Essays on the Term Structure of Interest Rates and Long Run Variance of Stock Returns

Dissertation

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

By

Ting Wu

Graduate Program in Economics

The Ohio State University

2010

Dissertation Committee:

Professor Pok-sang Lam, Adviser

Professor Paul Evans

Professor Huston McCulloch
Copyright by

Ting Wu

2010
ABSTRACT

My dissertation contains three chapters. It studies the bond market as well as the stock market. For the bond market I examine aspects of the term structure of interest rates using macro models with the goal of advancing our understanding of the pricing of bonds with different maturities from a macroeconomic perspective. For the stock market, I study the variance of stock returns over different investment horizons.

In Chapter 1, "Bond Pricing with Model Uncertainty", I propose and implement a term structure model based on risk-sensitive preferences. Following Hansen and Sargent (2008), I model a risk-sensitive consumer who shows aversion to uncertainties, and I evaluate his utility using the max-min utility function. He considers three types of uncertainties: (a) uncertainty of future states conditional on current states; (b) uncertainty about current states; and (c) uncertainty about the model generating the data. I use a parameter to represent his aversion to the each type of uncertainty. The max-min utility function implies multiplicative adjustments to the standard pricing kernel. A term premium results because these uncertainties figure more prominently in the pricing of long term bonds.

The pricing kernel is combined with the exogenous processes of consumption growth and inflation to price the bond yields. I specify two bivariate long run risk models to represent the model uncertainty. The two models share a common specification of consumption growth, but the inflation process differs across the models:
in one model inflation is non-stationary, while in the other it is stationary. The risk-sensitive consumer behaves as if the probability estimates for the models are tilted in favor of the one implying lower lifetime utility. The model with non-stationary inflation turns out to have the lower lifetime utility. As the probability estimates are tilted to favor the model with non-stationary inflation, both the short rate and yield spread increase. I show that this phenomenon helps to explain the high short rate and yield spread in the 1980's. More generally, I show that the model fits the observed shape of the yield curve, volatility of long yields, predictability of excess bond returns and correlation between yields of different maturities.

In Chapter 2, "Forecasting Bond Returns in a Macro Model", I consider the predictability of excess bond returns. Recent research has shown that a forecasting factor based on the forward rates has significant predictive power for excess bond returns at all maturities. In this chapter, I investigate the macroeconomic factors underlying those forward rates. I specify a rich stochastic general equilibrium model and use the Bayesian method to extract key macro variables such as habit, the government spending shock, the technology shock, the inflation target and the monetary policy shock. I then relate them to the forecasting factor and show that the forecasting factor is mostly capturing the effect of technology shock. Following the literature, I construct a forecasting factor based on a linear combination of extracted macro variables. This new factor predicts both excess bond returns and equity returns better than the forecasting factor based on forward rates.

In Chapter 3, "Decomposing the Variance-covariance Matrix: A Reinvestigation of Long Run Stock Variance", I reinvestigate the long run variance of stock returns following Pastor and Stambaugh (2009), who find that stock returns are riskier in
the long run. As in Pastor and Stambaugh (2009), I use a Bayesian approach to assess the risk. I find that their conclusion is likely to be sensitive to the prior of the correlation between innovation in expected returns and unexpected returns. The correlation plays a key role in determining the riskiness of stock returns in the long run through the mean reverting component. My analysis suggests that their result depends critically on a prior that is sufficiently uninformative. If the prior of a highly negative correlation is sufficiently informative, the result would be overturned. I also find that their conclusion is robust to the addition of dividend growth into the predictive system. By estimating \( \rho_{uu} \) with a sharp prior distribution, I show that the posterior draws of \( \rho_{uw} \) are sufficiently negative to generate a variance ratio smaller than 1 for 30 year stock returns.
I dedicate this dissertation to my parents for their continued encouragement and love throughout my life.
ACKNOWLEDGMENTS

Many individuals have contributed to the completion of my degree. I would like to give special thanks to the following:

My adviser, Dr. Pok-sang Lam, for providing continuous advice and support throughout my study.

My committee members: Dr. Paul Evans, Dr. Huston McCulloch, and Dr. Masao Ogaki for their expertise and time throughout my dissertation experience.

Dr. Bill Dupor for his useful comments and suggestions for my job market interview.
VITA

July 27, 1983 .......................... Born — China

June 30, 2005 .......................... BS, Applied Mathematics, Huazhong University of Science and Technology, China

June 30, 2005 .......................... BA, Economics, Huazhong University of Science and Technology, China

June 30, 2006 .......................... MA, Economics, The Ohio State University, USA

FIELDS OF STUDY

Major Field: Economics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Abstract</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
</tbody>
</table>

## Chapters:

1. Bond Pricing with Model Uncertainty | 1 |
   1.1 Introduction | 1 |
   1.1.1 Stylized Facts and Failures of Traditional Consumption Based Models | 1 |
   1.1.2 Introducing Hansen and Sargent’s Fragile Beliefs | 3 |
   1.1.3 Relation to Other Consumption Based Model of the Term Structure of Interest Rates | 5 |
   1.1.4 Organization | 7 |
   1.2 Pricing Theory and Data | 8 |
   1.3 Step 1: A Model with One Risk Sensitivity Operator | 9 |
   1.4 Step 2: A Model with Unobserved States and Another Risk Sensitivity Adjustment | 14 |
   1.5 Step 3: Model Selection and Another Risk Sensitivity Operator | 19 |
   1.5.1 Calibration | 23 |
   1.5.2 Short Rate and Yield Spread | 24 |
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Average Nominal Yield Curve (Step 1 Model)</td>
</tr>
<tr>
<td>1.2</td>
<td>Average Nominal Yield Curve (Step 2 Model)</td>
</tr>
<tr>
<td>1.3</td>
<td>Average Nominal Yield Curve: Step 3 Model vs PS Model</td>
</tr>
<tr>
<td>1.4</td>
<td>Correlation of Yields with Different Maturities</td>
</tr>
<tr>
<td>1.5</td>
<td>Forecasting Bond Returns Using Forecast Factor</td>
</tr>
<tr>
<td>1.6</td>
<td>Forecasting Bond Returns Using Forward Spread</td>
</tr>
<tr>
<td>2.1</td>
<td>Prior Distribution</td>
</tr>
<tr>
<td>2.2</td>
<td>Posterior Distribution</td>
</tr>
<tr>
<td>2.3</td>
<td>Regression of Forecasting Factor on Different Macro Variables</td>
</tr>
<tr>
<td>2.4</td>
<td>Regression of the Average One Year Excess Return of 2,3,4,5 Year Bond on Macro Variables and the Forecasting Factor</td>
</tr>
<tr>
<td>2.5</td>
<td>Correlations of Two Groups of Macro Variables</td>
</tr>
<tr>
<td>2.6</td>
<td>Forecasts of Excess Bond Return</td>
</tr>
<tr>
<td>2.7</td>
<td>Forecasting Stock Return Using Macro Factor</td>
</tr>
<tr>
<td>3.1</td>
<td>Posterior Distributions</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Short Rate and Yield Spread: Step 1 Model</td>
<td>13</td>
</tr>
<tr>
<td>1.2</td>
<td>Step 2 Model: Short Rate and Yield Spread</td>
<td>18</td>
</tr>
<tr>
<td>1.3</td>
<td>Probability Assigned to Model 0 with $p_0 = 0.5$</td>
<td>21</td>
</tr>
<tr>
<td>1.4</td>
<td>Probability Assignment vs Short Rate</td>
<td>25</td>
</tr>
<tr>
<td>1.5</td>
<td>Probability Difference vs Yield Spreads</td>
<td>27</td>
</tr>
<tr>
<td>1.6</td>
<td>The Unconditional Mean of Consumption Growth Rates for Two Models</td>
<td>28</td>
</tr>
<tr>
<td>1.7</td>
<td>Initial Responses of the Unconditional Mean of the Consumption Growth Rate</td>
<td>29</td>
</tr>
<tr>
<td>1.8</td>
<td>Short Rate Comparison</td>
<td>31</td>
</tr>
<tr>
<td>1.9</td>
<td>Yield Spread Comparison</td>
<td>32</td>
</tr>
<tr>
<td>2.1</td>
<td>Mark up Shork vs Forecasting Factors</td>
<td>60</td>
</tr>
<tr>
<td>2.2</td>
<td>The Technology Shock and the Business Cycle</td>
<td>63</td>
</tr>
<tr>
<td>2.3</td>
<td>Comparison of Two Groups of Macro Variables</td>
<td>66</td>
</tr>
<tr>
<td>2.4</td>
<td>Regression Coefficients of One-year Excess Returns on Macro Variables</td>
<td>68</td>
</tr>
<tr>
<td>2.5</td>
<td>Forecasting Factor vs Macro Factor</td>
<td>70</td>
</tr>
<tr>
<td>3.1</td>
<td>Variance Ratio for Different Values of $\rho_{uw}$</td>
<td>85</td>
</tr>
</tbody>
</table>
3.2 Comparison of Each Component for Different Values of $\rho_{uw}$ .......................... 85
3.3 Variance Ratios with Estimated $\rho_{uw}$ .......................... 88
3.4 Conditional Variances for Different Values of $\rho_{uw}$: Extended Model .......................... 89
3.5 Comparison of Each Component for Two Models, $\rho_{uw} = -0.8$ .......................... 90
CHAPTER 1

Bond Pricing with Model Uncertainty

1.1 Introduction

1.1.1 Stylized Facts and Failures of Traditional Consumption Based Models

Several important stylized facts should be accounted for when building a term structure model: 1) an upward sloping nominal yield curve, 2) positive and increasing excess returns on nominal bonds along maturity, 3) time varying term premia, which imply predictability of excess bond returns, and 4) high volatility of long yields. The failure of consumption based models to account for each of these stylized facts will be discussed in turn.

The first two facts are equivalent to each other according to Piazzesi and Schneider (2006); the average excess return on an \( n \)-period bond is approximately equal to the average spread between the \( n \)-period yield and the short rate over long enough samples. These two facts challenge the models with additively separable preferences and lead to the bond premium puzzle. This puzzle was first documented by Backus, Gregory and Zin (1989). They found that with additively separable preferences, the average excess returns of long bonds are negative and small for coefficients of relative
risk aversion below 10. To solve the puzzle, we may either use Epstein and Zin’s recursive utility function or Campbell and Cochrane’s habit formation model.

The second stylized fact goes against the expectation hypothesis, which states that the long-term interest rate is equal to the average of current and expected future short rates up to a constant. The constant term premium implies unpredictable excess bond returns. This conflicts with the empirical finding of Fama and Bliss (1987), Campbell and Shiller (1991) and Cochrane and Piazzesi (2005), each of which provides empirical evidence that we are able to predict excess bond returns using financial data.

Finally, the expectations hypothesis also leads to the "excess volatility puzzle" for long bond yields, which was first documented by Shiller (1979). The excess volatility puzzle tells us that according to the expectations hypothesis, long yields are conditional expected values of future short rates. As a result, we cannot reconcile the high volatility of long yields with observed persistence in short rate.

A close investigation shows that these models share some common features in the way they define the preferences of the consumer and the exogenous process for consumption and inflation, which are two important ingredients of a consumption based asset pricing model. They assume that the consumer knows the model that generates consumption and inflation data. As a result, there is no uncertainty about the model and no effect of model uncertainty when the utility function is evaluated. Although Piazzesi and Schneider (2006) consider the fact of unobservable states, they fail to adjust the utility function of the consumer who faces the unobserved states. This paper will show how model uncertainty, including unobserved states and not knowing the true data generating model, will affect consumer’s utility function and the resulting pricing kernel.
1.1.2 Introducing Hansen and Sargent’s Fragile Beliefs

The consumer knows neither the distribution of unobserved states nor the true model generating the consumption and inflation data. As a result, he has to estimate the mean and variance of the unobserved states and assign probabilities to the two candidate models. The consumer is endowed with a set of probability measures, near his approximating model, to represent his doubt about the approximating model (Hansen and Sargent (2008)). There is a two step evaluation of the consumer’s lifetime utility. In the first step, the consumer minimizes his lifetime utility over all possible probability measures. This represents the consumer’s aversion to model uncertainty, and thus he seeks the lowest bound of the continuation value function. The resulting probability measure is twisted in a pessimistic direction relative to what it would be in comparable rational expectations models. In the second step, under the twisted probability measure, the consumer acts as if he is a rational expectations investor. With the twisted probability measure, multiplicative adjustments are made to the traditional pricing kernel. Hansen and Sargent (2008) derive the new pricing kernel using the Radon-Nikodym derivative and use it to price the one period ahead asset. They find that the twisted probabilities give rise to model uncertainty premia that contribute a time-varying component to the market price of risk. To be more specific, a model selection problem together with the consumer’s concern about probability assignment help generate the countercyclical risk premium of equity.

In this paper we adopt Hansen and Sargent’s three step framework to show how the consumer’s aversion to different uncertainties would affect the pricing of nominal bonds. We consider the short rate as well as the 1 to 5 year yields. In this sense, this paper can be seen as an extension of the Hansen and Sargent (2008) model from
one period pricing to multiple period pricing. In addition, we investigate problem of model selection. Hansen and Sargent (2008) assume two models generating the consumption data: one takes Bansal and Yaron’s (2004) long run risk model, and the other takes the i.i.d form. In this paper, the process of inflation is a major concern since we study nominal bonds, which are closely related to the behavior of inflation rate. In model 0, we assume that the consumption growth rate and inflation growth rate follow a bivariate form of the long run risk model. Stock and Watson (2007) find that the univariate inflation process is well described by an integrated moving average process that can be nested as a special case in model 0. In model 1, a similar process is used to model the consumption growth rate and inflation rate. Inflation is assumed to be stationary in model 1. This specification is the same as the one in Piazzesi and Schneider (2006). The model selection problem is critical to generating a high volatile yield spread, which is defined as the difference between the 5 year yield and the short rate. The consumer does not know which model generates the observed consumption and inflation data, and assigns probabilities to both models each period based on all information up to that period. In the mean time, a set of distorted probability assignments are generated that are tilted towards the model that implies a lower lifetime utility due to consumer’s distrust of the ordinary probability. An important finding of this paper is that the model implied yield spread is highly correlated with the difference between distorted probability and ordinary probability assigned to model 0, the model that models the inflation growth rate as containing a persistent component and transitory component. The more the probability is twisted towards Model 0, the higher is the yield spread.
1.1.3 Relation to Other Consumption Based Model of the Term Structure of Interest Rates

Currently, there are two important papers on the term structure of interest rates. Wachter (2006) proposes a consumption-based model with external habit that can account for many properties of the nominal term structure of interest rates, including the above stylized facts. It’s success depends on a new preference parameter that represents a trade-off between the intertemporal substitution effect and the precautionary savings effect. The parameter is chosen to match the slope of the yield curve, and it produces reasonable volatilities of bond yields. The success of the habit formation model in fitting the yield data motivates our work on the long run risk model, because these two models belong to two different categories and can explain the stock market data equally well.

Similar work has been done by Piazzesi and Schneider (2006) who examine a representative agent asset pricing model with recursive utility preferences and exogenous consumption growth and inflation. Following Bansal and Yaron’s long run risk model, consumption growth and inflation are assumed to follow a bivariate process that contains both a persistent and transitory component. Their model implies an upward sloping nominal yield curve. With adaptive learning and parameter uncertainty, their model is able to generate time varying term premia and high volatility of long bonds.

This paper differs from Piazzesi and Schneider (2006) in the following ways. First, unlike their specification of a finite horizon model with a high discount factor larger than 1, which is uncommon in the whole literature, this paper uses the traditional infinite horizon model and restricts the discount factor to be less than 1. In addition, Piazzesi and Schneider use rolling estimation to estimate their model. That is,
every period the model is reestimated using all available data and with past data down weighted. This rolling estimation delivers a sequence of parameter estimates, and each period they plug the corresponding estimates into the ordinary pricing kernel derived from Epstein and Zin’s utility function. The changing parameters help generate high volatility of long term bonds and increase the goodness of fit for the yield spread. However, it is logically incoherent with the methodology. The changing parameters in their model are used to capture the structural change of the economy. If the consumer believes the existence of structure change, this concern should be reflected when they price the long term bond. However, their model fails to account for this concern. In this paper, we use a different methodology to capture the changing economic structure: model selection. Two different processes are specified for the fundamentals, consumption and inflation, of the economy, and each period the consumer is asked to assign a probability to each process based on past observations. There is a parameter that represents the consumer’s concern about the probability assignment. This parameter affects the pricing kernel directly, and thus the pricing of long term bonds. In other words, structural change is taken into consideration when long term bonds are priced. In this paper, we show that the model selection problem increases the goodness-of-fit for the yield spread. The large fluctuations in the yield spread are captured by the model. Finally, Piazzesi and Schneider (2006) introduce parameter uncertainty to their model. The representative consumer does not trust the estimates of the mean of fundamentals and treat them as random variables, like the persistent components of consumption growth and inflation. This treatment does not price the parameter uncertainty, and as a result, the changes in uncertainty do not have first order effects on interest rates. This paper contributes by incorporating
one risk sensitivity operator that takes into consideration consumer’s concern about
the distribution of states. This modification provides a tractable way to evaluate how
states uncertainty would affect the yields of bonds.

Besides the above differences, this paper also interprets Epstein and Zin’s prefer-
e nce used by Piazzesi and Schneider (2006) in a different way that is consistent with
Hansen and Sargent (2008). Epstein and Zin’s preference actually comes from the
minimization problem of the pessimistic consumer who minimizes his lifetime utility
with respect to all possible probability measures. In the literature, this paper can be
seen as an extension of Piazzesi and Schneider’s (2006) paper. Their model setup can
be nested as a special case of the Step 2 model with the coefficient of risk aversion
equal to 1.

1.1.4 Organization

The paper is organized as follows: The standard pricing theory and data descrip-
tion are given in Section 2. The pricing formula is applied to the subsequent sections
with different stochastic discount factors obtained from different models. In Section
3, we build a model that captures the consumer’s concern about the distribution of
future states conditional on current states and signals. The model is then used to
price bond returns. The model in this section corresponds to the Step 1 model in
Hansen and Sargent (2008). Section 4 extends the model by consider consumer’s
distrust of current states that are unobservable and obtained through learning. This
corresponds to the Step 2 model in Hansen and Sargent (2008). Model selection and
its implication for bond pricing is discussed in Section 5, which corresponds to the
1.2 Pricing Theory and Data

Before moving to the model, we discuss some standard results from asset pricing theory. The theory tells us that the time $t$ price of a real bond that pays 1 unit of consumption $n$ periods later, $P_t^{(n)}$, is determined by the representative consumer’s Euler equation with a stochastic discount factor $M$.

$$P_t^{(n)} = E_t \left( P_{t+1}^{(n-1)} M_{t+1} \right) = E_t \left( \Pi_{i=1}^{n} M_{t+i} \right) \tag{1.1}$$

We take the log on both sides and use the normality to get the equation for the log price of $n$-year bond.

$$p_t^{(n)} = E_t \left( \sum_{i=1}^{n} m_{t+i} \right) + \frac{1}{2} \text{var}_t \left( \sum_{i=1}^{n} m_{t+i} \right) \tag{1.2}$$

Usually, the stochastic discount factor is obtained from the consumer’s optimization problem, which is a function of the consumption growth rate. As a result, its conditional moments are determined by some particular process of the consumption growth rate, either endogenous or exogenous. In this paper, we are going to assume an exogenous process for consumption growth.

Price is converted to yield using the relation

$$y_t^{(n)} = -\frac{1}{n} p_t^{(n)} = -\frac{1}{n} E_t \left( \sum_{i=1}^{n} m_{t+i} \right) - \frac{1}{n} \frac{1}{2} \text{var}_t \left( \sum_{i=1}^{n} m_{t+i} \right) \tag{1.3}$$

To price the nominal bond, we should deflate the stochastic discount factor with inflation. As a result, the nominal yields can be computed as

$$y_t^{\$} = -\frac{1}{n} p_t^{\$} = -\frac{1}{n} E_t \left( \sum_{i=1}^{n} m_{t+i} - \pi_{t+i} \right) - \frac{1}{n} \frac{1}{2} \text{var}_t \left( \sum_{i=1}^{n} m_{t+i} - \pi_{t+i} \right) \tag{1.4}$$
This equation will be used later with different pricing kernels derived from different models.

The per capita consumption growth rate and the inflation rate are used for model estimation. The consumption growth rate is measured using quarterly NIPA data on nondurables and services. A constant population growth rate is assumed to avoid different sources of population data. The per capita consumption growth rate is obtained by deflating the consumption growth rate by 0.004, which is roughly the quarterly population growth rate. Inflation is constructed using quarterly NIPA data on the price index. The short rate and nominal yields for 1 to 5 year bonds are from the CRSP Fama risk free rate file and the Fama-Bliss discount bond files. The sample period goes from 1952:Q2 through 2005:Q4.

1.3 Step 1: A Model with One Risk Sensitivity Operator

In this step, we consider the consumer’s concern about the distribution of future state variables, conditional on current states and signals. That is, we assume the consumer observes current states, but distrusts the distribution of future states. The preference of the consumer is defined with a risk sensitivity operator, \( T^1 \):

\[
v (s_t, c_t) = (1 - \beta) c_t + T^1 [\beta v (s_{t+1}, c_{t+1})]
\]  

(1.5)
where $T^1$ is mathematically defined as

$$T^1 [\beta v (s_{t+1}, c_{t+1})]$$

$$= \frac{1}{1 - \gamma_1} \log E \left[ \exp \left( \beta (1 - \gamma_1) v (s_{t+1}, c_{t+1}) \right) \right] (s_t, c_t)$$

$$= \min_{m(e_{t+1}) \geq 0, Em(e_{t+1}) = 1} E \left\{ m(e_{t+1}) \left[ \beta v (s_{t+1}, c_{t+1}) - \frac{1}{1 - \gamma_1} \log m(e_{t+1}) \right] \right\} \left\{ s_t, c_t \right\}$$

where $c$ is the log of consumption and $v$ is the present value of the log consumption stream. $s$ is the state vector of this economy and its relation with consumption can be represented using the following state space representation:

$$s_{t+1} = A s_t + C e_{t+1}$$

$$[\Delta e_{t+1} \pi_{t+1}] = D s_t + G e_{t+1}$$

where $e_{t+1}$ is a 2 by 1 vector that is distributed as $N(0, I)$. $s_t$ is a 4 by 1 state vector, which contains the persistent component of consumption growth, unconditional mean of consumption growth, persistent component of inflation and unconditional mean of inflation. This state space representation is a bivariate setting of the long run risk model proposed by Bansal and Yaron (2004).

The minimization problem reveals the consumer’s concern about not knowing the true distribution of future states, and his aversion to this uncertainty makes him seek the probability measure that will lead to the worst case. The resulting probability measure is called the worst case probability, and it is useful to derive the pricing kernel. Details can be found in Hansen et al. (2002).

The resulting preference is the same as the recursive preference used by Piazzesi and Schneider (2006). They interpret $\gamma_1$ as the consumer’s attitude toward persistence.
of the consumption stream. The consumer shows aversion to consumption persistence if $\gamma_1 > 1$. If $\gamma_1 < 1$, the consumer likes consumption persistence. In their paper they set $\gamma_1 > 1$. Here we interpret $\gamma_1$ as the consumer’s attitude toward model uncertainty. If $\gamma_1 = 1$, his attitude is neutral and the preference degenerates to the traditional log utility function.

$$v(s_t, c_t) = (1 - \beta) c_t + E_t [\beta v(s_{t+1}, c_{t+1})]$$

(1.7)

If $\gamma_1 > 1$, then consumer shows aversion to model uncertainty.

Under this preference, we compute the stochastic discount factor using the perturbation method in Hansen and Sargent (2008) and get the following result.

$$m_{t+1} = \log \beta - \Delta c_{t+1} - \frac{1}{2} \beta^2 (1 - \gamma_1)^2 (\lambda' C + \phi_c G) (\lambda' C + \phi_c G)' + \beta (1 - \gamma_1) (\lambda' C + \phi_c G) e_{t+1}$$

(1.8)

where $\phi_c = [0, 1]$, $\lambda' = \beta \phi_c D (I - \beta A)^{-1}$. The last two terms represent the adjustment to consumer’s concern about distribution of future states. Appendix A gives the derivation of this stochastic discount factor.

With the stochastic discount factor, nominal yields are computed using equation 1.4.

Table 1.1 compares the mean and standard deviation of average nominal yields implied by the model with those in the data for different calibrations of $\gamma_1$. $\beta$ is set to be 0.999. A large value of $\gamma_1$ will help generate an upward sloping yield curve, but not as steep as in the data. This cannot be overcome by continually increasing $\gamma_1$, because eventually the yield curve will become hump shaped. When $\gamma_1$ is 100, the 5 year yield is smaller than the 4 year yield. Besides, as emphasized above, the
volatility puzzle appears here. The volatility decreases with maturities and reaches 1.119 for the 5 year bond, which is less than half of the real data. $\gamma_1$ does not have any effect on the volatility.

The nominal short rate and yield spread are shown in Figure 1.1. The value of $\gamma_1$ has a negligible effect on the shape of short rate and yield spread. We report the results with $\beta = 0.999$ and $\gamma_1 = 10$. The model gives a reasonable fit for the short rate, while it fails to generate high volatility of the yield spread, especially in the 1970’s and 1980’s.

The Step 1 model corresponds to the benchmark model in Piazzesi and Schneider (2006). They argue the upward sloping yield curve depends on both the preference specification and distribution of fundamentals. In their interpretation $\gamma_1 > 1$ represents the consumer’s aversion to consumption persistence. In this case, a premium is commanded by an asset when its payoff covaries more with news about future consumption growth. The estimated process of consumption growth and inflation implies

### Table 1.1: Average Nominal Yield Curve (Step 1 Model)

<table>
<thead>
<tr>
<th></th>
<th>1 quarter</th>
<th>1 year</th>
<th>2 year</th>
<th>3 year</th>
<th>4 year</th>
<th>5 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>mean</td>
<td>5.148</td>
<td>5.557</td>
<td>5.763</td>
<td>5.933</td>
<td>6.059</td>
</tr>
<tr>
<td></td>
<td>volatility</td>
<td>2.915</td>
<td>2.923</td>
<td>2.883</td>
<td>2.812</td>
<td>2.784</td>
</tr>
<tr>
<td>$\gamma_1 = 1$</td>
<td>mean</td>
<td>5.795</td>
<td>5.793</td>
<td>5.793</td>
<td>5.793</td>
<td>5.793</td>
</tr>
<tr>
<td></td>
<td>volatility</td>
<td>1.802</td>
<td>1.642</td>
<td>1.476</td>
<td>1.340</td>
<td>1.222</td>
</tr>
<tr>
<td>$\gamma_1 = 10$</td>
<td>mean</td>
<td>5.828</td>
<td>5.839</td>
<td>5.841</td>
<td>5.841</td>
<td>5.840</td>
</tr>
<tr>
<td></td>
<td>volatility</td>
<td>1.802</td>
<td>1.642</td>
<td>1.476</td>
<td>1.340</td>
<td>1.222</td>
</tr>
<tr>
<td>$\gamma_1 = 100$</td>
<td>mean</td>
<td>6.150</td>
<td>6.300</td>
<td>6.322</td>
<td>6.319</td>
<td>6.310</td>
</tr>
<tr>
<td></td>
<td>volatility</td>
<td>1.802</td>
<td>1.642</td>
<td>1.476</td>
<td>1.340</td>
<td>1.222</td>
</tr>
</tbody>
</table>

Note: Sample periods: 1952:Q2-2005:Q4
Figure 1.1: Short Rate and Yield Spread: Step 1 Model.
that inflation will bring down the future consumption growth rate, and acts as bad news for future consumption. High inflation makes the payment of the long term bond imprecise and also implies low consumption in the future. As a result, the long term bond is not an attractive asset because it pays off little when consumption is low, which is why it requires a premium over the short rate.

1.4 Step 2: A Model with Unobserved States and Another Risk Sensitivity Adjustment

In this step, we assume that the representative consumer does not observe the state variable \( \zeta_t \) at time \( t \). Instead, he observes a vector of signals, \( s_t = (c_t, \pi_t)' \). He uses the time \( t \) signals together with previous signals to make an inference about \( \zeta_t \). The initial state is assumed to follow a normal distribution with mean \( \hat{\zeta}_0 \) and covariance matrix \( \Sigma_0 \).

The state variables can be learned using the Kalman filter, which yields the following innovation representations:

\[
\hat{\zeta}_{t+1} = A\zeta_t + K(\Sigma_t)a_{t+1} \tag{1.9}
\]
\[
\Sigma_{t+1} = A\Sigma_t A' + CC' - k(\Sigma_t)(A\Sigma_t D' + CG')'
\]
\[
s_{t+1} - s_t = D\zeta_t + a_{t+1}
\]

where \( \hat{\zeta}_t = E[\zeta_t|s_t, s_{t-1}, ... s_0] \), \( a_{t+1} = s_{t+1} - E[s_{t+1}|s_t, s_{t-1}, ... s_0] \), \( \Sigma_t = E(\zeta_t - \hat{\zeta}_t)(\zeta_t - \hat{\zeta}_t)' \), \( \Delta_t \triangleq E[ a_{t+1}a_{t+1}' ] = D\Sigma_t D' + GG' \), and \( K(\Sigma) = (A\Sigma D' + CG')(D\Sigma D' + GG')^{-1} \).

To capture the consumer’s distrust of the distribution of unobserved states, we apply \( T^2 \) on both sides of equation 1.5.
\[
T^2 v(\zeta_t, c_t) = T^2 \left\{ (1 - \beta)c_t + T^1 [\beta v(\zeta_{t+1}, c_{t+1})] \right\} \rightarrow \\
T^2 v(\zeta_t, c_t) = (1 - \beta)c_t + T^2T^1 [\beta v(\zeta_{t+1}, c_{t+1})]
\]

(1.10)

where

\[
T^2 v(\zeta_t, c_t) = \frac{1}{1 - \gamma_2} \log \int \exp ((1 - \gamma_2) v(\zeta_t, c_t)) \phi(\zeta_t|\bar{\zeta}_t, \Sigma_t) d\zeta
\]

\[
= \min_{h(\zeta_t) \geq 0, f(\zeta_t)\phi(\zeta_t|\bar{\zeta}_t, \Sigma_t)} \int \left\{ v(\zeta_t, c_t) - \frac{1}{1 - \gamma_2} \log h(\zeta_t) h(\zeta_t) \phi(\zeta_t|\bar{\zeta}_t, \Sigma_t) \right\} \zeta_t, c_t
\]

where \( \phi(\zeta_t|\bar{\zeta}_t, \Sigma_t) \) is the normal density function with mean \( \bar{\zeta}_t \) and covariance matrix \( \Sigma_t \). \( \gamma_2 \) captures the consumer’s concern about the distribution of current states conditional on signals.

Following Hansen and Sargent, we compute the log of the stochastic discount factor, which takes the form

\[
m_{t+1} = \log \beta - \Delta c_{t+1}
\]

(1.11)

\[
= -\frac{1}{2} \left[ \beta (1 - \gamma_1) G (C' \lambda + G' \phi_c') + (1 - \gamma_2) D \Sigma_t \lambda' (GG' + D \Sigma_t D')^{-1} \right] \]

\[
\times [\beta (1 - \gamma_1) G (C' \lambda + G' \phi_c') + (1 - \gamma_2) D \Sigma_t \lambda]
\]

\[
+ [\beta (1 - \gamma_1) G (C' \lambda + G' \phi_c') + (1 - \gamma_2) D \Sigma_t \lambda'] (GG' + D \Sigma_t D')^{-1}
\]

\[
(D \zeta_t + G \epsilon_{t+1} - D \bar{\zeta}_t)
\]

Appendix B gives a brief derivation of the stochastic discount factor.

Compared with equation 1.8, the new pricing kernel discounts future payoffs more than the old one, because of the uncertainty about the unobserved states. If we set
$\gamma_2 = 1$ and $\Sigma_t = 0$, we will get exactly the same pricing kernel as in Step 1. The existence of the consumer’s concern aggravates the effect of unobserved states. Table 1.2 shows how the change in the consumer’s concern about unobserved states affects the nominal yield curve for $\beta = 0.999$ and $\gamma_1 = 50$. One interesting finding is that $\gamma_2$ has a negative impact on the slope of the yield curve. When $\gamma_2 = 1$, the yield curve is upward sloping, and the difference between the 5 year yield and the short rate is 0.831. This number falls to 0.732 when $\gamma_2 = 10$ and to 0.619 when $\gamma_2 = 50$. The intuition of the negative effect of $\gamma_2$ on the slope is as follows. By construction, $\gamma_2$ represents the consumer’s concern about robustness in state estimation which is governed by the variance-covariance matrix $\Sigma$. From equation 1.9, the recursive estimation of $\Sigma$ tells us that at time $t$, the consumer can predict future $\Sigma$ without knowing any new observations. They know with certainty that $\Sigma$ will decrease in the future, and state uncertainty is less severe in the future. Speaking of state uncertainty only, long term bonds are more safe than short term bonds, so they do not command a premium, which is why $\gamma_2$ has a negative effect on the slope of the yield curve. Table 1.2 also confirms the volatility puzzle for the Step 2 model. The volatility of long term bonds is much smaller than the that in the data.

Piazzesi and Schneider (2006) also consider parameter uncertainty in their model. However, they do not have a parameter that captures the consumer’s concern about this uncertainty. That is, in their model $\gamma_2$ is set to be 1.

The model implied nominal short rate and yield spread are plotted in Figure 1.2. The value of $\gamma_2$ has a negligible effect on the shape of short rate and yield spread. We report the results with $\beta = 0.999$, $\gamma_1 = 50$ and $\gamma_2 = 10$. Similar results can be drawn here: the model can fit the short rate well, but not the yield spread. The
high volatility of yield spread in the 1980’s suggests that the process for consumption growth and inflation may not be stable across the whole sample period. It is possible to have time varying parameters or model shifting.

Piazzesi and Schneider (2006) consider the case of time varying parameters. They reestimate the system given in equation 1.6 for every date $t$ using only data up to that date. Past observations are downweighted to accommodate the consumer’s concern about structural change. A sequence of parameters are generated, and different parameter values are used to price nominal bonds for every date $t$. However, it is logically incoherent with their methodology. The changing parameters in their model are used to capture structural changes in the economy. If the consumer believes the existence of structural changes, this concern should be reflected when they price the long term bond. But their model fails to account for this concern. In this paper, we use a different methodology to capture the changing economic structure: model selection. Two different processes are specified for the fundamentals, consumption

<table>
<thead>
<tr>
<th>$\gamma_2$</th>
<th>mean 1 quarter</th>
<th>mean 1 year</th>
<th>mean 2 year</th>
<th>mean 3 year</th>
<th>mean 4 year</th>
<th>mean 5 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.148</td>
<td>5.557</td>
<td>5.763</td>
<td>5.933</td>
<td>6.059</td>
<td>6.141</td>
</tr>
<tr>
<td>volatility</td>
<td>2.915</td>
<td>2.923</td>
<td>2.883</td>
<td>2.812</td>
<td>2.784</td>
<td>2.739</td>
</tr>
<tr>
<td>10</td>
<td>5.976</td>
<td>6.140</td>
<td>6.327</td>
<td>6.503</td>
<td>6.663</td>
<td>6.807</td>
</tr>
<tr>
<td>volatility</td>
<td>1.808</td>
<td>1.657</td>
<td>1.488</td>
<td>1.351</td>
<td>1.233</td>
<td>1.131</td>
</tr>
<tr>
<td>50</td>
<td>5.880</td>
<td>6.017</td>
<td>6.185</td>
<td>6.351</td>
<td>6.507</td>
<td>6.652</td>
</tr>
<tr>
<td>volatility</td>
<td>1.809</td>
<td>1.661</td>
<td>1.490</td>
<td>1.353</td>
<td>1.235</td>
<td>1.132</td>
</tr>
<tr>
<td>17</td>
<td>5.453</td>
<td>5.479</td>
<td>5.570</td>
<td>5.716</td>
<td>5.887</td>
<td>6.072</td>
</tr>
<tr>
<td>volatility</td>
<td>1.817</td>
<td>1.690</td>
<td>1.504</td>
<td>1.361</td>
<td>1.237</td>
<td>1.126</td>
</tr>
</tbody>
</table>

Note: Sample periods: 1952:Q2-2005:Q4

Table 1.2: Average Nominal Yield Curve (Step 2 Model)
Figure 1.2: Step 2 Model: Short Rate and Yield Spread.
and inflation, of the economy, and each period the consumer is asked to assign a probability to each process based on past observations. There is a parameter that represents the consumer’s concern about the probability assignment. This parameter directly affects the pricing kernel, and thus the pricing of long term bonds. In other words, structural change is taken into consideration when the long term bonds are priced. Details of the model selection problem are described in next section.

1.5 Step 3: Model Selection and Another Risk Sensitivity Operator

Two different processes are assumed for consumption growth and inflation. Model 1 is the one described above, and model 0 is specified as

\[
\begin{bmatrix}
\Delta c_{t+1} \\
\Delta \pi_{t+1}
\end{bmatrix} = D(0)c_t(0) + G(0)\epsilon_{t+1}
\]

\[
c_{t+1}(0) = A(0)c_t(0) + C(0)\epsilon_{t+1}
\]

Modeling inflation in first differences is sensible since Stock and Watson (2007) find that the univariate inflation process is well described by an integrated moving average process, which can be nested as a special case in our model.

Faced with two models, the consumer will assign probabilities to both models each period by comparing the log likelihood for two models. The log likelihood of model \(\nu\) (\(\nu = 0\) or \(1\)) is defined as

\[
L_T(\nu) = \sum_{i=1}^{T} l_i(\nu)
\]

where

\[
l_i(\nu) = -\frac{1}{2} \left[ \log(2\pi) + \log \det(\Delta_{t-1}(\nu)) + a_t(\nu)' \Delta_{t-1}(\nu)^{-1} a_t(\nu) \right]
\]
We compute $\hat{p}_t$ following Hansen and Sargent (2008).

$$\hat{p}_t = \frac{\hat{p}_{t-1} \exp(l_t(0) - l_t(1))}{1 - \hat{p}_{t-1} + \hat{p}_{t-1} \exp(l_t(0) - l_t(1))}$$

(1.15)

Figure 1.3 plots the time series for $\hat{p}_t$ with initial probability set to be 0.5. The plot shows that probability assigned to model 0 fluctuates over time, and it is difficult to distinguish between them using limited data. That is, the representative consumer faces the difficulty of disentangling the permanent and persistent transitory components of inflation. The changing of belief across models may explain why one model alone cannot fit the yield data well.

To capture the consumer’s distrust of his prior over two models, we apply the $T^3$ operator to continuation value of consumer derived in Step 2. This is slightly different from the one in Hansen and Sargent (2008) which uses the same $T^2$ here as in the Step 2 model. We make this modification to represent the consumer’s different attitude towards parameter uncertainty and model selection. The consumer’s aversion to model selection is the most important for bond pricing.

$$T^3 v(t) = \frac{1}{1 - \gamma_3} \log \{\hat{p} \exp [(1 - \gamma_3) v(0)] + (1 - \hat{p}) \exp [(1 - \gamma_3) v(1)]\}$$

(1.16)

The new stochastic discount factor can be obtained by assuming $G(0) = G(1) = G$, which is a good approximation based on our estimation. Details can be found in Hansen and Sargent (2008). Appendix C gives a brief derivation of the stochastic discount factor.
Figure 1.3: Probability Assigned to Model 0 with $p_0 = 0.5$. 
\[ m_{t+1} = \log \beta - \Delta c_{t+1} - (s_{t+1} - \tilde{\mu}_{s,t})' (GG')^{-1} (\tilde{\mu}_{s,t} - \tilde{\mu}_{s,t}) \]

\[ -0.5 (\tilde{\mu}_{s,t} - \tilde{\mu}_{s,t})' (GG')^{-1} (\tilde{\mu}_{s,t} - \tilde{\mu}_{s,t}) \]

where

\[ \tilde{\mu}_{s,t} = \tilde{p}_t D(0) \tilde{\zeta}_t (0) + (1 - \tilde{p}_t) [D(1) \tilde{\zeta}_t (1) - \phi'_e \pi_t] \]

\[ \tilde{\mu}_{s,t} = \tilde{p}_t [D(0) \tilde{\zeta}_t (0) + Gw(0) + D(0) u_t (0)] \]

\[ + (1 - \tilde{p}_t) [D(1) \tilde{\zeta}_t (1) - \phi'_e \pi_t + Gw(1) + D(1) u_t (1)] \]

\[ w(t) = \beta (1 - \gamma_1) (C(t)' \lambda (t) + G(t)' \phi'_e) , \text{ for } t = 0 \text{ and } 1 \]

\[ u_t (t) = (1 - \gamma_2) \Sigma_t (t) \lambda (t), \text{ for } t = 0 \text{ and } 1 \]

\[ \tilde{p}_t \propto \tilde{p}_t \exp [(1 - \gamma_3) v(0)] \] is the twisted model probability

The twisted probability will be shifted toward the model that has the worse value at each time \( t \) when we incorporate a risk sensitivity operator to adjust the consumer’s distrust of his prior over two models. Hansen and Sargent (2008) have shown that this risk sensitivity operator produces a countercyclical risk premium for equity. Multiple period pricing requires that we compute the conditional mean and variance of the sum of future pricing kernels from time \( t+1 \) till \( t+20 \). There is no analytical solution to this because of the nonlinearity introduced by \( \tilde{p}_t, \tilde{p}_{t+1}, \ldots \), so we simulate instead. The stochastic part of the pricing kernel comes from \( e_{t+1}, e_{t+2}, \ldots e_{t+20} \). Different values of the random variables are generated to compute the bond yields.

For this section, we report our results starting from year 1965 so that we can make a comparison with Piazzesi and Schneider (2006). The sample period used for estimation needs some explanation. Now the model implied yields depend on the mean and variance of state variables, \( \tilde{\zeta}_t \) and \( \Sigma_t \). In order to make them more accurate
for year 1965 and after, it is useful to start at an earlier date so that the estimates will converge to the true value after many updates. However, the following results rely on the high probability assigned to model 0 in the early 1980’s, which cannot be obtained if we start before 1960. As a compromise, we choose the year 1962 as the starting point.

1.5.1 Calibration

Before we give any result derived from The Step 3 model, let us first discuss the role played by each parameter. We define \( \Theta = (\beta, \gamma_1, \gamma_2, \gamma_3, p_0) \) as the vector of parameters and explain how to choose the value of each parameter. The choice of values aims to give a good fit of the model, including an upward sloping yield curve, high volatility of bond yields, and high fluctuation of the yield spread.

\( \gamma_3 \) has important effects on slope of the yield curve and the volatility of bond yields. \( \gamma_3 > 1 \) helps generate an upward sloping yield curve, because without any new information, the problem of model selection is more severe in the future. Thus, the long term bonds command a premium to account for the consumer’s aversion to this uncertainty. However, if \( \gamma_3 \) is increased beyond some point, a hump shaped yield curve will be generated. As for the volatility of bond yields, the larger the value of \( \gamma_3 \), the higher the volatility. To compromise between an upward sloping yield curve and high volatility, \( \gamma_3 \) is set to be 2.

\( \beta \) is the objective discount factor, and it governs the level of the yield curve. We set \( \beta = 0.9995 \), so that the level difference is around 0.2 percent between model implied nominal yield curve and the data. Although we can further increase the value of \( \beta \) to reduce the difference, there is a tradeoff to do it. For a given \( \gamma_3 \), the increase of
\( \beta \) will increase the impact of \( \gamma_3 \) on the slope of the yield curve. Hence the minimum value of \( \gamma_3 \) necessary to generate a hump shaped yield curve will decrease. Thus, a smaller value of \( \gamma_3 \) is necessary to get an upward sloping yield curve, while this smaller number will reduce the volatility of yields and fluctuation of the yield spread.

When model selection becomes the main uncertainty faced by the consumer, they tend to ignore other uncertainties, such as the distribution of future state variables and parameter uncertainty. In the presence of large uncertainty, model selection in our problem, the consumer tend to ignore other more subtle uncertainties. We set \( \gamma_1 = 1.05 \) and \( \gamma_2 = 1.02 \).

The initial probability assignment, \( p_0 \) is set to be 0.5 which puts equal weight on the two models described above.

### 1.5.2 Short Rate and Yield Spread

There is a positive correlation between the model implied short rate and the probability assignment. When inflation is high, such as in the early 1980’s, higher probability is assigned to Model 0, which specifies inflation as an \( I(1) \) process. In the meantime, the short rate is also high to count for the effect of high inflation. Figure 1.4 plots the data series for probability assignment and the short rate.

Another result is that the model implied yield spread is highly correlated with the difference between the twisted probability and the original one. The more the belief is twisted to Model 0, the higher is the yield spread. Figure 1.5 illustrates this point. The correlation coefficient between the probability difference and yield spread is as high as 0.966. The intuition goes as follows. Because we set \( \gamma_1 \) and \( \gamma_2 \) to be very close to 1, the valuation function is mostly affected by consumption growth rates. The lower utility
Figure 1.4: Probability Assignment vs Short Rate.
of Model 0 comes mostly from the lower unconditional mean of consumption growth rates, $u_c$, which is treated as a random variable. Figure 1.6 plots the unconditional mean of consumption growth rates for two models. The unconditional mean from Model 0 remains low from 1976 until recently. As a result, the distorted probability is tilted toward model 0. Initial responses of the unconditional mean of consumption growth rates from a real and nominal shock are plotted in Figure 1.7 for both models. The bottom panel gives the response differences for the two models. In the beginning, the unconditional mean of consumption growth increases less for positive real shocks and decreases more for positive nominal shocks for Model 0. The accumulation of the differences results in the lower $u_c$ for Model 0. Later the negative difference for real shocks decreases and the difference for nominal shocks becomes positive. As a result, $u_c$ for Model 0 catches up with that of Model 1. With the probability assignments twisted toward Model 0, the consumer will believe that in the future, low consumption growth rates will be accompanied with high inflation rates. Under this belief, long term bonds are unattractive because their payment will be deteriorated by the high inflation when consumption is low. Thus, long term bonds command a premium. The more the probability is twisted, the stronger is the belief and thus the higher the risk premium for long term bond.

1.5.3 Comparison with Piazzesi and Schneider’s Adaptive Learning Model

Table 1.3 summarizes the equilibrium yields for the Step 3 model and the adaptive learning model of Piazzesi and Schneider (2006). The level of the Step 3 model is
Figure 1.5: Probability Difference vs Yield Spreads.
Figure 1.6: The Unconditional Mean of Consumption Growth Rates for Two Models.
Figure 1.7: Initial Responses of the Unconditional Mean of the Consumption Growth Rate.
about 0.2 percent higher than the real data, while Piazzesi and Schneider’s model (PS model hereafter) does not suffer such a problem. To overcome the level problem, they set the value of $\beta$ to 1. As argued above, we cannot increase the value of $\beta$ to get a good fit of level, because it will affect the slope of the yield curve. However, the PS model fails to generate the curvature of the yield curve, which comes from the steep incline from the 3 month maturity to the 1 year maturity. The step 3 model can capture this feature. The model implied yield curve is steep at the short end and flat at the long end. Both models can generate high volatility of the yields.

Figure 1.8 and Figure 1.9 compare the short rate and the yield spread from the two models. The short rates look similar, while the yield spread from step 3 model captures the trend of real data better than PS model, especially after the 1990’s.

Piazzesi and Schneider’s results come mostly from the rolling estimation, which, as argued above, is logically incoherent. The comparisons show that even with the theoretically questionable rolling estimation, the performance of their model is similar to mine.
1.5.4 Investigation of Other Properties

The high correlation among nominal yields is an important stylized fact. Panel A of Table 1.4 shows the correlation of the implied yields from the Step 3 model. The correlation of yields in the data is given in panel B. Compared with the data, the model captures the property of high correlation among yields. The correlation coefficients among 1 to 5 year yields are very close to the data, while the correlations between the short rate and different long yields are lower than in the data.

Another stylized fact is the predictability of excess bond returns. Two experiments are conducted below.
Figure 1.9: Yield Spread Comparison.
### Panel A: Correlation of Model Implied Yields

<table>
<thead>
<tr>
<th></th>
<th>3 month</th>
<th>1 year</th>
<th>2 year</th>
<th>3 year</th>
<th>4 year</th>
<th>5 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>1.000</td>
<td>0.809</td>
<td>0.740</td>
<td>0.712</td>
<td>0.695</td>
<td>0.683</td>
</tr>
<tr>
<td>1 year</td>
<td>0.809</td>
<td>1.000</td>
<td>0.991</td>
<td>0.980</td>
<td>0.972</td>
<td>0.965</td>
</tr>
<tr>
<td>2 year</td>
<td>0.740</td>
<td>0.991</td>
<td>1.000</td>
<td>0.998</td>
<td>0.994</td>
<td>0.990</td>
</tr>
<tr>
<td>3 year</td>
<td>0.712</td>
<td>0.980</td>
<td>0.998</td>
<td>1.000</td>
<td>0.999</td>
<td>0.997</td>
</tr>
<tr>
<td>4 year</td>
<td>0.695</td>
<td>0.972</td>
<td>0.994</td>
<td>0.999</td>
<td>1.0000</td>
<td>0.999</td>
</tr>
<tr>
<td>5 year</td>
<td>0.683</td>
<td>0.965</td>
<td>0.990</td>
<td>0.997</td>
<td>0.999</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Panel B: Correlation of Yields in Data

<table>
<thead>
<tr>
<th></th>
<th>3 month</th>
<th>1 year</th>
<th>2 year</th>
<th>3 year</th>
<th>4 year</th>
<th>5 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>1.000</td>
<td>0.985</td>
<td>0.967</td>
<td>0.947</td>
<td>0.930</td>
<td>0.915</td>
</tr>
<tr>
<td>1 year</td>
<td>0.985</td>
<td>1.000</td>
<td>0.991</td>
<td>0.978</td>
<td>0.963</td>
<td>0.950</td>
</tr>
<tr>
<td>2 year</td>
<td>0.967</td>
<td>0.991</td>
<td>1.000</td>
<td>0.996</td>
<td>0.989</td>
<td>0.981</td>
</tr>
<tr>
<td>3 year</td>
<td>0.947</td>
<td>0.978</td>
<td>0.996</td>
<td>1.000</td>
<td>0.997</td>
<td>0.993</td>
</tr>
<tr>
<td>4 year</td>
<td>0.930</td>
<td>0.963</td>
<td>0.989</td>
<td>0.997</td>
<td>1.000</td>
<td>0.998</td>
</tr>
<tr>
<td>5 year</td>
<td>0.915</td>
<td>0.950</td>
<td>0.981</td>
<td>0.993</td>
<td>0.998</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Sample periods: 1964:Q1-2005:Q4

Table 1.4: Correlation of Yields with Different Maturities
In the first experiment, we follow the method proposed by Cochrane and Piazzesi (2005) to construct a forecasting factor, a linear combination of different forward rates, and use this factor to predict bond returns.

Forward rates and excess bond returns are constructed using the 1 to 5 year yields. Define the log price of a \( n \)-year bond as

\[
p^{(n)}_t = -ny^{(n)}_t
\]

where \( y^{(n)}_t \) is the log yield of the \( n \)-year bond that has been derived from the Step 3 model. A \( n \)-year forward rate is then defined as

\[
f^{(n)}_t = p^{(n-1)}_t - p^{(n)}_t
\]

The excess return of a \( n \)-year bond is defined as the one year holding period return of the \( n \)-year bond less the 1-year yield.

\[
rx^{(n)}_{t+1} = p^{(n-1)}_{t+1} - p^{(n)}_t - y^{(1)}_t
\]

With the constructed financial variables, we use the two stage least squares estimation used by Cochrane and Piazzesi (2005).

First, we estimate the \( \alpha 's \) by running a regression of the average excess return on all forward rates

\[
\frac{1}{4} \sum_{n=2}^{5} rx^{(n)}_{t+1} = \alpha_0 + \alpha_1 y^{(1)}_t + \alpha_2 f^{(2)}_t + \alpha_3 f^{(3)}_t + \alpha_4 f^{(4)}_t + \alpha_5 f^{(5)}_t + \varepsilon_{t+1}
\]

and define
Panel A (Model) \hspace{1cm} Panel B (Data)

<table>
<thead>
<tr>
<th>Model</th>
<th>( b^{(n)} )</th>
<th>( R^2 )</th>
<th>Data</th>
<th>( b^{(n)} )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_{xt+1}^{(2)} )</td>
<td>0.614 (25.320)</td>
<td>0.304</td>
<td>0.454 (63.027)</td>
<td>0.383</td>
<td></td>
</tr>
<tr>
<td>( r_{xt+1}^{(3)} )</td>
<td>0.898 (11.504)</td>
<td>0.211</td>
<td>0.864 (89.729)</td>
<td>0.417</td>
<td></td>
</tr>
<tr>
<td>( r_{xt+1}^{(4)} )</td>
<td>1.131 (8.267)</td>
<td>0.175</td>
<td>1.229 (94.841)</td>
<td>0.444</td>
<td></td>
</tr>
<tr>
<td>( r_{xt+1}^{(5)} )</td>
<td>1.357 (7.065)</td>
<td>0.157</td>
<td>1.453 (99.267)</td>
<td>0.420</td>
<td></td>
</tr>
</tbody>
</table>

F statistics are given in ( ).

Table 1.5: Forecasting Bond Returns Using Forecast Factor

\[
ff_t = \hat{\alpha}_0 + \hat{\alpha}_1 y_t^{(1)} + \hat{\alpha}_2 f_t^{(2)} + \hat{\alpha}_3 f_t^{(3)} + \hat{\alpha}_4 f_t^{(4)} + \hat{\alpha}_5 f_t^{(5)}
\]

\[
r_{xt+1}^{(n)} = b^{(n)} ff_t + z_{t+1}^{(n)}
\]

\[f_f_t\] is the forecasting factor we construct. Then we regress the excess return of each bond on this forecasting factor.

The estimation results are listed on Panel A of Table 1.5. As a comparison, the estimation results from real data are given in Panel B. The Hansen-Hodrick Correction with 4 lags is used to account for overlapping data. Table 1.5 shows the forecasting factor constructed using the model implied yield data forecasts model implied bond returns at all maturities. Bonds with a longer maturity have a heavier loading on this forecasting factor. Compared with the data, the model implied excess bond returns are less predictive. However, they do present certain predictability, with \( R^2 \)'s larger than 15%.
The other experiment comes from Fama and Bliss (1987) who use the spread between the $n$ year forward rate and 1 year yield to predict excess holding period returns of $n$ year bond.

\[ r x_{t+1}^{(n)} = \beta_0^{(n)} + \beta_1^{(n)} \left( f_t^{(n)} - y_t^{(1)} \right) + \varepsilon_{t+1}^{(n)} \quad (1.24) \]

The estimation results are listed on Panel A of Table 1.6. As a comparison, the estimation results from the data are given in Panel B. The Hansen Hodrick Correction with 4 lags is used to account for overlapping data. The predictability of excess bond returns in the data is higher than in the model, especially for long term bonds. But the model implied excess bond returns are still predictable using the model implied forward spread.

The predictability in yields comes from time variation in perceived risk which is related to the changing conditional variance of the unobserved states and the changing probability assignments.

1.6 Conclusion

This paper proposes a theoretical model for the term structure of interest rates. Three different kinds of uncertainties and their implications for bond pricing are discussed in turn. The main result is that uncertainty about the true model generating the observed data is the most important to consider when modelling the term structure of interest rates. The consumer’s aversion to model uncertainty implies a multiplicative adjustment of the traditional pricing kernel. A premium is commanded for the long term bonds, because the uncertainty is more severe in the future conditional on current information. The resulting pricing kernel is combined with the
Panel A (Model)

<table>
<thead>
<tr>
<th>$r_{x_1}^{(2)}$</th>
<th>$\beta_0^{(n)}$</th>
<th>$\beta_1^{(n)}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.054 (0.194)</td>
<td>1.130 (0.367)</td>
<td>0.186</td>
</tr>
<tr>
<td>$r_{x_1}^{(3)}$</td>
<td>-0.026 (0.404)</td>
<td>1.014 (0.595)</td>
<td>0.065</td>
</tr>
<tr>
<td>$r_{x_1}^{(4)}$</td>
<td>-0.020 (0.593)</td>
<td>0.983 (0.694)</td>
<td>0.037</td>
</tr>
<tr>
<td>$r_{x_1}^{(5)}$</td>
<td>-0.031 (0.768)</td>
<td>0.990 (0.842)</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Panel B (Data)

<table>
<thead>
<tr>
<th>$r_{x_1}^{(2)}$</th>
<th>$\beta_0^{(n)}$</th>
<th>$\beta_1^{(n)}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.010 (0.285)</td>
<td>1.005 (0.263)</td>
<td>0.153</td>
</tr>
<tr>
<td>$r_{x_1}^{(3)}$</td>
<td>-0.257 (0.493)</td>
<td>1.362 (0.337)</td>
<td>0.166</td>
</tr>
<tr>
<td>$r_{x_1}^{(4)}$</td>
<td>-0.458 (0.704)</td>
<td>1.501 (0.446)</td>
<td>0.155</td>
</tr>
<tr>
<td>$r_{x_1}^{(5)}$</td>
<td>0.030 (1.155)</td>
<td>0.995 (0.612)</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Standard deviations are given in ( ).

Table 1.6: Forecasting Bond Returns Using Forward Spread

exogenous processes for the consumption growth rate and inflation (growth rate) to price the bond yields. A bivariate version of the long run risk model is adopted to describe consumption and inflation with two different settings. The consumption growth rate is the same in the two specifications while inflation differs. In Model 0, inflation is assumed to follow an $I(1)$ process, while in Model 1, it is stationary. As a result, the consumer has to assign a probability to each model every period conditional on all available information. A distorted probability will be obtained by tilting the original probability to the model, which gives a lower lifetime utility. The probability assigned to Model 0 is high around the 1980’s, which helps explain the high interest rates during that period. The yield spread is found to be closely related
to the difference between the twisted probability and original probability. The more
the probability is twisted to Model 0, the higher the yield spread is.

With proper calibrations, this model is able to capture many important features of
the term structure of interest rates, including, (1) an upward sloping yield curve with
curvature at the short end, (2) high volatility of the long yields, (3) predictability
of excess holding period returns of 1 to 5 year bonds, and (4) the high correlations
of yields with different maturities. The model also fits the short rate and the yield
spread well.

In the literature, Wachter (2006) explains the behavior of interest rates using the
habit formation model. In my future work, I will conduct a horse race between the
habit formation model and the long run risk model using data from the bond market
to see which model explains the bond data better. The models belong to two different
categories and can explain the stock market equally well. Their performance in the
bond market is thus of great interest.
CHAPTER 2

Forecasting Bond Returns in a Macro Model

2.1 Introduction

Whether bond returns are forecastable is of great interest to both industry and academia. Several lines of work have been pursued on this topic over the last few decades. According to the expectation theory, which implies constant risk premia, excess bond returns are not predictable. However, some recent empirical work does suggest forecastable bond returns. For example, Fama and Bliss (1987) find that the one year excess return of a year bond can be predicted using the spread between the year forward rate and the one year yield. The $R^2$ is about 18%. Campbell and Shiller (1991) use yield spread to forecast yield changes. The evidence for forecastability is substantially strengthened by Cochrane and Piazzesi (2005). They find a tent-shaped forecasting factor, a linear combination of forward rates, which can predict excess returns on one to five year maturity bonds with an $R^2$ of about 35%.

Because the forecasting factor is obtained from forward rates, its economic interpretation is not clear. The economic meaning of the forecasting factor is of great interest for economists since no macro variable that strongly predicts bond returns has been found so far.
Duffee (2006) finds that only a small fraction of the variation in expected excess bond returns is associated with inflation, output growth or the short rate. Although bond returns are not correlated with traditional macro variables, we may expect them to be related to some latent macro variables, such as macro shocks. Based on this conjecture, this paper uses the Kalman filter to estimate the latent economic variables and investigate their relationship to bond risk premia.

Wachter (2006) studies the Campbell-Cochrane habit formation model and concludes that bond risk premia vary with the slow moving habit driven by shocks to aggregate consumption. However, habit is a latent macro variable, and there is no compelling empirical evidence for it.

Ludvigson and Ng (2007) use dynamic factor analysis for large datasets to investigate the possible empirical linkages between forecastable variation in excess bond returns and macroeconomic fundamentals. They find that real and inflation factors have important forecasting power for future excess returns, above and beyond the predictive power contained in the forward rates and yields spreads. When they combine the Cochrane and Piazzesi’s forecasting factor with their macro factor to forecast the one year excess bond return, the $R^2$ increases to 44%. The economic meaning of the macro factors found by Ludvigson and Ng is still vague since they are extracted from hundreds of data series. In addition, the high predictive power mostly comes from the forecasting factor. The predictive power of macro factors alone is not high.

This paper investigates the relationship between the forecasting factor and the latent macro variables obtained from a dynamic stochastic general equilibrium model. After we obtain the economic variables that have predictive power, a macro factor is constructed following Cochrane and Piazzesi (2005). Two important findings of
this paper are: (1) Cochrane and Piazzesi’s forecasting factor is mostly driven by the technology shock obtained from the dynamic stochastic general equilibrium (DSGE) model; and (2) the macro factor predicts better than Cochrane and Piazzesi’s forecasting factor.

Ludvigson and Ng (2007) explain why it is difficult to uncover a direct link between macroeconomic activity and bond risk premia.

“First, some macroeconomic driving variables may be latent and impossible to summarize with a few observable series. The Campbell-Cochrane habit may fall into this category. Second, macro variables are more likely than financial series to be imperfectly measured and less likely to correspond to the precise economic concepts provided by theoretical models. ... Third, the models themselves are imperfect descriptions of reality and may restrict attention to a small set of variables that fail to span the information sets of financial market participants.” (Ludvigson and Ng (2007), p. 1-2)

This paper tackles the above challenges in the following ways. First, the Kalman filter is used to extract the latent macro variables and the extracted series are used as data. Second, both macro data and financial data are used to estimate the DSGE model, and as a result, we may expect better estimation results. Third, we study a DSGE model with many shocks to avoid the limitation of variables.

The paper is organized as follows. Section 2 gives a brief description of the DSGE model and presents a linear approximation to model solution. Section 3 introduces the estimation procedure and Section 4 gives the estimation result. Various econometric analyses are conducted to characterize the forecasting factor in Section 5. Section 6 concludes.
2.2 Model

The macro model in this paper is an extension of the DSGE model proposed in Doh (2007). Two important modifications are made to Doh (2007).

The first one is about the utility function. Doh’s utility function takes a simple internal habit formation form,

\[
E_t \left[ \sum_{s=0}^{\infty} \beta^s \left( \frac{\left( C_{t+s}^u / A_{t+s} \right)^{1-\tau} - 1}{1 - \tau} - \frac{H_{t+s}^{1+1/\nu}}{1+1/\nu} \right) \right]
\]  

(2.1)

where \( C_{t+s}^u = C_{t+s} - he^{u_s}C_{t+s-1} \) is the consumption relative to the habit level, which is determined by the previous period consumption, \( h \) is the parameter governing the magnitude of habit persistence, \( \tau \) is the curvature of utility function, and \( \nu \) is the short-run (Frisch) labor supply elasticity, fixed at 0.5. This modification is made so that the utility function takes Campbell and Cochrane’s habit formation form. There are two important advantages in using Campbell and Cochrane’s habit formation. First, Campbell and Cochrane’s habit is found to be consistent with financial market data. Second, with this model, we can examine Wachter’s (2006) argument that bond risk premia vary with the slow moving habit driven by shocks to aggregate consumption.

Another modification is that we assume that the government consumes some of the total output. In Doh’s model, the government is inactive. This modification allows us to study the effect of government spending shocks on bond risk premia.

The model is a New Keynesian model that consists of five building blocks: a representative final good producer in a competitive market, a continuum of intermediate goods producers with imperfect competition, a representative household with a habit
formation utility function, and a monetary as well as a fiscal authority. In addition, a complete market is assumed for state-contingent claims.

2.2.1 Final Good Producer

There is a perfectly competitive final good producer in the economy that combines each intermediate good indexed by \( j \in [0, 1] \) using the following technology

\[
Y_t = \left( \int_0^1 Y(j)^{\frac{\varsigma_t - 1}{\varsigma_t}} \, dj \right)^{\frac{\varsigma_t}{\varsigma_t - 1}}
\] (2.2)

The representative firm in the final good sector maximizes its profit given output prices \( P_t \) and input prices \( P_t(j) \). This maximization behavior yields the following demand function

\[
Y_t(j) = \left( \frac{P_t(j)}{P_t} \right)^{-\varsigma_t} Y_t
\] (2.3)

\( \varsigma_t > 0 \) represents the elasticity of demand for each intermediate good.

2.2.2 Intermediate Goods Producers

There is a continuum of imperfectly competitive, intermediate goods producers. The production technology for the \( j \)-th intermediate goods producer takes the following simple form

\[
Y_t(j) = A_t N_t(j)
\] (2.4)

where \( A_t \) is an exogenously given common technology, and \( N_t(j) \) is the labor input for firm \( j \). The labor market is assumed to be perfectly competitive, and the real
wage is denoted as $W_t$. Firms face nominal rigidities in the form of quadratic price adjustment costs.

$$AC_t(j) = \frac{\phi}{2} \left( \frac{P_t(j)}{P_{t-1}(j)} - \pi_t^* \right)^2 Y_t(j)$$  \hspace{1cm} (2.5)

Here, $\phi$ is a parameter governing the degree of price stickiness in this economy, and $\pi_t^*$ is the target inflation of central bank in terms of the price of the final good.

Each firm chooses price level $P_t(j)$ and labor input $N_t(j)$ to maximize the present value of the profit stream:

$$E_t \left[ \sum_{s=0}^{\infty} \beta^s \frac{\lambda_{t+s}}{\lambda_t} \left( \frac{P_{t+s}(j)}{P_{t+s}} Y_{t+s}(j) - W_{t+s} N_{t+s}(j) - AC_{t+s}(j) \right) \right]$$  \hspace{1cm} (2.6)

where $\lambda_{t+s}$ is the marginal utility of a final good to the representative household at time $t + s$, which is exogenous to the firms.

### 2.2.3 Household

The representative household maximizes its utility by choosing consumption $C_t$ and labor supply $H_t$. The objective function of the household is defined below

$$E_t \left[ \sum_{s=0}^{\infty} \beta^s \left( \frac{(C_t - X_t)^{1-\tau}}{1-\tau} - \frac{H_t^{1+1/\nu}}{1+1/\nu} \right) \right]$$  \hspace{1cm} (2.7)

where $X_t$ is a reference point or the habit, which is assumed exogenous, and the surplus ratio is $S_t = \frac{C_t - X_t}{C_t}$.

Under the assumption of complete asset markets, the household is subject to the following budget constraint:
\[ P_t C_t + \sum_{s=0}^{\infty} P_{n,t}(B_{n,t} - B_{n+1,t-1}) + T_t = P_t W_t H_t + B_{1,t-1} + Q_t + \Pi_t \]  

(2.8)

where \( P_{n,t} \) is the price of an \( n \) month bond, \( B_{n,t} \), bond holding, \( T_t \), lump-sum tax, \( Q_t \), the net cash flow from participating in state-contingent security markets, and \( \Pi_t \), the aggregate profit.

### 2.2.4 Monetary Authority

The central bank follows a forward-looking Taylor rule with interest rate inertia. That is, the central bank sets the nominal interest rate according to expected inflation and the output gap in the following way:

\[
1 + i_t = (1 + i_t^*)^{1-\rho_i} (1 + i_{t-1})^{\rho_i} e^{\eta_i e_i t} \\
(1 + i_t^*) = (1 + r^*) \pi^* \left( \frac{E_t(\pi_{t+1})}{\pi_t^*} \right)^{\gamma_y} \left( \frac{Y_t}{A_t y_{t}^f} \right)^{\gamma_y}
\]

(2.9)

where \( r^* = \frac{u^*_a}{\beta} - 1 \) is the steady state real interest rate. \( \pi_t = \frac{P_{t+1}}{P_t} \) represents the actual inflation. \( \pi^* \) is the steady state inflation, and \( y_{t}^f \) is the steady state value of \( \frac{Y_t}{A_t} \) perturbed by a markup shock. The reason for using \( y_{t}^f \) instead of the steady state value of \( \frac{Y_t}{A_t} \) is, as in Doh (2007), to enhance the central bank’s ability to adjust the short term target interest rate in response to a markup shock. The time varying target inflation is assumed to be exogenously given. \( \rho_i \) is the parameter that captures the degree of interest rate inertia.

### 2.2.5 The Fiscal Authority

The fiscal authority consumes a fraction \( \chi_t \) of aggregate output \( Y_t \), where \( \chi_t \in [0, 1] \) and is governed by an exogenous process. The government finances its expenditure
by new issues of bonds and lump-sum tax. Its period by period budget constraint is
given by

\[ \sum_{n=1}^{\infty} P_{n,t}(B_{n,t} - B_{n+1,t-1}) + T_t = B_{1,t-1} + P_t G_t \] (2.10)

where \( G_t \) denotes government spending.

### 2.2.6 Exogenous Processes

The model economy is subject to five structural disturbances. Aggregate productivity evolves according to

\[ u_{a,t} = \frac{A_t}{A_{t-1}}, \ln u_{a,t+1} = (1 - \rho_u) \ln u_{a}^* + \rho_u \ln u_{a,t} + \eta_a \varepsilon_{a,t+1} \] (2.11)

The desired markups of the firm, \( \frac{s_t}{s_t - 1} \), evolves according to

\[ f_t = \frac{s_t}{s_t - 1}, \ln f_{t+1} = (1 - \rho_f) \ln f^* + \rho_f \ln f_t + \eta_f \varepsilon_{f,t+1} \] (2.12)

In Doh (2007), the markups are affected by a persistent shock that follows an \( AR(1) \) process with \( ARCH(1) \) structure. Here the \( ARCH \) structure is ignored since the \( ARCH \) effect is found to be insignificant in Doh (2007).

Monetary policy is subject to a serially uncorrelated transitory shock \( \varepsilon_{i,t} \) that appears in the Taylor rule, and a persistent target inflation shock that evolves according to the following equation:

\[ \ln \pi_{t+1}^* = (1 - \rho_{\pi^*}) \ln \pi^* + \rho_{\pi^*} \ln \pi_t^* + \eta_{\pi^*} \varepsilon_{\pi^*,t+1} \] (2.13)

As for the fiscal authority, we define \( g_t = \frac{1}{1 - \chi_t} \) and assume that
\[
\ln g_{t+1} = (1 - \rho_g) \ln g^* + \rho_g \ln g_t + \eta_g \varepsilon_{g,t+1} \tag{2.14}
\]

All five serially uncorrelated shocks \( \varepsilon_{a,t}, \varepsilon_{f,t}, \varepsilon_{i,t}, \varepsilon_{\pi^*,t}, \varepsilon_{g,t} \) are assumed to be independent of each other at all leads and lags. Each shock is assumed to follow a standard normal distribution.

Finally, we define the law of motion for habit as follows:

\[
\ln(X/A)_{t+1} = -xx + \psi \ln(X/A)_t + (1 - \psi) \ln(C/A)_t \tag{2.15}
\]

Here, we specify the law of motion for habit instead of the surplus ratio, as Campbell and Cochrane (1999) did. These two specifications are equivalent to each other. Note that an optimizing consumer facing this habit process will never choose a consumption level below habit.

### 2.2.7 Model Solution

The non-stationary technology trend is eliminated from the model by dividing the non-stationary variables by \( A \). Using the first order conditions and the constraints of the model, we can express the control variables, which are denoted as \( y_t = [\hat{\lambda}_{t}, \hat{\pi}_{t}, 1 + i_t, \hat{\lambda}_{t}^a, \hat{a}_t, \hat{f}_t, \hat{\pi}_{t}^*, \hat{\eta}_{t}] \), as functions of the state variables \( x_t = [1 + i_{t-1}, \hat{\pi}_{A_t}, \hat{u}_{a,t}, \hat{f}_{t}, \hat{\pi}_{t}^*, \hat{\eta}_{t}] \). We do so by solving the following system of equations:

\[
E_t f(y_{t+1}, y_t, x_{t+1}, x_t, \varepsilon_{t+1}) = 0 \tag{2.16}
\]
Here, a variable with a hat represents its deviation from its steady state value. For example, $\hat{X}_t = \ln\left(\frac{X_t}{X}\right)$, where $X_t$ represents the time variable, and $X$ is its steady state value.

The explicit form of the rational expectation system can be expressed as follows:

$$1 + i_t = (1 - \rho_t)\gamma_p(\hat{\pi}_{t+1} - \hat{\pi}_t^*) + (1 - \rho_t)\gamma_p\left(\frac{\hat{Y}_t}{A_t} + \frac{1}{\tau + 1/\nu} \hat{f}_t\right) + \rho_t \hat{u}_t + i_{t-1} + \eta_i \epsilon_{i,t} \quad (2.17)$$

$$\hat{u}_{a,t+1} = \rho_a \hat{u}_{a,t} \quad (2.18)$$

$$\hat{f}_{t+1} = \rho_f \hat{f}_t$$

$$\hat{\pi}_{t+1}^* = \rho_{\pi^*} \hat{\pi}_t^* \quad (2.19)$$

$$\hat{g}_{t+1} = \rho_g \hat{g}_t \quad (2.20)$$

$$0 = \frac{1}{g^*} \exp\left(\frac{\hat{C}_t}{A_t}\right) - \frac{1}{g^* \exp(\hat{g}_t)} \exp\left(\frac{\hat{Y}_t}{A_t}\right) + \frac{\phi}{2} \left[\exp(\hat{\pi}_t) - \exp(\hat{\pi}_t^*)\right]^2 (\pi^*)^2 \exp\left(\frac{\hat{Y}_t}{A_t}\right) \quad (2.21)$$

$$\exp(\hat{\lambda}_t^a) \left[1 - \exp\left(\frac{xx}{\psi - 1}\right)\right]^{-\tau} \left[1 - (1 - \psi) \beta \exp\left(\frac{xx}{\psi - 1}\right)\right] \quad (2.22)$$

$$= \left[\exp\left(\frac{\hat{C}_t}{A_t}\right) - \exp\left(\frac{xx}{\psi - 1} + \hat{X}_t\right)\right]^{-\tau}$$

$$- \beta \left[\exp\left(\frac{\hat{C}_{t+1}}{A_{t+1}}\right) - \exp\left(\frac{xx}{\psi - 1} + \hat{X}_{t+1}\right)\right]^{-\tau} \exp\left(\frac{xx}{\psi - 1}\right)$$

$$= \exp(\psi \frac{\hat{X}_t}{A_t})(1 - \psi) \exp(-\psi \frac{\hat{C}_t}{A_t})$$

48
\begin{equation}
\tilde{\lambda}_{t-1}^a - \tilde{\lambda}_t^a - \hat{u}_{a,t+1} + \hat{\theta}_t - \hat{\pi}_{t+1} = 0
\end{equation}

\begin{equation}
1 = \frac{\exp(\hat{f}_t) f^*}{\exp(\hat{f}_t) f^* - 1} \left[ 1 - \frac{\exp(\hat{\lambda}_t^a)}{\exp(\lambda_t^a) f^*} \right]
+ \phi \exp(\hat{\pi}_t) (\pi^*)^2 [\exp(\hat{\pi}_t) - \exp(\hat{\pi}_t^*)] \\
- \frac{\phi}{2} \frac{\exp(\hat{f}_t) f^*}{\exp(\hat{f}_t) f^* - 1} [\exp(\hat{\pi}_t) \pi^* - \exp(\hat{\pi}_t^*) \pi^*] \\
- \phi \beta \exp(\hat{\lambda}_{t-1}^a - \hat{\lambda}_t^a + \hat{Y}_{t+1}^a - \hat{Y}_t^a + \hat{\pi}_{t+1}^a) (\pi^*)^2 \exp(\frac{\hat{\pi}_{t+1}^a}{\hat{\pi}_{t+1}^*})
\end{equation}

\begin{equation}
\exp(\hat{S}_t) - \exp(\hat{S}_t + \frac{xx}{\psi - 1}) = 1 - \exp(\frac{\hat{X}_t}{A_t} + \frac{xx}{\psi - 1} - \frac{\hat{C}_t}{A_t})
\end{equation}

Linear approximation is used to solve the rational expectations system. The advantage of this method is that it will lead to a state-space representation of the DSGE model that can be analyzed by Kalman filter. The linear solution of the DSGE model can be expressed as follows:

\begin{equation}
y_t = g_x x_t
\end{equation}

\begin{equation}
x_{t+1} = h_x x_t + \eta \varepsilon_{t+1}
\end{equation}

The computational algorithm of Schmitt-Grohe and Uribe (2004) is used to obtain $g_x$ and $h_x$.

The complete set of asset pricing implications cannot be studied in the pure log-linear approximation since there is no consideration of risk in the log-linearized macro dynamics. Doh (2007) obtains the risk premium term by combining the conditional
log normality of the model implied pricing kernel with the log-linear approximation of macro dynamics. The log bond price of maturity of \( n \) month, \( \hat{p}_{n,t} \), has the following representation under the no-arbitrage assumption:

\[
\hat{p}_{n,t} = a_n + b_n x_t
\]  
(2.27)

\( a_n \) and \( b_n \) take the same form as those in Doh (2007).

With the bond price, the yield of \( n \) month bond \( \hat{y}_{n,t} \) can be computed as \( -\frac{\hat{p}_{n,t}}{n} \).

### 2.2.8 Measurement Equations for Macro Variables and Bond Yields

To estimate the DSGE model, 8 measurement equations are used to relate macro state variables to observations of macro and term structure variables.

\[
\begin{align*}
GDP_t &= \ln A_t + \ln \left( \frac{Y}{A} \right)^* + \frac{\hat{Y}_t}{A_t} + \xi_{y,t} \\
INF_t &= \ln \pi^* + \hat{\pi}_t + \xi_{\pi,t} \\
INT_t &= \ln (1 + i^*) + 1 + \hat{i}_t + \xi_{i,t} \\
YIE1 &= \ln (y_1^*) + \hat{y}_{1,t} + \xi_{y1,t} \\
YIE2 &= \ln (y_2^*) + \hat{y}_{2,t} + \xi_{y2,t} \\
YIE3 &= \ln (y_3^*) + \hat{y}_{3,t} + \xi_{y3,t} \\
YIE4 &= \ln (y_4^*) + \hat{y}_{4,t} + \xi_{y4,t} \\
YIE5 &= \ln (y_5^*) + \hat{y}_{5,t} + \xi_{y5,t}
\end{align*}
\]  

where \( \xi_{y,t}, \xi_{\pi,t}, \xi_{i,t}, \xi_{y1,t}, \xi_{y2,t}, \xi_{y3,t}, \xi_{y4,t} \) are measurement errors for per capita GDP, month to month inflation rates, 3-month Treasury bill rates, 1-year yield, 2-year yield,
3-year yield, 4-year yield and 5-year yield whose standard deviations are $\sigma_{my}$, $\sigma_{m\pi}$, $\sigma_{mi}$, $\sigma_{my1}$, $\sigma_{my2}$, $\sigma_{my3}$, $\sigma_{my4}$ and $\sigma_{my5}$ respectively.

### 2.3 Econometric Methodology

The Bayesian method is used to estimate the model. The likelihood function is constructed using Kalman filter and then combined with the prior density to obtain the posterior density. Because the posterior density cannot be derived analytically, multiple MCMC chains are used to generate posterior draws. Unobserved macro factors are estimated using Kalman filter.

#### 2.3.1 Construction of the Likelihood Function Using Kalman Filtering

With the model solutions and measurement equations, it is easy to construct the following linear state space model.

\[
x_t = \Gamma_1(\vartheta)x_{t-1} + \eta \xi_t
\]

\[
z_t = \alpha_0(\vartheta) + \alpha_1(\vartheta)x_t + \xi_t \text{ where } \xi_t \sim N(0, H)
\]

\[
z_t = [\ln Y_t, \ln (1 + \pi_t), \ln (1 + i_t), y_{1,t}, y_{2,t}, y_{3,t}, y_{4,t}, y_{5,t}]'
\]

\[
\vartheta = [\tau, \beta, \nu, \ln f^*, \phi, \ln u^*_a, \gamma_p, \gamma_y, \rho_f, \rho_i, \rho_{a*}, \rho_g, \eta_f, \eta_i, \eta_{a*}, \eta_{g*}, \ln A_0, \ln \pi^*, xx, \psi, g^*]'
\]

where $z_t$ is a vector of observed variables and $\xi_t$ is a vector of measurement errors. The standard deviation of measurement errors are fixed as 20% of that of the observation in the sample period for simplicity. Since the above state space model is linear and Gaussian, Kalman filter is appropriate for constructing the likelihood function. For details, see Hamilton (1994, Ch 13).
2.3.2 Posterior Simulation

After we obtain the likelihood function, we combine it with the prior density to get the posterior distribution of the parameters using \( p(\theta | z^T) \propto p(\theta) L(\theta | z^T) \). The MCMC method is used to generate random draws from the posterior distribution, as the analytical form of the posterior density is not available. The MCMC method is the Metropolis-Hastings algorithm, which is summarized below:

Step 1 (Selection of the Starting Point): Compute the log-likelihood for 100 draws from the prior distribution. Select the point that gives the highest log likelihood value \( \theta^* \).

Step 2: (Proposal): Starting from \( \theta^* \), generate a new draw by the following random-walk proposal density. The scaling matrix \( c \) is chosen by multiplying a small positive real number to the prior covariance matrix. \( \theta^{N+1} = \theta^{j} + cN(0, I) \), \( j = 0, \ldots, R - 1 \).

Step 3: (Accept/Reject): Compute the acceptance rate \( \alpha = \min \left\{ \frac{p(\theta^{N+1} | z^T)}{p(\theta^{j} | z^T)}, 1 \right\} \) and accept or reject according to the value of \( u \) that is drawn from the uniform distribution over the unit interval \([0, 1]\). \( \theta^{j+1} = \theta^{N+1} \), if \( u < \alpha \); \( \theta^{j+1} = \theta^{j} \), otherwise.

Step 4 (Burn In): For the purpose of the posterior inference, discard the initial \( B \) draws and use the remaining draws.

For details, see Doh (2007).

After we get the posterior draws for each parameter, we take their mean values as the estimates of the parameters. The estimation results are given in Section 4.
2.4 Estimation Results

The estimation method described above is applied to US data. The dataset is described in section 2.4.1. Section 2.4.2 describes the prior distributions for the parameters and Section 2.4.3 gives the parameter estimates.

2.4.1 Data

Monthly macro and financial data are used in the estimation. Macro data are taken from the Federal Reserve Database (FRED) St. Louis. Because monthly real GDP data is not available, we use real personal income instead, which is obtained by deflating personal income (PI) by the consumption price index (CPIAUCSL). Then real personal income is divided by total population (POP) to get per-capita real GDP. The inflation rate is the log difference of the consumption price index (CPIAUCSL). The nominal interest rate is extracted from the Fama CRSP risk free rate file, and the average quote of 3-month Treasury bill rate is used. Bond yields are obtained from bond prices (1, 2, 3, 4, and 5 years), which are obtained from the Fama CRSP discount bond yields files. The sample period is 1964-2003, as in Cochrane and Piazzesi (2005).

2.4.2 Prior Distributions of Parameters

Most prior distributions of parameters follow Doh (2007) with minor modifications to convert the quarterly setting into the monthly setting. The curvature of utility function $\tau$, follows a gamma distribution with mean 2 and standard deviation 0.5, which reflects the range frequently used in literature. The prior mean of the average technology progress $u_a^*$ is set to be 0.2% at the monthly frequency, which is roughly one third of Doh’s quarterly setting. The prior mean of the discount factor $\beta$ is
chosen such that the steady state annual real interest rate is 2.8% in the annualized percentage. The elasticity of labor supply $\nu$ is fixed at 0.5 as in Doh (2007). Because we do not have data for hours worked in the estimation, the estimate of may not be reliable. There are five new parameters. Three of them come from equation (14), which defines the law of motion for government spending. We adopt the same prior distributions for these three parameters as An (2007). The other two come from Campbell and Cochrane’s habit formation. The serial correlation parameter $\psi$ is chosen to match the serial correlation of the log price-dividend ratios in their original paper. We fixed the value of $\psi$ at 0.99, which is the calibrated value in Campbell and Cochrane (1999). The prior distribution for $xx$ is chosen to mimic the estimated surplus ratio in Campbell and Cochrane (1999).

Table 2.1 summarizes the prior distribution for each parameter.

### 2.4.3 Estimation Results

The estimation procedure follows An (2007). The Metropolis-Hasting algorithm is used to obtain the posterior draws. We use the scaling factor 0.3 for the estimation to target an acceptance ratio of about 40%.

The estimation results of the linear DSGE model are given in Table 2.2 together with the 90% confidence intervals.

### 2.5 Econometric Analysis

So far we have discussed the model. When we obtain the parameter values of the model, the latent macro variables are estimated using the Kalman filter. This
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Domain</th>
<th>Density</th>
<th>Para(1)</th>
<th>Para(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>( \beta )</td>
<td>[0, 1]</td>
<td>Beta</td>
<td>0.997</td>
<td>0.001</td>
</tr>
<tr>
<td>( \ln f^* )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>( \phi )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>( \ln u_a^* )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>( \gamma_p )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>( \gamma_y )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>( \rho_a )</td>
<td>[0, 1]</td>
<td>Beta</td>
<td>0.9</td>
<td>0.05</td>
</tr>
<tr>
<td>( \rho_f )</td>
<td>[0, 1]</td>
<td>Beta</td>
<td>0.93</td>
<td>0.05</td>
</tr>
<tr>
<td>( \rho_i )</td>
<td>[0, 1]</td>
<td>Beta</td>
<td>0.79</td>
<td>0.1</td>
</tr>
<tr>
<td>( \rho_{\pi^*} )</td>
<td>[0, 1]</td>
<td>Beta</td>
<td>0.93</td>
<td>0.05</td>
</tr>
<tr>
<td>( \eta_a )</td>
<td>( R^+ )</td>
<td>Inverse Gamma</td>
<td>0.004</td>
<td>4</td>
</tr>
<tr>
<td>( \eta_f )</td>
<td>( R^+ )</td>
<td>Inverse Gamma</td>
<td>0.01</td>
<td>4</td>
</tr>
<tr>
<td>( \eta_i )</td>
<td>( R^+ )</td>
<td>Inverse Gamma</td>
<td>0.003</td>
<td>4</td>
</tr>
<tr>
<td>( \eta_{\pi^*} )</td>
<td>( R^+ )</td>
<td>Inverse Gamma</td>
<td>0.02</td>
<td>4</td>
</tr>
<tr>
<td>( \ln \pi^* )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>( xx )</td>
<td>( R^+ )</td>
<td>Gamma</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td>( \ln A_0 )</td>
<td>( R^+ )</td>
<td>Normal</td>
<td>7.8</td>
<td>0.2</td>
</tr>
<tr>
<td>( \rho_g )</td>
<td>[0, 1]</td>
<td>Beta</td>
<td>0.9</td>
<td>0.05</td>
</tr>
<tr>
<td>( \eta_g )</td>
<td>( R^+ )</td>
<td>Inverse Gamma</td>
<td>0.002</td>
<td>4</td>
</tr>
<tr>
<td>( 1/g^* )</td>
<td>[0, 1]</td>
<td>Beta</td>
<td>0.85</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Note: para(1) and para(2) list means and stds for Beta, Gamma, Normal distributions. \( s \) and \( v \) for Inverse Gamma, where \( PIG \propto \sigma^{-v-1}e^{-vs^2/2\sigma^2} \)

Table 2.1: Prior Distribution
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>Std</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>2.564</td>
<td>0.470</td>
<td>(1.797, 3.332)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.998</td>
<td>0.000</td>
<td>(0.997, 0.998)</td>
</tr>
<tr>
<td>$\ln f^*$</td>
<td>0.356</td>
<td>0.059</td>
<td>(0.258, 0.452)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>239.793</td>
<td>5.862</td>
<td>(231.815, 250.098)</td>
</tr>
<tr>
<td>$\ln u_n$</td>
<td>0.001</td>
<td>0.000</td>
<td>(0.001, 0.002)</td>
</tr>
<tr>
<td>$\gamma_p$</td>
<td>1.530</td>
<td>0.153</td>
<td>(1.281, 1.781)</td>
</tr>
<tr>
<td>$\gamma_y$</td>
<td>0.013</td>
<td>0.004</td>
<td>(0.006, 0.019)</td>
</tr>
<tr>
<td>$\rho_a$</td>
<td>0.898</td>
<td>0.008</td>
<td>(0.883, 0.912)</td>
</tr>
<tr>
<td>$\rho_f$</td>
<td>0.991</td>
<td>0.001</td>
<td>(0.988, 0.992)</td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>0.907</td>
<td>0.011</td>
<td>(0.888, 0.927)</td>
</tr>
<tr>
<td>$\rho_{\pi^*}$</td>
<td>0.684</td>
<td>0.041</td>
<td>(0.616, 0.749)</td>
</tr>
<tr>
<td>$\eta_a$</td>
<td>0.002</td>
<td>0.000</td>
<td>(0.001, 0.002)</td>
</tr>
<tr>
<td>$\eta_f$</td>
<td>0.017</td>
<td>0.002</td>
<td>(0.014, 0.020)</td>
</tr>
<tr>
<td>$\eta_i$</td>
<td>0.001</td>
<td>0.000</td>
<td>(0.001, 0.001)</td>
</tr>
<tr>
<td>$\eta_{\pi^*}$</td>
<td>0.002</td>
<td>0.000</td>
<td>(0.002, 0.002)</td>
</tr>
<tr>
<td>$\ln \pi^*$</td>
<td>0.003</td>
<td>0.000</td>
<td>(0.002, 0.004)</td>
</tr>
<tr>
<td>$xx$</td>
<td>0.002</td>
<td>0.001</td>
<td>(0.001, 0.003)</td>
</tr>
<tr>
<td>$\ln A_0$</td>
<td>8.072</td>
<td>0.146</td>
<td>(7.839, 8.315)</td>
</tr>
<tr>
<td>$\rho_g$</td>
<td>0.991</td>
<td>0.002</td>
<td>(0.988, 0.994)</td>
</tr>
<tr>
<td>$\eta_g$</td>
<td>0.005</td>
<td>0.000</td>
<td>(0.004, 0.005)</td>
</tr>
<tr>
<td>$1/g^*$</td>
<td>0.785</td>
<td>0.127</td>
<td>(0.602, 0.987)</td>
</tr>
</tbody>
</table>

Table 2.2: Posterior Distribution
provides the data series of the latent macro variables. Using those series, we conduct various econometric analyses to investigate the relationship between bond risk premia and macro variables.

### 2.5.1 Characterizing the Forecasting Factor

Unobserved macro variables, such as the marginal rate of substitution and macro shocks, are extracted and used as data series in econometric analysis. Doh (2007) uses the same method to interpret the term structure factors, level, slope and curvature by regressing the term structure factors on the estimated data series. He finds that the level of the yield curve is accounted for by a persistent monetary policy shock, while the slope and curvature are driven by a markup shock and a transitory monetary policy shock, respectively.

In this paper, we characterize the forecasting factor of Cochrane and Piazzesi (2005), using the same approach as Doh (2007). First, we follow Cochrane and Piazzesi (2005) and derive the forecasting factor as a linear combination of forward rates. Specifically, given the log yield data, we derive the log price of n year discount bond at time $t$ as $p_t^{(n)} = -n y_t^{(n)}$. Then the log forward rate at time $t + n$ for loans between time $t + n - 1$ and $t + n$ is $f_t^{(n)} = p_t^{(n-1)} - p_t^{(n)}$, and the log holding period return from buying an $n$ year bond at time $t$ and selling it as an $n - 1$ year bond at time $t + 1$ is $r_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)}$. We denote the excess log returns by $r_{x_{t+1}}^{(n)} = r_{t+1}^{(n)} - y_t^{(1)}$. Vector forms are used for convenience. We use the letters without $n$ index to denote vectors across maturity.
With all these constructed data, we derive the forecasting factor by running the following regression

\[
\frac{1}{4} \sum_{n=2}^{5} rx_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_{t}^{(1)} + \gamma_2 y_{t}^{(2)} + \gamma_3 y_{t}^{(3)} + \gamma_4 y_{t}^{(4)} + \gamma_5 y_{t}^{(5)} + \varepsilon_{t+1} \tag{2.30}
\]

where \(\gamma' f_t\) is the forecasting factor we try to construct.

The forecasting factor can be understood by regressing the factor on macro variables. The macro variables of interest include the marginal rate of substitution, habit \(\tilde{X}_t\), surplus ratio \(\tilde{S}_t\), technology shock \(\tilde{u}_{a,t}\), markup shock \(\tilde{f}_t\), transitory monetary shock \(\varepsilon_{it}\), permanent monetary shock \(\tilde{\pi}_{t}^{*}\), and government spending shock \(\tilde{g}_t\). Because we are considering the 1-year excess return, marginal rate of substitution is defined as the difference between the marginal utility today and the marginal utility one year ago. That is \(\overline{MRS}_t = \tilde{\lambda}_t - \tilde{\lambda}_{t-12}\). We regress the forecasting factor on combinations of the macro variables and report the results with significant coefficients in Table 2.3. The standard test for significance is not accurate due to serial correlations caused by overlapping data, so the Hansen-Hodrick correction is used to conduct inference.

Table 2.3 shows that, although almost all the variables can individually explain the variation of the forecasting factor, the technology shock explains the most variation
Table 2.3: Regression of Forecasting Factor on Different Macro Variables

<table>
<thead>
<tr>
<th></th>
<th>$\widehat{MRS}_t$</th>
<th>$\widehat{S}_t$</th>
<th>$\widehat{X}_t$</th>
<th>$\widehat{u}_{a,t}$</th>
<th>$\widehat{f}_t$</th>
<th>$\varepsilon_{it}$</th>
<th>$\widehat{\pi}_t$</th>
<th>$\widehat{g}_t$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$FF_t$</td>
<td>0.056</td>
<td>2.012</td>
<td>2.609</td>
<td>-7.942</td>
<td>0.568</td>
<td>-0.001</td>
<td>-2.545</td>
<td>1.151</td>
<td>0.664</td>
</tr>
<tr>
<td>$FF_t$</td>
<td>-0.236</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.182</td>
</tr>
<tr>
<td>$FF_t$</td>
<td></td>
<td>-0.867</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.118</td>
</tr>
<tr>
<td>$FF_t$</td>
<td></td>
<td></td>
<td>-4.551</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.539</td>
</tr>
<tr>
<td>$FF_t$</td>
<td></td>
<td></td>
<td></td>
<td>0.096</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.247</td>
</tr>
<tr>
<td>$FF_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.009</td>
<td></td>
<td></td>
<td></td>
<td>0.048</td>
</tr>
<tr>
<td>$FF_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-4.623</td>
<td></td>
<td></td>
<td>0.130</td>
</tr>
<tr>
<td>$FF_t$</td>
<td>0.081</td>
<td>-0.095</td>
<td>-4.863</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.640</td>
</tr>
</tbody>
</table>

Note: Sample periods: 1964-2003

of the forecasting factor. The $R^2$ is 53.9%. When combined with the technology shock, variables such as the surplus ratio and the markup shock remain significant. Many variables, however, become insignificant and technology shock drives them out completely. The marginal rate of substitution is an exception. When regressed on the variable alone, the marginal rate of substitution is insignificant in explaining the forecasting factor. However, when combined with the technology shock, it turns out to be significant. This result leads us to use the marginal rate of substitution together with the technology shock to construct a macro factor in the next section.

Although the markup shock is less important than the technology shock in explaining the forecasting factor for the 1-year excess return, it is interesting that as the horizon of the excess bond return increases, the markup shock becomes increasingly important. The increasing explanatory power of the markup shock can be seen from Figure 2.1. In Figure 2.1, we plot the markup shock versus the forecasting factors for 1, 2 and 4-year excess returns. The plot shows that as the return horizon increases, the correlation of the markup shock and the forecasting factor also increases. The
result is confirmed in regression analysis. As the return horizon increases from 1 year to 2 year and to 4 year, the $R^2$ goes up from 25% to 82%, and to 86%.

2.5.2 Robustness Check

The technology shock is estimated from the DSGE model, so some robustness check is necessary to guarantee that it indeed captures the variation of technology change.

We start with labor productivity, a variable that is believed to vary closely with changes in technology. Labor productivity is obtained by dividing real GDP (GDPC1) by the civilian labor force (CLF16OV). The data series are obtained from the Federal
Reserve Database (FRED) St. Louis. The correlation between the detrended labor productivity and model implied technology shock is around 0.4.

A VAR method is also implemented to estimate the technology shock following Gali (1999) with the long run restriction that only the technology shock has a permanent effect on labor productivity. However, the result implies low negative correlation between the VAR-based technology shock and the technology we extracted from the DSGE model.

One possible reason for this might be the interaction of the structural shocks and measurement errors. In this model economy, we assume five structural shocks, and eight measurement errors. Boivin and Giannoni (2005) argue that

“...when no more than one indicator is used for any concept of the model..., both the structural shocks and the unobserved variables have to be identified entirely from the restricted dynamics of the DSGE model,... In that case, having more structural shocks in the model limits the number of independent sources of measurement errors that can be contemplated and it is difficult to formally test whether the resulting model is properly identified or not. Typically, researchers avoid these problems by assuming either no measurement error or few structural shocks,... (However,) measurement error or conceptual differences between the measured indicators and the theoretical variables might be quite prevalent, and if so, ignoring them would lead to biased inference.” (p. 14-15, Boivin and Giannoni (2005))

One tentative solution is estimating the DSGE model in a data rich environment. With more than one data series to indicate a variable, it is possible to identify the latent variables using the cross-section of macroeconomic indicators alone. Enough information is provided to identify the latent variables. As a result, we can allow
for a large number of measurement errors without the concern about the number of structural shocks that can be identified.

In addition, using more than one data series for each macro variable may correct for the conceptual differences between the model variables and the data used to measure them. For example, this paper uses personal income data to proxy for GDP. The two variables are quite different conceptually. Future research may focus on this issue, estimating the DSGE model in a data rich environment.

Although the technology shock implied by this model is different from the one derived by Gali (1999), it is closely related to business cycle expansions and contractions. We plot the series for model implied technology shock and compare it with the dates of peaks and troughs in economic activity documented by NBER. The second panel of Figure 2.2 shows that the booms coincide with the peaks of the technology shock, and the third panel of Figure 2.2 shows that the recessions also coincide with the troughs of the technology shock. This perfect match convinces us that what we derived from the macro model has the same property as the technology shock.

2.5.3 The Macro Factor versus the Forecasting Factor

Characterizing the forecasting factor is an indirect way to determine which macro variables will have forecasting power. Another way is to study the macro variables directly. Panel A of Table 2.4 gives the regression of the average 1-year excess return of 2, 3, 4 and 5 year bonds on macro variables and the forecasting factor. Regressions with insignificant coefficients and low $R^2$ are omitted. The Hansen-Hodrick correction is used to account for serial correlations caused by overlapping data.
Figure 2.2: The Technology Shock and the Business Cycle

Panel A

<table>
<thead>
<tr>
<th></th>
<th>$c$</th>
<th>$\widehat{MRS}_{t-12}$</th>
<th>$\left(\hat{X}_{t-12}^A\right)$</th>
<th>$\hat{u}_{a,t-12}$</th>
<th>$FF_{t-12}$</th>
<th>$\overline{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ARX_t$</td>
<td>0.013</td>
<td>0.255</td>
<td>-0.805</td>
<td>-5.810</td>
<td></td>
<td>0.388</td>
</tr>
<tr>
<td>$ARX_t$</td>
<td>0.008</td>
<td>0.188</td>
<td>-0.871</td>
<td>-2.437</td>
<td>0.617</td>
<td>0.438</td>
</tr>
<tr>
<td>$ARX_t$</td>
<td>0.019</td>
<td>-1.511</td>
<td></td>
<td></td>
<td>0.125</td>
<td>0.235</td>
</tr>
<tr>
<td>$ARX_t$</td>
<td>0.008</td>
<td>-5.115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B

<table>
<thead>
<tr>
<th></th>
<th>$c$</th>
<th>$\widehat{MRS}^*_{t-12}$</th>
<th>$\left(\hat{X}^*_{t-12}^A\right)$</th>
<th>$\hat{u}^*_{a,t-12}$</th>
<th>$FF_{t-12}$</th>
<th>$\overline{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ARX_t$</td>
<td>0.011</td>
<td>0.298</td>
<td>-0.440</td>
<td>-5.980</td>
<td></td>
<td>0.437</td>
</tr>
<tr>
<td>$ARX_t$</td>
<td>0.007</td>
<td>0.230</td>
<td>-0.424</td>
<td>-3.294</td>
<td>0.527</td>
<td>0.477</td>
</tr>
</tbody>
</table>

Note: Sample periods: 1964-2003

Table 2.4: Regression of the Average One Year Excess Return of 2,3,4,5 Year Bond on Macro Variables and the Forecasting Factor
The message is clear. The technology shock plays an important role in forecasting average 1-year excess bond return. If we forecast today’s excess bond return using the technology shock 1 year ago, we can obtain an of $R^2$ about 23%. This result confirms our previous finding that the forecasting factor is mostly driven by the technology shock.

We also find two additional macro variables that are important for forecasting bond returns. They are the marginal rate of substitution and habit. The adjusted $R^2$ increases to 39% with these three macro variables, which is about 4% higher than the one when using the Cochrane and Piazzesi’s forecasting factor. This result is positive. It demonstrates that macro variables do have forecasting power on bond returns. Although they are latent instead of traditional observed macro variables, their economic interpretations are clear.

One may argue that, since those macro variables are estimated from the model, small changes of parameters might cause different estimates of variables and thus resulting in different forecasting performance. To address this concern, we conduct the following exercise to investigate the relationship between the parameter values and forecasting power of macro variables.

The Bayesian estimation gives us the point estimates of the parameters as well as the 90% interval. We can examine the forecasting performance of macro variables under different parameter values within the interval. Considering the high dimensionality caused by a large number of parameters, we will only consider the parameters one at a time, and fix other parameters at their point estimates. For each parameter, we break the interval into nine subintervals and obtain 10 different values, arranged in an ascending order. The forecasting performance is examined for each of the 10
different values and for all the 21 parameters. Then for each parameter, we plot the $R^2$’s of the resulting macro variables against the parameter values within the 90% interval (Not shown here to save space). Although the $R^2$’s vary with the parameter values, they remain a high level for all the parameter changes. All of them stay above the level of 35%, the one obtained by Cochrane and Piazzesi. For most of parameters, the changes of $R^2$’s are monotonic, either decreasing or increasing. Because of this monotonic pattern, it is straightforward to locate the parameter values that produce even higher $R^2$. By examining combinations of different parameter values, we are able to reach an $R^2$ of 44%. An asterisk is used to denote the new macro variables that produce an $R^2$ of 44%. A comparison of these two groups of variables is given in Figure 2.3. New variables are plotted against the old ones and they look similar to each other. The close similarity is confirmed by the high correlation coefficients, which we report in Table 2.5.

The new variables are after all derived under different parameter values. The plausibility of new parameter values is necessary to justify the use of new variables. The result of the likelihood ratio test shows the model with new parameters cannot be rejected.

In Panel B of Table 2.4, we report the regression results of the average 1-year excess return on $\overline{MRS}_{t-12}^*$, $\overline{(\frac{X}{A})}_{t-12}^*$, and $\widehat{u}_{a,t-12}^*$. All three variables are significant in
Figure 2.3: Comparison of Two Groups of Macro Variables
explaining the average excess return and remain significant even when combined with the forecasting factor. The highest is 48%, obtained by combining the macro variables and the forecasting factor, which is very large relative to the rest of the literature.

Following Cochrane and Piazzesi (2005), we also construct a single macro factor by forming linear combinations of $\widehat{MRS}_{t-12}^*, \widehat{\left(\frac{X}{A}\right)}_{t-12}^*$ and $\widehat{u}_{a,t-12}^*$.

We run regressions of excess returns on all macro variables.

$$rx_t^{(n)} = \beta_0^{(n)} + \beta_1^{(n)} \widehat{MRS}_{t-12}^* + \beta_2^{(n)} \widehat{\left(\frac{X}{A}\right)}_{t-12}^* + \beta_3^{(n)} \widehat{u}_{a,t-12}^* + \varepsilon_t^{n} \quad (2.31)$$

The top panel of Figure 2.4 graphs the slope coefficients $[\beta_1^{(n)}, \beta_2^{(n)}, \beta_3^{(n)}]$ as a function of maturity $n$. We reach a similar conclusion as Cochrane and Piazzesi (2005), namely that the same function of macro variables predicts holding period returns at all maturities. A single macro factor can be constructed to describe expected excess returns of all maturities as follows:

$$rx_t^{(n)} = b_n (\gamma_0 + \gamma_1 \widehat{MRS}_{t-12}^* + \gamma_2 \widehat{\left(\frac{X}{A}\right)}_{t-12}^* + \gamma_3 \widehat{u}_{a,t-12}^*) + \varepsilon_t^{n} \quad (2.32)$$

A normalization is made such that $\frac{1}{4} \sum_{n=2}^{5} b_n = 1$.

The two stage least squares estimation in Cochrane and Piazzesi (2005) is used. First, we estimate the $\gamma$’s by running a regression of the average excess return on all macro variables.

$$\frac{1}{4} \sum_{n=2}^{5} rx_t^{(n)} = (\gamma_0 + \gamma_1 \widehat{MRS}_{t-12}^* + \gamma_2 \widehat{\left(\frac{X}{A}\right)}_{t-12}^* + \gamma_3 \widehat{u}_{a,t-12}^*) + \varepsilon_t \quad (2.33)$$

We write it in vector form as follows:
Figure 2.4: Regression Coefficients of One-year Excess Returns on Macro Variables
$$\bar{x}_t = \Gamma' MF_{t-12} + \varepsilon_t$$  \hspace{1cm} (2.34)$$

where $\Gamma = [\gamma_0, \gamma_1, \gamma_2, \gamma_3]'$. $MF_{t-12} = [1, \overline{MRS}_{t-12}; (\overline{X})^*_t_{t-12}; \overline{u}_{a,t-12}]'$. 

Then we estimate by running the four regressions:

$$rx_t^{(n)} = b_n \Gamma' MF_{t-12} + \varepsilon_t^n$$  \hspace{1cm} (2.35)$$

The single factor model is a restricted model. The coefficients of individual bond expected returns on macro variables implied by restricted model is plotted in the bottom panel of Figure 2.4. Although a formal statistical test has not been conducted to determine whether the restricted model is rejected, visual inspection suggests that the restricted model captures most of the information in the unrestricted model.

Table 2.6 shows that the macro factor constructed forecasts bond returns at all maturities. A bond with a longer maturity has a heavier loading on this macro factor. The results of Cochrane and Piazzesi’s (2005) forecasting factor are also provided for comparison. The $R^2$’s are more than 8% higher for the macro factor. The Macro factor thus predicts better than the forecasting factor. Another interesting finding is that the $R^2$’s increase with the time horizon, which is consistent with the view that the return for long term bonds is more stable and thus more predictable. To gain additional insights, we also plot the macro factor against the forecasting factor. Figure 2.5 shows that the macro factor and the forecasting factor are highly correlated. They capture almost the same variation of the 1-year excess bond return. However, the forecasting factor seems to be less volatile, and that perhaps is why it does not forecast as well as the macro factor.
Figure 2.5: Forecasting Factor vs Macro Factor
Panel A

<table>
<thead>
<tr>
<th></th>
<th>Macro factor</th>
<th>$\bar{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r \times_t^{(2)}$</td>
<td>0.466 (7.646)</td>
<td>0.397</td>
</tr>
<tr>
<td>$r \times_t^{(3)}$</td>
<td>0.857 (7.984)</td>
<td>0.416</td>
</tr>
<tr>
<td>$r \times_t^{(4)}$</td>
<td>1.211 (8.341)</td>
<td>0.445</td>
</tr>
<tr>
<td>$r \times_t^{(5)}$</td>
<td>1.464 (8.719)</td>
<td>0.449</td>
</tr>
<tr>
<td>$ARX_t$</td>
<td>1.000 (8.419)</td>
<td>0.441</td>
</tr>
</tbody>
</table>

Panel B

<table>
<thead>
<tr>
<th></th>
<th>Forecasting Factor</th>
<th>$\bar{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r \times_t^{(2)}$</td>
<td>0.465 (8.077)</td>
<td>0.310</td>
</tr>
<tr>
<td>$r \times_t^{(3)}$</td>
<td>0.866 (7.775)</td>
<td>0.336</td>
</tr>
<tr>
<td>$r \times_t^{(4)}$</td>
<td>1.235 (7.476)</td>
<td>0.370</td>
</tr>
<tr>
<td>$r \times_t^{(5)}$</td>
<td>1.433 (6.832)</td>
<td>0.343</td>
</tr>
<tr>
<td>$ARX_t$</td>
<td>1.000 (7.448)</td>
<td>0.351</td>
</tr>
</tbody>
</table>

Note: Sample periods: 1964-2003
Numbers in () are t-values

Table 2.6: Forecasts of Excess Bond Return

Since the macro factor and the forecasting factor are highly correlated, we should not expect much higher predictive power when using both factors to forecast bond return. The result is given as follows:

$$ARX_t = -0.002 + 0.429FF_{t-12} + 0.735MF_t + e_t \ , \ \bar{R}^2 = 0.472 \quad (2.36)$$

The Hansen-Hodrick correction with 12 lags is used here to account for the serial correlations caused by overlapping data. The significance of the forecasting factor decreases dramatically. The predictive power of the forecasting factor is somewhat driven out by that of the macro factor. The adjusted $R^2$ only increases a little beyond the one with only macro factor, which is around 44%. Thus, we only need to use the macro factor for the purpose of forecasting. This result is stronger than Ludvigson.
and Ng (2007). Although they finally reach a high $R^2$ of 44% using the forecasting factor and their macro factor derived from dynamic factor analysis with hundreds of data series, they rely significantly on the forecasting factor. They did not produce an $R^2$ larger than 30% with their macro factors alone.

2.5.4 What is Causing the Forecastability of Bond Returns?

The forecastability of bond returns implies time-varying term premia that are caused by the presence of habit. In this paper the expressions of bond yields are derived in a log-linearized DSGE model. Terms that contribute to the time-varying term premia are completely missing due to the first order approximation. That is although the model itself implies time-varying term premia, we lose this property by taking a first order approximation. The model implied bond returns should not be predictable. With this concern, we move to a non-linear approximation of the model that is compatible with time-varying term premia.

Following Schmitt-Grohe and Uribe (2004), we compute second order accurate approximations to the solution of the DSGE model. The closed form solution for yields is also derived using the method proposed by Doh (2007). As emphasized by Doh (2007), the higher order approximation brings in the time-varying component for term premia.

The non-linear model can be estimated using a Bayesian method. The particle filter is used to construct the likelihood function and estimate the unobservable macro variables. Details can be found in Doh (2007).

To construct the likelihood function with the particle filter, we have to determine the number of particles. In Doh (2007), the number of particles is chosen based on
evaluation of log-likelihood across 40 different seeds. They choose 60,000 particles because the change of standard deviation of likelihood values is very small. The number of observations will affect the choice of particles ceteris paribus. Intuitively, more observations will require a larger number of particles. Monthly data is used in this paper, and the time span is from 1964 through 2003; as a result, the number of observations is 6 times as large as the one in Doh (2007). A much larger number is required. There is a tradeoff between computational cost and accuracy of estimation. We use 100,000 particles to estimate the parameters of the model and the results are still not reliable. As a compromise, the parameters are estimated in the log-linear model and we extract the unobservable macro variables using the particle filter. It is appropriate to do this, because Doh (2007) shows the parameters estimated from the linear model and nonlinear model do not differ much. The same econometric analysis is conducted using these macro variables and similar results are obtained.

One argument can be made based on this exercise: the linearized model captures most of the properties of the original one, and the macro variables obtained from the linear model are good approximations, which explains why we can use these variables to forecast bond returns.

2.5.5 Forecasting Stock Returns

Stocks can be viewed as a long-term bond plus cash flow risk (Cochrane and Piazzesi (2005)). Therefore, any variable forecasting bond returns also forecasts stock returns. The first row of Table 2.7 presents results on forecasting excess stock returns with the macro factor. The coefficient is 1.544 and statistically significant. This is larger than the coefficient for five year bond return, which is 1.464, so the stock return
\begin{table}
\centering
\begin{tabular}{lcccc}
 & $c$ & $FF_t$ & $MF_t$ & $R^2$ \\
$SXR_{t+12}$ & 0.029 & 1.544 & 0.079 & \\
$SXR_{t+12}$ & 0.028 & 1.663 & 0.072 & \\
$SXR_{t+12}$ & 0.025 & 0.874 & 1.007 & 0.088 \\
\end{tabular}
\caption{Forecasting Stock Return Using Macro Factor}
\end{table}

Note: Sample periods: 1964-2003

corresponds to a very long-term bond, as one would expect. The 0.078 $R^2$ is lower than for bond returns, because of the cash flow and discount rate risks.

Regressions 2 and 3 show that both the macro factor and the forecasting factor predict stock returns with similar predictive power. We may conjecture that the predictable part of average bond returns helps forecast stock returns. Both the forecasting factor and the macro factor are in fact the fitted value of the average excess bond returns. They both capture the similar predictable part of bond returns, that’s why they have similar predictability on stock return.

\section{Conclusion}

We use an estimated DSGE model to interpret the forecasting factor proposed by Cochrane and Piazzesi (2005). Unobserved macro variables are extracted using the Kalman filter. Solutions for bond prices are obtained by combining the conditional log normality of the model implied pricing kernel with the log-linear approximation of macro dynamics.

The main conclusion of this paper is that the technology shock captures most variations in the forecasting factor for 1-year excess returns. As the return horizon increases, the markup shock becomes increasingly important. Following Cochrane
and Piazzesi (2005), we also construct a single macro factor by forming a linear combination of macro variables. By choosing parameters judiciously, we successfully constructed a macro factor that has a higher prediction power than the forecasting factor constructed by Cochrane and Piazzesi. This macro factor also forecasts stock returns.
CHAPTER 3

A Reinvestigation of Long Run Stock Variance

3.1 Introduction

The long run volatility of stock returns is of great interest for investors. Various empirical studies tend to support the view that stock returns are less volatile over longer investment horizons. For example, Siegel (2008) points out that the unconditional variances of stock returns are substantially lower in the long horizon. Campbell and Viceira (2002, 2005) find that the conditional variances of stock returns also decrease with investment horizons. However, recent work by Pastor and Stambaugh (2008) shows that stocks are more volatile over long horizons from the viewpoint of an investor. They decompose the conditional variance of stock returns into five components. Although the mean reversion component contributes negatively to long horizon variance, this contribution is more than offset by the effects of uncertainties faced by the investor. These uncertainties include uncertainty about future expected stock returns, uncertainty about current expected stock returns and parameter uncertainty. They apply the Bayesian approach to a predictive system (Pastor and Stambaugh, (2009)) and find that for quarterly stock returns, the 30-year predictive variance is more than double the 1 year variance.
According to their work, whether stocks are more or less volatile in the long run depends on the relative contribution of each component. If the mean reversion component dominates other components, then stocks will be less risky in the long run. There is a key parameter that controls the magnitude of the mean reversion component, the negative correlation between unexpected returns and innovation in expected returns, $\rho_{uw}$. Pastor and Stambaugh (2008) consider different priors for $\rho_{uw}$, ranging from the noninformative prior, which is flat between $-0.9$ to $0.9$, to more informative prior, which has $99.9\%$ of its mass below $-0.71$, and show that across different priors, the 30-year predictive variance ratios range from $2.02$ to $2.44$. This paper further investigates the role played by $\rho_{uw}$ in determining the variance ratios. We use Cholesky decomposition to decompose the variance covariance matrix of all innovations, $\Sigma = L'L$ and estimate all elements in $L$ using MCMC. Explicit restrictions on $\rho_{uw}$ are made by reducing the number of elements by 1. We investigate different values of $\rho_{uw}$ and obtain different results. For example, the variance ratio is smaller than one when $\rho_{uw} = -0.8$ or $-0.6$, while it is bigger than one when $\rho_{uw}$ reduces to $-0.4$.

We also incorporate the dividend growth rate and present value relation into the predictive system to investigate their effects on variance ratios. There are several reasons to model the dividend growth rate and stock returns together. First, there are many studies modeling returns and the dividend growth rate together. Research along this line includes Binsbergen and Kojien (2008), Cochrane (2008), and Rytchkov (2008). It is interesting to see how dividend growth rate and present value relation will affect the variance ratios. It provides an additional channel to extract information on expected returns from predictors, through the dividend growth rate. Thus
the incorporation of the dividend growth rate may reduce the uncertainty of current expected returns. However, the result is that the dividend growth rate reduces the uncertainty of current expected stock returns only for horizons short enough. Second, the property of the dividend growth rate itself can be investigated through the predictive system. For example, We find that the expected dividend growth rate is time varying and it can predict the future dividend growth rate with an $R^2$ over 30%.

The paper proceeds as follows. In Section 1, We introduce the predictive system proposed by Pastor and Stambaugh (2009) and shows the variance decomposition. In Section 2, We discuss the estimation procedure and the empirical results. In Section 3, We incorporate the dividend growth rate into the model and shows how it affects the variance ratios and Section 4 concludes.

3.2 Predictive System of Pastor and Stambaugh

In this section, We introduce the predictive system proposed by Pastor and Stambaugh (2009). The key equations of the system are summarized as follows:

\begin{align}
 r_{t+1} & = \mu_t^r + U_{t+1}^r \\
 \mu_t^{r} - \mu_r & = \phi_r (\mu_{t+1}^r - \mu_r) + W_{t+1}^r \\
 X_{t+1} - X & = A(X_t - X) + V_{t+1}
\end{align}

where $r_{t+1}$ denotes the log stock returns from time $t$ to $t+1$. It contains two component, $\mu_t^r$, the expected stock returns, which is unobservable at time $t$, and $U_{t+1}^r$, the unexpected return, which has mean zero conditional on all information available at time $t$. The expected stock returns are assumed to follow an $AR(1)$ process with
autoregressive coefficient $\phi_r$. $X_{t+1}$ is a vector of predictors often used to predict stock returns, like dividend yield and yield spread. It also follows an $AR(1)$ process. These equations form a general $VAR$ process which is stationary. The residuals in the system are assumed to distributed identically and independently across $t$ as

$$
\begin{bmatrix}
U_{t+1}^r \\
V_{t+1}^r \\
W_{t+1}^r
\end{bmatrix} \sim N (0, \Sigma) 
$$

(3.2)

The predictive system allows predictors, $X_{t+1}$ to be imperfectly correlated with the expected stock returns, $\mu_t^r$. In this case, the correlation between $U_{t+1}^r$ and $W_{t+1}^r$ plays an important role in estimating expected stock returns. A negative correlation is imposed as a prior in their Bayesian analysis. This negative correlation helps reduce the predictive variance of long horizon stock returns through the mean reversion component. The increase in unexpected stock returns will be offset by the decrease in future expected stock returns. As a result, investment over a long horizon will have lower volatility. Besides, the imperfect predictors add another uncertainty about current expected stock returns, $\mu_T^r$. The predictors deliver imperfect information about $\mu_T^r$, and conditional on all the available data, the variance of $\mu_T^r$ contributes positively to long horizon variance. They show that the uncertainty of current expected stock returns, together with other uncertainties, make stocks riskier in the long run from the perspective of investors.

To be more explicit, they define the $k$-period return from period $T + 1$ through period $T + k$ as

$$
r_{T,T+k} = r_{T+1} + r_{T+2} + \ldots + r_{T+k}
$$

(3.3)
They compute the conditional variance of $r_{T,T+k}$, given all available data up to $T$, $D_T$ and decompose it into five different components:

1) i.i.d. uncertainty
2) mean reversion
3) uncertainty about future expected returns
4) uncertainty about current expected return
5) estimation risk

\[
\begin{align*}
\text{Var}(r_{T,T+k} \mid D_T) &= E \left\{ k \sigma_u^2 \mid D_T \right\} + E \left\{ 2k \sigma_u^2 d \rho_{uw} A(k) \mid D_T \right\} \\
&\quad + E \left\{ k \sigma_u^2 d^2 B(k) \mid D_T \right\} + E \left\{ \left( \frac{1 - \phi_r^k}{1 - \phi_r} \right)^2 q_T \mid D_T \right\} \\
&\quad + \text{Var} \left\{ k E_r + \frac{1 - \phi_r^k}{1 - \phi_r} (b_T - E_r) \mid D_T \right\}
\end{align*}
\]  

(3.4)

where $\sigma_u^2 = \text{var} \left( U_{t+1}^r \right), \rho_{uw} = \rho \left( U_{t+1}^r, W_{t+1}^r \right), b_T = E \left( \mu_T \mid D_T, \theta \right), q_T = \text{var} \left( \mu_T \mid D_T, \theta \right), A(k) = 1 + \frac{1}{k} \left( -1 - \phi_r \frac{1 - \phi_r^k}{1 - \phi_r} \right), B(k) = 1 + \frac{1}{k} \left( -1 - 2 \phi_r \frac{1 - \phi_r^k}{1 - \phi_r} + \phi_r^2 \frac{1 - \phi_r^{2(k-1)}}{1 - \phi_r^2} \right),
\]

\[
d = \left[ \frac{1 + \phi_r}{1 - \phi_r} \frac{R^2}{1 - R^2} \right], \theta \text{ is the set of all parameters and } R^2 \text{ is the ratio of the variance of } \mu_t \text{ to the variance of } r_{t+1}.
\]

The i.i.d. uncertainty is associated with the variance of unexpected stock returns. Because $\{U_{t+i}^r\}$ are independent of each other, this uncertainty is linear in the investment horizon $k$. The mean reversion component comes from the negative correlation between $U_{t+1}^r$ and $W_{t+1}^r$. This component decreases more than linearly in $k$ and contributes negatively to the long run stock variance. Pastor and Stambaugh (2009) also show that future $\mu_{T+i}$ uncertainty, current $\mu_{T+i}$ uncertainty, and estimation increase...
more than linearly in $k$. Overall, in the predictive system, the positive contributions outweigh the negative ones and make stocks riskier in the long run.

3.3 Estimation

In this section, we estimate the model using a different procedure than Pastor and Stambaugh (2009). With the new procedure, quantitative analysis can be made on how different values of $\rho_{uw}$ will have different implications for variance ratios.

3.3.1 Data

The model is estimated using quarterly data from 1952 to 2007. Three predictors are used for estimation as in Pastor and Stambaugh (2009). The first one is market wide dividend yield, which is equal to total dividends paid over the previous 12 months divided by current total market capitalization. Following Pastor and Stambaugh (2009), I compute the dividend yield from the with-dividend and without-dividend monthly returns on the value-weighted portfolio of all New York Stock Exchange (NYSE), Amex, and Nasdaq stocks, which can be obtained from Center for Research in Security Prices (CRSP). The second one is the cointegration residual $cay_t$ of consumption, dividends and labor income. The data series can be obtained from Martin Lettau’s web site. The third predictor is the yield spread, which is defined as the difference between the yield of the 5 year bond and that of the 1 year bond. The yield data can be obtained from the CRSP Fama-Bliss discount bond files. The log dividend growth rate can be constructed from with-dividend and without-dividend quarterly returns on NYSE, Amex and Nasdaq stocks. All variables are deflated by the CPI, which is obtained from the website of Federal Reserve Bank of St Louis.
Figure 3.1 plots the data series for each variable in quarterly frequency. One interesting thing is that quarterly dividend growth rates are highly negatively autocorrelated. This explains why dividend growth rates can be predicted by a negatively correlated data series, i.e. expected dividend growth rates.

### 3.3.2 Estimation Procedure

Pastor and Stambaugh (2009) adopt a Bayesian approach to estimate their model. They obtain posterior distributions using Gibbs sampling, a Markov Chain Monte Carlo (MCMC) technique. In each step of the MCMC chain, they first draw the parameters conditional on the current draw of expected stock returns, \( \{\mu_t^r\} \). Then they use the forward filtering, backward sampling algorithm to draw the time series of \( \{\mu_t^r\} \) conditional on the current draw of parameters. (Details can be found in Pastor and Stambaugh (2009)). The analytical form of posterior distributions can be obtained because of the linear system.
The key prior distribution is the one on $\rho_{uw}$. They use three different priors on $\rho_{uw}$, ranging from noninformative to more informative. The noninformative prior is flat between $-0.9$ and $0.9$. The less informative prior has $99.9\%$ of its mass below zero. The more informative prior has $99.9\%$ of its mass below $-0.71$. They develop a hyperparameter approach to change the prior on $\rho_{uw}$. For all three prior specifications, they find that the variance ratio is larger than 1.

We adopt a different procedure to estimate the model. Instead of putting different priors on $\rho_{uw}$, we fix $\rho_{uw}$ at different values. Because of the nonlinear restrictions on $\Sigma$ introduced by different values of $\rho_{uw}$, an analytical form of the posterior distributions can not be obtained. The MCMC method is used here. We decompose $\Sigma$ into two parts using Cholesky decomposition.

$$\Sigma = L \ast L' \quad (3.5)$$

where $L$ is a lower triangular matrix. The decomposition is unique after fixing the sign of diagonal elements of the $L$, either positive or negative. With this decomposition, We can impose different restrictions on $\Sigma$ through the elements of $L$. We control the elements of $L$ to assign different values to $\rho_{uw}$. With the decomposition, We use MH algorithm to generate the posterior draws for all the parameters based on the likelihood constructed using the Kalman filter. (Details can be found in Appendix D and E). The posterior draws are based on the log likelihood only, and no specific prior distributions are assumed for those parameters, because there is not much prior information about these parameters. It is equivalent to assume that the each prior distribution follows a uniform distribution over $(-\infty, +\infty)$. After I get posterior draws for $L$, $\Sigma$ can be constructed accordingly.
3.3.3 Empirical Results

With the posterior draws, each component in equation (4) can be easily computed, which gives us the conditional variance. The variance ratio for different investment horizon $k$ is defined as

$$VR(k) = \frac{Var(r_{T+1} | D_T)}{k Var(r_{T+1} | D_T)}$$

(3.6)

and the ratios for different values of $\rho_{uw}$ are plotted in Figure 3.1. For $\rho_{uw} = -0.8$, all variance ratios are smaller than 1. For the 30 investment horizons, the ratio is around 0.73, which is very different from Pastor and Stambaugh (2009) who argue that this ratio is more than double. The results change a little when $\rho_{uw}$ is set to be $-0.6$, which is indicated by the black line in figure 3.1. Compared to previous case, the variance ratio under $\rho_{uw} = -0.6$ increases for all horizons and is lightly larger than 1 for the 30 year horizon. When $\rho_{uw}$ further increases to $-0.4$, different results stand out. The variance ratio keeps increasing with horizons and reach 8.17 for the 30 year horizon. It is consistent with Pastor and Stambaugh (2009), but with an even larger variance ratio.

Figure 3.2 plots each component of the conditional variance for $\rho_{uw} = -0.8$ and $-0.6$ as in (3.4). Each component is deflated by investment horizon $k$ to get the per-period value. This plot shows the importance of the mean reversion component in determining the variance ratio. Although the components of uncertainties about current and future expected returns decrease as $\rho_{uw}$ increase from -0.8 to -0.6, these effects are more than offset by the increase in the mean reversion component. As a result, the variance ratios increase as $\rho_{uw}$ increase from -0.8 to -0.6. The mean reversion component is critical to determine the variance ratios.
Figure 3.1: Variance Ratio for Different Values of $\rho_{uw}$

Figure 3.2: Comparison of Each Component for Different Values of $\rho_{uw}$
3.3.4 Plausibility of Large Negative Value of $\rho_{uw}$

From the above analysis, we can see that the value of $\rho_{uw}$ plays an important role in determining the variance ratios. If $\rho_{uw}$ is sufficiently negative, a variance ratio smaller than 1 can be obtained for a 30 year investment horizon. Is it plausible to assume a large negative value for $\rho_{uw}$? Pastor and Stambaugh (2009) justify the plausibility of their more informative prior by listing various empirical evidence: "... the evidence of Campbell (1991) implies estimates of $\rho_{uw}^2$ ranging from 0.5 to 0.74 across three different specifications in his full sample period 1927 to 1988. The estimates of $\rho_{uw}^2$ implied by the post-war evidence range from 0.84 to 0.88 in 1952 to 1988, and the estimates of Campbell and Ammer (1993) in their Table III range from 0.86 to 0.91." (See Pastor and Stambaugh (2009), P1606). However, the large negative value is just assumed for the prior distribution in their paper. If the prior is not informative enough to determine the posterior distribution, smaller negative values of $\rho_{uw}$ may result, which fall into the range that delivers a higher variance in the long run. For example, if the prior on $\Sigma$ is assumed to follow an inverted Wishart distribution with a small number of degrees of freedom $\nu = 7$, $\Sigma^{-1}IW(\Sigma_0, \nu)$, and $\rho_{uw}$ is restricted to be -0.8 in $\Sigma_0$, then the posterior mean of $\rho_{uw}$ is only -0.1, which is not negative enough to generate a variance ratio smaller than 1 for long investment horizons.

To get more information about the posterior distribution of $\rho_{uw}$, we estimate $\rho_{uw}$ together with other parameters using MCMC. Unlike other parameters that are assumed to follow uniform distributions over $(-\infty, +\infty)$, we assume a sharp prior for $\rho_{uw}$, that is, $\rho_{uw} \sim N(-0.8, 0.001)$. The prior belief holds that $\rho_{uw}$ will be centering around -0.8. The estimates of parameters are listed in Table 3.1.
<table>
<thead>
<tr>
<th>para</th>
<th>mean</th>
<th>std</th>
<th>para</th>
<th>mean</th>
<th>std</th>
<th>para</th>
<th>mean</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_r$</td>
<td>0.335</td>
<td>0.168</td>
<td>$l_{51}$</td>
<td>0.001</td>
<td>0.001</td>
<td>$a_{23}$</td>
<td>-0.080</td>
<td>0.133</td>
</tr>
<tr>
<td>$l_{11}$</td>
<td>0.081</td>
<td>0.008</td>
<td>$l_{52}$</td>
<td>0.002</td>
<td>0.002</td>
<td>$a_{31}$</td>
<td>-0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>$l_{22}$</td>
<td>0.029</td>
<td>0.012</td>
<td>$l_{53}$</td>
<td>0.002</td>
<td>0.002</td>
<td>$a_{32}$</td>
<td>0.0145</td>
<td>0.049</td>
</tr>
<tr>
<td>$l_{31}$</td>
<td>-0.026</td>
<td>0.006</td>
<td>$l_{54}$</td>
<td>0.004</td>
<td>0.004</td>
<td>$a_{33}$</td>
<td>0.838</td>
<td>0.045</td>
</tr>
<tr>
<td>$l_{32}$</td>
<td>0.002</td>
<td>0.012</td>
<td>$l_{55}$</td>
<td>0.008</td>
<td>0.006</td>
<td>$E_r$</td>
<td>0.016</td>
<td>0.007</td>
</tr>
<tr>
<td>$l_{33}$</td>
<td>0.029</td>
<td>0.006</td>
<td>$a_{11}$</td>
<td>0.987</td>
<td>0.015</td>
<td>$X_1$</td>
<td>-4.607</td>
<td>0.825</td>
</tr>
<tr>
<td>$l_{41}$</td>
<td>-0.004</td>
<td>0.001</td>
<td>$a_{12}$</td>
<td>-0.188</td>
<td>0.200</td>
<td>$X_2$</td>
<td>0.026</td>
<td>0.035</td>
</tr>
<tr>
<td>$l_{42}$</td>
<td>0.002</td>
<td>0.005</td>
<td>$a_{13}$</td>
<td>-0.089</td>
<td>0.169</td>
<td>$X_3$</td>
<td>0.061</td>
<td>0.043</td>
</tr>
<tr>
<td>$l_{43}$</td>
<td>-0.003</td>
<td>0.002</td>
<td>$a_{21}$</td>
<td>-0.003</td>
<td>0.006</td>
<td>$\rho_{uw}$</td>
<td>-0.765</td>
<td>0.051</td>
</tr>
<tr>
<td>$l_{44}$</td>
<td>0.002</td>
<td>0.004</td>
<td>$a_{22}$</td>
<td>0.996</td>
<td>0.084</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Table 3.1: Posterior Distributions |

We recompute the variance ratios and plot them in Figure 3.3. The variance ratio decreases from 1 to around 0.52 for $k = 13$ and then increases to 0.989 for $k = 120$. Long run returns are less risky than short run returns. This result is largely based on the posterior mean of $\rho_{uw}$, -0.765, which is sufficiently negative to generate a variance ratio smaller than 1. This exercise serves as empirical evidence to support a large negative $\rho_{uw}$.

### 3.4 Incorporating Dividend Growth Rate

In this section, we extend the PS model by incorporating the dividend growth rate into the predictive system and explore the present value relation in the new system. Following Rytchkov (2008) and Binsbergen and Koijen (2008), we assume that log dividend growth rate follows a similar process as stock returns. It contains the expected dividend growth rate which follows an $AR(1)$ process, and the unexpected dividend growth rate, which has mean zero conditional on all information available at time $t$. The extended model can be written as follows:
Two new equations are added to the predictive system, the equation of log dividend growth rate and the expected dividend growth rate. Now the residuals of the new system are assumed to distributed identically and independently across $t$ as

\[
\begin{align*}
\Delta d_{t+1} &= \mu_t^d + U_{t+1}^d \\
\rho_{t+1} - \mu_r &= \phi_r (\mu_{t+1}^r - \mu_r) + W_{t+1}^r \\
\rho_{t+1}^d - \mu_d &= \phi_d (\mu_{t+1}^d - \mu_d) + W_{t+1}^d \\
X_{t+1} - X &= A(X_t - X) + V_{t+1}
\end{align*}
\]
Figure 3.4: Conditional Variances for Different Values of $\rho_{uw}$: Extended Model.

\[
\begin{bmatrix}
U^r_{t+1} \\
U^d_{t+1} \\
V_{t+1} \\
W^r_{t+1} \\
W^d_{t+1}
\end{bmatrix}
\sim N(0, \Sigma^*)
\]  

(3.8)

Restrictions are implied by the present value relation on $\Sigma^*$. Rytchokov (2008) shows that the present value relation implies the following decomposition for $W^d_{t+1}$

\[
W^d_{t+1} = \frac{1 - \rho \phi_d}{\rho} \left( U^r_{t+1} + \frac{\rho}{1 - \rho \phi_r} W^r_{t+1} - U^d_{t+1} \right)
\]  

(3.9)

where $\rho$ is a constant smaller than 1. In this paper it is fixed at 0.9865.

Figure 3.4 plots the conditional variances for different values of $\rho_{uw}$ obtained from the extended system, and similar results can be drawn here. The conditional variance of 30 year stock returns is less than the variance of 1 year return for $\rho_{uw} = -0.8$. It keeps increasing until exceeding the variance of 1 year stock return as $\rho_{uw}$ increases.
To further investigate the difference for both models, Figure 3.5 plots each component of the conditional variance for two models with $\rho_{uw} = -0.8$ as in (3.4). Each component is deflated by investment horizon $k$ to get the per-period value. Several conclusions can be drawn here. For $k < 120$ (i.e. 30 years), 1) the PS model delivers a more negative mean reversion component; 2) the inclusion of the dividend growth rate helps reduce the uncertainty of future expected stock returns; 3) for uncertainty of current expected stock returns, the model with dividend growth rate delivers a smaller value in short horizon while a larger value in long horizons; 4) larger estimation risk results with the dividend growth rate due to more parameters to be estimated. Overall, the model with the dividend growth rate cannot reduce the volatility of stock returns in the long run.
The last finding of this paper is that when incorporated into the predictive system, the dividend growth rate becomes highly predictable by the expected dividend growth rate, with an average $R^2$ of 36%. This finding is consistent with Rytchkov (2008) and Binsbergen and Koijen (2008).

3.5 Conclusion

Pastor and Stambaugh (2009) found that stock returns are riskier in the long run. This paper reinvestigates this issue. As in Pastor and Stambaugh (2009), we use a Bayesian approach to assess the risk. We find that their conclusion is likely to be sensitive to the prior of the correlation between innovation in expected returns and unexpected returns. The parameter plays a key role in determining the riskiness of stock returns in the long run through the mean reverting component. My analysis suggests that their result depends critically on a prior that is sufficiently uninformative. If the prior of a highly negative correlation is sufficiently informative, the result would be overturned. We also find that their conclusion is robust to the addition of dividend growth into the prediction system.

By estimating $\rho_{uw}$ with a sharp prior distribution, we show that the posterior draws of $\rho_{uw}$ are sufficiently negative to generate a variance ratio smaller than 1 for 30 year stock returns.
BIBLIOGRAPHY


APPENDIX A

Derivation of the Stochastic Discount Factor for Step 1 Model

Following Hansen and Sargent (2008), the stochastic discount factor is derived using the distribution of $e_{t+1}$ under the approximating model and its distorted distribution under the distorted model.

Under the approximating model, $e_{t+1} \sim N(0, I)$, while under the distorted model, $e_{t+1} \sim N(\beta (1 - \gamma_1) (\lambda'C + \phi_c G) , I)$.

The ratio of the distorted distribution and the approximating distribution is

\[
R = \frac{\exp - \frac{1}{2} (e_{t+1} - \beta (1 - \gamma_1) (\lambda'C + \phi_c G))' I^{-1} (e_{t+1} - \beta (1 - \gamma_1) (\lambda'C + \phi_c G))}{\exp - \frac{1}{2} (e_{t+1} - 0)' I^{-1} (e_{t+1} - 0)}
\]

\[
= -\frac{1}{2} \beta^2 (1 - \gamma_1)^2 (\lambda'C + \phi_c G)' (\lambda'C + \phi_c G) + \beta (1 - \gamma_1) (\lambda'C + \phi_c G) e_{t+1}
\]

This ratio serves as an adjustment to the traditional stochastic discount factor, and the new discount factor takes the form,

\[
m_{t+1} = \log \beta - \Delta e_{t+1} - \frac{1}{2} \beta^2 (1 - \gamma_1)^2 (\lambda'C + \phi_c G)' (\lambda'C + \phi_c G) + \beta (1 - \gamma_1) (\lambda'C + \phi_c G) e_{t+1}
\]
APPENDIX B

Derivation of the Stochastic Discount Factor for Step 2 Model

Following Hansen and Sargent (2008), the stochastic discount factor is derived using the distribution of \[
\begin{bmatrix}
\Delta c_{t+1} \\
\pi_{t+1}
\end{bmatrix}
\]
under the approximating model and its distorted distribution under the distorted model.

Under the approximating model, \[
\begin{bmatrix}
\Delta c_{t+1} \\
\pi_{t+1}
\end{bmatrix}
\sim N(D\zeta_t + \beta (1 - \gamma_1) G (C' \lambda + G' \phi') + (1 - \gamma_2) D\Sigma_t \lambda, \Sigma_t),
\]
while under the distorted model, \[
\begin{bmatrix}
\Delta c_{t+1} \\
\pi_{t+1}
\end{bmatrix}
\sim N(D\tilde{\zeta}_t, \Sigma_t).
\]

The ratio of the distorted distribution and the approximating distribution is

\[
R = \exp\left\{ -\frac{1}{2} \left( \begin{bmatrix}
\Delta c_{t+1} \\
\pi_{t+1}
\end{bmatrix} - D\zeta_t - \beta (1 - \gamma_1) (\lambda C + \phi, G) - (1 - \gamma_2) D\Sigma_t \lambda \right)' \Sigma_t^{-1} \left( \begin{bmatrix}
\Delta c_{t+1} \\
\pi_{t+1}
\end{bmatrix} - D\zeta_t - \beta (1 - \gamma_1) (\lambda C + \phi, G) - (1 - \gamma_2) D\Sigma_t \lambda \right) \right\}
\]

\[
\exp -\frac{1}{2} \left( \begin{bmatrix}
\Delta c_{t+1} \\
\pi_{t+1}
\end{bmatrix} - D\zeta_t \right)' \Sigma_t^{-1} \left( \begin{bmatrix}
\Delta c_{t+1} \\
\pi_{t+1}
\end{bmatrix} - D\zeta_t \right)
\]

This ratio serves as an adjustment to the traditional stochastic discount factor, and the new discount factor takes the form
\begin{align*}
  m_{t+1} \\
  = \log \beta - \Delta c_{t+1} \\
  - \frac{1}{2} [\beta (1 - \gamma_1) G (C' \lambda + G' \phi'_C) + (1 - \gamma_2) D \Sigma_t \lambda]' (G G' + D \Sigma_t D')^{-1} \\
  \times [\beta (1 - \gamma_1) G (C' \lambda + G' \phi'_C) + (1 - \gamma_2) D \Sigma_t \lambda] \\
  + [\beta (1 - \gamma_1) G (C'' \lambda + G'' \phi''_C) + (1 - \gamma_2) D \Sigma_t \lambda]' (G G' + D \Sigma_t D')^{-1} \\
  (D \varsigma_t + G \epsilon_{t+1} - D \zeta_t)
\end{align*}
APPENDIX C

Derivation of the Stochastic Discount Factor for Step 3 Model

The stochastic discount factor is derived under the continuous time model. Hansen and Sargent (2008) show that in the continuous time limit, for both the distorted and approximating models, \( \begin{bmatrix} \Delta c_{t+1} \\ \Delta \pi_{t+1} \end{bmatrix} \) is normally distributed. With the approximation that \( G(0) = G(1) = G \), the conditional mean of the fundamentals is

\[
\tilde{\mu}_{s,t} = \tilde{p}_t D(0) \tilde{\zeta}_t (0) + (1 - \tilde{p}_t) [D(1) \tilde{\zeta}_t (1) - \phi' \pi_t]
\]

under the approximating model and

\[
\tilde{\mu}_{s,t} = \tilde{p}_t [D(0) \tilde{\zeta}_t (0) + Gw(0) + D(0) u_t (0)] + (1 - \tilde{p}_t) [D(1) \tilde{\zeta}_t (1) - \phi' \pi_t + Gw(1) + D(1) u_t (1)]
\]

under the distorted model, while the variance is \( G \) for both models.

To sum up, under the approximating model, \( \begin{bmatrix} \Delta c_{t+1} \\ \Delta \pi_{t+1} \end{bmatrix} \sim N(\tilde{\mu}_{s,t}, G) \), while under the distorted model, \( \begin{bmatrix} \Delta c_{t+1} \\ \Delta \pi_{t+1} \end{bmatrix} \sim N(\bar{\mu}_{s,t}, G) \).

The ratio of the distorted distribution and the approximating distribution is
This ratio serves as an adjustment to the traditional stochastic discount factor, and the new discount factor takes the form

\[
m_{t+1} = \log \beta - \Delta c_{t+1} - (s_{t+1} - \tilde{\mu}_{s,t})' (G G')^{-1} (\tilde{\mu}_{s,t} - \tilde{\mu}_{s,t})
\]

\[
-0.5 (\tilde{\mu}_{s,t} - \tilde{\mu}_{s,t})' (G G')^{-1} (\tilde{\mu}_{s,t} - \tilde{\mu}_{s,t})
\]
APPENDIX D

Construction of Log Likelihood

The log likelihood function is constructed using standard Kalman Filter proposed by Hamilton (1994, Chapter 13).

Let \( y_t = [r_t, \Delta d_t, X_t - X] \), \( \zeta_t = [\mu_{t-1} - \mu_r, \mu_{t-1} - \mu_d, U_t^r, W_t^r, V_t^1, V_t^2, V_t^3, U_t^d] \) and \( X_t \) is the vector of predictors at time \( t \). Note, that \( \zeta_t \) is not observable at time \( t \).

\[
\zeta_{t+1} = F \zeta_t + \Gamma e_{t+1}
\]

where

\[
F = \begin{pmatrix}
\phi_r & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & \phi_d & \frac{1 - \rho \phi_d}{\rho} & 1 - \rho \phi_d & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\[
\Gamma = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\]

\[
e_{t+1} = \begin{pmatrix}
U_t^r \\
W_t^r \\
V_t^1 \\
V_t^2 \\
V_t^3 \\
U_t^d
\end{pmatrix} \sim N(0, \Omega)
\]
\[ \Omega = L \times L', \quad L = \begin{pmatrix} l_{11} & l_{21} & l_{31} & l_{41} & l_{51} & l_{61} \\ l_{21} & l_{22} & l_{32} & l_{42} & l_{52} & l_{62} \\ l_{31} & l_{32} & l_{33} & l_{43} & l_{53} & l_{63} \\ l_{41} & l_{42} & l_{43} & l_{44} & l_{54} & l_{64} \\ l_{51} & l_{52} & l_{53} & l_{54} & l_{55} & l_{65} \\ l_{61} & l_{62} & l_{63} & l_{64} & l_{65} & l_{66} \end{pmatrix} \]

The observables \( y_t \) are

\[ y_t = B' (X_{t-1} - X) + H' \zeta_t \]

where \( H' = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad B' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \]

Let \( \theta \) denote the vector of parameters to be estimated. \( \theta = [\phi_r, \phi_d, l_{11}, l_{21}, l_{22}, \\ l_{31}, l_{32}, l_{33}, l_{41}, l_{42}, l_{43}, l_{44}, l_{51}, l_{52}, l_{53}, l_{54}, l_{55}, l_{61}, l_{62}, l_{63}, l_{64}, l_{65}, l_{66}, a_{11}, a_{12}, a_{13}, a_{21}, \\ a_{22}, a_{23}, a_{31}, a_{32}, a_{33}, \mu_r, \mu_d, X] \).

The restriction \( \rho (U_{t+1}^r, W_{t+1}^r) = -0.8 \) implies that \( l_{21} = -\frac{4}{3} \times |l_{22}|. \)

The \( VAR \) system has been written in the state space representation and the Kalman Filter can be used to construct the log likelihood function. (Details can be found in Hamilton (1994), Chapter 13).
APPENDIX E

Posterior Simulation

After we obtain the likelihood function, we use the MCMC method to generate random draws from the posterior distribution, as the analytical form of the density is not available.

Step 1 (Selection of Starting Point): We use MLE estimates as the starting points $\theta^*$. 

Step 2 (Proposal): Starting from $\theta^*$, we generate a new draw by the following random walk proposal density. The scaling matrix $c$ is chosen by multiplying a small positive real number to the prior covariance matrix. I use the covariance matrix from MLE as the prior covariance matrix.

$$\theta^{j+1} = \theta^j + cN(0, I), \ j = 0, \ldots, R - 1$$

Step 3 (Accept/Reject): Accept the new draws if

$$\min\left(\frac{p(\theta^{j+1} \mid D_T)}{p(\theta^j \mid D_T)}, 1\right) > u,$$

where $u \rightarrow U(0, 1)$. 

Step 4 (Burn In): Discard the initial $n$ draws.