013 and its standard deviation is 0.0012)
Figure 4.5: Probability density function for log consumption growth rate and the stochastic parameter - Skewness and kurtosis for the NIG are -0.001 and 3.01 respectively while the mean and variance for the IG are 0.0013 and 0.0001 each. 

a) Probability density function of the NIG distribution for consumption growth with skewness -0.001 and kurtosis 3.01 in comparison with the Gaussian distribution 

b) Probability density function of the Inverse Gaussian distribution from which the stochastic variance parameter for the NIG in Figure 4.5.1 is drawn (the mean of the stochastic variance parameter is 0.0013 and its standard deviation is 0.0001)
CHAPTER 5

STOCK MARKET VOLATILITY AND MACROECONOMIC RISK: ASSET PRICING UNDER TWO-FACTOR PRODUCTION ECONOMY

5.1 Introduction

The U.S. economic activity has been much less volatile since the early 1980’s – the volatilities of output and consumption, for example, have been reduced significantly during the period.¹ This “great moderation” in macroeconomic risk, however, has not led to moderation in stock market volatility in contrast to the prediction made by the standard consumption-based asset pricing model (C-CAPM), which implies that declining consumption risk should lead to declining stock market risk. Figure 5.1 shows quarterly per-capita real consumption growth rates for nondurable goods and services, and Figure 5.2 shows quarterly real returns of the aggregate stock market. The volatility movements of the two time series are hard to explain relying on the standard C-CAPM. This apparent contradiction between the empirical evidence and the C-CAPM is resolved in our model which incorporates a two-factor (stock market traded and non-traded factor) production function into the standard model.

The relative importance of stock wealth in the U.S. economy has been changing over time as agents change their relative equity holdings in their wealth portfolio. This

feature is captured by Figure 5.3 which shows ratios of stock market capitalization to nondurable consumption expenditures over the post-war period. As the ratio of stock-wealth to total wealth changes, it may influence the correlation and covariance between consumption growth and stock returns – the covariance indicates consumption risk due to equity-holding, which tells why holding equity is risky. We incorporate this ratio of stock to total wealth in our model and show that the correlation and covariance respond positively to the change in the stock share of total wealth. Our result is consistent with the empirical finding in Duffee (2005).

Our model explicitly specifies consumption as the sum of payouts from stock and non-stock assets based on the marginal contribution of each asset to the production of the economy. Thus, we can distinguish stock price and dividend from total wealth and consumption - this difference is often ignored in consumption-based asset pricing models, in which the income claims to a single-asset portfolio (the Lucas one-tree) are assumed to be traded in a stock market. Alternative modeling strategy often appearing in literature is to specify a stock dividend process separately from consumption, but they do not explicitly show how consumption and stock dividend are determined in equilibrium of the economy. See Campbell & Cochrane (1999) and Cochrane (2004) for the discussion.

Stock wealth is a part of total wealth in our model. This realistic feature is one of the key ingredients which allows the model to generate the historically high volatility of stock returns, low consumption risks, and the weak correlation between them.

We extend the standard Lucas 1-tree model by introducing a two-factor production function which drives consumption process of the economy. Each factor of production pays out to its owner its marginal product and their sum is equal to
consumption in equilibrium. Assets employed as factors of production in this economy are divided into stock-market traded and non-traded ones based on tradability of their ownership through a stock market. The representative consumer-investor holds an economy-wide wealth portfolio consisting of the two types of assets and receives consumption from the portfolio as a sum of payouts from each asset every period. This is how we distinguish stock price and dividend from total wealth (the price of the portfolio) and consumption. The other key ingredient of our model is stochastic productivities associated with each factor of production. That is, two sources of shocks drive asset pricing features in our model economy in contrast to one source of shock in the Lucas 1-tree model. This is how our model can generate the weak correlation between consumption growth and stock returns – Campbell & Cochrane (1999) point out that the weak correlation is at the heart of empirical failures of the C-CAPM.

One of the most essential features of the standard C-CAPM is that the Stochastic Discount Factor (SDF) is given as a power function of gross consumption growth rate because the CRRA utility function is employed for the consumer-investor’s optimization problem. That is, the SDF \( M_{t+1} = \beta(C_{t+1} / C_t)^{-\eta} \) in the canonical form of stock-asset pricing model \( P_t = E_t[M_{t+1}(P_{t+1} + D_{t+1})] \), where \( \eta \) is the risk aversion coefficient, \( \beta \) is a time-discount factor, and \( D \) is stock dividends. Another important feature of the standard model is that a log-normal distribution with known (ex-post) parameters is used to compute the expectation operator, \( E_t[\cdot] \). Our model preserves these essential features of the standard model while specifying consumption as a sum of stochastic payouts from stock and non-stock factors of production. That is, our model is pretty standard except
that consumption is decomposed into stock-dividend and non-stock asset payouts. By
taking this direction of modification, we can explicitly show that which shortcomings of
the standard model are overcome and what else still remain unsolved.

As pointed out by Mehra & Prescott (1985), Epstein & Zin (1989), and Campbell
& Cochrane (1999), the standard consumption-based asset pricing models are at conflict
with the empirically high equity premium and low riskless rate. Since our model also
keeps the essential features of the standard model, it can not be free from the puzzles.
However, we do not attempt to solve the equity premium puzzle in this paper but to
answer the following questions raised in the beginning - why the volatility of the stock
market returns has not declined during the great moderation era, how market
capitalization may influence the consumption risk of equity-holding, and how to match
the empirically high volatility of stock returns, low variance of consumption growth and
the weak correlation between them. Even though we do not aim to overcome some
puzzling features of the standard model, they might be solved by modifying either the
SDF or the underlying distributional assumption to compute the expectation operator,
\[ E_t[\cdot] \]
as pointed out by Cecchetti, Lam & Mark (2001). For example, Campbell &
Cochrane (1999) take the direction of modifying the SDF while Barsky & Delong (1993),
Cecchetti, Lam & Mark (2000), and Weitzman (2007) take the direction of modifying the
underlying distributional assumptions retreating from the rational expectation. In this
paper, we extend the standard model by introducing a two-factor production function and
attempt to answer the questions discussed above.

Contents of this paper are organized in the following order. Section 5.2 describes
the model economy with its important features and the solution to the investor’s problem
in stock market. Section 5.3 then shows asset pricing features of the model economy with its important statistical properties. Section 5.4 discusses a potential link between macroeconomic risk and the stock market volatility and thus its implication for the impact of the great moderation on stock market volatility. Section 5.5 shows the mechanism of how the stock market capitalization may influence consumption risk of holding equity. Section 5.6 concludes.

5.2 The model

5.2.1 Background of the model economy

A representative consumer-investor holds an aggregate wealth portfolio which consists of a stock market traded asset and a non-traded one. Each type of asset is employed as a factor of production, and their owner receives their marginal contribution to production as payouts from each asset. The sum of payouts from each asset determines the amount of consumption for the investor each period. That is, the investor with the wealth portfolio enjoys consumption whose income source is decomposed into stock dividend and non-stock payout. Note that stock dividends are a small part of consumption in equilibrium of the model economy. The investor then decides how many shares of stock asset to hold to maximize her expected lifetime utility. The solution to this optimization problem along with the equilibrium condition gives rise to stock price and return.

As a simple example of this model economy, we may consider two sources of income: property and labor. These two types of resources, or assets are employed as factors of production so that owners of each asset receive their income in proportion to each asset’s marginal productivity and enjoy consumption. Note that ownership of property asset is traded in this economy whereas ownership of labor - i.e. ownership of
human capital - is not. Thus, the consumer investor chooses how many shares of property ownership to carry as a saving vehicle while she is stuck with payouts from labor in this economy. The sum of payouts from the property and labor assets determines the amount of consumption for the investor each period. We may interpret consumption as a payout from an economy-wide basket portfolio which contains two types of assets: property and labor. Thus, the payout from property asset is a small part of consumption while the price of property ownership is a part of total wealth in this economy. In this paper, we divide all resources of the economy into stock market traded asset and non-traded asset so that we distinguish stock dividend from consumption as stock price is different from the total wealth of the economy.

5.2.2 Production of the economy

Production process of the economy determines payouts from each asset and thus the amount of consumption for the investor holding the aggregate wealth portfolio that consists of the two types of assets. Here we assume the CES technology which combines the two factors and transforms them into consumption goods as below. Another important character of the production function is that each factor combined with stochastic productivity shocks contributes to the amount of production.

\[ Y_t = A \left\{ \theta (\beta_t S_t)^\gamma + (1 - \theta) (\gamma_t N_t)^\rho \right\}^{1/\rho} \]

where \( A \) defines the factor-neutral productivity of consumption goods. \( \theta \) characterizes the relative significance of the stock factor for production, while \( \rho = (\sigma - 1)/\sigma \) with \( \sigma \) the elasticity of substitution between factors of production. \( S \) denotes stock factor of production (stock market traded asset) while \( N \) denotes non-stock factor of production (stock market non-traded asset). Note that all factors of production in this economy are
divided into two types of production factors, stock factor $S$ and non-stock factor $N$ based on whether their ownership is traded in a market. For example, we may consider a hypothetical model economy which utilizes both physical and human asset (capital) to produce consumption goods. In this simple economy, the claim for the future payouts from the physical asset is traded while the claim for the stream of future payouts from human asset is not. We may interpret the payout from stock asset as stock dividend while interpreting those from human asset as labor income for this example. Productivity shocks to each factor are assumed jointly log-normally distributed with independent and identical distributional assumption.

\[
\begin{pmatrix}
\log \beta_i \\
\log \gamma_i
\end{pmatrix}
\sim iid \mathcal{N}\left(\begin{pmatrix}
\mu_\beta \\
\mu_\gamma
\end{pmatrix}, \begin{pmatrix}
\sigma_\beta^2 & \sigma_{\beta\gamma} \\
\sigma_{\beta\gamma} & \sigma_\gamma^2
\end{pmatrix}\right)
\]

Since the production function is characterized by Constant Returns to Scale (CRS), each factor is paid its marginal contribution to production. That is,

\[
Y = \frac{\partial Y}{\partial S} S + \frac{\partial Y}{\partial N} N = \epsilon_{c,s} Y + \epsilon_{c,n} Y = D_s + D_n
\]

where $\epsilon_{c,s} \equiv \frac{\partial Y}{\partial S} S$; stock-factor elasticity of output

$\epsilon_{c,n} \equiv \frac{\partial Y}{\partial N} N$; non-stock-factor elasticity of output

Thus, the owner of stock asset receives $D_s$ while the owner of non-stock asset receives $D_n$. Payouts $D_s$ and $D_n$ combined together determine the level of consumption for the portfolio holder. Note that claims to a stream of stock payouts, $D_s$ are traded in the stock market whereas the claims to a flow of non-stock payouts, $D_n$ are not.
5.2.3 Investor’s problem in equity market

The representative investor with a time-separable CRRA utility function chooses her consumption $C_t$ and the amount of stock shares, $x_t \in [0,1]$ each period to maximize her expected lifetime utility. The consumer-investor carries the stock asset as a saving vehicle to maximize her expected lifetime utility.

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

s.t. $C_t + x_t P_t \leq Y_t + x_{t-1} (P_t + D_{s,t})$

where the CRRA utility function $U(C) = C^{1-\eta} / (1 - \eta)$ is the key ingredient of the standard consumption-based asset pricing model (C-CAPM). The consumer investor’s time preference is given by the time-discount factor $\beta \equiv e^{-\delta}$ with $\delta$ the time discount rate. The investor is also assumed to know the underlying parameters of shocks to this economy. Note that this rational expectation assumption along with Gaussianity of shocks is another key ingredient of the standard C-CAPM. We keep these key ingredients of the standard C-CAPM in this paper.

The first order conditions from the optimization problem lead to the following Euler equation.

$$U'(C_t) P_t = e^{-\delta} E_t[U'(C_{t+1})(P_{t+1} + D_{s,t+1})]$$

Hence, the stock price is determined by

$$P_t = E_t[\beta \frac{U'(C_{t+1})}{U'(C_t)} (P_{t+1} + D_{s,t+1})] = E_t[e^{-\delta} \left( \frac{C_{t+1}}{C_t} \right)^{-\eta} (P_{t+1} + D_{s,t+1})]$$

130
where $\beta U'(C_{t+1})/U'(C_t)$ in the equation is the Inter-temporal Marginal Rate of Substitution (IMRS) used by the investor as a pricing kernel to discount future pay-offs. Note that the investor discounts a stream of future stock dividends ($D_{s,t+j}$) which is part of consumption ($C = D_s + D_n$); thus, stock price $P_t$ is distinguished from the total wealth of the economy in contrast to the Lucas 1-tree model, where the investor discounts the stream of future consumption – i.e. $P_t = E_t[e^{-\delta} (C_{t+1} / C_t)^{-\eta} C_{t+1}] = E_t[e^{-\delta} C_{t+1}^{1-\eta} / C_t^{-\eta}]$ because $D_s = C$ each period at the equilibrium of the Lucas 1-tree endowment economy.

5.3 Asset pricing features

5.3.1 Stock returns and its variance

One-period gross stock returns is given by $R_{t+1} = (P_{t+1} + D_{s,t+1})/P_t$. Note that the price-dividend ratio is constant because shocks to this economy are assumed independent and identical across time (see Appendix D.1). Letting $P_t / D_{s,f} = \alpha$, 

$$R_{t+1} = \frac{P_{t+1} + D_{s,t+1}}{P_t} = \frac{P_{t+1} + D_{s,t+1} D_{s,t} D_{s,t+1}}{P_t D_{s,t}} = \frac{1 + \alpha D_{s,t+1}}{\alpha D_{s,t}} = \frac{1 + \alpha \epsilon_{cs,t+1} C_{t+1}}{\alpha \epsilon_{cs,t} C_t}$$

The one-period log-return from holding stock asset is then given by

$$\log R_{t+1} = \log(1 + \alpha) - \log \alpha + \log(\epsilon_{cs,t+1} C_{t+1}) - \log(\epsilon_{cs,t} C_t)$$

The conditional variance of the log-stock returns is then

$$Var_t(\log R_{t+1}) = Var_t(\log \epsilon_{cs,t+1} + \log C_{t+1})$$

$$= Var_t(\log \epsilon_{cs,t+1}) + Var_t(\log C_{t+1}) + 2Cov_t(\log C_{t+1}, \log \epsilon_{cs,t+1}) \quad (5.1)$$

Based on the stochastic production function above, we can compute each component of the conditional variance of log stock return $Var_t(\log R_{t+1})$ as below (see Appendix D.2).
\[
\begin{align*}
Var_i(\log C_{i+1}) & = Var_i(\log C_{i+1} / C_i) = K^2 \sigma_\beta^2 + H^2 \sigma_\gamma^2 + 2KH \sigma_{\rho \gamma}
\end{align*}
\]

\[
\begin{align*}
Var_i(\log \epsilon_{ct,i+1}) & = \rho^2 H^2 (\sigma_\beta^2 + \sigma_\gamma^2 - 2\sigma_{\rho \gamma})
\end{align*}
\]

\[
\begin{align*}
Cov_i(\log C_{i+1}, \log \epsilon_{ct,i+1}) & = \rho H \{K \sigma_\beta^2 - H \sigma_\gamma^2 + (H - K) \sigma_{\rho \gamma}\}
\end{align*}
\]

where \( K \equiv (\theta S^\rho e^{\mu \beta}) / \{\theta S^\rho e^{\mu \beta} + (1 - \theta) N^\rho e^{\rho \gamma}\} \) is the elasticity of output with respect to stock factor evaluated at the mean of productivity shocks and

\( H \equiv ((1 - \theta) N^\rho e^{\rho \gamma}) / \{\theta S^\rho e^{\mu \beta} + (1 - \theta) N^\rho e^{\rho \gamma}\} \) is the elasticity of output with respect to non-stock factor evaluated at the mean of productivity shocks. Note that \( K \) and \( H \) satisfy \( K + H = 1 \).

Because the supply of stock and non-stock factor available in this economy is assumed fixed respectively, \( S \) and \( N \) may be fixed at a unit (\( S = N = 1 \)); hence, further simplified as

\[
K = (\theta e^{\rho \mu \beta}) / \{\theta e^{\rho \mu \beta} + (1 - \theta) e^{\rho \gamma}\}
\]

and

\[
H = ((1 - \theta) e^{\rho \gamma}) / \{\theta e^{\rho \mu \beta} + (1 - \theta) e^{\rho \gamma}\}.
\]

As implied by equation (5.1), the variance of stock returns can be much larger than the variance of consumption growth depending on the underlying parameters of the production function and shocks to each factor. If we assume the investor’s wealth portfolio is composed of only stock asset, \( \theta = 1 \), then stock return and consumption growth are equally volatile as predicted by the standard Lucas 1-tree model because \( K = 1 \) and \( H = 0 \) for \( \theta = 1 \). That is, our model includes the standard model as an extreme case in which consumption equals stock dividends. Thus, the observed high volatility of stock return and smooth consumption growth can be explained by the fact that the stock dividend composes only a small part of consumption – i.e. stock asset is a small part of the investor’s wealth portfolio – while it is volatile.
5.3.2 Covariance between consumption growth and stock returns

- Systematic risk of holding stock asset

Holding equity is risky to the consumer investor because it affects the level of consumption which again affects utility of the investor. Hence, we may use the covariance between consumption growth and stock returns as a measure of systematic risk associated with holding equity. The covariance between consumption growth, $g_{t+1} \equiv \log(C_{t+1} / C_t)$ and stock return is computed as below.

$$
Cov_t[g_{t+1}, \log R_{t+1}] = Cov_t[g_{t+1}, \log D_{s,t+1}] = Cov_t[g_{t+1}, g_{t+1} + \log \epsilon_{cs,t+1}]
$$

$$
= Var_t[g_{t+1}] + Cov_t[g_{t+1}, \log \epsilon_{cs,t+1}]
$$

$$
= \{K^2 \sigma^2_{\beta} + H^2 \sigma^2_{\gamma} + 2KH\sigma_{\beta\gamma}\} + \rho H \{K\sigma^2_{\beta} - H\sigma^2_{\gamma} + (H - K)\sigma_{\beta\gamma}\}
$$

$$
= (\rho + (1-\rho)K)K\sigma^2_{\beta} + (1-\rho)H^2 \sigma^2_{\gamma} + (\rho + 2(1-\rho)K)H\sigma_{\beta\gamma}
$$

As shown in equation (5.2), the covariance between consumption growth and stock return may differ from the variance of consumption growth in contrast to the result from the standard Lucas 1-tree model. In other words, the consumption risk of holding equity, represented by its covariance with stock returns may differ from consumption risk itself. However, the difference between them disappears if consumption equals stock dividends, $\theta = 1$ – i.e. the case the investor’s portfolio consists of only stock asset – because the stock elasticity of output is then one. Thus, our two-factor model results in the standard model as an extreme case in which all consumption claims are traded in the stock market.

Note that Beta is measured as the covariance between market return and individual stock returns divided by the variance of market return. We may also divide the covariance by the variance of consumption growth and call this ratio the systematic risk of equity-holding in consumption-CAPM. This is similar to the concept of Beta, which is also a measure of systematic risk in the CAPM.
The covariance between consumption growth and stock returns is so important in consumption-based asset pricing model that it determines the equity premium combined with the relative risk aversion. That is, the equity premium is determined by a product of risk-aversion and the consumption risk of holding equity. In a standard C-CAPM with the time-separable CRRA utility function as in our model, risk-aversion is given simply by the CRRA coefficient – which also equals the reciprocal of the Inter-temporal Elasticity of Substitution (IES); thus, leading to the equity premium puzzle because the risk aversion coefficient in a plausible range can not explain the high equity premium. To overcome the equity premium puzzle associated with the standard model, Campbell and Cochrane (1999) introduce a consumption-surplus ratio in the utility function so that they are able to generate a high risk-aversion even with the small curvature parameter of the utility function. We do not adopt this approach in this paper; hence, leading to the equity premium puzzle. However, we rather focus on how the covariance and correlation between consumption growth and stock returns are determined in a two-factor production economy in which stock asset is a small part of the portfolio in contrast to the standard one-tree model. This aspect is further discussed in one of the following sections.

5.3.3 Matching the historical variance and covariance with the model counterpart

The variance of consumption growth and stock returns and their covariance are expressed in terms of the underlying parameters that characterize the production of the economy.

\[ \sigma_g^2 \equiv Var_t(g_{t+1}) = K^2 \sigma_{\beta}^2 + H^2 \sigma_{\gamma}^2 + 2KH\sigma_{\rho y} \]

\[ \sigma_r^2 \equiv Var_t(\log R_{t+1}) = \{\rho + (1-\rho)K\}^2 \sigma_{\beta}^2 + (1-\rho)^2 H^2 \sigma_{\gamma}^2 + 2\{\rho + (1-\rho)K\}(1-\rho)H\sigma_{\rho y} \]

\[ \sigma_{g,r} \equiv Cov_t(g_{t+1}, \log R_{t+1}) = \{\rho + (1-\rho)K\}K\sigma_{\beta}^2 + (1-\rho)H^2 \sigma_{\gamma}^2 + \{\rho + 2(1-\rho)K\}H\sigma_{\rho y} \]
The above three equations can be expressed in a matrix form as below.

$$
\begin{bmatrix}
\sigma_g^2 \\
\sigma_r^2 \\
\sigma_{g,r}
\end{bmatrix}
= \begin{bmatrix}
\sigma_\beta^2 \\
\sigma_\gamma^2 \\
\sigma_{\beta\gamma}
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
\sigma_g^2 \\
\sigma_r^2 \\
\sigma_{g,r}
\end{bmatrix}
= A^{-1}
\begin{bmatrix}
C_{11} & C_{12} & C_{13} \\
C_{21} & C_{22} & C_{23} \\
C_{31} & C_{32} & C_{33}
\end{bmatrix}
\begin{bmatrix}
\sigma_g^2 \\
\sigma_r^2 \\
\sigma_{g,r}
\end{bmatrix}
$$

where $|A| = -\rho^3 H^3$

$$
C_{11} = -\rho(1-\rho)^2 H^3; \quad C_{12} = -\rho H^3; \quad C_{13} = 2\rho(1-\rho)H^3
$$

$$
C_{21} = -\rho(\rho + (1-\rho)K)^2 H; \quad C_{22} = -\rho K^2 H; \quad C_{23} = 2\rho(\rho + (1-\rho)K)KH
$$

$$
C_{31} = \rho(1-\rho)(\rho + (1-\rho)K)H^2; \quad C_{32} = \rho KH^2; \quad C_{33} = -\rho(\rho + 2(1-\rho)K)H^2
$$

As shown above, the observed variance and covariance of consumption growth and stock returns can be matched by appropriate choice of their underlying parameter values: $\sigma_\beta^2, \sigma_\gamma^2, \sigma_{\beta\gamma}, \rho$ and $K$.

The unconditional standard deviation of and correlation between consumption growth and stock returns over the postwar period are estimated at an annual rate as below.

$$
\sigma(g) = 0.01 (1\%); \quad \sigma(\log R) = 0.15 (15\%) \quad \text{and} \quad Corr(g, \log R) = 0.2
$$

See Figure 5.4 and Figure 5.5 for the historical pattern of those variables - Campbell (1998) also reports stylized facts about asset pricing features for the industrialized economies.

From the above relationship, we can recover the underlying parameter values which may have generated the historical consumption growth and stock returns. They are shown in Table 5.1 for various degrees of substitutability $\rho = (\sigma - 1)/\sigma$ between factors assuming $K = 0.03$. Note that $K = \theta$ if the fixed supply of factors is normalized ($S = N = 1$) and the average log-productivity shocks are assumed zero, $\mu_\beta = \mu_\gamma = 0$. 

135
That is, $K$ approximates the stock share of total wealth under this condition. Hence, $K$ might be 3 - 5 % if consumption were treated as a payment from perpetual bonds (consols) given real interest rate of 3 - 5% because the ratio of stock-wealth to consumption is close to one on average during the postwar period as shown in Figure 5.3.\(^3\)

We then try to pick a plausible range of substitutability between factors – represented by $\rho = (\sigma - 1)/\sigma$ – based on the behavior of $\text{Corr}(g, \log R)$ and $\text{Cov}(g, \log R)$ over a feasible range of $K$ for the given sets of underlying parameter values. Note that stock and non-stock assets are substitutes for $\rho > 0$, i.e. $\sigma > 1$ while they are complementary for $\rho < 0$, i.e. $0 < \sigma < 1$. The behavior of these statistics are displayed in Graph 5.1 and 5.2 for two qualitatively different degrees of substitutability between factors – i.e. $\rho > 0$ v.s. $\rho < 0$. Note that $\sigma^2_\beta$, $\sigma^2_r$ and $\sigma_{\rho_r}$ are chosen corresponding to the value of $\rho$ from Table 5.1 for the benchmark case of $K = 0.03$. In Graph 5.1 and 5.2, we analyze how $\text{Corr}(g, \log R)$ and $\text{Cov}(g, \log R)$ change as $K$ - i.e. the proxy for the stock share of total wealth - changes at around the benchmark case of $K = 0.03$. Note that $\text{sig}(g)$, $\text{sig}(\log R)$ and $\text{Corr}(g, \log R)$ are set so that they match the historical estimates 0.01, 0.15 and 0.2 respectively at the benchmark case $K = 0.03$.

Graph 5.1 shows that the systematic risk of holding equity - measured by $\text{Corr}(g, \log R)$ and $\text{Cov}(g, \log R)$ - increases as equity share of the wealth portfolio - as captured by $K$ -

\[^3\] The ratio of stock wealth $W_s$ to total wealth $W$ could be roughly approximated as the following:

$$\frac{W_s}{W} = \frac{W_s}{C} \times \frac{C}{W} = \frac{W_s}{C} \times 0.03 = 1 \times 0.03 = 0.03.$$

If we consider the recent period of around year 2000, the ratio of stock wealth to consumption ($W_s/C$) goes up even close to 3; thus, the stock share of total wealth ($W_s/W$) might be even close to 10%. Still, the stock asset is a small part of the economy so that the main results of this paper do not change much.
increases if the two factors are substitutable ($\rho = 0.8$); hence, consistent with our intuition. In contrast, the systematic risk declines as the stock share of total wealth increases at around $K = 0.03$ if the two factors are complementary ($\rho = -2.5$).

Based on the behavior of covariance and correlation between consumption growth and stock returns for various values of $\rho$, we find the following relationships in general. If $\rho$ is (+), $\text{Cov}(g, \log R)$ and $\text{Corr}(g, \log R)$ increase as $K$ increases at around 0.03. If $\rho$ is (-), $\text{Cov}(g, \log R)$ and $\text{Corr}(g, \log R)$ decline as $K$ increases at around 0.03.

Which one is correct? Duffee (2005) reports that both $\text{Cov}(g, \log R)$ and $\text{Corr}(g, \log R)$ increase as the ratio of stock-wealth to consumption increases; hence, supporting the substitutability ($\rho > 0$; $\sigma > 1$) between factors. Based on the empirical evidence, we may conclude that $\rho > 0$ would be plausible as also consistent with our intuition. That is, it might be correct to assume that stock and non-stock factors are quite substitutable in their contribution to production. Hence, under close substitutability between factors, i.e. $\rho > 0$, the systematic risk of holding equity increases as the wealth portfolio contains more of stock asset because the investor’s consumption relies more on the increasing portion of stock asset.

### 5.4 Great moderation in consumption but no moderation in stock returns

Consumption growth has been moderated since the early 1980s as shown in Figure 5.1 and as documented by Kim & Nelson (1999) and Stock & Watson (2002). The standard consumption-based asset pricing model implies that declining consumption risk should be accompanied by moderated volatility of stock returns. However, the overall volatility of

---

4 In our model, consumption equals output. If we analyze the variance of output growth, the moderation is more dramatic around the early 1980s than that of consumption growth of non-durable goods and services.
stock returns has not been moderated during the “great moderation” era of the 1980s; thus, inconsistent with the implication of the standard C-CAPM.

In contrast to the standard model, our model can explain why moderated consumption risk doesn’t necessarily have to be followed by moderated risk of stock returns. In our model, consumption is composed of payouts from stock asset and non-stock asset in return for their contribution to production. Each factor of production is associated with their corresponding productivity shocks in our model. As productivity shocks to the non-stock factor are moderated while the volatility of shocks to the stock factor barely changes, consumption risk may decline while stock market risk stays unchanged under some plausible conditions.

5.4.1 Model implication - sensitivity analysis
We can explicitly show how the two sources of shocks drive the volatility of consumption growth and stock returns below. For simplicity, we assume the two productivity shocks are uncorrelated, $\sigma_{\beta} = 0$ without much loss of generality. Then, the variance of consumption growth and stock returns are computed in terms of their underlying parameter values as below.

\[
Var_t[g_{t+1}] = K^2 \sigma_\beta^2 + H^2 \sigma_\gamma^2
\]

\[
Var_t[\log R_{t+1}] = (K + \rho H)^2 \sigma_\beta^2 + (1 - \rho)^2 H^2 \sigma_\gamma^2
\]

Note that $K^2 \approx 10^{-3}$, $H^2 \approx 1$ and $\sigma_\gamma^2 \ll \sigma_\beta^2$ as discussed above. Given reduction in volatility of shocks to non-stock asset $\sigma_\gamma$, consumption volatility $Var_t[g_{t+1}]$ drops almost by the same amount whereas the risk of equity returns $Var_t[\log R_{t+1}]$ barely changes for $\rho$ close to 1 according to the equations. Table 5.2 shows the impact of a
25% reduction in the standard deviation of shocks to non-stock asset $\sigma_r$ on both consumption risk $\sigma(g)$ and stock market risk $\sigma(\log R)$ for various degrees of substitutability between factors, $\rho$. The values of $\sigma_\beta$ and $\sigma_r$ in Table 5.2 correspond to various values of the substitutability, $\rho$ from Table 5.1, to be consistent with the small variance of consumption growth and the large variance of stock returns. As the volatility of shocks to non-equity factor (asset) $\sigma_r$ drops by 25%, we see from Table 5.2 that the standard deviation of consumption growth $\sigma(g)$ declines by 18-20% while stock market risk $\sigma(\log R)$ barely changes.

This result from the sensitivity analysis shows that each factor-specific productivity shock plays significantly different roles for the volatility of consumption growth and stock returns respectively. That is, productivity shocks associated with the stock factor play a significant role for the risk of stock returns whereas productivity shocks associated with non-stock factor are the primary determinant of consumption risk, mainly because the non-stock asset composes the major share of the wealth portfolio. Thus, we may conclude that moderated productivity shocks associated with the factor of production whose ownership (income claim) is not traded must have led to the great moderation in consumption, while its impact on the volatility of stock returns has been relatively minor.

5.4.2 Empirical evidence

The above decomposition-analysis of macroeconomic risk based on our model implies that volatility of productivity shocks behind non-stock factor must be the main source which drives the volatility of consumption (output) growth. Given the divergence of
volatility movement between output (consumption) growth and stock market returns during the 1980s of the so called “great moderation” era, the model suggests that volatility of shocks underlying non-stock asset should decline substantially while shocks associated with stock asset barely change. If this is actually the case, we should observe according to the model that the growth rate of payouts from non-stock asset should also be greatly moderated while the volatility of stock dividend growth barely changes.

Our model allows the source of income (consumption) to be decomposed into stock payouts and non-stock payouts. For simplicity, we may assume that all assets in the model economy are divided into physical assets and human assets each of which provide stock payouts and non-stock payouts respectively in proportion to their marginal contribution to production as discussed above.\(^5\) Based on this interpretation of the model, stock payout is stock dividend or corporate profits, while non-stock payout is labor income. Thus, we are motivated to analyze whether the volatility movement of the growth rate of each payout – i.e. stock dividend and labor income growth – is consistent with what the model predicts. In response to the declining volatility of shocks associated with non-stock factor as discussed above, we should observe according to the model that volatility of the growth rate of stock dividends or corporate profit is barely moderated while the volatility of labor income growth is significantly moderated during the “great moderation” era of the 1980s. This prediction made by the model is actually supported empirically as shown in Figures 5.6 to 5.10.

Figures 5.6 and 5.7 show annualized per capita quarterly real growth rate of net corporate dividends and net corporate profits after tax respectively. According to these

\(^5\) Jaganathan & Wang (1996) also decompose the economy-wide wealth portfolio into two distinct groups: stock asset and human capital.
graphs, the growth rate of payouts from stock asset is barely moderated during the moderation period of the 1980s. Figures 5.8, 5.9 and 5.10 show annualized quarterly real growth rate of average hourly earnings, unit labor cost, and per capita compensation of employees respectively as gauges of the growth rate of payouts from non-stock asset, or human capital. These graphs all together confirm that volatility of the growth rate of non-stock asset payout is significantly moderated during the great moderation era of the 1980s. Thus, the prediction of our model is confirmed by the empirical evidence. In other words, the declining volatility of shocks underlying the non-stock factor of production may be the main source which must have brought out the great moderation of the 1980s so that barely moderated risk of stock asset could coexist with the significantly declining macroeconomic risk.

5.5 Stock market capitalization and the systematic risk of equity-holding

The covariance between consumption growth and stock returns defines the risk associated with holding equity in the consumption-based asset pricing model – also called consumption risk of equity-holding. That is, equity-holding is risky because it is closely linked to consumption against consumption-smoothing motives of the investor, not because its return is volatile in itself. Combined with the degree of risk aversion, the risk of equity-holding determines the equity premium over the riskless real interest rate. This is also true even for the modified C-CAPM such as that based on habit formation in Campbell & Cochrane (1999) – the important difference of the equity premium between the standard model and the habit-based model is the degree of risk-aversion, not the covariance between consumption growth and stock return – Cochrane (2005) characterizes the failure of the standard model as a quantitative one, not as a qualitative
one. In this section, we show how the consumption risk of holding equity might be influenced by the relative importance of the stock asset over the total wealth portfolio.

As Figure 5.3 shows, the stock-wealth-to-consumption ratio has been changing over time. This varying ratio reflects the change in the relative importance of stock asset over the total wealth portfolio – the ratio is captured by \( K \) in our model as discussed above. The relative importance of stock-asset over the portfolio, \( K \) influences the covariance and correlation between consumption growth and stock returns. Graphs 5.3 to 5.6 show how the change in \( K \) around the benchmark value of 0.03 affects the consumption risk of equity-holding under some plausible conditions. The graphs show that the increase in the stock share of the total wealth portfolio raises the covariance and correlation between consumption growth and stock returns; thus, the overall consumption risk due to equity-holding increases as consistent with Duffee (2005).

5.6 Conclusion

We have analyzed an economy where an investor holds a wealth portfolio that consists of two types of assets (stock-market traded and non-traded one). Each asset is used as a factor of production and then pays dividends to the investor in return for their marginal contribution to the production which determines consumption every period. In addition, each factor is associated with productivity shocks which ultimately characterize the statistical features of the economy. These ingredients combined with the standard C-CAPM are able to match the variance of and covariance between consumption growth and stock market returns during the postwar period. Our model implies that the great moderation in consumption growth may have originated from the declining productivity shocks to the factor of production whose ownership is not traded in stock market; thus,
the volatility of stock returns is barely moderated during the great moderation era. We also showed the mechanism of how the stock share of total wealth may have influenced the consumption risk of holding equity, consistent with empirical findings in literature.
Table 5.1: Variance and covariance of productivity shocks corresponding to the historical variance and covariance of stock returns and consumption growth for different degrees of substitutability $\sigma$ between factors of production at $K = 0.03$. Note that $\rho = (\sigma - 1) / \sigma$.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\sigma^2_p$</th>
<th>$\sigma^2_\gamma$</th>
<th>$\sigma_{p,\gamma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1500</td>
<td>0.0104</td>
<td>-0.2472</td>
</tr>
<tr>
<td>0.8</td>
<td>0.1870</td>
<td>0.0108</td>
<td>-0.3557</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2982</td>
<td>0.0126</td>
<td>-0.5934</td>
</tr>
<tr>
<td>0.2</td>
<td>0.7430</td>
<td>0.0238</td>
<td>-0.9032</td>
</tr>
<tr>
<td>0.001</td>
<td>148.3253</td>
<td>4.5860</td>
<td>-1.0000</td>
</tr>
<tr>
<td>-0.2</td>
<td>1.4819</td>
<td>0.0483</td>
<td>-0.9773</td>
</tr>
<tr>
<td>-0.8</td>
<td>0.1843</td>
<td>0.0122</td>
<td>-0.5369</td>
</tr>
<tr>
<td>-1.5</td>
<td>0.0980</td>
<td>0.0108</td>
<td>-0.3119</td>
</tr>
<tr>
<td>-2.5</td>
<td>0.0588</td>
<td>0.0104</td>
<td>-0.1411</td>
</tr>
<tr>
<td>-5</td>
<td>0.0300</td>
<td>0.0102</td>
<td>0.1116</td>
</tr>
<tr>
<td>-10</td>
<td>0.0167</td>
<td>0.0101</td>
<td>0.4380</td>
</tr>
<tr>
<td>-100</td>
<td>0.0099</td>
<td>0.0100</td>
<td>0.9883</td>
</tr>
</tbody>
</table>

Table 5.2: Change in the standard deviation of consumption growth and stock returns in response to 25% decline of productivity shocks to the non-stock factor (asset)

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\sigma_p$</th>
<th>$\sigma_\gamma$</th>
<th>$\sigma(g_1)$</th>
<th>$\sigma(\log R)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1500</td>
<td>0.0104</td>
<td>20.3% drop</td>
<td>no change</td>
</tr>
<tr>
<td>0.8</td>
<td>0.1870</td>
<td>0.0108</td>
<td>18.8% drop</td>
<td>0.004% drop</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2982</td>
<td>0.0126</td>
<td>18.8% drop</td>
<td>0.064% drop</td>
</tr>
</tbody>
</table>

Table 5.1: Variance and covariance of productivity shocks corresponding to the historical variance and covariance of stock returns and consumption growth for different degrees of substitutability $\sigma$ between factors of production at $K = 0.03$. Note that $\rho = (\sigma - 1) / \sigma$. 

Table 5.2: Change in the standard deviation of consumption growth and stock returns in response to 25% decline of productivity shocks to the non-stock factor (asset)
Graph 5.1: Standard deviation of consumption growth $\sigma(g)$, standard deviation of stock returns $\sigma(\log R)$, their correlation $\text{corr}(g, \log R)$ and the covariance $\text{cov}(g, \log R)$ (with the benchmark case of $K = 0.03$ indicated by the vertical line) for $\rho = 0.8$

Graph 5.2: Standard deviation of consumption growth $\sigma(g)$, standard deviation of stock returns $\sigma(\log R)$, their correlation $\text{corr}(g, \log R)$ and the covariance $\text{cov}(g, \log R)$ (with the benchmark case of $K = 0.03$ indicated by the vertical line) for $\rho = -2.5$
Graph 5.3: The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 1$, $\sigma_\rho = 0.15$ and $\sigma_\gamma = 0.01$ (assuming productivity shocks to each factor are not correlated $\sigma_{\beta\gamma} = 0$)

Graph 5.4: The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 0.5$, $\sigma_\rho = 0.15$ and $\sigma_\gamma = 0.01$ (assuming productivity shocks to each factor are not correlated $\sigma_{\beta\gamma} = 0$)
Graph 5.5: The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 0.8$, $\sigma_\beta = 0.12$, $\sigma_\gamma = 0.012$ and $\sigma_{\beta\gamma} = 0.12 \times 0.012 \times 0.05$ (assuming productivity shocks to each factor are positively correlated $\sigma_{\beta\gamma} > 0$)

Graph 5.6: The behavior of statistical properties for consumption growth and stock returns in response to the change of $K$ at around $K = 0.03$ for $\rho = 0.8$, $\sigma_\beta = 0.12$, $\sigma_\gamma = 0.012$ and $\sigma_{\beta\gamma} = 0.12 \times 0.012 \times 0.2$ (assuming productivity shocks to each factor are positively correlated $\sigma_{\beta\gamma} > 0$)
Figure 5.1: Growth rate of quarterly per capita real personal consumption expenditure for nondurable goods and services

Figure 5.2: Quarterly New York Stock Exchange real stock returns
Figure 5.3: Log ratio of quarterly stock wealth to nondurable consumption expenditures

Figure 5.4: Quarterly real stock return and quarterly growth rate of per capita real personal consumption expenditure for nondurable goods and services
Figure 5.5: Scatter-plot of quarterly real stock return and growth rate of per capita real personal consumption expenditure for nondurable goods and services

Figure 5.6: Quarterly per capita real net corporate dividend growth at an annual rate
Figure 5.7: Quarterly per capita real corporate profits after tax at an annual rate

Figure 5.8: Quarterly real growth rate of unit labor cost of non-farm business at an annual rate
Figure 5.9: Quarterly HP-filtered average hourly earnings growth at an annual rate

Figure 5.10: Quarterly per capita real compensation of employees at an annual rate
CHAPTER 6

CONCLUDING REMARKS

In this dissertation, I discuss and address two main issues regarding asset returns: a new econometric modeling framework for asset returns and puzzling features of the standard consumption-based asset pricing model. In chapter 2 and 3, I provide a new empirical guidance for modeling a skewed and thick-tailed error distribution along with GARCH effects that characterize U.S. stock data based on the theoretical derivation for the GARCH-skew-t model. In chapter 4 and 5, puzzling features of the consumption-based asset pricing model are discussed and possible solutions are provided. Chapter 4 shows that the high equity premium and low riskfree rate can be explained by parameter uncertainty underlying the Data Generating Process (DGP) that leads to a negatively skewed and thick-tailed perceived distribution of the DGP. Chapter 5 develops a new consumption-based asset pricing model which explains a possible link between stock market volatility and macroeconomic risk.

In chapter 2, the skew-Student distribution is derived as a normal variance-mean mixture with the inverted chi-square as a mixing distribution. The distribution turns out to be the same - but with different parameterization - as that derived as a special limiting case of the Generalized Hyperbolic (GH) distribution, which Aas & Haff (2006) call the GH skew Student’s t distribution. The moments, cumulants and MGF are also derived
based on the mixture-of-distributions setting and their properties are discussed. The skew-t is then extended to a dynamic model to represent volatility clustering as well as the skewed and leptokurtic properties of financial data. The dynamic skew-t model is derived by incorporating observation-driven modeling approach such as GARCH-type specifications so that the stochastic latent variance is updated upon arrival of new observations. Hence, the mixture-of-distributions setting is preserved over time while the scale parameter of the conditionally inverted-chi-squared variance changes as the information set is updated.

The skew-Student distribution is then applied to U.S. excess stock market returns and is shown to be much better than other familiar distributions such as Gaussian and Student t based on goodness of fit both conditionally and unconditionally – i.e. with and without GARCH effects. The equity premium is then computed based on each estimated distributional model and compared with one another. The MGF of the Student t explodes due to the heavy-tail density; hence, leading to an infinite equity premium as noted by Weitzman (2007). The skew-t may also lead to an infinite equity premium unless the skewness is sufficiently negative. However, the substantially negative skewness inherent in U.S. stock market returns results in a finite equity premium which is close to that of Gaussian both conditionally and unconditionally. The cumulant-based decomposition of the equity premium under the Gaussian and the skew-t implies that the impact of skewness and kurtosis on the equity premium is relatively trivial compared with the first two central moments for both conditional and unconditional distributional models.

In chapter 3, I find that the distribution of monthly U.S. stock market returns conditional on the Realized Volatility (RV) is normal based on the "variance-mean"
mixture specification with $RV$ as the mixing random variance $N(\mu + \beta RV, RV)$; hence, confirming the basic tenet of the mixture-of-distribution hypothesis originally proposed by Clark (1973). It is also found that the distribution of the realized volatility is well represented by the inverted-chi-square both conditionally and unconditionally. These empirical findings serve as the building blocks underlying the skew-Student distribution both conditionally and unconditionally. Hence, the empirical finding of the building blocks suggest according to the mixing rule that the distribution of returns should be well approximated by the skew-Student that is derived as a normal inverted-chi-square mixture distribution in chapter 2. I find that the implied skew-Student distribution is actually well supported by the U.S. stock data both conditionally and unconditionally – i.e. with and without GARCH effects. Thus, this paper provides a new empirical support for the distributional assumption underlying the GARCH-skew-t model. Moreover, I find a direct analytical connection between the realized volatility and the latent stochastic variance in the normal variance-mean mixture modeling of U.S. stock returns.

Chapter 4 shows that parameter uncertainty that leads to a negatively skewed and thick-tailed error distribution for the perceived DGP of the endowment process can provide a solution to asset return puzzles associated with the standard consumption-based asset-pricing model such as high equity premium and low risk free rate as been noted in macro-finance literature. The consumer-investor may perceive the endowment process as non-Gaussian as he/she is uncertain about the (variance) parameter value underlying the DGP of consumption growth. In this case, the investor may compute or derive the consequent perceived distribution for the consumption growth as a mixture of Normal and the probability distribution from which the stochastic parameter is drawn. For
tractability as well as simplicity, this paper assumes that the investor draws the uncertain (variance) parameter from the Inverse Gaussian (IG) distribution so that he/she perceives the resulting distribution for the endowment process as a skewed and thick-tailed NIG distribution. We find that parameter uncertainty can lead to a negatively skewed and fat-tailed distribution for the perceived consumption growth so that most of high equity premium and low riskfree rate can be explained even with the degree of risk-aversion below 10; thus, providing a possible solution to asset return puzzles as been paid much attention in literature.

Chapter 5 analyzes an economy where an investor holds a wealth portfolio that consists of two types of assets (stock-market traded and non-traded ones). Each asset is employed as a factor of production and then distributes payouts to the investor in return for their marginal contribution to production that determines the amount of consumption every period. In addition, each factor is associated with productivity shocks which ultimately characterize statistical features of the economy. These ingredients combined with the standard C-CAPM are able to match the variance of and covariance between consumption growth and stock market returns during the postwar period. The model implies that the great moderation in consumption growth may have originated from declining productivity shocks to the factor of production whose ownership is not traded in stock market; thus, the volatility of stock return is barely moderated during the great moderation era. We also show the mechanism of how the ratio of stock wealth to total wealth may have influenced the consumption risk of holding equity, consistent with empirical findings in literature.
APPENDIX A

DERIVATIONS FOR CHAPTER 2

A.1 Derivation of the skew-Student density function

\[ f_{\tilde{r}}(r \mid \nu, \beta, h, \mu) = f_{\tilde{r}}(r - \mu \mid \nu, \beta, h) \]

\[ = \int_{0}^{\infty} f_{\tilde{r}}(r - \mu \mid \beta, \theta) f_{\tilde{\theta}}(\theta \mid \nu, h) d\theta \]

Setting \( r' = r - \mu \),

\[ = \int_{0}^{\infty} f_{\tilde{r}}(r' \mid \beta, \tilde{\theta}) f_{\tilde{\theta}}(\theta \mid \nu, h) d\theta \]

\[ = \frac{h^\nu(\nu/2)^{\nu/2}}{\sqrt{2\pi \Gamma(\nu/2)}} \int_{0}^{\infty} \theta^{(\nu+1)/2-1} \exp\left\{ -\frac{1}{2} \left( \theta(r' - \beta/\theta)^2 + h^2 \nu \theta \right) \right\} d\theta \]

\[ = \frac{h^\nu(\nu/2)^{\nu/2}}{\sqrt{2\pi \Gamma(\nu/2)}} \exp(\beta r') \int_{0}^{\infty} \theta^{(\nu+1)/2-1} \exp\left\{ -\frac{1}{2} \left( \theta(r'^2 + h^2 \nu) + \beta^2 / \theta \right) \right\} d\theta \]

Utilizing the following integration in Gradshteyn & Ryzhik (1965),

\[ \int x^{\gamma-1} \exp\left\{ -\frac{1}{2} (ax + bx^{-1}) \right\} dx = 2K_{\gamma}(a^{1/2}b^{1/2}) \left( \frac{b^{1/2}}{a^{1/2}} \right)^{\gamma} \text{ for } a > 0, b > 0, \quad (2.18) \]

where \( K_{\gamma}(\cdot) \) is the modified Bessel function of the second type of order \( \gamma \).

The above integration is computed as below.

\[ \int \theta^{(\nu+1)/2-1} \exp\left\{ -\frac{1}{2} ((r'^2 + h^2 \nu)\theta + \beta^2 / \theta) \right\} d\theta = 2K_{(\nu+1)/2}(\beta | \sqrt{r'^2 + \nu}) \left( \frac{\beta}{\sqrt{r'^2 + \nu}} \right)^{(\nu+1)/2} \]

Continuing from above,
A.2 Derivations for the mean, variance, skewness and kurtosis of the skew-Student distribution

First, note that $\bar{r} = \mu + \beta \theta^{-1} + \theta^{-1/2} z$

where $z$ is drawn from $N(0,1)$ and is assumed independent of $\theta$.

$E[\bar{r}] = \mu + \beta E[\theta^{-1}]$ where $E[\theta^{-1}]$ is computed utilizing the gamma function as below.

$$E[\theta^{-1}] = \frac{\Gamma(\nu/2)}{\nu} \int_0^\infty \frac{\theta^{\nu/2-2} \exp(-h^2 \nu \theta/2) \, d\theta}{\Gamma(\nu/2)}$$

Setting $x = -h^2 \nu \theta / 2$, $dx = -(h^2 \nu / 2) \, d\theta$,

$$= \frac{h^2 \nu}{\Gamma(\nu/2)} \int_0^\infty x^{(\nu-2)/2-1} \exp(-x) \, dx$$

Note that the integral gives rise to $\Gamma(\frac{\nu-2}{2})$.

$$= \frac{h^2 \nu \Gamma((\nu-2)/2)}{\Gamma(\nu/2)}$$

$$= \frac{h^2 \nu}{(\nu-2)}$$

$Var[\bar{r}] = Var[\mu + \beta \theta^{-1} + \theta^{-1/2} z]$ where $z$ is the standard normal distribution

$$= \beta^2 Var[\theta^{-1}] + Var[\theta^{-1/2} \, z] + 2 \beta Cov[\theta^{-1}, \theta^{-1/2} \, z]$$

$$= \beta^2 Var[\theta^{-1}] + E[\theta^{-1}]$$

This is because $\theta$ and $z$ are assumed to be independent of each other.
\[ \text{Var}[\bar{\theta}^{-1}] = E[\bar{\theta}^{-2}] - (E[\bar{\theta}^{-1}])^2 \] in the above.

Utilizing the gamma function again as above, the following equation can be derived.

\[ E[\bar{\theta}^{-2}] = \frac{\nu^2 h^4}{(\nu - 2)(\nu - 4)} \]

Thus, we can derive \( \text{Var}[\bar{r}] \) as (2.5).

\[ \text{Skw}[\bar{r}] = \frac{E[(\bar{r} - E[\bar{r}])^3]}{(\text{Var}[\bar{r}])^{3/2}} \]

where \( E[(\bar{r} - E[\bar{r}])^3] = E[\{\beta(\bar{\theta}^{-1} - E[\bar{\theta}^{-1}]) + \bar{\theta}^{-1/2}z\}^3] \)

Note that \( E[\bar{\theta}^{-3}] \) can be computed as above utilizing the gamma function.

That is, \( E[\bar{\theta}^{-3}] = \frac{h^6 \nu^3}{(\nu - 2)(\nu - 4)(\nu - 6)} \)

Thus, we can derive \( \text{Skw}[\bar{r}] \) as (2.6).

\[ \text{Kts}[\bar{r}] = \frac{E[(\bar{r} - E[\bar{r}])^4]}{(\text{Var}[\bar{r}])^2} \]

\[ E[(\bar{r} - E[\bar{r}])^4] = E[\{\beta(\bar{\theta}^{-1} - E[\bar{\theta}^{-1}]) + \bar{\theta}^{-1/2}z\}^4] \]

Expansion of this equation consists of \( E[\bar{\theta}^{-1}] \), \( E[\bar{\theta}^{-2}] \), \( E[\bar{\theta}^{-3}] \) and \( E[\bar{\theta}^{-4}] \).

Note that \( E[\bar{\theta}^{-4}] \) can be computed as above utilizing the gamma function.

That is, \( E[\bar{\theta}^{-4}] = \frac{h^8 \nu^4}{(\nu - 2)(\nu - 4)(\nu - 6)(\nu - 8)} \)

Thus, we can derive \( \text{Kts}[\bar{r}] \) as (2.7).
A.3 Derivation of the MGF of the skew t distribution

The MGF exists provided \( t \geq 0 \) with \( \beta < -t/2 \) or \( t \leq 0 \) with \( \beta > -t/2 \).

\[
M_f(t \mid \nu, \beta, h, \nu) = E[e^{t \bar{X}}]
\]
\[
= \int_{-\infty}^{\infty} e^{t'r} f_r(r \mid \nu, \beta, h, \mu) dr
\]
\[
= \int_{-\infty}^{\infty} e^{t'r} f_r(r - \mu \mid \nu, \beta, h) dr
\]

Setting \( r' = r - \mu \), \( dr' = dr \),

\[
e^{t'\mu} \int_{-\infty}^{\infty} e^{t'r} \int_{\bar{\theta}, \beta} f_{\bar{\theta}}(r' \mid \bar{\theta}, \beta) f_{\bar{\theta}}(\theta \mid \nu, h) d\theta dr'
\]
\[
e^{t'\mu} \int_{-\infty}^{\infty} e^{t'r} \left[ \frac{\theta^{1/2}}{\sqrt{2\pi}} \right] \exp\left\{ -\frac{1}{2} \left( \theta(r' - \frac{\beta}{\theta})^2 \right) \right\} \frac{h^\nu (\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \theta^{\nu/2-1} \exp\left( -\frac{1}{2} h^2 \nu \theta \right) d\theta dr'
\]
\[
e^{t'\mu} \frac{h^\nu (\nu/2)^{\nu/2}}{\sqrt{2\pi} \Gamma(\nu/2)} \int_{-\infty}^{\infty} e^{t'r} \left[ \theta^{(\nu-1)/2} \exp\left\{ -\frac{1}{2} \left( \theta(r'^2 + h^2 \nu + \beta^2 / \theta) \right) \right\} d\theta dr'
\]
\[
e^{t'\mu} \frac{h^\nu (\nu/2)^{\nu/2}}{\sqrt{2\pi} \Gamma(\nu/2)} \int_{-\infty}^{\infty} \theta^{(\nu-1)/2} \exp\left\{ -\frac{1}{2} \left( \theta^2 \nu + \beta^2 / \theta \right) \right\} \int_{-\infty}^{\infty} \exp\left\{ (t + \theta)r' - \frac{1}{2} \theta r'^2 \right\} dr' d\theta
\]

Using the result of Gaussian integration,

\[
\int_{-\infty}^{\infty} \exp\left\{ (t + \theta)r' - \frac{1}{2} \theta r'^2 \right\} dr' = \exp\left( \frac{(t + \beta)^2}{2\theta} \right) \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2} \left( \theta (r' - \frac{t + \beta}{\theta})^2 \right) \right\} dr'
\]
\[
= \sqrt{2\pi} \theta^{-1/2} \exp(\frac{(t + \beta)^2}{2\theta})
\]

Continuing from above,

\[
e^{t'\mu} \frac{h^\nu (\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \int_{0}^{\infty} \theta^{(\nu-1)/2} \exp\left\{ -\frac{1}{2} \left( \theta^2 \nu - \frac{t(t + 2\beta)}{\theta} \right) \right\} d\theta
\]
Using the integration (2.18) in Appendix A.1,

\[
2e^{\mu} h^{\nu/2} (\nu/4)^{\nu/4} (-t(t + 2\beta))^{\nu/4} \frac{K_{\nu/2}(h\sqrt{-\nu t(t + 2\beta)})}{\Gamma(\nu/2)}
\]

A.4 Derivation of the CDF and numerical approximation for the skew-Student distribution

\[
F_\tau (r) = \int_0^\infty F_{r|\theta} (r | \theta) F_\theta (d\theta)
\]

Setting \( s = F_\theta (\theta) \); \( ds = F_\theta (d\theta) = f_\theta (\theta) d\theta \),

\[
F_\tau (r) = \int_0^1 F_{r|\theta} (r | F_\theta^{-1}(s)) ds
\]  

(2.19)

where \( F_{r|\theta} (r | \theta) \) is the conditional Gaussian CDF, and \( F_\theta (\theta) \) is the CDF of the chi-square distribution. That is, \( F_{r|\theta} (r | \theta) = \Phi((r - \mu - \frac{\beta}{\theta})\theta^{1/2}) \) and \( F_\theta (\theta) = C_\nu (h^2 \nu \theta) \)

where \( \Phi(\cdot) \) is the standard Gaussian CDF, and \( C_\nu (\cdot) \) is the standard chi-square CDF with \( \nu \) DOF. For a numerical integration, the unconditional CDF, equation (2.19) is approximated as the following.

\[
F_\tau (r) \approx \sum_{i=1}^n w_i \Phi((r - \mu - \frac{\beta}{\theta_i})\theta_i^{1/2}) \text{ where } \theta_i = C_\nu^{-1}(s_i) / h^2 \nu \text{ from above}
\]

To utilize the ternary integration\(^1\), \( w \) and \( s \) are set in the following manner.

\[
s_i = \begin{cases} 
(i - 2/3)/n & \text{if } i \text{ is odd} \\
(i - 1/3)/n & \text{if } i \text{ is even} 
\end{cases}
\]

\[
w_i = 1/n
\]

\(^1\) The ternary integration for the numerical computation is suggested by J. Huston McCulloch.
A.5 The MGF, CGF, moments, and cumulants in general

The Moment Generating Function (MGF) is defined as

\[ M_X(t) = E[\exp(tX)] = 1 + \sum_{n=1}^{\infty} \frac{E[X^n]}{n!} t^n \]

where \( E[X^n] \) is the \( n \)th raw moment. This definition of the MGF implies that the equity premium is simply a log of the MGF evaluated at \( t = 1 \). That is \( EP = \log\{M_X(1)\} \). The Cumulant Generating Function (CGF) is defined as a log of the MGF.

\[ K_X(t) = \log M_X(t) = \sum_{n=1}^{\infty} \frac{k_n}{n!} t^n = \sum_{n=1}^{\infty} \frac{K^{(n)}(0)}{n!} t^n \]

where \( k_n \) is the \( n \)th cumulant and \( K^{(n)}(0) \) is the \( n \)th derivative of the CGF at zero.

Thus, the equity premium is simply \( K_X(1) \). From the definition of the MGF and CGF,

\[ 1 + \sum_{n=1}^{\infty} \frac{E[X^n]}{n!} t^n = \exp(\sum_{n=1}^{\infty} \frac{k_n}{n!} t^n) \]

Expanding as a sum of polynomials and then comparing coefficients to both sides, the following relations between moments and cumulants can be derived:

\[ k_1 = E[X] \]
\[ k_2 = E[(X - EX)^2] = Var[X] \]
\[ k_3 = E[(X - EX)^3] = k_2^{3/2} Skw[X] \]
\[ k_4 = E[(X - EX)^4] - 3k_2^2 = k_2^2 (Kts[X] - 3) \]
\[ k_5 = E[(X - EX)^5] - 10k_2k_3 \]

……

where \( Skw[X] \) is skewness and \( Kts[X] \) is kurtosis of the random variable \( X \).
APPENDIX B

INVERTED CHI-SQUARE DISTRIBUTION AND THE CDF

The reciprocal of the variance is drawn from the scaled chi-square distribution in (3.2).

Defining $\tilde{\theta} \equiv \tilde{V}^{-1}$, called the precision,

$$\tilde{\theta} \sim f_{\tilde{\theta}}(\theta \mid h, \nu) = \frac{h^{\nu} (\nu / 2)^{\nu/2}}{\Gamma(\nu/2)} \theta^{\nu/2-1} \exp(-\frac{1}{2} h^2 \nu \theta)$$

where $\Gamma(\cdot)$ is the gamma function. If a random variable $x$ is drawn from a density function $f(x)$, then it can be shown that its reciprocal $y(\equiv x^{-1})$ is drawn from $g(y) = y^{-2} f(y^{-1})$. Utilizing this relationship, the probability density function of the variance is given by (3.3). The inverted chi-square is also called the scale-inverse-chi-square distribution, which we use interchangeably in this paper. Note that the scale-inverse-chi-square distribution of a random variable $\tilde{V}$ is related to the inverse-gamma distribution by $\tilde{V} \sim Inv\_Gamma(\nu/2, h^2 \nu / 2)$. The inverse-gamma distribution is again related to the gamma distribution by $\tilde{V}^{-1} \sim Gamma\left(\frac{\nu}{2}, \frac{2}{h^2 \nu}\right)$.

The CDF of the scale-inverse-chi-square distribution is then given by
\[ F_p(V_i | h, \nu) = \int_0^{\nu_i} f_p(x | h, \nu) dx \]

\[ = \frac{1}{\Gamma(\nu/2)} \{ y^{\nu/2-1} \exp(-y) dy - \int_0^{(h^2\nu)/(2V_i)} y^{\nu/2-1} \exp(-y) dy \} \]

\[ = 1 - \frac{\Gamma\left(\frac{h^2\nu}{2V_i}, \frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \]

where \( \Gamma(\cdot, \nu/2) \) is the incomplete gamma function and \( \Gamma(\nu/2) \) is the gamma function.
C.1 Cumulant Generating Function (CGF) and cumulants

The CGF is simply the log of the MGF and can be expanded as the summation of polynomials just as the Taylor series expansion. In addition, it is very useful to analyze properties of economic models in relation to the size of cumulants with different orders.

If \( m(\theta) \) is the MGF, the CGF given by \( k(\theta) = \log \{ m(\theta) \} \). It can be expanded as the summation of polynomials with cumulants as below.

\[
\begin{align*}
  k(\theta) &= \log \{ E[\exp(\theta y)] \} \\
  &= \sum_{n=1}^{\infty} \kappa_n \theta^n / n! \\
  &= \kappa_1 \theta + \kappa_2 \theta^2 / 2! + \kappa_3 \theta^3 / 3! + \kappa_4 \theta^4 / 4! + \ldots
\end{align*}
\]

where \( \kappa_n \) is the n-th order cumulant, which is the n-th derivative of the CGF evaluated at zero. That is \( \kappa_n = k^{(n)}(0) \). The central moments can be computed as a combination of cumulants. The first few cumulants are expressed in terms of central moments as below.

\[
\begin{align*}
  \kappa_1 &= E[y] \\
  \kappa_2 &= Var[y] \\
  \kappa_3 &= (\kappa_2)^{3/2} Skw(y) \\
  \kappa_4 &= (\kappa_2)^2 (Kts(y) - 3)
\end{align*}
\]

Note that higher order cumulants above the 4th are a little more complicated than this. A Gaussian distribution gives rise to only the first two cumulants – higher cumulants beyond the 2nd are zero. This property of the Gaussian distribution implies that asset-
pricing features are determined only by mean and variance of the underlying distribution – i.e. higher moments don’t contribute to the asset-pricing features.

C.2 The MGF (CGF) for Gaussian distribution

If the log consumption growth rate $g$ were drawn from the normal distribution, its MGF is given by

$$m(\theta) = E[\exp(\theta g)] = \exp\{\theta E[g] + \theta^2 \text{Var}[g]/2\}.$$ 

Since the CGF is the log of the MGF, the CGF is given by

$$k(\theta) = \theta E[g] + \theta^2 \text{Var}[g]/2.$$ 

Note that the CGF of the Gaussian distribution has only the first and second cumulants which are the mean and variance of the log consumption growth rate respectively.
APPENDIX D

DERIVATIONS FOR CHAPTER 5

D.1 Price-dividend ratio under iid shocks to each pay-out (stock and non-stock)

\[ U'(C_t)P_t = e^{-\delta} E_t[U'(C_{t+1})(P_{t+1} + D_{t+1})] \]

\[ P_t = e^{-\delta} E_t\left[ \frac{U'(C_{t+1})}{U'(C_t)}(P_{t+1} + D_{t+1}) \right] \]

Under the CRRA preference, the growth of marginal utilities is given as a power function of gross consumption growth rate.

\[ P_t = e^{-\delta} E_t\left[ \left( \frac{C_{t+1}}{C_t} \right)^{-\eta} (P_{t+1} + D_{t+1}) \right] \]

\[ P_t = e^{-\delta} E_t\left[ \left( \frac{C_{t+1}}{C_t} \right)^{-\eta} (P_{t+1} + D_{t+1}) \right] \]

\[ = E_t\left[ e^{-\delta} \left( \frac{C_{t+1}}{C_t} \right)^{-\eta} D_{t+1} + e^{-2\delta} \left( \frac{C_{t+2}}{C_t} \right)^{-\eta} D_{t+2} + e^{-3\delta} \left( \frac{C_{t+3}}{C_t} \right)^{-\eta} D_{t+3} + \ldots \right] + \lim_{T \to \infty} e^{-T\delta} \left( \frac{C_{t+T}}{C_t} \right)^{-\eta} P_T \]

Under the TVC, \( \lim_{T \to \infty} e^{-T\delta} \left( \frac{C_{t+T}}{C_t} \right)^{-\eta} = 0 \)

\[ P_t = E_t\left[ \sum_{j=1}^{\infty} e^{-j\delta} \left( \frac{C_{t+j}}{C_t} \right)^{-\eta} D_{t+j} \right] \]

\[ \frac{P_t}{D_t} = E_t\left[ \sum_{j=1}^{\infty} e^{-j\delta} \left( \frac{C_{t+j}}{C_t} \right)^{-\eta} D_{t+j} \right] \]
\[
\sum_{j=1}^{\infty} e^{-j\beta} E_t \left[ \left( \frac{C_{t+j}}{C_t} \right)^{-\eta} \left( \frac{D_{t+j}}{D_t} \right) \right]
\]

\[
\sum_{j=1}^{\infty} e^{-j\beta} E_t \left[ \left( \frac{C_{t+1} C_{t+2} \ldots C_{t+j}}{C_t C_{t+1} \ldots C_{t+j-1}} \right)^{-\eta} \left( \frac{D_{t+1} D_{t+2} \ldots D_{t+j}}{D_t D_{t+1} \ldots D_{t+j-1}} \right) \right]
\]

\[
\sum_{j=1}^{\infty} e^{-j\beta} E_t \left[ \exp \{-\eta \left( g_{t+1} + g_{t+2} + \ldots + g_{t+j} \right) \} \right]
\]

where \( g_{t+j} \) is log consumption growth rate while \( g_{t+j}^D \) is log stock-dividend growth rate.

\[
\frac{P_t}{D_t} = \sum_{j=1}^{\infty} e^{-j\beta} E_t \left[ \exp \{-\eta \left( g_{t+1} + g_{t+2} + \ldots + g_{t+j} \right) \} \right]
\]

\[
= \sum_{j=1}^{\infty} e^{-j\beta} E_t \left[ \exp \{-\eta g_{t+1} \} \exp \{-\eta g_{t+2} \} \ldots \exp \{-\eta g_{t+j} \} \right]
\]

Note that \( \exp \{-\eta g_{t+j} + g_{t+j}^D \} \) is independently and identically distributed because shocks to each payouts are iid. Thus,

\[
\frac{P_t}{D_t} = \sum_{j=1}^{\infty} e^{-j\beta} E_t \left[ \exp \{-\eta g_{t+1} + g_{t+1}^D \} \right] E_t \left[ \exp \{-\eta g_{t+2} + g_{t+2}^D \} \right] \ldots E_t \left[ \exp \{-\eta g_{t+j} + g_{t+j}^D \} \right]
\]

Thus, we can conclude that the price-dividend ratio is constant under the iid assumption for productivity shocks and the TVC.

D.2 Computing \( Var_t (\log C_{t+1}) \), \( Var_t (\log \epsilon_{ct,t+1}) \) and \( Cov_t (\log C_{t+1}, \log \epsilon_{ct,t+1}) \)

\[
C = A \{ \theta(\beta S)^{\rho} + (1 - \theta) (\gamma N)^{\rho} \}^{1/\rho}
\]

\[
\log C = \log A + \left( 1/\rho \right) \log \{ \theta(\beta S)^{\rho} + (1 - \theta) (\gamma N)^{\rho} \}
\]

\[
= \log A + \left( 1/\rho \right) \log \{ \theta S^{\rho} e^{\rho \log \beta} + (1 - \theta) N^{\rho} e^{\rho \log \gamma} \}
\]

Letting \( y \equiv \log \beta \) and \( z \equiv \log \gamma \),
\[ \log C = \log A + (1/\rho) \log \{ \theta S^\rho e^{\rho y} + (1 - \theta) N^\rho e^{\rho x} \} \]

We consider \( \log \{ \theta S^\rho e^{\rho y} + (1 - \theta) N^\rho e^{\rho x} \} \) as a function of random variables \( y(= \log \beta) \) and \( z(= \log \gamma) \). Let \( f(y, z) \equiv \log \{ \theta S^\rho e^{\rho y} + (1 - \theta) N^\rho e^{\rho x} \} \). To the first order approximation,

\[ f(y, z) \approx f(\mu_y, \mu_z) + f_y(\mu_y, \mu_z)(y - \mu_y) + f_z(\mu_y, \mu_z)(z - \mu_z) \]

where

\[ f_y = f_y(\mu_y, \mu_z) = \rho \frac{\theta S^\rho e^{\rho \mu_y}}{\theta S^\rho e^{\rho \mu_y} + (1 - \theta) N^\rho e^{\rho \mu_y}} = \rho K \]

\[ f_z = f_z(\mu_y, \mu_z) = \rho \frac{(1 - \theta) N^\rho e^{\rho \mu_z}}{\theta S^\rho e^{\rho \mu_y} + (1 - \theta) N^\rho e^{\rho \mu_y}} = \rho H \]

Utilizing the above equations, the conditional variance of consumption growth is approximated as below.

\[ \text{Var}_t(\log C_{t+1}) = \text{Var}_t[\log A + (1/\rho) \log \{ \theta (\beta_{t+1} S_{t+1})^\rho + (1 - \theta) (\gamma_{t+1} N_{t+1})^\rho \} ] \]

\[ = (1/\rho^2) \text{Var}_t[\log \{ \theta S_{t+1}^\rho e^{\rho \log \beta_{t+1}} + (1 - \theta) N_{t+1}^\rho e^{\rho \log \gamma_{t+1}} \} ] \]

\[ \approx (1/\rho^2) \text{Var}_t[f(\mu_y, \mu_z) + f_y(\mu_y, \mu_z)(y - \mu_y) + f_z(\mu_y, \mu_z)(z - \mu_z)] \]

\[ = (1/\rho^2) \{ f_y^2 \sigma_y^2 + f_z^2 \sigma_z^2 + 2f_y f_z \sigma_y \} \]

\[ = (1/\rho^2) \{ \rho^2 K^2 \sigma_y^2 + \rho^2 H^2 \sigma_z^2 + 2\rho^2 KH \sigma_y \} \]

The elasticity of output with respect to the stock asset \( \varepsilon_{c,s} \) is given by

\[ \varepsilon_{c,s} = \frac{\partial C}{\partial S} \frac{S}{C} = \frac{\theta S^\rho e^{\rho y}}{\theta S^\rho e^{\rho y} + (1 - \theta) N^\rho e^{\rho x}} \]

The conditional variance of the stock elasticity of output \( \varepsilon_{c,s} \) is then approximated as below.
\[ \text{Var}_t(\log \varepsilon_{c,t+1}) = \text{Var}_t(\rho y - \log \{\theta S^\rho e^{\rho y} + (1 - \theta)N^\rho e^{\rho y}\}) \\
\approx \text{Var}_t(\rho y - f_y(y - \mu_y) - f_z(z - \mu_z)) \\
= \text{Var}_t(\rho y - \rho Ky - \rho Hz) \\
= \rho^2(1 - K)^2\text{Var}_t(y) + \rho^2 H^2\text{Var}_t(z) - 2\rho^2(1 - K)HCov_t(y, z) \\
= \rho^2(1 - K)^2\sigma^2_\beta + \rho^2 H^2\sigma^2_\gamma - 2\rho^2(1 - K)H\sigma_{\rho_y} \\
= \rho^2 H^2(\sigma^2_\beta + \sigma^2_\gamma - 2\sigma_{\rho_y}) \]

\[ \text{Cov}_t(\log C_{t+1}, \log \varepsilon_{c,t+1}) \\
= \text{Cov}_t((1/ \rho) \log \{\theta S^\rho e^{\rho y} + (1 - \theta)N^\rho e^{\rho y}\}, \rho y - \log \{\theta S^\rho e^{\rho y} + (1 - \theta)N^\rho e^{\rho y}\}) \\
= \text{Cov}_t(\log \{\theta S^\rho e^{\rho y} + (1 - \theta)N^\rho e^{\rho y}\}, y) - (1/ \rho)\text{Var}_t(\log \{\theta S^\rho e^{\rho y} + (1 - \theta)N^\rho e^{\rho y}\}) \\
= \text{Cov}_t(f(y, z), y) - (1/ \rho)\text{Var}_t(f(y, z)) \\
\approx \text{Cov}_t(y, \rho K(y - \mu_y) + \rho H(z - \mu_z)) - (1/ \rho)\text{Var}_t(\rho K(y - \mu_y) + \rho H(z - \mu_z)) \\
= \rho H\{K\sigma^2_\beta - H\sigma^2_\gamma + (H - K)\sigma_{\rho_y}\} \]

We can also compute the growth rate of pay-offs from non-stock asset as below.

\[ D_n = \frac{\partial C}{\partial N} N = \frac{\partial C}{\partial N} C = \varepsilon_{c,n} C \]

where \( \varepsilon_{c,n} = \frac{\partial C}{\partial N} C \) is the elasticity of output with respect to non-stock asset

\[ D_n = A(1 - \theta)\gamma^\rho N^\rho \{\theta S^\rho \beta^\rho + (1 - \theta)\gamma^\rho N^\rho\}^{1/\rho - 1} \]

The conditional variance of the log growth rate of the non-stock asset pay-out is given by

\[ \text{Var}_t[\log(D_{n,t+1}/D_{n,t})] = \text{Var}_t[\log(D_{n,t+1})] \\
= \text{Var}_t[\rho \log \gamma_{t+1} + ((1 - \rho)/ \rho) \log \{\theta S^\rho \beta_{t+1}^\rho + (1 - \theta)\gamma_{t+1}^\rho N^\rho\}] \\
= \text{Var}_t[\rho x + ((1 - \rho)/ \rho) f(y, z)] \\
\approx \text{Var}_t[\rho x + ((1 - \rho)/ \rho)(f_y y + f_z z)] \\
= \text{Var}_t[\rho((1 - \rho)/ \rho)f_y y + \rho + ((1 - \rho)/ \rho)f_z z] \\
= \text{Var}_t[(1 - \rho)Ky + \rho + (1 - \rho)H]z \\
= (1 - \rho)^2 K^2\sigma^2_\beta + (\rho + (1 - \rho)H)^2\sigma^2_\gamma + 2K\rho(1 - \rho) + (1 - \rho)^2 H\sigma_{\rho_y} \]
BIBLIOGRAPHY


