Title
Essays on the predictability and volatility of returns in the stock market

Permalink
https://escholarship.org/uc/item/41k6k7h9

Author
Wu, Ruojun

Publication Date
2008

Peer reviewed|Thesis/dissertation
Essays on the Predictability and Volatility of Returns in Stock Market

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Economics by Ruojun Wu

Committee in charge:
Professor Allan Timmermann, Chair
Professor Graham Elliot
Professor James Hamilton
Professor Jun Liu
Professor Rossen Valkanov

2008
The dissertation of Ruojun Wu is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2008
DEDICATION

To my parents and my husband.
## TABLE OF CONTENTS

Signature Page ........................................ iii  
Dedication ............................................. iv  
Table of Contents ...................................... v  
List of Figures ......................................... vii  
List of Tables .......................................... viii  
Acknowledgements ...................................... ix  
Vita and Publications .................................. x  
Abstract ............................................... xi  

**Chapter 1**  
Stock Return Variability, Forecast Revisions, and Investors’ Learning  
1.1 Introduction ........................................ 1  
1.2 The Modelling Framework ........................... 5  
1.2.1 Expected returns and unexpected returns ....... 6  
1.2.2 Gross returns and excess returns ............... 7  
1.2.3 The Variance Decomposition ..................... 8  
1.2.4 The VAR Approach and perfect information ... 9  
1.2.5 Empirical Results: Perfect Information ....... 11  
1.3 Parameter Uncertainty and Learning  ............... 15  
1.3.1 Why Learning? .................................. 16  
1.3.2 Learning in Real-time .......................... 17  
1.3.3 A Naive Learning Scheme ....................... 20  
1.3.4 A Sophisticated Learning Scheme: ............... 22  
1.4 Modeling Bayesian VAR ............................ 24  
1.4.1 Independent Normal-Wishart prior: .............. 24  
1.4.2 Prior on Stationarity: .......................... 26  
1.4.3 Posterior belief .................................. 28  
1.4.4 Posterior Evaluation ............................ 29  
1.4.5 Prior elicitation .................................. 31  
1.4.6 Empirical Results: Sophisticated Learning with BVAR ............................... 32  
1.5 Conclusion ......................................... 33  
1.6 Tables and Figures ................................. 38
Chapter 2 On the Robustness of Variance Decompositions ........................................ 49
  2.1 Introduction ......................................................................................... 49
  2.2 A Brief on Variance Decomposition .................................................... 51
  2.3 Persistent Predictors and the Sensitivity of Variance Decomposition .............................. 54
    2.3.1 Persistent Predictors in Stock Return Prediction .............................. 55
    2.3.2 Experiments with Real Data .......................................................... 56
    2.3.3 Simulated Confidence Intervals ....................................................... 58
  2.4 Conclusion .......................................................................................... 60
  2.5 Tables and Figures .............................................................................. 62

Chapter 3 Return Predictability Under Equilibrium Constraints on the Equity Premium Returns ............................................................ 71
  3.1 Introduction .......................................................................................... 71
  3.2 Methodology ........................................................................................ 77
    3.2.1 Constraints on the Return Forecasting Model .................................. 77
    3.2.2 Accounting For Constraints through Investors’ Prior Beliefs .................. 80
    3.2.3 Choice of Priors ............................................................................ 84
    3.2.4 Posterior Distributions ................................................................... 86
    3.2.5 The Gibbs Sampler ....................................................................... 89
  3.3 Empirical Results ................................................................................ 91
    3.3.1 Data ............................................................................................... 91
    3.3.2 Effect of Constraints on coefficient estimates .................................... 92
  3.4 Out-of-sample Forecasts of Stock Returns ............................................. 96
    3.4.1 Evaluation of Forecasts .................................................................. 99
  3.5 Stambaugh Bias ................................................................................. 102
    3.5.1 Simulation Results ....................................................................... 105
    3.5.2 Forecasts Based on Valuation Ratios: Results from Longer Samples ................. 107
    3.5.3 Parameter Estimation Error ......................................................... 108
  3.6 Conclusion ......................................................................................... 109
  3.7 Acknowledgement .............................................................................. 110
  3.8 Tables and Figures: .......................................................................... 111

Bibliography ................................................................................................. 127

Appendix A Log-linearized Present Value Model ........................................... 133

Appendix B Data Summary .......................................................................... 136
LIST OF FIGURES

Figure 1.1: OLS In-sample, Unsmoothed ........................................ 43
Figure 1.2: OLS In-sample, Smoothed ........................................... 44
Figure 1.3: Naive Learning with Expanding Window ....................... 45
Figure 1.4: Naive Learning with Rolling Window .............................. 46
Figure 1.5: Maximum Eigenvalues .................................................. 47
Figure 1.6: Sophisticated Learning .................................................. 48

Figure 2.1: Persistent State Variables vs. Variance Decomposition .... 68
Figure 2.2: Persistent State Variables vs. Variance Decomposition – R2
adjusted .......................................................... 69
Figure 2.3: Histogram of Variance Decomposition Ratios based on Simulation 70

Figure 3.1: Posterior Distribution of Slope Coefficients, Part I .......... 120
Figure 3.2: Posterior Distribution of Slope Coefficients, Part II ......... 121
Figure 3.3: Posterior Distribution of Slope Coefficients, Part III ........ 122
Figure 3.4: Posterior Distribution of Slope Coefficients, Part IV ........ 123
Figure 3.5: Out-of-sample Forecasts at 1-month frequency ............... 124
Figure 3.6: The Effect of Constraints on Regression Parameters......... 125
Figure 3.7: Out-of-sample Forecasts with different frequency .......... 126
LIST OF TABLES

| Table 1.1: | Descriptive Statistics of the VAR Variables | 38 |
| Table 1.2: | VAR parameter estimates by OLS in-sample | 39 |
| Table 1.3: | Variance Decomposition with Perfect Information | 40 |
| Table 1.4: | Variance Decomposition with Naive Learning | 41 |
| Table 1.5: | Variance Decomposition with Sophisticated Learning | 42 |
| Table 2.1: | Conventional Wisdom on Variance Decomposition | 63 |
| Table 2.2: | Summary Statistics on Persistent State Variables | 64 |
| Table 2.3: | Summary Statistics on Benchmark case | 65 |
| Table 2.4: | Sensitivity of Variance Decomposition on Variable Persistency | 66 |
| Table 2.5: | Confidence Interval of Variance Decomposition | 67 |
| Table 3.1: | Full-sample estimates of slope coefficients for the individual return forecasting models: Part 1 | 112 |
| Table 3.2: | Full-sample estimates of slope coefficients for the individual return forecasting models: Part 2 | 113 |
| Table 3.3: | Forecasting performance of individual return models: Part 1 | 114 |
| Table 3.4: | Forecasting performance of individual return models: Part 2 | 115 |
| Table 3.5: | Estimates of slope coefficient and forecast performance for the dividend yield model, Part I | 116 |
| Table 3.6: | Estimates of slope coefficient and forecast performance for the dividend yield model, Part II | 117 |
| Table 3.7: | Out-of-sample R2 of individual return forecasting models | 118 |
| Table 3.8: | Compare Model C and Model D | 119 |
ACKNOWLEDGEMENTS

I am indebted to Allan Timmermann for his generous advice and constant encouragement. I specially thank Graham Elliot, James Hamilton, Jun Liu, and Rossen Valkanov for their insights and patience. Helpful comments from participants at the Econometrics workshops, Applied Economics Lunch seminars at Economics Department, and Finance seminars at Rady School of Management, UC-San Diego, the Federal Reserve Bank – St. Louis, Hong Kong University of Science and Technology, and University of Hong Kong are also gratefully acknowledged. All remaining errors are mine.

Chapter 3, in full, is a reprint of the material as it appears in the Working paper by Davide Pettenuzzo, Allan Timmermann, Rossen Valkanov, and Rosalin Wu. The dissertation author was one of the primary authors of this paper.
VITA

2001  Bachelor of Arts, International Finance, Fudan University, China

2002  Master of Arts, Economics
       University of California, Davis

2006  Master of Science, Statistics
       University of California, San Diego.

2008  Doctor of Philosophy, Economics
       University of California, San Diego.
ABSTRACT OF THE DISSERTATION

Essays on the Predictability and Volatility of Returns in Stock Market

by

Ruojun Wu

Doctor of Philosophy in Economics

University of California San Diego, 2008

Professor Allan Timmermann, Chair

This dissertation studies the effect of parameter uncertainty on the return predictability and volatility of the stock market. The first two chapters focus on the decomposition of market volatility, and the third chapter studies the return predictability.

When facing imperfect information, the investors tend to form a learning scheme that encompasses both historical data and prior beliefs. In the variance decomposition framework, the introducing of learning directly impacts the way that return forecasts are revised and consequently the relative component of market volatility based on these forecasts, namely the price movements from revision on future discount rates and those from future cash flows. According to the empirical study in Chapter 1, the former is not necessarily the major driving force of market volatility, which provides an alternative view on what moves stock prices. Learning
is modeled and estimated by Bayesian method. Chapter 2 follows the topic in Chapter 1 and studies the role of persistent state variables in return decomposition in order to provide more robust inference on variance decomposition. In Chapter 3 we propose to utilize theoretical constraints to help predict market returns when in sample data is very noisy and creates model uncertainty for the investors. The constraints are also incorporated by Bayesian method. We show in the out-of-sample forecast experiment that models with theoretical constraints produce better forecasts.
Chapter 1

Stock Return Variability, Forecast Revisions, and Investors’ Learning

1.1 Introduction

Understanding what moves stock prices is a key question in finance in both interpretation and forecasting. The Present Value model is one of the simple models that addresses this issue: movements in stock prices, and thus returns, are due to investors’ change in expectations on future risk premium (the "discount rate news" thereafter), and their change in expectations on future dividends (the "cash flow news" thereafter). Investors’ expectations are not directly observed, yet they could be estimated from the observed return series with assumptions on the dynamics in expected returns and cash flows.

Campbell and Shiller (1988) and others log-linearizes PV model, and use
estimates from full-sample vector autoregression (VAR) with constant parameters to simulate investors’ expectations and construct measures of the discount rate news and cash flow news based on the estimated expectations. Their results suggest that roughly 2/3 of variation in stock returns comes from discount rate news, and 1/3 from cash flow news. The weakness of the above analysis is that it implicitly assumes investors have perfect knowledge about the underlying return process and thus does not account for model and parameter uncertainty. However, in real time, investors don’t have perfect information on the true parameters and must learn about the unknown process with certain learning scheme.

Uncertainty about return process could play a big role for the relative importance of cash-flow news and discount rate news. There is a simple explanation. When investors form rational expectation under full certainty, a positive (negative) dividend innovation will lead to a proportional upward (downward) adjustment in prices. On the other hand, if investors form forecasts based on estimates of the underlying model, the same innovation will have an additional effect on stock prices under the assumption that the (in-sample) dividend growth estimate applies to future dividend payments. Investors are more optimistic (pessimistic) on the future since they revise their growth rate estimates upwards (downwards) based on changes in the estimated parameter values. The latter effect could even dominate the former one depending on the forecasting horizon.

This paper relaxed the full information assumption and focuses on the effect of learning on investors’ forecast revisions and its indirect effect on the variance
components of stock returns. Specifically, I examines two learning schemes. The first is a naive learning scheme where investors rely merely on historical data. This type of learning is represented by real-time OLS estimates of the VAR. The second learning scheme studied is more sophisticated where investors incorporate prior beliefs with observed data by the Bayes rule. Accordingly I propose to use Bayesian VAR to simulate investors’ forecast revisions from period to period and back out the variance components for this type of learning.

Bayesian method gives a natural way to model investors’ real-time decision making when learning problems are hard and they face updated signals every period. The key feature of this approach is to summarize the uncertainty of parameters by a prior distribution. The foundation for the Bayesian approach was provided by Savage (1954). It is also a well established method in finance: Kandel and Stambaugh (1996), Brennan (1998), Barberis (2000) show that parameter uncertainty can affect significantly investors’ portfolio choice. Pastor (2000) and Pastor and Stambaugh (2000), in addition to parameter uncertainty, also consider model uncertainty. With this new approach, I address the following questions: What priors would investors have to make the observed return series consistent with the VAR process for cash flow and discount rate? Moreover, given these priors, what do they say about the relative importance of cash flow and discount rate news?

Statistically, Bayesian methods also have the advantage of providing more accurate estimates and inferences. As discussed in Stambaugh (1999), classical
OLS estimates of the parameters can be biased. Bayesian techniques instead provide us with exact small sample distributions for the decompositions, which are highly nonlinear functions of the parameters of the VARs. Second, statistical tests indicate that some of the variables in the VAR system are borderline non-stationary even though we have theoretical reasons for believing them to be stationary. Bayesian methods instead allow such theoretical restrictions to be imposed through the prior.

In the empirical part, I use U.S. aggregate stock market data to study the effect of learning. Results of investors under full information and under parameter uncertainty with different learning schemes are presented. As expected, learning increases the variation of investors’ forecasts from time to time. Higher variations of forecast revisions are summed up to higher variance components of stock returns. Moreover, as opposed to the conventional wisdom, cash flow news contributes more than discount rate news.

The results of this paper could be a significant contribution to literature. Not only do they exploit the learning effect in the stock return variance literature, they also alert the common usage of in-sample VAR in variance decomposition. In-sample variance decompostion with a VAR is in common use among economists. Some of the more recent applications and extensions include: Campbell and Vuolteenaho (AER 2004) apply the method to cross sectional stock return volatilities and construct a "cash-flow" beta to explain cross-sectional variation in average returns. Lustig and Van Nieuwerburgh (2006) apply the same framework
to the decomposition of innovations to consumption into returns on human wealth and returns on financial assets, and they conclude that these two components are negatively correlated. Bernanke and Kuttner (2005) explain why (and how) stock prices respond as they do to monetary policy by extending the system and relate the proxies for expectations to the news about the path of monetary policy embodied in the surprises derived from Federal funds futures. They conclude that the reaction of equity prices to monetary policy is, for the most part, not directly attributable to policy effect on the real interest rate. Instead, it comes through its effect on expected future excess returns and expected future dividends.

The plan of the paper is as follows. The next section provides the basic framework, including both the asset pricing framework and statistical setup for a general variance decomposition through VAR. As a baseline case, empirical results based on full information OLS estimation are also presented in Section 2. Section 3 introduces learning in real-time to the basic framework. Different learning schemes are discussed. Section 4 focuses on the Bayesian learning and its effect on stock return movements. Formal Bayesian set-ups are introduced, followed with estimation methods and empirical results. Section 5 concludes.

1.2 The Modelling Framework

In this section I first introduce the log-linearized present value model of Campbell and Shiller (1988a) and Campbell (1991) to express unexpected stock
returns as a function of forecast revisions about future dividend growth rates and stock returns. I then explain how to use VAR systems to obtain empirical proxies for the forecast revisions that leads to stock return movements ex post. Empirical results are presented when investors have perfect information, as a benchmark case of this paper.

1.2.1 Expected returns and unexpected returns

The basic equation for stock returns relates the unexpected stock return in period \( t + 1 \) to revisions in investors’ forecasts. I write \( r_{t+1} \) for the log return on a stock held from the end of period \( t \) to the end of period \( t + 1 \), and write \( d_{t+1} \) for the log dividend paid during period \( t + 1 \). Using log linearization developed by Campbell and Shiller (1988), I can express the period \( t + 1 \) unexpected return on equity in terms of the revision of the expectation of discounted future dividends and future returns:

\[
r_{t+1} - E_t r_{t+1} = (E_{t+1} - E_t)[\sum_{i=0}^{\infty} \rho^i \Delta d_{t+1+i} - \sum_{i=1}^{\infty} \rho^i r_{t+1+i}] \tag{1.1}
\]

Here \( E_t \) denotes an expectation formed at the end of period \( t \), conditional on an information set that includes the history of stock prices and dividends at the least, while \( \Delta \) denotes a one-period backward difference. The parameter \( \rho \) comes out of the log-linear approximation procedure; it is a number a little bit smaller than one. Following Campbell and Ammer (1993), it is set to 0.9962.
Equation (1.1) is the basic equation used in this paper. It is best thought of as a consistency condition for expectations. If the unexpected stock return is positive, then either expected future dividend growth must have become higher, meaning higher future cash flows, or expected future stock returns must be lower, meaning lower future discount rates or both. Formally, equation (1.1) is derived by taking a first-order Taylor approximation of the equation relating the log stock return to log stock returns and dividends, and solving the approximate equation forward with proper terminal condition imposed. (A sketch of the derivation can be found in Appendix I.)

1.2.2 Gross returns and excess returns

Sometimes it is more natural to work with excess stock returns over some short-term interest rate. Define the excess return as the difference between the gross return and the short-term interest rate, $\psi_{t+1} \equiv r_{t+1} - r_{f,t+1}^f$. Equation (1.1) can be further decomposed as

$$\psi_{t+1} - E_t\psi_{t+1} = (E_{t+1} - E_t)[\sum_{i=0}^{\infty} \rho^i \Delta d_{t+1+i} - \sum_{i=0}^{\infty} \rho^i r_{f,t+1+i}^f - \sum_{i=1}^{\infty} \rho^i \psi_{t+1+i}] \quad (1.2)$$

If the first two terms on the right hand side of (1.2) are treated as a composite residual, then (1.2) has exactly the same form as (1.1). As found in the literature, variation in forecast revisions of future short-term interest rate doesn’t count a
significant amount of the total stock return volatility\(^1\). So in this paper I am going to ignore the effect of short-term interest rate and focus on the study of gross stock return as in equation (1.1). Nevertheless, the framework and methodology discussed in the rest of this paper are general enough to be easily extended to study the variability of excess stock returns as in equation (1.2) and other extensions as well.

1.2.3 The Variance Decomposition

It will be convenient to simplify the notation in equation (1.1). Let us define \(\eta_{t+1}\) to be the unexpected component of the stock return \(r_{t+1}\), \(\eta_{cf,t+1}\) to be the term in equation (1.1) that represents investors’ forecast revision on future cash flows, and \(\eta_{dr,t+1}\) to be the term representing their forecast revision on future discount rates. Then equation (1.1) can be written as:

\[
\eta_{t+1} = \eta_{cf,t+1} - \eta_{dr,t+1}
\]

\(^1\)For example, in Bernanke and Kuttner (2005), the variance of revisions in expectations of real interest rates accounts for 0.6% to 1.4% of the total excess stock return volatility, depending on the sample period used.
where

\[ \eta_{t+1} \equiv r_{t+1} - E_t r_{t+1} \]  \hspace{1cm} (1.4) \\
\[ \eta_{cf,t+1} \equiv (E_{t+1} - E_t) \left[ \sum_{i=0}^{\infty} \rho^i \Delta d_{t+1+i} \right] \]
\[ \eta_{dr,t+1} \equiv (E_{t+1} - E_t) \left[ \sum_{i=1}^{\infty} \rho^i r_{t+1+i} \right] \]

From equation (1.3) it is easily seen that the variability of stock return comes from three sources: the variation in investors’ forecast revision of future cash flows, the variation in forecast revision of future discount rates, and their comovements:

\[ \text{var}(\eta_{t+1}) = \text{var}(\eta_{cf,t+1}) + \text{var}(\eta_{dr,t+1}) - 2\text{cov}(\eta_{cf,t+1}, \eta_{dr,t+1}) \]  \hspace{1cm} (1.5)

1.2.4 The VAR Approach and perfect information

In practice, revisions to investors’ expectations are not directly observable. Following Campbell and Ammer (1993), I use a vector autoregression approach to obtain empirical proxies for investors’ expectational revisions. The VAR system involves the variables of interest (stock returns in the context of this paper) along with any other state variables that might be helpful in forecasting those variables. Specifically, express a p-lag, n-variable VAR as a first-order system:

\[ z_t = A + B z_{t-1} + u_t \]  \hspace{1cm} (1.6)
where $z_t$ is a properly stacked $np \times 1$ vector containing the variable of interest and other state variables. The corresponding demeaned process is then:

$$z_t - \tilde{A} = B(z_{t-1} - \tilde{A}) + u_t$$

(1.7)

where $\tilde{A} = (I - B)^{-1}A$ is the unconditional mean vector.

With perfect information on the VAR process as specified above, the investors revise their forecasts only when they observe realized innovation $u_t$, and the ingredients of (1.4) are given by:

$$\eta_{t+1} = e_1' u_{t+1}$$

(1.8)

$$\eta_{dr,t+1} = e_1' (I - \rho B)^{-1} \rho B u_{t+1}$$

$$\eta_{cf,t+1} = \eta_{t+1} + \eta_{dr,t+1}$$

where $e_1$ is the appropriate selection vector.

Two features of this VAR method deserve further comment. First, this analysis is based on the assumption that the VAR captures the exact dynamics of the expected return process. It might seem that this approach does not take proper account of extra information that market participants may have. To reproduce the market’s forecast without knowing what the market might know about the future paths of cash flows and discount rates, I include the lag of log dividend-price ratio itself as one variable in the vector autoregression. Intuitively, even though we do
not observe everything that market participants do, we do observe the log dividend-price ratio, and that variable summarizes the market’s relevant information.

Second, this analysis estimates forecast revisions about discount rates directly from the VAR, and attributes any remaining component of the unexpected return to forecast revisions about dividends. Alternatively, one could instead include the dividend growth rates into the VAR system, estimating forecast revisions about cash flows directly and leaving some other terms as residual. However there is some doubt as to whether dividends follow a linear time-series model with constant coefficients\(^2\). The current method avoids this potentially serious problem.

1.2.5 Empirical Results: Perfect Information

With perfect information regarding the return process, the only source of forecast revisions comes from realized innovations over time, which could be adequately captured by a full-sample VAR estimation. In this session I report empirical results under a full-sample VAR method. This is also the method accepted by the conventional wisdom.

\(^2\)See, for example, Lehmann (1991) "The conventional practice of assuming a particular dividend policy and computing its present value is fraught with hazard... since managers have no obvious incentive to adopt or maintain a consistent dividend policy."
The empirical version of (1.4) under a full-sample VAR is:

\[ \hat{\eta}_{t+1} = e'_1 \hat{u}_{t+1} \]  
\[ \hat{\eta}_{dr,t+1}|F_T = e'_1 (I - \rho \hat{B}_T)^{-1} \rho \hat{B}_T \hat{u}_{t+1} \]  
\[ \hat{\eta}_{cf,t+1}|F_T = \hat{\eta}_{t+1} + \hat{\eta}_{dr,t+1}|F_T \]

where $T$ is the sample length; $F_T$ denotes a filtration with respect to information up to period $T$; $\{\hat{\eta}_{dr,t+1}\}_{t=0}^{T-1}$ and $\{\hat{\eta}_{cf,t+1}\}_{t=0}^{T-1}$ denote estimated forecast revisions of future returns and dividends respectively; $\hat{B}_T = E(B|F_T)$ is the expectation of $B$ based on filtration $F_T$. In the context of perfect information, it is the estimated coefficient matrix from regression (1.6) using full sample data (0 to T); $\{\hat{u}_t\}_{t=1}^T$ is the series of estimated innovation, i.e., the in-sample fit of regression errors.

In this paper I study the variability of monthly market return with a five-variable VAR system. The market return ($R_m$) is measured as the value-weighted S&P 500 index returns. The other four state variables included: the dividend growth rate ($Gd$), the smoothed price-earnings ratio ($PE_{10}$), the dividend yield ($YLD$), and inflation ($INFL$). Details of the variables are described in Appendix II. These four predictors are commonly used in the stock return prediction literature. High price-earnings ratios ($PE_{10}$) will necessarily imply low long-run expected returns, if expected earnings growth is constant, see Campbell and Shiller (1988a, 1998), and Lamont (1998). The dividend yield is included because it should reflect any changes that may occur in future expected returns.
Inflation is also a widely recognized predictor that measures the macroeconomic risk, for example, in Lintner (1975), Fama (1981), and Campbell and Vuolteenaho (2004).

Table 1 presents a summary of the sample characteristics of the variables. The sample runs from January 1926 to December 2005. The gross stock return over 1926 to 2005 is about .8% per month, corresponding to an annualized return of about 10%. The period 1926.1 – 1939.12 involves more volatile market returns, and contains more extreme return cases, such as the market crash in Oct. 1929 with a monthly return of -20%, the great bear market in Sep. 1931 with a record of -34% monthly return, and the recovery starting from the second half of 1932 when monthly return topped 30% in a couple months.

To get estimates of innovations under perfect information, I use full sample ordinary least squares method to estimate the VAR system. This is legitimate and efficient given the perfect information assumption and that the system is stationary. Table 2 reports VAR estimates. Each row of the table corresponds to a separate equation of the VAR. Our interest is mainly in the first equation, namely, the return prediction regression. Standard errors are provided in parentheses. All coefficients in this regression are significantly different from zero, indicating more or less of predicting power to stock returns. The last column of Table 2 contains R² for each regression in the VAR system. The return prediction regression has 2.1-percent R², which is reasonable for a monthly return model.

The remaining rows of Table 2 summarize the dynamics of the explanatory
variables. As highlighted in the table, of the four predicting variables, two of them (the smoothed price-earnings ratio, and the dividend yield) are highly persistent AR(1) processes with autoregressive coefficients around 0.95 and $R^2$ above 90%. The near non-stationarity regressors in the VAR lead to the concern of problematic inference in VAR parameters\(^3\) and its enlarged effect on the inference in variance components led by the infinite sum.

Table 3 summarizes the behavior of the implied forecast revisions in cash-flows and in discount-rates under perfect information assumption. Following Campbell and Vuolteenaho (2004), I set the discount factor $\rho = 0.95^{1/12}$. An annualized $\rho$ of 0.95 corresponds to an average dividend-price or consumption-wealth ratio of 5.2%, which is reasonable to a long-term investor with 5-percent consumption of total wealth per year\(^4\). I present both the full forecasting period result (1950.1 – 2005.12, "overall" period) and the sub-sample periods results decade by decade. The first three columns are variances of the discount-rate news, of the cash-flow news, and of the total return innovations. The next three columns show the proportion of the variance components. As is consistent with the finding of Campbell (1991) and Campbell and Vuolteenaho (2004), the overall forecast revisions in discount-rate is the dominant component of the market return, and explains about 77% of total stock return variations. The cash flow revisions contribute about

\(^3\)See, for example, Elliot and Stock (1994), Phillips (1995), and Cavanagh, Elliot and Stock (1995).

\(^4\)This interpretation could be thought of as follows: a mutual fund investor finances his/her consumption by redeeming a fraction of the fund shares every period, and this mutual fund reinvests all the dividends paid by the stocks it holds.
22%, and these two components are not significantly correlated. From the point of view of time-variation, total stock return varies relatively more during the 1970s and 1980s, together with the increase of variations in forecast revisions. The ratio of these two doesn’t change much cross decades however.

Figure 1 plots the total return variation and the discount rate revision from full-sample OLS. From the plot we see that the 1970s and 1980s experience high stock return volatility, and the volatility goes up again since late 1990s. To get a visually clearer pattern of the series overtime, Figure 2 plots the smoothed version of Figure 1. Besides the observation in time-varying volatility, we could also see that the revision in discount rate moves in opposite direction to stock return innovation which is expected by the theory, and cash flow revisions are far less volatile than the other two.

### 1.3 Parameter Uncertainty and Learning

In this section I introduce parameter uncertainty and investors’ learning about the return process. Two learning schemes are discussed. First is a naive scheme where investors merely rely on historical data for routine updating; and the second is a more sophisticated scheme where investors incorporate prior beliefs with observed data by Bayes rule.
1.3.1 Why Learning?

It is widely recognized that financial market is full of noise. Predictive regressions and in-sample analysis as described in Section 2 implicitly use information that investors do not have when portfolio decisions are made, since they use an entire time series of data. As shown in Lewellen and Shanken (2002), prices are backward looking and are based on previous dividends and ex ante future dividends. In this paper, we instead look forward from the point of view of a real-time investor. Real-time investors make inference based on the information available. More accurate inference is possible with more information. Although we assume the underlying return process is constant over time as in (1.6), the investors, on the other hand, view the parameters as random variables. They incorporate new information every period to make better inference on the underlying return process, i.e. they update parameter values of the VAR system and form expectations of future discount rates and cash flows based on the most updated parameter values. Thus investors’ revisions of their expectations not only come from "news" per se, but from their revision on the return process as well.

By setting out in advance a set of rules for observation windows and variable selection, estimation and modification of the econometric model, real-time method reproduces investors’ expectations recursively, providing a way to reduce the effects of data-snooping and facilitates learning from the performance of a given model when applied to a historical data set.
In the empirical part, I present results on both rolling window and expanding window estimations. Rolling regressions allow for more general structures such as smooth regime changes; on the other hand, they throw away some of the sample at each point in time and the sample size used at time $t$ is chosen in a deterministic way. Regressions with expanding window is more coherent with information updating framework but might be less efficient if the true model is time varying.

1.3.2 Learning in Real-time

Rather than utilizing information for the whole period retrospectively, the out-of-sample method only uses the investor’s real-time information set in forming his/her one-step ahead expectation from period to period. This method thus takes the view of an investor who builds asset pricing models and uses econometric methods to solve the model and guide decisions in real time.

For simplicity, I keep the assumption that the parameters of the underlying VAR process are constant over time as in equation (1.6). Formally, assume at each time $t$ investors forms estimates of $A, B$ conditioning on $F$: $\hat{A}_t \equiv E(A|F_t)$, and $\hat{B}_t \equiv E(B|F_t)$. Notice that the time subscript here denotes their inference/estimation of the true parameters from learning at period $t$, which could be different from period to period. The real-time version of equation (1.6) becomes:

$$z_t = \hat{A}_t + \hat{B}_t z_{t-1} + \hat{u}_t$$
where $\hat{u}_t$ is the fitted error vector. The conditional expectation of the $(t + 1 + j)^{th}$ period return is

$$E[r_{t+1+j}|r_t, \hat{A}_t, \hat{B}_t] = e_1'(\hat{A}_t + \hat{B}_t^{j+1}(z_t - \hat{A}_t))$$  \hspace{1cm} (1.10)$$

where $\bar{A} = (I - B)^{-1}A$ is the unconditional mean of $z_t$. Hereafter, I ignore the hat on all estimated parameters and use subscript in $E_t$ to indicate conditional expectation based on information and parameter estimates up to $t$, for notation simplification.

Next period, $t + 1$, new information arrives, which includes realized values of asset return and state variables. The investor replaces the $t^{th}$ period estimate on $z_{t+1}$ with realized values, and updates his/her estimates on the VAR parameters as well, and forms $A_{t+1}, B_{t+1}$:

$$z_{t+1} = A_{t+1} + B_{t+1}z_t + u_{t+1}$$  \hspace{1cm} (1.11)$$

This leads to an update of all future discount rates

$$E_{t+1}r_{t+1+j} = e_1'(\bar{A}_{t+1} + \bar{B}_{t+1}^{j}(z_{t+1} - \bar{A}_{t+1}))$$  \hspace{1cm} (1.12)$$
Combining (1.11) and (1.12) gives

\[ E_{t+1}r_{t+1+j} = e_1'(\tilde{A}_{t+1} + B_{t+1}^j(\tilde{A}_{t+1} + B_{t+1}(z_t - \tilde{A}_{t+1}) + u_{t+1} - \tilde{A}_{t+1})) \]
\[ = e_1'(\tilde{A}_{t+1} + B_{t+1}^j(z_t - \tilde{A}_{t+1}) + B_{t+1}^j u_{t+1}) \]

The difference between (1.12) and (1.10) is

\[ (E_{t+1} - E_t)r_{t+1+j} = e_1'(\tilde{A}_{t+1} - \tilde{A}_t) + e_1'(B_{t+1}^j(z_t - \tilde{A}_{t+1}) - B_{t}^1 z_t - \tilde{A}_t) + e_1'B_{t+1} u_{t+1} \]

(1.13)

In general, \( \tilde{A}_t \neq \tilde{A}_{t+1} \) and \( B_t \neq B_{t+1} \). Clearly (??) is nested as a special case of (1.13) when \( \tilde{A}_t = \tilde{A}_{t+1} \) and \( B_t = B_{t+1} \). The news component of the future discount rate is an infinite sum of discounted revisions to future returns. Denote the parameters of interest \( \theta = (A, B) \),

\[ \eta_{dr,t+1} \equiv (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} = g_{1,t+1}(\theta) + g_{2,t+1}(\theta) + g_{3,t+1}(\theta) \]

(1.14)

where

\[ g_{1,t+1}(\theta) \equiv e_1'(\frac{\rho}{1 - \rho})(\tilde{A}_{t+1} - \tilde{A}_t) \]
\[ g_{2,t+1}(\theta) \equiv e_1'(I - \rho B_{t+1})^{-1} \rho B_{t+1}^2(z_t - \tilde{A}_{t+1}) - (I - \rho B_t)^{-1} \rho B_t^2(z_t - \tilde{A}_t) \]
\[ g_{3,t}(\theta) \equiv e_1'(I - \rho B_{t+1})^{-1} \rho B_{t+1} u_{t+1} \]

These three components of the investor’s change in expectations of future discount
rate could be interpreted as follows: $g_1$ stands for change coming from the investor’s adjusting the unconditional mean of excess stock returns; $g_2$ stands for the part coming from investor’s re-gauging last period stock return’s deviation from unconditional mean. The effect of the previous return deviation would be amplified in the future due to the persistent structure embedded in VAR; The last part, $g_3$, is the effect from innovations of the return process. Under the assumption that the return process is stable and investors have full information on the process, $g_1$ and $g_2$ are equal to zero and are thus ignored in the traditional in-sample framework which assumes perfect information.

1.3.3 A Naive Learning Scheme

Under parameter uncertainty, investors’ forecasts are based on limited information and the way they learn about the true return process. Let us start from a very simple learning scheme under which the investors merely rely on historical data to do routine updating, and I call it a "naive" learning scheme. In each period these type of investors take all the data available at that point and run OLS estimation. They do forecasts of both short and long horizons based on OLS estimates, although knowing that they are going to change the whole set of parameter estimates next period with the arrival of new information. Obviously, this is not an efficient way of forecasting.

Empirically this naive learning scheme can be captured by forecast revisions based on recursive OLS estimation. As in the previous section we estimate a five-
state-variable VAR, which contains 30 mean parameters alone, not yet including parameters in the variance-covariance structure. To mitigate potential errors from small sample, in the expanding window exercise I use the first 24 years of data (1926-49) to obtain initial parameter estimates, so the real-time updating begins in 1950; in the rolling window exercise I chose a 30 years window.

Table 4 reports results under recursive OLS estimations. Focus on the expanding window regressions first: there is significant difference compared with the results under perfect information presented in Table 3. The estimated unexpected return volatility is about the same in the two tables, varying from 17% to 18%. However, the volatility of forecast revisions on discount-rates is more than 1.5 times more volatile than that under perfect information. This is not a surprise, as I discuss at the beginning of this paper, because learning would lead to more volatile forecasts through investors’ adjustment of future prospects. As to the contribution of each revision component, under naive learning with expanding window the forecast revision of future discount rates has about the same volatility as that of the stock returns and in some decades it even exceeds the latter. At the same time the contribution from the revision of cash flows is also much bigger, increasing from around 23% under perfect information to an overall level of 50%. The second panel of Table 4 contains rolling regression results. In brief, rolling regression yields even more dramatic difference with the revision components each contributes more than 100% to the stock return innovation. Figure 3.1 and 3.2 plot the smoothed series under naive learning.
If the underlying return process is stable over time and the OLS method is reliable, we should see similar results from the in-sample OLS estimation and the recursive one. The huge difference between the results based on perfect information and those based on naive learning indicate that investors don’t behave so naively as to merely rely on the data especially when they form long horizon forecasts where small variations in estimates of the return process could lead to huge swings of forecasts.

In Figure 4 I plot the maximum eigenvalue of the estimated $B_t$ matrix in the VAR system. A couple comments on the figure: first, both rolling window and expanding window regressions indicate an upward sloping trend of the maximum eigenvalue of B, which indicates a more persistent VAR system; second, the estimates based on rolling regression are more volatile than those based on expanding window regression. Especially in the early 1980s and early 2000 the estimated maximum eigenvalue exceeds one which indicates an explosive system. "Wise" investors shouldn’t accept such an estimated value at all since it violates the underlying assumption that the process is stationary. Long horizon forecast under non-stationary process is not meaningful. This example illustrates the potential problem of the "naive" investors.

1.3.4 A Sophisticated Learning Scheme:

The second learning scheme I study is a relatively sophisticated one under which investors don’t simply accept the results from recursive OLS estimation, but
rather have their own prior beliefs regarding the return process and incorporate observed data with their prior beliefs according to the Bayes rule; they don’t put all weights on one point estimate which is subject to severe potential estimation problem, but consider all possible values of the return process and weigh them accordingly in forecasting future returns.

The Bayesian method is a natural and quantifying way to model the investors’ real-time decision making when learning problems are hard and he/she faces updated signals every period. The key feature of this approach is to summarize the uncertainty of parameters by a prior distribution. The foundation for the Bayesian approach was provided by Savage (1954). Under this framework, Kandel and Stambaugh (1996), Brennan (1998), Barberis (2000) show that parameter uncertainty can affect significantly investors’ portfolio choice. Pastor (2000) and Pastor and Stambaugh (2000), in addition to parameter uncertainty, consider also model uncertainty.

The Bayesian method is also particularly useful in estimating the decomposition of stock returns. First, as discussed in Stambaugh (1999), classical OLS estimates of the parameters can be biased. Bayesian techniques instead provide us with exact small sample distributions for the decompositions, which are highly nonlinear functions of the parameters of the VARs. Second, statistical tests indicate that some of the variables in the VAR system are almost non-stationary even though we have theoretical reasons for believing them to be stationary. Bayesian methods allow such theoretical restrictions to be imposed through the prior.
In the next section I describe the Bayesian technique in detail. Empirical results are presented afterwards.

### 1.4 Modeling Bayesian VAR

The Bayesian version of (1.14) is

$$ (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j \epsilon_{t+1+j} = \int (g_{1,t+1}(\theta) + g_{2,t+1}(\theta) + g_{3,t+1}(\theta)) f(\theta|z^{t+1}) d\theta \quad (1.15) $$

In the above expression, $f(\theta|z^{t+1})$ denotes the investor’s posterior belief on the parameters of interest, given information until $t+1$. In the following sub-sections, we focus on the details of real-time forecasting error decomposition.

#### 1.4.1 Independent Normal-Wishart prior:

We introduce a couple extra notations before formal analysis. Stacking the vectors $z'_t, [1, z'_{t-1}]$, and $u_t$ into $Z, X$, and $U$, let’s express (1.6) as

$$ Z = X \Gamma + U \quad (1.16) $$

Further stack the columns of $Z, \Gamma$, and $U$ into $z, \gamma$, and $u$

$$ z = (I_k \otimes X) \gamma + u $$
where \( u \sim N(0, \Sigma \otimes I_T) \). We assume no serial correlation in error terms, but they may be correlated across equations. In this model the unknown parameters are \( \gamma \) and \( \Sigma \). Treating the initial observation \( z_0 \) as a known constant, the likelihood is given by

\[
L(\gamma, \Sigma) \propto |\Sigma|^{-T/2} \exp\{-tr[(Z - X\hat{\Gamma})'\Sigma^{-1}(Z - X\hat{\Gamma})]/2\}
\]

\[
\propto N(\gamma|\hat{\gamma}, \Sigma \otimes (X'X)^{-1})
\]

\[
\times IW(\Sigma|(Z - X\hat{\Gamma})'(Z - X\hat{\Gamma}), T - k - m - 1) \tag{1.17}
\]

where the "hat" \( \hat{\gamma} \) denotes OLS estimates of the parameters, \( \hat{\Gamma} = (X'X)^{-1}X'Z \), and \( \hat{\Sigma} = T^{-1}(Z - X\hat{\Gamma})'(Z - X\hat{\Gamma}) \). The likelihood is a product of an inverse Wishart density for \( \Sigma \), and a Normal density for \( \gamma \) conditional on \( \Sigma \).

The first prior we consider is an independent conjugate prior, denoted \( p_1(\gamma, \Sigma^{-1}) \)

\[
p_1(\gamma, \Sigma^{-1}) \sim p(\gamma), p(\Sigma^{-1})
\]

where

\[
\gamma \sim N(\gamma, H_\gamma^{-1}), \text{ and } \Sigma^{-1} \sim W(\alpha, H) \tag{1.18}
\]

The Wishart distribution can be thought of as the multivariate analog of the chi-square distribution. There are four hyperparameters that help shape the prior beliefs: the coefficient vector \( \gamma \) obeys multivariate Normal distribution, with \( \gamma \) the prior mean and \( H_\gamma \) the prior precision matrix that reflects the investor’s confidence.
on his/her belief of prior mean $\gamma$. The variance covariance matrix $\Sigma$ obeys Inverse Wishart distribution, where $a$ specifies the degrees of freedom, and $H$ reflects prior belief on the variance of innovations in the original return process as in Equation (1.6).

As discussed in Kadiyala and Karlsson (1997), the Normal-Wishart prior avoids the two main shortcomings of the Minnesota prior: the forced posterior independence between equations and the fixed residual variance-covariance matrix. It is also quite flexible in the sense that when the prior hyperparameters $H$, $a$, and $H$ all converge to zero, the informative prior as in (1.18) converges to non-informative Jeffreys’ prior. When only the hyperparameters $a$ and $H$ coverge to zero, we get Normal-Diffuse prior, a Normal prior for the regression parameters $\gamma$, and a diffuse prior for the precision matrix $\Sigma^{-1}$.

1.4.2 Prior on Stationarity:

One premise of the variance decomposition method is that the return process is stationary, so that expected future return and its variance are both bounded. The Normal-Wishart prior discussed in the last sector impose a normal distribution on each coefficient including the persistency parameters in every equation of the VAR system. We could narrow most of the probability mass on persistency parameters within a small range, but would still have positive mass on any segment.

---

5This is also a commonly used prior in the literature, for example, in Stambaugh (1999) and Barberis (2000).

6Wachter and Warusawitharana (2005) used a conditional version of the Normal-Diffuse prior.
of the whole real line. For this reason, it is necessary to impose stationarity as part of prior belief.

At the same time, some state variables that we choose are well recognized as having high persistency such as term spread and price-earnings ratio. They usually have an auto-correlation around 0.95 depending on the sample period and model specification. The "discounted infinite sum" effect with a discount rate close to 1 basically would exaggerate any small variation in the auto-correlation either due to different sample period, or model misspecification, or pure randomness. This is another reason that a prior on stationarity is beneficial: it helps get more robust results and inference.

The need to impose a prior of non-explosive behavior is well documented in the Bayesian literature (e.g. Bauwens, Lubrano and Richard, 1999). A prior on stationarity is equivalent to limiting the auto-correlation to a subset of real-line under one-equation regression. For VAR systems as we have in (1.6), stationarity is equivalent to restricting the absolute value of eigenvalues of matrix B to be less than one. To incorporate this, we set up a more general prior based on (1.18) that could put constraints on parameters

$$p_2(\gamma, \Sigma^{-1}) \sim p(\gamma).p(\Sigma^{-1}).I(\gamma \in S)$$

(1.19)

Without prior stationarity, $$S = R^{k(k+1)}$$, where k is the number of variables in the
VAR system. With stationarity imposed,

\[ S = \{ \gamma : \max\{\text{eigenvalue of } B\} < \lambda \text{ and } \gamma \in R^{k(k+1)} \} \]

where \( \lambda \) is a number very close to 1. I set \( \lambda = 1 - 0.001 \) to avoid significant distortion of the distribution.

### 1.4.3 Posterior belief

Investors adjust their prior beliefs after observing data. Combining the prior belief (1.19) and likelihood (??), we obtain the joint posterior for \((\gamma, \Sigma^{-1})\) as

\[
p(\gamma, \Sigma^{-1}|Z) \propto p(\gamma, \Sigma^{-1}).p(Z|\gamma, \Sigma^{-1})
\]

(1.20)

which is not a regular distribution. However, we could obtain conditional posterior distributions of \(\gamma\) and \(\Sigma^{-1}\) separately. Let the upper bar indicate posterior parameters. The conditional posterior distribution of precision matrix \(H\) is

\[
H \equiv \Sigma^{-1}|Z, X, \Gamma \sim W(\bar{a}, \bar{H}),
\]

(1.21)

with \(\bar{a} = T + a\), and \(\bar{H} = ((Z - X\Gamma)'(Z - X\Gamma) + H^{-1})^{-1}\). The conditional posterior distribution of coefficients is

\[
\gamma|Z, X, \Sigma \sim N(\bar{\gamma}, \bar{H}^{-1}_{\gamma}).I(\gamma \in S)
\]

(1.22)
with $H = +H$, and $\gamma = H^{-1}[(H \otimes X'X)\hat{\gamma} + H\gamma]$, where $\hat{\gamma} = [X'(H \otimes IT)X]^{-1}X'(H \otimes IT)Y$ is the OLS estimator of $\gamma$.

### 1.4.4 Posterior Evaluation

Analytical results are available for model 1 that do not impose any constraints on the parameter estimates and thus preserve the untruncated distribution. Unfortunately, this is not the case for constrained models. To estimate the parameters of these models requires evaluating the posterior distribution of the parameters, $\delta = (\gamma, \Sigma^{-1})$, given the data up to time $t$, $F_t$, denoted $\pi(\gamma, \Sigma^{-1} | F_t)$. This in turn requires repeatedly drawing from this distribution which is not always feasible in our context. So we cannot use Monte Carlo integration methods to simulate posterior moments of functions on our parameters of interest. Instead, we implement Gibbs Sampler and importance sampling techniques. These work as follows.

The Gibbs sampler begins with a partition on parameters of interest. A natural and convenient way of partition in our context is to set two blocks:

$$\delta_{(1)} = \Sigma^{-1}, \delta_{(2)} = \gamma \quad (1.23)$$

Given a single draw $\delta^{(0)} = (\delta_{(1)}^{(0)}, \delta_{(2)}^{(0)})$ from $\pi(\delta | F_t)$, we successively make drawings

$$\delta_{(b)}^{(m)} \sim \pi\left(\delta_{(b)} \mid \delta_{< (b)}^{(m)}, \delta_{< (b)}^{(m-1)}, F_t\right), b = 1, 2; m = 1, 2, \ldots \quad (1.24)$$
Notice that the above random sampling involves the two conditional densities we’ve illustrated, the posterior distribution equations (1.22), (1.21). The sequence thus produced, \( \{\delta^{(m)}\} \), is a realization of a Markov chain. Roberts and Smith (1994) proved the convergence of this Markov chain. Under certain conditions, any single iterate \( \delta^{(m)} \) retains the property that it is drawn from the joint density \( \pi (\gamma, \Sigma^{-1}|F_t) \).

The remaining question is how to make draws from (1.22) and (1.21) to form the Markov chain as in (1.24). The latter is trivial and could be done directly by many statistic packages. The former is non-standard, and importance sampling is applied here to achieve the drawings. Suppose that random draws \( \gamma^i, i = 1, ..., I \) can be generated from a simpler density \( q(\gamma) \). This density, \( q(\gamma) \), is called the importance function and by appropriately weighing the random draws from \( q(\gamma) \), the moments computed from the draws of the importance function, \( \gamma^i \), converge to the moments obtained from the (unknown) posterior distribution \( \pi (\gamma|F_t) \). This property makes use of the result proved by Geweke (1989) that if \( \gamma^i, i = 1, ..., I \) is a random sample from \( q(\gamma) \), then under weak conditions

\[
\hat{g}_I = \frac{\sum_{i=1}^{I} \omega(\gamma^i)g(\gamma^i)}{\sum_{i=1}^{I} \omega(\gamma^i)} \rightarrow E[g(\gamma)|\Omega_I],
\]

The weights \( \omega(\gamma^i) = \frac{\pi(\gamma=\gamma^i|\Omega_t)}{q(\gamma=\gamma^i)} \) are known as the importance function.

For importance sampling to be feasible, \( q(\gamma) \) needs to approximate \( \pi (\gamma|\Omega_t) \) quite well, otherwise cases can be found where \( \omega(\gamma^i) \) is equal to zero for virtually
every draw and the weighted average involves very few draws. Thus, importance sampling may be inaccurate unless \( q(\gamma) \) is chosen carefully. However, the problem of finding an accurate importance function is easily resolved for the standard linear regression model subject to inequality constraints. By setting the importance function equal to the unconstrained posterior distribution, the weights can be computed as

\[
\omega(\beta^i) = I(\beta^i \in S).
\]

In this fashion, the weights are either one (if \( \beta^i \in S \)) or zero (if \( \beta^i \notin S \)) and this strategy simply involves drawing from the unrestricted posterior in (1.18) and discarding draws that violate the relevant inequality restrictions. Hence this approach is very simple to implement in practice.

### 1.4.5 Prior elicitation

In the Normal-Wishart prior, there are four hyperparameters that help shape the prior beliefs. I set the prior mean of regression coefficients in front of the state variables equal to zero, and the constant equal to historical mean \( \bar{\gamma} = vec[\bar{\gamma}, 0_{k \times k}] \), where \( vec \) denotes the vectorization operator and \( \bar{\gamma} \) denotes the vector of historical mean; the prior precision matrix that reflects the investor’s confidence on his/her belief of prior mean \( \bar{\gamma} \) is set as \( H_{\gamma} = sc^{-1}.I(k^2) \). A smaller value of \( c \) implies a more diffuse prior; we set \( sc = 0.01 \). As to the prior on variance covariance matrix \( \Sigma \), the degree of freedom \( a = 8 \); and \( H = d^{-1}.I(k^2) \) with \( d = 1 \).
implies a prior mean with the covariance matrix being Identity.

1.4.6 Empirical Results: Sophisticated Learning with BVAR

In this section we show the results using real-time Bayesian estimation. Again, the estimation period starts in 1926, and the forecasting period starts in 1950. Table 5 presents the variance of the forecast revisions and the variance ratios under sophisticated learning. The first row shows the overall statistics from 1950.1 to 2005.12. The next three rows after that contain results during subperiods as indicated in the table.

Comparing the results under sophisticated learning (captured by real-time BVAR) with those under perfect information (by in-sample OLS) and those under naive learning scheme (by real-time OLS), we can see that the magnitude of the unexpected stock returns are about the same under these different assumptions. In Table 5, the overall unexpected return volatility is 0.002, and it varies from as low as 0.0013 in the 1960s to as high as 0.0024 in the 1980s. The average discount rate revisions under BVAR is 0.0013, which again has a similar magnitude compared with that by in-sample OLS.

However, the variation in revisions of cash flows is much bigger under sophisticated learning, going from an average of 0.0004 in Table 3 to 0.0012 in Table 5. Correspondingly, under sophisticated learning assumption, a significant proportion of stock variance comes from cash flow revisions which contradicts the conventional wisdom that discount rate revisions dominate the stock return inno-
vation. On average the cash flow revisions count for 60% of stock return variance, and varies from more than 150% in the 1950s to about 33% in the 1980s. We also see an increase of the cash flow revisions component in the most recent 15 years. The smoothed series are plot in Figure 5.

Correlation between the discount rate revisions and cash flow revisions component also varies across decades. It is positive before the 1970s, is not significantly different from zero in the 1980s, and turns to negative since the 1990s. This result also contradicts the conventional wisdom that cash flow news and discount rate news don’t correlate with each other.

Overall, learning in real-time with Bayesian method explores large time-variation both in terms of the parameter values, and in terms of the relation between news components. As for how investors update information, the results show that Bayesian learning matches the data better than naive OLS updating.

1.5 Conclusion

In this paper I examines the variability of stock returns and its variance components. The framework is based on a dynamic accounting model and time-series econometrics methods to break the unexpected stock returns into forecast revisions on future discount rates and those on future cash flows. I relax the conventional assumption of perfect information and study this issue from a new perspective: the effect of parameter uncertainty and investors’ learning on stock
return movements. Learning affects investors’ forecast on future discount rates and cash flows through parameter estimation, and thus has an indirectly impact on stock return movements. Specifically I study two different learning schemes: a naive learning scheme captured by recursive OLS estimation and a sophisticated learning scheme captured by real-time Bayesian method.

Since I use an accounting framework rather than a behavioral model, the statements I made in this paper are only about proximate causes rather than fundamental causes of asset price movements. Furthermore, the statements are made upon the validity of assumptions regarding investors’ learning schemes. Nevertheless, the empirical results of this paper shed light on several issues that have long been ignored in the variance decomposition literature.

The first important finding is that learning does have big impacts on investors’ forecast revisions. Under perfect information, the parameter value is constant when the underlying return process is stable. When investors are not certain about the process and have to infer it from limited information, we observe distinguishable time-variation in the estimated parameters. Specifically, investors tend to revise their inference on parameters more often when they follow naive learning schemes and also when they apply recursive estimation with a rolling window. On the other hand, investors with sophisticated learning schemes tend to be more cautious when updating their beliefs and revising forecasts.

Variance components of the unexpected stock return movements are discounted infinite sum of forecast revisions. As a consequence of the first finding,
I find difference in forecast revisions leads to difference in the absolute value and in the relative contribution of variance components as well. The literature led by Campbell and Ammer (1993) and others raised the "excess volatility" puzzle where they find, under perfect information, the returns of the postwar U.S. stock market have a standard deviation two or three times greater than the standard deviation of revisions about future cash flows. However, the empirical results in this paper show that if we relax the perfect information assumption, revisions about future cash flows contribute an un-negligible proportion. Moreover, real-time BVAR reveals that the variation coming from cash flow news is as important as the variation from discount rate news, if not more important. This finding implies that investors revise their forecasts regarding future cash flows much more than we usually think they do. The potential reasons for these big adjustments from period to period might be that cash flow growth is hard to predict out-of-sample, or the cash flow process is more persistent than what researchers think and thus a small innovation in cash flow growth may cumulate over time and have a big effect on the stock price, or both. Since I don’t directly model the cash flow process in this paper, the above hypothesis could not be tested. However, this is a very interesting direction for future research as an extension of this paper.

The third finding is regarding the correlation between the two variance components. From the results presented, I reach another conclusion that differs from the conventional wisdom. In the variance decomposition literature people agreed that these two components on average move independently. However, under
uncertainty investors’ revision of the two components tends to correlate with each other. In most of the cases I study, they tend to move in the same direction. Is it because news regarding future discount rate shed lights on future cash flows? Is it because they share some common predictive variable? This finding provides a first stage empirical evidence for modeling cash flow process, and some guidance for theoretical asset pricing models on the correlation of innovations regarding future discount rate and cash flows.

Various robustness checks are done for the empirical part of the paper, regarding the state variables included in the VAR, the estimation windows, and the prior beliefs in the Bayesian method that I carried out in this paper. They are available upon request.
1.6 Tables and Figures
Table 1.1: Descriptive Statistics of the VAR Variables

<table>
<thead>
<tr>
<th></th>
<th>Rm</th>
<th>Gd</th>
<th>PE10</th>
<th>YLD</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.008</td>
<td>0.004</td>
<td>2.758</td>
<td>-3.284</td>
<td>0.003</td>
</tr>
<tr>
<td>median</td>
<td>0.013</td>
<td>0.005</td>
<td>2.766</td>
<td>-3.252</td>
<td>0.002</td>
</tr>
<tr>
<td>std</td>
<td>0.056</td>
<td>0.013</td>
<td>0.395</td>
<td>0.438</td>
<td>0.005</td>
</tr>
<tr>
<td>min</td>
<td>-0.339</td>
<td>-0.094</td>
<td>1.717</td>
<td>-4.525</td>
<td>-0.021</td>
</tr>
<tr>
<td>max</td>
<td>0.348</td>
<td>0.072</td>
<td>3.789</td>
<td>-1.913</td>
<td>0.057</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.078</td>
<td>0.598</td>
<td>0.990</td>
<td>0.988</td>
<td>0.548</td>
</tr>
</tbody>
</table>

Note: This table presents descriptive statistics for the variables included in the VAR: Stock return, Dividend growth rate, Smoothed Price-earnings Ratio, Dividend yield, and Inflation. They are all monthly data. The statistics include: mean, median, standard deviation, minimum, maximum, and autocorrelation.
Table 1.2: VAR parameter estimates by OLS in-sample

<table>
<thead>
<tr>
<th></th>
<th>constant</th>
<th>Rm(t-1)</th>
<th>Gd(t-1)</th>
<th>PE10(t-1)</th>
<th>YLD(t-1)</th>
<th>INFL(t-1)</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rm(t)</td>
<td>0.0359</td>
<td>0.082</td>
<td>0.1893</td>
<td>-0.0281</td>
<td>-0.0153</td>
<td>-0.741</td>
<td>2.1%</td>
</tr>
<tr>
<td>t-stat of Rm</td>
<td>(0.0137)</td>
<td>(0.032)</td>
<td>(0.1422)</td>
<td>(0.0092)</td>
<td>(0.0083)</td>
<td>(0.3551)</td>
<td></td>
</tr>
<tr>
<td>Gd(t)</td>
<td>-0.0053</td>
<td>0.0112</td>
<td>0.5779</td>
<td>-0.0013</td>
<td>-0.003</td>
<td>0.1981</td>
<td>37.3%</td>
</tr>
<tr>
<td>PE10(t)</td>
<td>0.0111</td>
<td>0.5095</td>
<td>0.0734</td>
<td>0.9788</td>
<td>-0.0143</td>
<td>-1.132</td>
<td>99.2%</td>
</tr>
<tr>
<td>YLD(t)</td>
<td>0.006</td>
<td>-0.9856</td>
<td>0.5826</td>
<td>-0.0023</td>
<td>0.9986</td>
<td>0.2023</td>
<td>99.9%</td>
</tr>
<tr>
<td>INFL(t)</td>
<td>0.0013</td>
<td>0.0041</td>
<td>0.0376</td>
<td>-0.0037</td>
<td>-0.0031</td>
<td>0.4903</td>
<td>32.8%</td>
</tr>
</tbody>
</table>

Note: This table presents estimation results by in-sample OLS. The first six columns list estimated coefficients of the VAR. Standard errors of the return equation (first row) are listed in parenthesis. The last column lists in-sample R-square for each equation in the VAR system.
Table 1.3: Variance Decomposition with Perfect Information

<table>
<thead>
<tr>
<th></th>
<th>$\text{var} (\eta_{dr})$</th>
<th>$\text{var} (\eta_{cf})$</th>
<th>$\text{var} (\eta)$</th>
<th>$\text{var} (\eta_{dr}) / \text{var} (\eta)$</th>
<th>$\text{var} (\eta_{cf}) / \text{var} (\eta)$</th>
<th>$-2 \text{cov} (\eta_{dr}, \eta_{cf}) / \text{var} (\eta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall</td>
<td>0.0013</td>
<td>0.0004</td>
<td>0.0017</td>
<td>0.7717</td>
<td>0.2236</td>
<td>0.0046</td>
</tr>
<tr>
<td>1950.1 – 1959.12</td>
<td>0.0009</td>
<td>0.0003</td>
<td>0.0011</td>
<td>0.8152</td>
<td>0.2436</td>
<td>-0.0588</td>
</tr>
<tr>
<td>1960.1 – 1969.12</td>
<td>0.0010</td>
<td>0.0002</td>
<td>0.0012</td>
<td>0.8253</td>
<td>0.1834</td>
<td>-0.0087</td>
</tr>
<tr>
<td>1970.1 – 1979.12</td>
<td>0.0016</td>
<td>0.0004</td>
<td>0.0021</td>
<td>0.7885</td>
<td>0.2127</td>
<td>-0.0011</td>
</tr>
<tr>
<td>1980.1 – 1989.12</td>
<td>0.0016</td>
<td>0.0005</td>
<td>0.0023</td>
<td>0.7144</td>
<td>0.2162</td>
<td>0.0694</td>
</tr>
<tr>
<td>1990.1 – 2005.12</td>
<td>0.0013</td>
<td>0.0004</td>
<td>0.0017</td>
<td>0.7653</td>
<td>0.2539</td>
<td>-0.0192</td>
</tr>
</tbody>
</table>

Notes: This table presents the variance decomposition results based on OLS in-sample estimation. Both whole forecasting sample (1950.1 – 2005.12, denoted "overall") and subsample results are presented. The first three columns show the variance of the overall stock innovation and that of the two news components, discount-rate news $\eta_{dr}$ and cash flow news $\eta_{cf}$. The next three columns present variance ratios of each component to the overall variance, and the proportion that comes from the covariance of the two components.
Table 1.4: Variance Decomposition with Naive Learning

<table>
<thead>
<tr>
<th></th>
<th>$\text{var}(\eta_{dr})$</th>
<th>$\text{var}(\eta_{cf})$</th>
<th>$\text{var}(\eta)$</th>
<th>$\frac{\text{var}(\eta_{dr})}{\text{var}(\eta)}$</th>
<th>$\frac{\text{var}(\eta_{cf})}{\text{var}(\eta)}$</th>
<th>$-2\text{cov}(\eta_{dr}, \eta_{cf})$ \left/ \text{var}(\eta) \right.$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expanding Window</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>0.0019</td>
<td>0.0010</td>
<td>0.0018</td>
<td>1.0876</td>
<td>0.5455</td>
<td>-0.6331</td>
</tr>
<tr>
<td>1950.1 – 1959.12</td>
<td>0.0015</td>
<td>0.0010</td>
<td>0.0011</td>
<td>1.3027</td>
<td>0.9014</td>
<td>-1.2041</td>
</tr>
<tr>
<td>1960.1 – 1969.12</td>
<td>0.0015</td>
<td>0.0004</td>
<td>0.0012</td>
<td>1.2973</td>
<td>0.3564</td>
<td>-0.6537</td>
</tr>
<tr>
<td>1970.1 – 1979.12</td>
<td>0.0020</td>
<td>0.0009</td>
<td>0.0020</td>
<td>1.0112</td>
<td>0.4745</td>
<td>-0.4857</td>
</tr>
<tr>
<td>1980.1 – 1989.12</td>
<td>0.0024</td>
<td>0.0011</td>
<td>0.0018</td>
<td>1.3447</td>
<td>0.6306</td>
<td>-0.9753</td>
</tr>
<tr>
<td>1990.1 – 2005.12</td>
<td>0.0019</td>
<td>0.0012</td>
<td>0.0020</td>
<td>0.9328</td>
<td>0.5731</td>
<td>-0.5060</td>
</tr>
<tr>
<td><strong>Rolling Window</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>0.0054</td>
<td>0.0033</td>
<td>0.0018</td>
<td>3.0851</td>
<td>1.8542</td>
<td>-3.9393</td>
</tr>
<tr>
<td>1956.1 – 1959.12</td>
<td>0.0021</td>
<td>0.0012</td>
<td>0.0012</td>
<td>1.8481</td>
<td>1.0773</td>
<td>-1.9254</td>
</tr>
<tr>
<td>1960.1 – 1969.12</td>
<td>0.0018</td>
<td>0.0008</td>
<td>0.0014</td>
<td>1.2810</td>
<td>0.5842</td>
<td>-0.8652</td>
</tr>
<tr>
<td>1970.1 – 1979.12</td>
<td>0.0058</td>
<td>0.0025</td>
<td>0.0020</td>
<td>2.8603</td>
<td>1.2374</td>
<td>-3.0976</td>
</tr>
<tr>
<td>1980.1 – 1989.12</td>
<td>0.0083</td>
<td>0.0064</td>
<td>0.0023</td>
<td>3.6041</td>
<td>2.7681</td>
<td>-5.3722</td>
</tr>
<tr>
<td>1990.1 – 2005.12</td>
<td>0.0069</td>
<td>0.0040</td>
<td>0.0017</td>
<td>4.1021</td>
<td>2.4126</td>
<td>-5.5146</td>
</tr>
</tbody>
</table>

Notes: This table presents the variance decomposition results under naive learning through OLS real-time estimation. Results on two different estimation window are presented: an expanding window and a fixed 30-year rolling window. For the expanding window case, forecast period starts from 1950.1 as before. For the rolling window case, forecast period starts from 30 years after 1926.1, which is 1956.1
<table>
<thead>
<tr>
<th></th>
<th>$\text{var}(\eta_{dr})$</th>
<th>$\text{var}(\eta_{cf})$</th>
<th>$\text{var}(\eta)$</th>
<th>$\frac{\text{var}(\eta_{dr})}{\text{var}(\eta)}$</th>
<th>$\frac{\text{var}(\eta_{cf})}{\text{var}(\eta)}$</th>
<th>$-2\text{cov}(\eta_{dr}, \eta_{cf})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall</td>
<td>0.0013</td>
<td>0.0012</td>
<td>0.0020</td>
<td>0.6437</td>
<td>0.6242</td>
<td>-0.2679</td>
</tr>
<tr>
<td>1950.1 – 1959.12</td>
<td>0.0014</td>
<td>0.0025</td>
<td>0.0015</td>
<td>0.9626</td>
<td>1.6748</td>
<td>-1.6373</td>
</tr>
<tr>
<td>1960.1 – 1969.12</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0013</td>
<td>0.6219</td>
<td>0.7036</td>
<td>-0.3254</td>
</tr>
<tr>
<td>1970.1 – 1979.12</td>
<td>0.0010</td>
<td>0.0009</td>
<td>0.0021</td>
<td>0.4793</td>
<td>0.4456</td>
<td>0.0751</td>
</tr>
<tr>
<td>1980.1 – 1989.12</td>
<td>0.0013</td>
<td>0.0008</td>
<td>0.0024</td>
<td>0.5396</td>
<td>0.3332</td>
<td>0.1273</td>
</tr>
<tr>
<td>1990.1 – 2005.12</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0018</td>
<td>0.6279</td>
<td>0.5890</td>
<td>-0.2169</td>
</tr>
</tbody>
</table>

Notes: This table presents the variance decomposition results under sophisticated learning through real-time Bayesian VAR estimation. Results are based on expanding window regression. Forecast period starts from 1950.1.
Figure 1.1: OLS In-sample, Unsmoothed

Notes: This figure plots the *unsmoothed* unexpected stock return ($\eta$), and the part of $\eta$ that comes from forecast revisions of future discount rates ($\eta_{\text{dr}}$). The series are calculated under the perfect information assumption through OLS in-sample estimation method. The estimation sample is from 1926.1 – 2005.12. The results presented in the figure are of the forecasting period, from 1950.1 – 2005.12.
Notes: This figure plots the *smoothed* series: the unexpected stock return ($\eta$), the part of $\eta$ that comes from forecast revisions of future discount rates ($\eta_{-dr}$), and the part that comes from cash flows ($\eta_{-cf}$). The original series are estimated under OLS in-sample and smoothed with a trailing exponentially weighted moving average according to $MA_t(N) = 0.08N_t + (1 - 0.08)MA_{t-1}(N)$.
Figure 1.3: Naive Learning with Expanding Window
Figure 1.4: Naive Learning with Rolling Window

Notes: Figure 1.3 and Figure 1.4 plot the smoothed series under Naive Learning: the unexpected stock return ($\eta$), the forecast revisions of future discount rates ($\eta_{dr}$), and those of cash flows ($\eta_{cf}$). The series are calculated by OLS real-time. Figure 3.1 presents the series using an expanding window, and Figure 3.2 presents those using a rolling 30-year window.
Figure 1.5: Maximum Eigenvalues

Notes: This figure plots the estimated maximum eigenvalue of AR(1) coefficient matrix $B$ in the VAR system (1.6) under parameter uncertainty. Results under both learning schemes are presented: 1. rolling regression under naive learning; 2. expanding window regression under naive learning; and 3. BVAR under sophisticated learning.
Figure 1.6: Sophisticated Learning

Notes: This figure plots the smoothed series under Sophisticated Learning: the unexpected stock return ($\eta$), the forecast revisions of future discount rates ($\eta_{dr}$), and those of cash flows ($\eta_{cf}$). The series are estimated by real-time Bayesian VAR.
Chapter 2

On the Robustness of Variance Decompositions

2.1 Introduction

The market volatility decomposition of stock returns into news about discount rates (DR) and news about cash flows (CF) was first proposed by Campbell and Shiller (1988). One popular approach in the study of the DR news and CF news is to directly model the DR news by a vector autoregression process and back out the CF news as the residual component.

Despite the large number of influential articles published using this method and their important academic and policy implications, few studies question the robustness of this approach. One attempt is by Chen and Zhao (2008) who question the conventional wisdom by pointing out the possibility of model misspecification
due to a missing variable problem. They conclude that the variance of CF news is larger than the DR news, and the relative importance of the two news components to the market volatility is sensitive to the choice of state variables. In this chapter I tackle the robustness of variance decomposition from a new angle: estimation bias resulting from the inclusion of persistent state variables in the predictive return model.

It is well known that the simple linear regression model most often used for stock return predictability in fact raises some very tough econometric issues. The high degree of persistence found in many predictor variables, especially in those valuation ratios such as the earnings-price ratio or dividend-price ratio in typical return forecasting regressions, is at the root of most econometric problems associated with predictive regressions. The high persistence of the regressors, coupled with a strong contemporaneous correlation between the innovations in the regressor and the regressand causes standard OLS estimates to suffer from a small sample bias, and normal t-tests to have the wrong size; this is the so-called Stambaugh (1999) bias in predictive regressions.

Although the Stambaugh bias effect is widely studied and accepted in single-equation regressions, its effect on multi-variate VAR systems is unknown, partly due to the complication introduced by the co-linearity of VARs. The impact of the bias is even enlarged in the variance decomposition framework where the news components are basically infinite sum of those forecasts based on VARs.

In this chapter I study the effect of persistent variables on the relative
importance of news components with respect to market volatility. The plan of this chapter is as follows: Section 2 briefly presents the variance decompositions framework. Section 3 outlines the role of the persistency of state variables in predictive regressions and in variance decompositions. Experiments and Simulation results are presented. Section 4 concludes the chapter.

2.2 A Brief on Variance Decomposition

Understanding what moves stock prices is a key question in finance in both interpretation and forecasting. The Present Value (PV) model is one of the simplest models that addresses this issue: movements in stock prices, and thus returns, are due to investors’ change in expectations about the future risk premium (the "discount rate news" thereafter), and their change in expectations about future dividends (the "cash flow news" thereafter). Investors’ expectations are not directly observed, yet they could be estimated from the observed return series with assumptions on the dynamics in expected returns and cash flows.

Campbell and Shiller (1988) and others log-linearizes the PV model, and use estimates from full-sample vector autoregression (VAR) with constant parameters to simulate investors’ expectations and construct measures of the discount rate news and cash flow news based on the estimated expectations.

The basic equation for stock returns relates the unexpected stock return in period \( t + 1 \) to revisions in investors’ forecasts. I can express the period \( t + 1 \) unex-
pected return on equity \((r_{t+1} - E_t r_{t+1})\) in terms of the revision of the expectation of discounted future dividends and future returns:

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

where \(\Delta d_t\) denotes the dividend growth rate at period \(t\), and \(\rho\) is a constant discount factor. The above equation can be re-written as

\[
\eta_{t+1} = \eta_{cf,t+1} - \eta_{dr,t+1}
\]  

(2.1)

where

\[
\eta_{t+1} \equiv r_{t+1} - E_t r_{t+1}
\]

\[
\eta_{cf,t+1} \equiv (E_{t+1} - E_t) \left[ \sum_{i=0}^{\infty} \rho^i \Delta d_{t+1+i} \right]
\]

\[
\eta_{dr,t+1} \equiv (E_{t+1} - E_t) \left[ \sum_{i=1}^{\infty} \rho^i r_{t+1+i} \right]
\]

In practice, revisions to investors’ expectations are not directly observable. To conquer this problem, a vector autoregression (VAR) approach is usually used to obtain empirical proxies for revisions to investors’ expectations. The VAR system involves the variables of interest (stock returns in the context of this paper) along with any other state variables that might be helpful in forecasting those variables.
Specifically, we express a p-lag, n-variable VAR as a first-order system:

\[ z_t - \tilde{A} = B(z_{t-1} - \tilde{A}) + u_t \]  

(2.2)

where \( z_t \) is a properly stacked \( np \times 1 \) vector containing the variable of interest and other state variables, \( B \) captures the dynamics of the variables, and \( \tilde{A} \) is the unconditional mean vector. With perfect information on the VAR process as specified above, the investors revise their forecasts only when they observe the realized innovation \( u \), and the components of (2.1) are given by:

\[
\eta_{t+1} = e_1' u_{t+1} \\
\eta_{dr,t+1} = e_1'(I - \rho B)^{-1} \rho B u_{t+1}
\]

\[
\eta_{cf,t+1} = \eta_{t+1} + \eta_{dr,t+1}
\]

where \( e_1 \) is the appropriate selection vector. Based on the model above, it is easily seen that the variability of stock return comes from three sources: the variation in investors’ forecast revision of future cash flows, the variation in forecast revisions of future discount rates, and their comovements:

\[
var(\eta_{t+1}) = var(\eta_{cf,t+1}) + var(\eta_{dr,t+1}) - 2cov(\eta_{cf,t+1}, \eta_{dr,t+1}).
\]  

(2.4)

Each component is a linear function of the VAR model parameters \( B \) and the
variability of innovations \( \sigma_u \).

Table (2.1) lists the leading results in the current literature. The conventional wisdom is, as shown in the table, that discount rate news is the dominant factor in moving the stock price. It counts for about \( 2/3 \) of the total market volatility. The cash flow news, on the other hand, counts for around \( 20\%-30\% \), and there is only a negligible effect from the comovement of these two news components.

### 2.3 Persistent Predictors and the Sensitivity of Variance Decomposition

As discussed in the previous section, the empirical calculation of Variance Decompositions essentially relies on a VAR model that consists of the return prediction regression and the evolution of all other state variables over time. In order to get a reliable estimate of the variance ratios, it is thus important to study the property of the VAR estimates.

Stock returns are assumed to be predictable, based on lagged instrumental variables, in the current conditional asset pricing literature. Standard lagged variables include the levels of short-term interest rates, payout-to-price ratios for stock market indexes, and yield spreads between low-grade and high-grade bonds or between long- and short-term bonds. Many of these variables behave as persistent, or highly autocorrelated, time series. It is shown by many studies in the stock return literature that some persistent variables have predictive power over
future returns while at the same time from a statistical point of view, they might cause estimation bias.

2.3.1 Persistent Predictors in Stock Return Prediction

Among the popular candidates of state variables that have been shown to have some predictive power on future stock returns, there are two groups of predictors that possess high autocorrelation.

The first group contains interest rate related variables, including the T-Bill rate: the 3-Month Treasury Bill rate served as short-term interest rate. The other is the Term Spread (tms) defined as the difference between the long term yield on government bonds and the T-bill. The second group of persistent variables include valuation ratios. A valuation ratio is a measure of how cheap or expensive a security is, compared to some measure of profit or value, and is calculated by dividing a measure of price by a measure of value, or vice-versa. The most widely used valuation ratio is the PE ratio which compares the cost of a share to the profits made for shareholders per share. Other commonly used measures include the dividend-price ratio, dividend yield, book-to-market (BM) ratio, etc. They are widely recognized as having predictive power in forecasting stock returns.

The potential problem of using them as state variable in a prediction regression is that the measure of value, either earnings per share in PE ratio, dividend per share in dividend-price ratio, or the book value in BM ratio, are often times calculated as the moving average over a longer time and subject to management
manipulation. The consequent high persistency character dominates that of the ratio. At the same time the embedded measures of the market value of the asset in these ratios cause significant correlation with the predictor in a regression.

Table (2.2) lists some statistics of a set of commonly used persistent variables: 10-year smoothed PE ratio, BM ratio, log dividend yield, term spread, and T-bill rate. The data covers the period from 1926.1 to 2005.12. The third row "autocorr" shows the high persistency for all these series with the first order autocorrelation equal to 0.983 – 0.991. At the same time the lower panel of Table 1 shows that on average the valuation ratios are highly correlated with each other, with correlation ranging from 0.80 to 0.89.

To illustrate how severely the variance decomposition is affected by the high persistency of some predictive variables, I conducted several experiments based on real data.

2.3.2 Experiments with Real Data

I experiment with a four-variable VAR system, where the first one stands for stock returns with the rest three labeled $v_1$, $v_2$, and $v_3$ standing for state variables that has the potential to predict stock returns. Table (2.3) summaries the variables included in the experiment. Notice that the first state variable $v_1$ is the possible "trouble-maker" with a high autocorrelation of 0.99. The third panel in Table (2.3) is the parameter value of this VAR estimated by OLS. As listed in the last row of the table, in-sample estimation of the four-variable VAR and variance decomposition
yields the ratio of DR news to total market volatility around 60%, and that of CF news is 18%. These numbers are quite close to the convention wisdom.

Notice that variance ratios are calculated upon parameter matrix $B$ and the variance innovations only. To see the effect of persistency on variance decomposition, I change the persistency of the highly persistent variable $b_{22}$, keeping the other parameters constant, and track the response of the corresponding variance ratios afterwards. Specifically, I vary $b_{22}$ within a $[1 \pm 0.5\%]$ range of the OLS estimate, 0.9926, with each increment equal to 0.01%. The actual persistency varies from $[0.9877, 0.9975]$. Figure 2.1 plots the variance components vs. the value of $b_{22}$. From the figure we see that the variance components swing dramatically with small changes in the persistency parameter. In Table (2.4) I list the range of variations. At the lower end, a 0.5% decrease in $b_{22}$ caused a 44% decrease in the variance ratio of DR news, a 44% increase in that of the CF news and a more than 100% increase in that of the covariance between the two news components. The corresponding numbers at the higher end are 116%, 33%, and $-500\%$.

Changing the persistency of a state variable also changes the $R^2$-value of the prediction regression. To see how persistency affects the variance decomposition per se, the above experiment is adjusted as follows. Let

$$R_{v1,r} = \frac{(b_{12}\sigma_{v1})^2}{\sum (b_{1i}\sigma_{vi})^2 + \sigma^2_e}$$

(2.5)

\(^1b_{ij}\) denotes the $i^{th}$ row and $j^{th}$ column element of matrix $B$
where $R_{v1,r}$ is the contribution of $v_1$ to the total R-square of the return prediction regression, $R_r$. If the state variables are uncorrelated with each other, fixing $R_{v1}$ equals fixing $R_r$. There are several ways to maintain a constant $R_{v1}$. I reach this goal through the balancing effect between the predictability and the persistency of $v_1$. Specifically,

$$b_{12}\sigma_{v1} = b_{12} \frac{\sigma_{v1}^2}{1 - (b_{22})^2}$$

For a given experimental value $b_{22}$, a corresponding $b_{12}$ is calculated such that $b_{12}\sigma_{v1}$ and $R_{v1,r}$ is fixed, and thus is $R_r$. Figure 2.2 plots the change of variance ratios with persistency of $v_1$ given the total predictability of stock return is fixed. The sensitivity of the variance ratios is mitigated, given the adjustment of predictability. However, within the range of a 1% change in $b_{22}$, the variance ratios result vary about 20%.

### 2.3.3 Simulated Confidence Intervals

As illustrated in the previous section, small sample bias caused by persistent state variables and the non-linearity in the variance decomposition make it hard to draw analytical inference on the variance ratios. To get a more concrete idea as to how robust the ratios based on OLS estimates are, I present in this section the simulated confidence intervals.

The model is simulated based on a five-variable VAR including the monthly market return ($Rm$) and the other four state variables, all at monthly frequency:
the dividend growth rate \((Gd)\), the smoothed price-earnings ratio \((PE10)\), the dividend yield \((YLD)\), and inflation \((INFL)\). Detailed data description is provided in Appendix A. This is the same set of variables in the empirical part of Chapter 1, "Stock Return Variability, Forecast Revisions, and Investors’ Learning". For simplicity, all series are demeaned.

Assume the true data generating process is

\[
\tilde{z}_t = \tilde{B}\tilde{z}_{t-1} + \tilde{u}_t, \tag{2.6}
\]

where the relevant coefficient matrix \(\tilde{B}\) and innovations \(\tilde{u}_t\) are estimated from real data using OLS. Based on (2.6) I generate 1000 sets of simulated series with length \(T = 960\) corresponding to the small sample case, which is about the same data length we could get in reality; and \(T = 10000\) corresponding to the large sample case which is not available to investors or researchers but may reveal some asymptotic properties. The innovations are bootstrapped using random draws from \(\tilde{u}_t\) with replacement. For each set of simulated data, I perform OLS estimation of the VAR model and the objects of variance decomposition are calculated according to (2.3) and (2.4). They are then treated as the empirical distribution of the ratios based on real data. Figure 2.3 plots the histogram of the ratios in the small sample case with \(T = 960\). The three panels are \(\frac{\text{var}(\eta_{dr})}{\text{var}(\eta)}, \frac{\text{var}(\eta_{cf})}{\text{var}(\eta)}, \) and \(\frac{-2\text{cov}(\eta_{dr},\eta_{cf})}{\text{var}(\eta)}\) respectively. It is easily seen that the ratios are very disperse. In the top panel, the distribution of discount rate news ratio is roughly symmetric, and centered
around 75%. The tails of the histogram extends as far as 20% and 140%. The next two panels in the figure are the histograms of cash flow news ratio which skews to the right, and the contribution of the correlation of these two news components to the total market volatility which skews to the left. The majority of cash flow news ratio fall between 10% to 40%. The contribution from the correlation between the news component varies across zero. Table (2.5) presents the variance ratios from real data and two-sided 95% confidence intervals constructed from the empirical distribution of these ratios. Compared with the OLS result, the ratio of volatility from the DR news is pitched at 70.1% while having a confidence interval [39.6%, 119.2%]; the confidence interval for the CF news is [12.1%, 51.5%] and that of the covariance between the two news components is [−55.8%, 34.0%]. The message delivered from the table and the figure is that, it is hard to draw any qualitative conclusion regarding the relative importance of the two news components and thus the conventional wisdom which states the dominant role of DRs in the market may not be reliable without further robustness checks.

2.4 Conclusion

In this chapter I studied the robustness of variance decomposition based on OLS estimations of VAR. Both experiments with real data and simulations show that OLS tends to yield a very wide empirical confidence interval and thus conclusions made in the literature regarding the relative importance of news components
to the overall market is not reliable without further examination.
2.5 Tables and Figures
Table 2.1: Conventional Wisdom on Variance Decomposition

<table>
<thead>
<tr>
<th>Artical</th>
<th>Sample Period</th>
<th>(\text{var}(\eta_{dr})/\text{var}(\eta))</th>
<th>(\text{var}(\eta_{cf})/\text{var}(\eta))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernanke &amp; Kuttner (2005)</td>
<td>1973–2002</td>
<td>77.40%</td>
<td>24.50%</td>
</tr>
<tr>
<td>Campbell &amp; Vuolteenaho (2004)</td>
<td>1929 – 2001</td>
<td>80.66%</td>
<td>19.34%</td>
</tr>
<tr>
<td>Campbell (1991)</td>
<td>1927 – 1988</td>
<td>66.67%</td>
<td>33.33%</td>
</tr>
</tbody>
</table>

Note: This table lists several leading articals in the variance decomposition literature and their conclusion on the variance ratios.
Table 2.2: Summary Statistics on Persistent State Variables

<table>
<thead>
<tr>
<th>1926.1–2005.12</th>
<th>PE10</th>
<th>BtoM</th>
<th>YLD</th>
<th>TSP</th>
<th>T_Bill</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>2.758</td>
<td>0.605</td>
<td>-3.284</td>
<td>0.016</td>
<td>0.038</td>
</tr>
<tr>
<td>std</td>
<td>0.395</td>
<td>0.263</td>
<td>0.438</td>
<td>0.013</td>
<td>0.031</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.990</td>
<td>0.983</td>
<td>0.988</td>
<td>0.956</td>
<td>0.991</td>
</tr>
</tbody>
</table>

Correlation

<table>
<thead>
<tr>
<th></th>
<th>PE10</th>
<th>BtoM</th>
<th>YLD</th>
<th>TSP</th>
<th>T_Bill</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE10</td>
<td>1</td>
<td>-0.899</td>
<td>-0.856</td>
<td>-0.057</td>
<td>-0.126</td>
</tr>
<tr>
<td>BtoM</td>
<td>-0.899</td>
<td>1</td>
<td>0.805</td>
<td>-0.029</td>
<td>0.072</td>
</tr>
<tr>
<td>YLD</td>
<td>-0.856</td>
<td>0.805</td>
<td>1</td>
<td>-0.074</td>
<td>-0.160</td>
</tr>
<tr>
<td>TSP</td>
<td>-0.057</td>
<td>-0.029</td>
<td>-0.074</td>
<td>1</td>
<td>-0.365</td>
</tr>
<tr>
<td>T_Bill</td>
<td>-0.126</td>
<td>0.072</td>
<td>-0.160</td>
<td>-0.365</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: This table presents some statistics of persistent predictive variables for stock returns. There are two groups of variables: the first three (smoothed price-earnings ratio (PE10), book to market ratio (BtoM), and dividend yield (yld)) are valuation ratios; the rest two (term spread (TSP), and treasury bill rate (T_Bill)) are interest related variables.
Table 2.3: Summary Statistics on Benchmark case

<table>
<thead>
<tr>
<th>Benchmark case</th>
<th>r</th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.008</td>
<td>2.758</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>median</td>
<td>0.013</td>
<td>2.766</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td>std</td>
<td>0.056</td>
<td>0.395</td>
<td>0.013</td>
<td>0.005</td>
</tr>
<tr>
<td>min</td>
<td>-0.339</td>
<td>1.717</td>
<td>-0.094</td>
<td>-0.021</td>
</tr>
<tr>
<td>max</td>
<td>0.348</td>
<td>3.789</td>
<td>0.072</td>
<td>0.057</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.078</td>
<td>0.990</td>
<td>0.598</td>
<td>0.548</td>
</tr>
</tbody>
</table>

Correlation

<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>1</td>
<td>-0.008</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>v1</td>
<td>-0.008</td>
<td>1</td>
<td>0.088</td>
<td>-0.122</td>
</tr>
<tr>
<td>v2</td>
<td>0.004</td>
<td>0.088</td>
<td>1</td>
<td>0.147</td>
</tr>
<tr>
<td>v3</td>
<td>0.003</td>
<td>-0.122</td>
<td>0.147</td>
<td>1</td>
</tr>
</tbody>
</table>

OLS Estimates

<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>0.078</td>
<td>-0.013</td>
<td>0.195</td>
<td>-0.550</td>
</tr>
<tr>
<td>v1</td>
<td>0.505</td>
<td>0.993</td>
<td>0.079</td>
<td>-0.954</td>
</tr>
<tr>
<td>v2</td>
<td>0.010</td>
<td>0.002</td>
<td>0.579</td>
<td>0.236</td>
</tr>
<tr>
<td>v3</td>
<td>0.003</td>
<td>-0.001</td>
<td>0.039</td>
<td>0.528</td>
</tr>
</tbody>
</table>

Variance dectp.  \frac{\text{var}(\eta_{dr})}{\text{var}(\eta)} \frac{\text{var}(\eta_{cf})}{\text{var}(\eta)} \frac{-2\text{cov}(\eta_{dr},\eta_{cf})}{\text{var}(\eta)}

64.8% 19.0% 16.1%

Note: This table presents the benchmark case in simulation, where the VAR contains stock return (r) and three predictive variables (v1, v2, v3). The first panel contains summary statistics, followed by the second panel with correlations between these variables. The third panel lists VAR coefficients estimated by OLS. The last panel reports variance decomposition results based on OLS estimates.
Table 2.4: Sensitivity of Variance Decomposition on Variable Persistency

<table>
<thead>
<tr>
<th></th>
<th>lower bound</th>
<th>center</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{22}$</td>
<td>0.9877</td>
<td>0.9926</td>
<td>0.9975</td>
</tr>
<tr>
<td>$\Delta$ (%)</td>
<td>99.5%</td>
<td>100%</td>
<td>100.5%</td>
</tr>
<tr>
<td>$\frac{\text{var}(\eta_{dr})}{\text{var}(\eta)}$</td>
<td>36.6%</td>
<td>64.8%</td>
<td>140.1%</td>
</tr>
<tr>
<td>$\Delta$ (%)</td>
<td>56.4%</td>
<td>100.0%</td>
<td>216.1%</td>
</tr>
<tr>
<td>$\frac{\text{var}(\eta_{cf})}{\text{var}(\eta)}$</td>
<td>27.5%</td>
<td>19.0%</td>
<td>25.3%</td>
</tr>
<tr>
<td>$\Delta$ (%)</td>
<td>144.4%</td>
<td>100.0%</td>
<td>133.0%</td>
</tr>
<tr>
<td>$\frac{-2\text{cov}(\eta_{dr},\eta_{cf})}{\text{var}(\eta)}$</td>
<td>35.9%</td>
<td>16.1%</td>
<td>-65.4%</td>
</tr>
<tr>
<td>$\Delta$ (%)</td>
<td>222.7%</td>
<td>100.0%</td>
<td>-405.4%</td>
</tr>
</tbody>
</table>

Note: This table reports the range of experiment values of the persistency of $v2$, the corresponding changes in the variance ratios, and the changes in percentage, $\Delta$ (%).
Table 2.5: Confidence Interval of Variance Decomposition

<table>
<thead>
<tr>
<th></th>
<th>( \text{var}(\eta_{\text{dr}}) )</th>
<th>( \text{var}(\eta_{\text{cf}}) )</th>
<th>( \text{var}(\eta) )</th>
<th>( \frac{\text{var}(\eta_{\text{dr}})}{\text{var}(\eta)} )</th>
<th>( \frac{\text{var}(\eta_{\text{cf}})}{\text{var}(\eta)} )</th>
<th>( \frac{-2\text{corr}(\eta_{\text{dr}}, \eta_{\text{cf}})}{\text{var}(\eta)} )</th>
<th>\text{corr}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>0.0021</td>
<td>0.0007</td>
<td>0.0030</td>
<td>70.1%</td>
<td>23.6%</td>
<td>6.4%</td>
<td>-0.875</td>
</tr>
<tr>
<td>Simulation with data length = 960</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>lower bound</strong></td>
<td>0.0012</td>
<td>0.0004</td>
<td>0.0025</td>
<td>39.6%</td>
<td>12.1%</td>
<td>-55.8%</td>
<td>-0.943</td>
</tr>
<tr>
<td><strong>upper bound</strong></td>
<td>0.0037</td>
<td>0.0016</td>
<td>0.0037</td>
<td>119.2%</td>
<td>51.5%</td>
<td>34.0%</td>
<td>-0.727</td>
</tr>
<tr>
<td>Simulation with data length = 10000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>lower bound</strong></td>
<td>0.0018</td>
<td>0.0005</td>
<td>0.0029</td>
<td>59.3%</td>
<td>17.1%</td>
<td>-12.3%</td>
<td>-0.912</td>
</tr>
<tr>
<td><strong>upper bound</strong></td>
<td>0.0027</td>
<td>0.0009</td>
<td>0.0032</td>
<td>87.3%</td>
<td>30.0%</td>
<td>18.0%</td>
<td>-0.838</td>
</tr>
</tbody>
</table>

Note: This table reports simulated confidence intervals for variance of news components and their relative importance to the total market volatility. The row "Data" presents estimates of the variance and variance ratios based on OLS method. The results in the rest of the table are based on 1000 simulations.
Figure 2.1: Persistent State Variables vs. Variance Decomposition
Figure 2.2: Persistent State Variables vs. Variance Decomposition – R2 adjusted
Figure 2.3: Histogram of Variance Decomposition Ratios based on Simulation
Chapter 3

Return Predictability Under Equilibrium Constraints on the Equity Premium Returns

3.1 Introduction

Over the last twenty years, the stock return predictability literature has influenced a broad range of areas such as performance evaluation, asset pricing and asset allocation.\footnote{Papers on time-series predictability of stock returns include Campbell (1987), Campbell and Shiller (1988), Fama and French (1988, 1989), Ferson and Harvey (1991), Keim and Stambaugh (1986) and Pesaran and Timmermann (1995). Examples of asset allocation studies under return predictability include Ait-Sahalia and Brandt (2001), Barberis (2000), Brennan, Schwartz and Lagnado (1997), Campbell and Viceira (1999), Kandel and Stambaugh (1996) and Xia (2001). Avramov and Wermers (2006) and Ferson and Schadt (1996) consider mutual fund performance under time-varying investment opportunities.} This influence is largely due to studies by Campbell (1987),
Campbell and Shiller (1988), Fama and French (1988, 1989), Ferson and Harvey (1991), and Keim and Stambaugh (1986) who provided convincing economic arguments and in-sample empirical results that some of the fluctuations in returns are predictable because of persistent time variation in expected returns. The in-sample evidence for predictability is accumulating as various new variables have been suggested as predictors of excess returns (Hodrick (1992), Pontiff and Schall (1998), Lamont (1998), Baker and Wurgler (2000), Lettau and Ludvigson (2001), Polk, Thompson, and Vuolteenaho (2006), among others). The out-of-sample predictability evidence, however, has been much less conclusive. Recent studies by Paye and Timmermann (2006) and Lettau and Van Nieuwerburgh (2007) argued that predictability weakened or disappeared during the 1990s. Bossaerts and Hillion (1999) and Goyal and Welch (2003, 2007) provide an even sharper critique by arguing that predictability was largely an in-sample or \textit{ex-post} phenomenon which disappears once the forecasting models are used to guide forecasts on new, out-of-sample, data.

A shortcoming of the forecasting models used throughout the finance literature is that, while the common state variables are broadly guided by theoretical considerations, finance theory provides little guidance for the choice of functional form of the forecasting model. Largely as a consequence, linear forecasting models are used almost exclusively. As pointed out by Campbell and Thompson (2007, CT henceforth), a problem with these models is that the implied conditional equity premium often turns negative. It is difficult to imagine an equilibrium setting
where risk-averse investors would hold stocks if their expected compensations were negative. CT argue that the out-of-sample or ex-ante forecasting performance of return prediction models can be improved by imposing theoretical constraints such as non-negativity of the conditional equity premium or a sign constraint on the coefficient of a given predictor variable. Whenever any of these constraints is violated, CT impose their constraints by truncating the return forecast at the unconditional estimate of the equity premium, i.e., the prevailing mean. While this can be viewed as a first approximation to imposing moment or parameter constraints, the approach fails to make full use of the information in the theoretical constraints.

In this paper we propose a new method that optimally incorporates theoretical constraints including, but not limited to, those proposed by CT. Our approach is based on Bayesian techniques which make imposing an arbitrary number of constraints computationally feasible. We show how to efficiently update the estimates of the restricted forecasting model every time new observations on returns and the predictor variable become available. Theoretical constraints on the conditional equity premium can indeed have a big impact on parameter estimates of the return forecasting model. To see this, suppose, for example, that a new observation on returns and the predictor variable becomes available that, under the previous parameter estimates, imply a negative conditional equity premium at some point in the sample. Since our approach makes use of this observation to inform the updated parameter estimate, the old parameter estimates would need
to be revised so as to ensure that the conditional equity premium is always positive. The theoretical constraint therefore allow investors to more efficiently update their beliefs about the parameters of the forecasting model. We argue that this is a highly attractive feature of our method since the constraints proposed by Campbell and Thompson—such as non-negativity of the conditional equity premium—are overwhelmingly supported by equilibrium arguments and hence should be fully exploited.

When implemented along the lines proposed in our paper, the economically motivated constraints turn out to be highly informative and lead to far more precise estimates of the parameters of the return forecasting model. Intuition for this surprising finding is that every time a new pair of observations on the predictor variable and returns becomes available, the non-negativity constraint on the conditional equity premium is used to rule out values of the parameter that are infeasible given the sign constraint. Since the conditional equity premium must be non-negative at each point in time, in a sample of $T$ observations, we have $T$ constraints rather than just a single constraint.

Through a set of Monte Carlo simulations we show that the better performance of our new forecasting approach can be understood in terms of a reduction in the bias of the slope coefficient known as the Stambaugh bias (Stambaugh (1986, 1999)) as well as smaller parameter estimation errors, i.e. a reduction in estimation uncertainty. The bias shifts estimates of the coefficient on variables such as the dividend yield away from zero. To see how this bias will be reduced (and
eventually removed as the sample size increases) in our context, suppose that the true coefficient on the dividend yield is zero but that the bias is such that the coefficient estimate is positive in the absence of any constraints on the equity premium. As new observations of the dividend yield below its sample average emerge, the effect of imposing a sign constraint on the conditional equity premium is to shrink the distribution of the estimated coefficient towards zero - otherwise the predicted value would become negative. The smaller the value of the dividend yield, the stronger this effect is likely to be and so the approach ensures faster learning in the sense that the dispersion of the distribution of the estimated parameter gets reduced more rapidly than in the absence of any constraints. Individual observations—particularly those at odds with the theoretical constraints—can therefore lead to large (and instantaneous) shifts in the entire distribution of the parameter estimates. In contrast, approaches that ignore theoretical constraints when updating the parameter estimates will repeatedly make the same mistakes (i.e. predict negative stock returns).^2

^2The reduced bias associated with the constrained forecasting models also means that our approach also provides a new way to handle spurious predictability, a phenomenon that could well explain the difference between the apparently strong in-sample predictability and weak out-of-sample predictability of returns (see Ferson et al. (2003))

Predictability from individual forecasting variables in monthly return regressions is likely to be relatively weak and the literature on predictability of asset returns has documented predictability from several regressors. We therefore next consider making use of multivariate forecasting models. Return predictability in the context of multivariate regressions poses complications since it is difficult to
impose sign restrictions on the coefficients of the individual predictor variables. Yet, this is often the type of restriction that economic theory implies. To simultaneously deal with sign restrictions on the individual predictor variables and incorporate information from several predictor variables, we propose to use combination methods that combine forecasts from several univariate return forecasting models each of which imposes such sign restrictions and also do not allow the conditional equity premium to be negative. We implement this strategy using Bayesian Model Averaging, a technique that has also been used in the return forecasting literature by Avramov (2004). This approach succeeds in both preserving the individual sign restrictions and in imposing that the conditional equity premium be non-negative. We find that there are considerable gains from forecast combinations that satisfy these sign constraints and use multivariate information.

The plan of the paper is as follows. Section 2 shows how to efficiently incorporate theoretical constraints on the forecasting models and outlines our proposed methodology. Section 3 presents empirical estimation results for a range of predictor variables while Section 4 studies the forecasting performance of both unconstrained and constrained return models. Section 5 considers the effect of the constraints on the bias and estimation error in the model parameters while Section 6 concludes.
3.2 Methodology

This section describes our new methodology to estimate the return forecasting model subject to a set of constraints motivated by finance theory. These take the form of inequality constraints on the conditional equity premium or constraints on the sign of coefficients relating state variables to the equity premium. Constraints on the signs of state variables are best understood in a univariate context since many of the predictor variables proposed in the literature are strongly correlated with each other and their signs can change in multivariate regressions.

3.2.1 Constraints on the Return Forecasting Model

The literature on predictability of stock returns is extensive. Early studies such as Campbell and Shiller (1988) and Fama and French (1988) found evidence that stock returns could be predicted by means of valuation ratios, while Fama and Schwert (1977), Keim and Stambaugh (1986) and Campbell (1987) found predictability from the T-bill rate or yields on long-term corporate and government bonds. Subsequent studies have explored information in corporate financing (Baker and Wurgler (2000)), consumption-wealth ratios (Lettau and Ludvigsson (2001)) and the value of high versus low beta stocks (Polk, Thompson and Vuolteenaho (2006)).

Almost invariably, return predictability has been explored in the context of the following simple unconstrained linear forecasting model for the stock return at
time \( t \), \( r_t \), measured in excess of a risk-free rate:

\[
\begin{align*}
  r_t &= \mu + \beta x_{t-1} + \epsilon_t. \\
\end{align*}
\]  

(3.1)

Here \( x_{t-1} \) is the lagged value of the predictor variable and \( \epsilon_t \) has zero mean and variance \( \sigma^2 \).

This model is attractive since it is simple to interpret and only requires estimating two mean parameters, \( \mu \) and \( \beta \). Finance theory generally does not restrict the functional form of the mapping from the state variable, \( x_{t-1} \), to the excess return, \( r_t \), so the use of the linear specification in (3.1) should be viewed as an approximation. Campbell and Thompson (2007) argue that finance theory can be used to improve on the model. In particular, the conditional equity premium should be non-negative since it is difficult to imagine that markets for stocks can clear while the conditional equity premium is negative. They implement this insight by proposing a truncated forecast which is simply the largest of the unconstrained OLS forecast and zero:

\[
\hat{r}_t = \max(0, \hat{\mu} + \hat{\beta} x_{t-1}),
\]

(3.2)

where \( \hat{\mu} \) and \( \hat{\beta} \) are the OLS estimates from (3.1). While this truncation prevents the predicted equity premium from becoming negative, the theoretical constraint is not used to obtain improved estimates of \( \mu \) and \( \beta \). While potentially an improvement over the simple unconstrained model, this approach therefore does not make efficient use of the theoretical constraints.
To efficiently exploit the information embedded in the constraint that the conditional equity premium is non-negative, the parameters $\mu$ and $\beta$ should be estimated subject to the conditional equity premium constraint that $\mu + \beta x_{\tau - 1} \geq 0$ for $\tau = 1, \ldots, t$:

$$r_{\tau} = \mu + \beta x_{\tau - 1} + \epsilon_{\tau}.$$  

(3.3)

$$\mu + \beta x_{\tau - 1} \geq 0 \ (\tau = 1, \ldots, t)$$

Although the conditional equity premium constraint is not directly a constraint on the model parameters, $\theta = (\mu, \beta)$, it clearly affects these parameters which have to be selected so as to be consistent with $\mu + \beta x_{\tau - 1} \geq 0$ for $\tau = 1, \ldots, t$. Note that the conditional equity premium constraint has to hold at each point in time, so the number of constraints grows in proportion with the length of the sample size. The seemingly simple equity premium constraint therefore potentially yields a very powerful way to tie down the parameters of the return forecasting model and obtain more precise estimates.

In many situations finance theory is informative about the sign of the slope coefficient, $\beta$, relating returns to the state variable, $x$. To cover such cases, we also consider a specification that estimates the forecasting model by first imposing a
sign constraint on $\beta$ and then imposing the conditional equity premium constraint:

$$ r_\tau = \mu + \beta x_{\tau - 1} + \epsilon_\tau $$  \hspace{1cm} (3.4) \\
$$ \beta I_x \geq 0, \mu + \beta x_{\tau - 1} \geq 0 \quad (\tau = 1, ..., t) $$  \hspace{1cm} (3.5)

Here $I_x$ is an indicator function that is either $+1$ or $-1$, depending on the sign of the constraint.

We next explain how the models are estimated and how the constraints are imposed.

### 3.2.2 Accounting For Constraints through Investors’ Prior Beliefs

The theoretical constraints incorporated in the models (3.3) and (3.4) are naturally interpreted as reflecting the forecaster’s prior beliefs on return predictability. Viewed this way, they can best be imposed using Bayesian techniques. Estimation of return forecasting models subject to these constraints therefore requires specifying priors for the regression coefficients and introducing inequality constraints on the model parameters as specified by models (3.3) and (3.4) through the priors. To this end we study two sets of priors: a set of Normal-Gamma priors and a set of conditional Normal-Jeffreys’ priors. We consider both types of priors to establish the robustness of our results. In what follows we first introduce the priors without constraints and then demonstrate how to incorporate the constraints.
Basic Priors

The first prior we consider for the unconstrained univariate return prediction models is the standard Normal-Gamma prior. Under this prior the parameters of the return model, i.e. the mean parameters $\theta \equiv (\mu, \beta)$ and the precision parameter $h = \sigma^{-2}$, follow independent Normal-Gamma distributions:

$$
\theta \sim N(\theta, V) \quad (3.6)
$$

$$
h \sim G(v, s^2).
$$

The first two moments of the parameters of interest are: $E(\theta) = \theta$, $\text{var}(\theta) = V$, $E[h] = v \times s^{-2}$ and $\text{var}(h) = v \times s^{-4}$ and so the parameters $\theta$, $V$, $s^2$ and $v$ fully characterize the priors. Moreover, it follows from the independence assumption that the joint prior distribution is simply the product of each part

$$
P(\theta, h) = P(\theta)P(h). \quad (3.7)
$$

Following Wachter and Warusawitharana (2007), the second prior is a conditional Normal-Jeffreys prior. We assume the investor holds non-informative (Jeffreys prior) beliefs on the intercept and precision parameter, while the prior concerning return predictability, measured by the slope coefficient, $\beta$, is allowed to depend on the other parameters:

$$
\beta|\mu, h \sim N(\beta, \varphi/h). \quad (3.8)
$$
Letting $P(\mu, h)$ be the prior probability density on $(\mu, h)$, the joint prior distribution on the whole set of parameters is

$$P(\theta, h) = P(\beta|\mu, h)P(\mu, h).$$  \hspace{1cm} (3.9)

We start by deriving a limiting Jeffreys prior on the full set of parameters $(\theta, h)$. Following Stambaugh (1999),

$$P(\theta, h) \propto (\det I(\theta, h))^{1/2}$$

where $I(\theta, h)$ is the Fisher information matrix. Lastly, combining this part with the conditional Normal density, we get

$$P(\theta, h) \propto (\varphi)^{-1/2}h^{1/2}\exp\left(-\frac{h(\beta - \hat{\beta})^2}{2\varphi}\right)h^{-1/2} = \varphi^{-1/2}\exp\left(-\frac{(\beta - \hat{\beta})^2}{2\varphi}\right)h. \hspace{1cm} (3.10)$$

Next we show how to incorporate constraints on these priors.

**Prior Beliefs under Constraints**

Both the constraint on the sign of the slope coefficient, $\beta$, and the conditional equity premium constraint restricts the mean parameters $\theta = (\mu, \beta)$ and hence take the following form:

$$\theta \sim F_0 \times I(\theta \in A),$$  \hspace{1cm} (3.11)
where $F_0$ is the prior distribution of $\theta$ without constraints. In the context of this paper, under the Normal-Gamma prior, $F_0 = N(\theta, V)$, while under the Normal-Jeffreys prior, $F_0 = \mathcal{N}(0, \varphi/h)$. $I(\theta \in A)$ is an indicator function that equals unity if $\theta \in A$ and is zero otherwise. $A$ is the admissible region for the regression coefficients as reflected in the theoretical constraints on the models (3.3) and (3.4). Specifically, under (3.3) $A$ is the set of parameter values satisfying that the predicted return, $\hat{r}_\tau \geq 0$, $\tau = 1, ..., t$. Hence the conditional equity premium constraint is equivalent to restricting $\theta$ to lie in the set $A$:

$$A = \{ \mu + x_{\tau-1} \cdot \beta \geq 0, \text{ for } \tau = 1, 2, ..., t \}$$

$$= \{ \Lambda \theta \geq 0 \}, \quad (3.12)$$

where $\Lambda = \begin{pmatrix} 1 & \max(X_{\tau-1}) \\ 1 & \min(X_{\tau-1}) \end{pmatrix}, \tau = 1, 2, ..., t$

Similarly, letting $\text{sign}(\beta) = \pm 1$ denote the prior belief on the sign of the slope coefficient, for the model that imposes constraints on both the sign of $\beta$ and on
the conditional equity premium, we have:

\[ A = \{ \mu + x_{\tau - 1} \cdot \beta \geq 0, \text{for } \tau = 1, 2, \ldots, t; \text{ and } sign(\beta) \cdot \beta \geq 0 \} \]

\[ = \{ \Psi \theta \geq 0 \}, \quad (3.13) \]

where \( \Psi = \begin{pmatrix} 1 & f(X) \\ 0 & sign(\beta) \end{pmatrix} \),

and \( f(X) = \begin{cases} \min(X_{\tau - 1}) & \text{if } sign(\beta) = 1 \\ \max(X_{\tau - 1}) & \text{if } sign(\beta) = -1 \end{cases}, \tau = 1, 2, \ldots, t. \]

\[ \text{3.2.3 Choice of Priors} \]

So far we have explained how we impose the constraints implied by the priors that the conditional equity premium is non-negative and/or the sign of the slope coefficient. We next explain our specific choice of prior parameters \((\theta, V, \varphi, \varphi^2)\) in the Normal-Gamma case and \((\beta, \varphi)\) in the Normal-Jeffreys case. We assume, first, that investors hold prior beliefs that stock returns are not predictable and, second, that investors hold diffuse priors about the remaining parameters.

Starting with the Normal-Gamma case, our prior reflects the “no predictability” view that the best predictor of the stock return is the historical average. At each point in time \( t \), we therefore center the prior intercept, \( \mu \), on the prevailing mean of historical excess returns, while the prior slope coefficient is centered on
zero, $\beta = 0$. The prior precision for the mean parameters is $V = \psi \times I_k$ where $I_k$ denotes the k-dimensional identity matrix (we only consider univariate models so $k = 2$) and $\psi$ is a scaling factor that controls the tightness of the prior. We consider values $\psi = 0.1$ and $\psi = 1$. For the prior belief on $h$, we set $\gamma = 1$ and $\sigma^2 = 4$ so the Gamma distribution reduces to an exponential distribution which has a significant probability mass near zero and reflects a diffuse view.$^3$

Turning to the Normal-Jeffreys prior, we continue to assume “no predictability”, i.e. $\beta = 0$. Besides, it is reasonable to let investors’ priors on $\beta$ depend on the variation in the predictor variable ($x$): A high variance of the predictor variable (captured by $\sigma_x$ ) might lower the spread of the prior on $\beta$. Thus, we rewrite $\varphi$ in terms of $\sigma_x$ and a prior scale parameter $\sigma_\beta$, that is common across all predictor variables irrespective of their variance: $\varphi \equiv \sigma_\beta^2 \times \sigma_x^{-2}$. We consider values $\sigma_\beta = 0.02$ and $\sigma_\beta = 0.2$ and, at each point in time, set $\sigma_x$ equal to the historical standard deviation of the explanatory variable, $x$. This conditional prior on $\beta$ also reflects a prior on the population $R^2$, since $R^2 = \beta^2 \sigma_x^2 / (\beta^2 \sigma_x^2 + \sigma_h^2)$, see Wachter and Warusawitharana (2007).

$^3$To see this, consider the probability density function (pdf) of the gamma distribution: $f(x; \nu, s^2) = x^{\nu-1} \frac{e^{-s^2 x}}{\Gamma(\nu)}$, for $x > 0$. In general, this pdf has a hump shape, but for $\nu = 1$ it reduces to $f(x; s^2) = s^2 \exp(-s^2 x)$, which is the pdf of an exponential distribution. The value of this function at zero is infinite and it exhibits an exponential decay thereafter.
3.2.4 Posterior Distributions

We next derive the posterior distributions under the two priors considered so far. Assuming that returns are normally distributed, we can use standard results from Bayesian analysis to obtain formulas for the posterior distribution of the mean and precision parameters.

First we introduce some notations needed for our analysis. Note that the return model can be written in matrix notations

\[
Y_t = X_t \theta + \varepsilon_t,
\]

where

\[
\begin{align*}
Y_t &= [r_1, r_2, \ldots, r_t]' \\
X_t &= \begin{bmatrix} 1, 1, 1, \ldots, 1 \\
x_0, x_1, x_2, \ldots, x_{t-1} \end{bmatrix}' \\
Z_t &= \{Y_t, X_t\} \\
\varepsilon_t &= [e_1, e_2, e_3, \ldots, e_t]' \sim N(0, \sigma^2 I_t)
\end{align*}
\]

In the following, for simplicity we ignore the subscript \( t \).

Under the Normal-Gamma prior, conditional on \( h \) and the data up to time \( t \), the posterior density of \( \theta \) is given by:

\[
\theta|h, Z \sim N(\overline{\theta}, H^{-1}) \times I(\theta \in A),
\]  
(3.14)
where $\bar{\theta}$ and $\bar{H}^{-1}$ is the posterior mean and covariance of the parameters. These moments of the posterior distribution are given by $\bar{H} = H + hX'X$ and $\bar{\theta} = \bar{H}^{-1}(H\theta + hX'Y) = \bar{H}^{-1}hX'Y$.\(^4\)

Conditional on $\theta$ and the data, $Z$, the posterior density of $h$ is given by:

$$h|\theta, Z \sim G(\bar{v}, \bar{S}^2)$$ \hfill (3.15)

where $\bar{v} = T + 2$ and $\bar{S}^2 = s^2 + (Y - X\theta)'(Y - X\theta)$. The unconstrained model (3.1) is nested as a special case when $A = R^2$. Analytical results are available for the unconstrained model (3.1) which does not impose any restrictions on the parameter estimates and thus preserves the full (non-truncated) distribution so that $I(\theta \in A)$ is always Identity.

To derive similar results under the Normal-Jeffreys prior, we start from the joint kernel of the posterior distribution:

$$P(\theta, h|Z) \propto P(Z|\theta, h)P(\theta, h) \times I(\theta \in A).$$ \hfill (3.16)

$$\propto h^{T/2} \exp\{-\frac{h}{2}(Y - X\theta)'(Y - X\theta)\} \varphi^{-1/2} \exp\{-\frac{\beta^2}{2\varphi} h\} \times I(\theta \in A).$$

If we interpret the above formula as a function of $h$ only, then it is a posterior

\(^4\)The priors are, as discussed earlier, $\theta = (\bar{Y}, 0)$ and $H = \psi I_2$. 
kernel for $h$ conditional on $\theta$:

$$P(h|\theta, Z) \propto h^{T/2} \exp\{-\frac{h}{2} (Y - X\theta)'(Y - X\theta)\} \exp(-\frac{\beta^2}{2\varphi} h)$$

$$= h^{((T+2)/2)\exp\{-\frac{h}{2} [\frac{\beta^2}{2\varphi} + (Y - X\theta)'(Y - X\theta)]\}.}$$

This is the kernel of a Gamma distribution:

$$h|\theta, Z \sim G(\overline{v}, \overline{S^2}),$$

where $\overline{v} = T + 2$ and $\overline{S^2} = \frac{\beta^2}{\varphi} + (Y - X\theta)'(Y - X\theta)$.

Conversely, if (3.16) is interpreted as a function of $\theta$ only, then it is a posterior kernel for $\theta$ conditional on $h$:

$$P(\theta|h, Z) \propto h^{T/2} \exp\{-\frac{h}{2} (Y - X\theta)'(Y - X\theta)\} \exp(-\frac{\beta^2}{2\varphi} h) \times I(\theta \in A).$$

$$= h^{T/2} \exp\{-\frac{1}{2} [(\theta - \overline{\theta})' \overline{H}(\theta - \overline{\theta}) + h(Y - X\theta)'(Y - X\theta)]\} \cdot I(\theta \in A),$$

This is the kernel of a Normal distribution:

$$\theta|(h, Z) \sim N(\overline{\theta}, \overline{H}^{-1}),$$

whose posterior moments are $\overline{H} = \overline{H} + hX'X$ and $\overline{\theta} = \overline{H}^{-1}(\overline{H}\overline{\theta} + hX'Y) = \overline{H}^{-1}hX'Y.$

$^5$For this case the priors are $\theta = [0, 0]$ and $H = \begin{bmatrix} 0 & 0 \\ 0 & h\varphi^{-1} \end{bmatrix}$. 

Comparing the posterior distributions under the two sets of priors, we see that they could be unified under a single Normal-Gamma prior framework, but with different specifications of the priors $[\theta, H]$.

### 3.2.5 The Gibbs Sampler

Unfortunately, closed-form expressions are not available for the constrained models (3.3) and (3.4). Estimating the parameters of these models requires evaluating the posterior distribution of the parameters given the data up to time $t, Z$, denoted $\pi(\theta, \sigma^{-2}|Z)$. This in turn requires repeatedly drawing from the distribution $\pi$ which is not always feasible in our context. Hence we cannot use Monte Carlo integration methods to simulate posterior moments of functions of the parameters. Instead we implement the Gibbs sampler and use importance sampling techniques which we next describe.

To implement the Gibbs sampler, we partition the parameters $\delta = (\theta, \sigma^{-2})$ into two blocks:

$$\delta_{(1)} = \sigma^{-2}, \delta_{(2)} = \theta. \quad (3.17)$$

Given an initial draw $\delta^{(0)} = (\delta_{(1)}^{(0)}, \delta_{(2)}^{(0)})$ from $\pi(\delta|Z)$, we successively draw new parameters

$$\delta_{(b)}^{(s)} \sim \pi \left( \delta_{(b)}^{(s)} | \delta_{(< (b), s)}^{(s-1)}, \delta_{(> (b), Z)}^{(s-1)} \right), \quad b = 1, 2; \quad s = 1, 2, \ldots. \quad (3.18)$$
The resulting sequence, \{s^{(s)}\}, is a realization of a Markov chain. Under well-known conditions the Markov chain converges (Roberts and Smith (1994)) and any single iterate \(s^{(s)}\) retains the property that it is draw from the joint density \(\pi(\theta, \sigma^{-2} | Z)\).

The Markov chain (3.18) requires sampling from the two conditional densities (3.14) and (3.15). Drawing from (3.15) is straightforward and can be carried out by many statistical packages. Drawing from (3.14), however, is non-standard and requires using importance sampling techniques. Suppose that random draws \(\theta^s, s = 1, ..., S\) can be generated from a density, \(q(\theta)\), the so-called importance function. By appropriately weighting the random draws from \(q(\theta)\), the moments computed from the draws of the importance function, \(\theta^s\), converge to the moments obtained from the (unknown) posterior distribution \(\pi(\theta | Z_t)\).\(^6\)

For importance sampling to work, \(q(\theta)\) needs to approximate \(\pi(\theta | Z_t)\) quite well. Otherwise cases can be found where \(\omega(\theta^s)\) is equal to zero for virtually every draw and the weighted average involves very few draws. Thus, importance sampling may become inaccurate unless \(q(\theta)\) is chosen carefully. Fortunately, the problem of finding an accurate importance function is easily resolved for the

\(^6\) This property makes use of the result in Geweke (1989) that if \(\theta^s, s = 1, ..., S\) is a random sample from \(q(\theta)\), then under weak conditions

\[
\frac{\sum_{s=1}^{S} \omega(\theta^s) g(\theta^s)}{\sum_{s=1}^{S} \omega(\theta^s)} \rightarrow E[g(\theta) | F_t],
\]

where the weights of the importance function, \(\omega(\theta^s)\), are given by

\[
\omega(\theta^s) = \frac{\pi(\theta = \theta^s | F_t)}{q(\theta = \theta^s)}.
\]
linear regression model that is subject to inequality constraints. By setting the importance function equal to the unconstrained posterior distribution, the weights can be computed as

$$\omega(\theta^s) = I(\theta^s \in A).$$

(3.19)

Hence the weights are either one (if $\theta^s \in A$) or zero (if $\theta^s \notin A$) and this strategy simply involves drawing from the unrestricted posterior distribution and discarding draws that violate the relevant inequality restrictions. Hence our approach is very simple to implement in practice.

### 3.3 Empirical Results

In this section we present empirical results from applying the methods described in the previous section to forecast stock returns.

#### 3.3.1 Data

Our empirical analysis uses the data on monthly stock returns along with a set of sixteen predictor variables analyzed in Goyal and Welch (2007). Stock returns are measured by the S&P500 index and include dividends. A short T-bill rate is subtracted from stock returns in order to capture excess returns. Data samples vary considerably across the individual predictor variables. To be able to compare —and later combine— results across the individual predictor variables.

---

7We are grateful to Amit Goyal for providing this data.
variables, we use the longest common sample which goes from 1940-2005.\footnote{One variable, the cross-sectional premium, only has data up to the end of 2003.}

The identity of the predictor variables is listed in Table 1. Most variables fall into three broad categories, namely (i) valuation ratios capturing some measure of ‘fundamentals’ to market value such as the dividend price ratio, the dividend yield, the earnings-price ratio, the 10-year earnings-price ratio or the book-to-market ratio; (ii) measures of bond yields capturing level effects (the three-month T-bill rate and the yield on long term government bonds), slope effects (the term spread), and default risk effects (the default yield spread defined as the yield spread between BAA and AAA rated corporate bonds, and the default return spread defined as the difference between the yield on long-term corporate and government bonds); (iii) estimates of equity risk such as the cross-sectional equity premium (the relative valuations of high- and low-beta stocks), long term return and stock variance (a volatility estimate based on daily squared returns). Finally, two corporate finance variables, namely the dividend payout ratio (the log of the dividend-earnings ratio), and net equity expansion (the ratio of 12-month net issues by NYSE-listed stocks over the year-end market capitalization) and a macroeconomic variable, inflation (the rate of change in the consumer price index) are considered.

### 3.3.2 Effect of Constraints on coefficient estimates

Before turning to the forecasts of stock returns, we consider the posterior distribution of the coefficient estimates based on the full data sample available
at the end of 2005. These contain interesting information about the economic significance of the various predictor variables.

Towards this end, the first two columns of Table 1 report OLS estimates of the slope coefficient $\beta$ along with the associated $t$-statistics for the unconstrained model. Roughly half of the predictor variables generate coefficient estimates that are significant at the 5% level. However, the $t$–statistics for the valuation ratios should not be taken at face value given the well-known biases in their estimates (see, e.g., Stambaugh (1999)).

Columns 3-10 of Table 1 report the posterior means of $\beta$ under the unconstrained and constrained Bayesian models using four combinations of uninformative Normal-Gamma priors with $\psi = 0.1$ or $\psi = 1$ and Normal-Jeffreys’ priors with $\sigma_{\beta} = 0.02$ or $\sigma_{\beta} = 0.2$. Since the results are very similar under the two constrained models (3.3) and (3.4), we only report estimates for the latter. First consider the results for the unconstrained models. Under Normal-Gamma priors the posterior means of the slope coefficient tend to be very similar to the OLS estimates—more so as $\psi$ is increased from 0.1 to 1 and less weight is put on the prior. Under the Jeffreys’ priors, the posterior means of the slope coefficients remain quite close to the OLS estimates—although of course slightly closer to zero, the center of the priors—when $\sigma_{\beta}$ is set to 0.2. Lowering $\sigma_{\beta}$ to 0.02, and thus using a prior more strongly concentrated on zero, has the effect of shrinking the posterior mean of $\beta$ more towards zero and so the absolute value of the posterior means are generally much smaller under this prior.
Turning to the constrained models, under either set of priors, the constraints have a clearly identifiable effect on the posterior means of the coefficients which tend to fall between the OLS estimates and the posterior means under the corresponding unconstrained models. The posterior means of the slope coefficients can vary significantly depending on whether the unconstrained or constrained model is adopted and on the choice of prior. For example, in the case of the net equity expansion variable, the OLS coefficient is -0.197 which is very close to the posterior mean of the unconstrained model under Jeffreys priors with $\sigma_\beta = 0.2$, but is somewhat smaller than the value (-0.12) obtained under the constrained model with $\sigma_\beta = 0.2$ and the values (-0.05 to -0.06) obtained when $\sigma_\beta = 0.02$.

One of the advantages of our methodology is that it treats the coefficients on the predictor variables as random variables. Hence we can study the entire distribution of the coefficients of the predictor variables ($\beta$) conditional on the data and any restrictions that may have been imposed on the forecasting model. This provides insights into the effect on the forecasting model of imposing constraints on the equity premium or on the sign of $\beta$. Figure 1 plots the posterior distribution of $\beta$ under Normal-Gamma prior with $\varphi = 0.1$ for each of the models described in section 2 (i.e. the unconstrained model (3.1), the model (3.3) that imposes non-negative equity premia, $\hat{r}_1, \ldots, \hat{r}_t \geq 0$ and the model (3.4) that further imposes a sign constraint on the slope coefficient).

Several points stand out from these plots. First, in many cases, imposing the constraint that the conditional equity premium cannot be negative has a
very significant impact on the distribution of the slope coefficients in the return equation. In comparison, the sign constraint on $\beta$ generally tends to have a much smaller additional effect on the posterior distribution of $\beta$.

Imposing the constraints on the forecasting model has separate effects on the location and dispersion of the slope parameter, $\beta$. First, the distribution of the slope coefficient, $\beta$, tends to be less dispersed with higher peaks under the constrained models. For example, whereas the distribution of $\beta$ in the unconstrained return model based on the dividend-price ratio is concentrated between -0.01 and 0.025, it lies in a much more narrow band between zero and 0.01 under the constrained model. Even larger effects of imposing the constraints can be observed for the slope coefficients of variables such as the T-bill rate, long-term return, stock variance and inflation.

The second effect is related to the location or center of the distribution of the slope coefficient, $\beta$. For most variables (with the possible exception of the default yield spread), Figure 1 shows that the distribution of $\beta$ is shifted towards zero. For example, in the case of the dividend-price ratio, the mean of $\beta$ changes from 0.008 under the unconstrained model to 0.004 under the constrained model. In general, predictor variables whose coefficient estimates are predominantly positive (such as the valuation ratios) therefore see their distributions shift to the left, while conversely variables such as the T-bill rate or the rate of inflation whose unconstrained coefficient estimates are centered on negative values are shifted to the right when the conditional equity premium is required to be non-negative.
Finally, Figure 1 shows the effect of imposing the double constraint that the conditional equity premium at each point in time is non-negative and that the sign of $\beta$ must be positive for the valuation ratios, spread and interest rate variables and negative for inflation and the T-bill rate. The result of the sign constraint is to further narrow the distribution of the slope coefficients. Unsurprisingly the effect of the additional constraint is largest for those models whose slope coefficient has a distribution that is centered near zero with considerable probability mass on both positive and negative values. In practice the constraint has a much smaller effect for variables such as the T-bill rate or the dividend yield whose unconstrained slope coefficients are largely distributed on one side of zero. Because of the asymmetric nature of the slope restriction, imposing sign constraints on $\beta$ generally pushes the mean of $\beta$ further away from zero compared to when only the conditional equity premium is constrained to be non-negative.

### 3.4 Out-of-sample Forecasts of Stock Returns

The key question in the literature on return predictability is whether stock returns can be predicted ex ante. To address this issue, we next study the out-of-sample forecasting performance of the models under consideration here. Some studies (e.g. Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Goyal and Welch (2003, 2007)) have addressed out-of-sample or ex-ante predictability by accounting for the effect of parameter estimation error associated with investors’
updating of their models based only on historically available ("real time") data.

Our analysis uses the first 10 years of data (1940-49) to obtain initial parameter estimates so the forecasts begin in 1950. More specifically, we use our approach to compute recursive parameter estimates both for the unconstrained model (3.1) and the constrained model (3.4). In order to avoid look-ahead bias in the parameter estimates we only use data up to the month prior to that for which return is being predicted.\footnote{Forecasts from the model that only constrains the equity premium (3.3) are very similar to those from the model (3.4) that constrains both the equity premium and the sign of the slope coefficient $\beta$ and are thus not reported separately.} For example, the forecast of returns in January 1950 is based on an estimate that uses data up to and including December 1949. To forecast excess returns for February 1950, we extend the data set by one observation (i.e. up to January 1950), re-estimate the parameters of the forecasting model subject to any constraints and then compute a new forecast. We continue recursively with this estimation and forecasting procedure up to the end of the sample in 2005:12.

Figure 2 plots the out-of-sample forecasts associated with the unconstrained and constrained forecasting models over the period from 1950-2005. To preserve space we only show the prevailing mean and the forecasts based on the dividend yield and the T-bill rate. The unconstrained and constrained forecasts from the prevailing mean model are almost identical since the sample estimate of the prevailing mean is always positive so the equity premium constraint has the rather modest effect of truncating a very small part of the left tail of the distribution of
the parameter \( \mu \) that controls mean excess returns, .

In contrast, the sequence of return forecasts based on the T-bill rate (shown in the middle window) is an example where imposing constraints on the equity premium makes a very big difference. The unconstrained forecasts based on this model are very volatile and negative most of the time from 1970-1985, whereas the constrained forecasts are far smoother and, by construction, never take negative values. The constrained forecasts are mostly higher than the unconstrained values although this is not the case during the last three years of the sample.

The unconstrained and constrained forecasts based on the model that uses the dividend yield as a predictor variable (shown in the bottom window) clearly share a common trend. This is a reflecting of the persistent movements in the dividend yield. However, the unconstrained forecasts are generally smaller than those based on the constrained model and, moreover, turn negative in 1987 and from 1992 onwards. As we shall later see, this explains the poor forecasting performance of this model since on average stock returns were quite high during this period. Imposing that the equity premium is non-negative leads the dividend yield model to perform much better as the restricted model predicts large and positive excess returns that even drift slightly upwards after the mid-1990s.

We conclude from these findings that imposing basic equilibrium restrictions on the conditional equity premium can have large effects on the ex-ante predicted return. Comparing the forecasts from the unconstrained and constrained models, the predicted returns can in some cases differ by more than 200 basis points per
3.4.1 Evaluation of Forecasts

From the substantial difference in the time series of forecasts shown in Figure 2, we would expect that the unconstrained and constrained models produce quite different out-of-sample forecasting performance. Table 2 confirms that this is indeed the case. Following Campbell and Thompson (2007) this table reports out-of-sample $R^2$-values computed as

$$ R^2_{\text{oos}, i} = 1 - \frac{\sum_{t=1}^{T} e_{t,i}^2}{\sum_{t=1}^{T} e_t^2}, $$

where $e_{t,i} = r_t - \hat{r}_{t,i}$ is the forecast error from the $i$th forecasting model, $\bar{e}_t = r_t - \bar{r}_t$ is the forecast error from the prevailing mean model (which assumes no predictability) and $t = 1, \ldots, T$ is the out-of-sample period. The out-of-sample $R^2$-value is one minus the (squared) ratio of the root mean squared error (RMSE) of the $i$th forecasting model measured relative to that of the no predictability (prevailing mean) model with recursively updated parameter estimates. Forecasting models with smaller out-of-sample RMSE-values than the prevailing mean model generate positive values of $R^2_{\text{oos}, i}$, while conversely models with greater out-of-sample RMSE-values produce negative values of $R^2_{\text{oos}, i}$. Because of this one-to-one mapping between RMSE and out-of-sample $R^2$, there is no need for us to separately report the RMSE values.
As a benchmark the first column in the table shows the full-sample $R^2$—value obtained under the simple least-squares model. This varies significantly across predictor variables and is quite high (0.6%) for the dividend-price ratio, the dividend yield, the smoothed earnings price ratio, the T-bill rate and long-term return. It is even higher in the case of inflation (1.5%) and the cross-sectional premium (1.14%), but is quite low for many of the remaining predictor variables.

The rest of the table reports out-of-sample $R^2$ value under different methods. Consistent with the findings reported by Campbell and Thompson (2007) and Goyal and Welch (2007), the OLS results in Panel A show that there is only weak evidence of out-of-sample predictability based on the unconstrained forecasting models. In fact, nine of the sixteen forecasting models produce negative $R^2$—values and the average $R^2$—value (computed across all 16 models) is also negative. More troubling, perhaps, is that for many predictor variables the out-of-sample $R^2$ is quite large with negative values that exceed the corresponding positive in-sample $R^2$—values listed in the first column. Basing investment strategies on such forecasts would therefore in all likelihood lead to underperformance compared to a simple model with no predictability.

Similar findings hold for the unconstrained Bayesian models. Across all priors, close to half of these models produce negative out-of-sample $R^2$—values.

Following Campbell and Thompson (2007) we also considered the truncated OLS forecasts (3.2). The results are listed in the first column under "Constrained Models". For 14 of the 16 models under consideration, the out-of-sample $R^2$—value
is improved by imposing the equity premium constraints. This is consistent with Campbell and Thompson’s finding that such constraints can improve forecasting performance.

The truncated OLS forecasts proposed by Campbell and Thompson (2007) do not revise the parameter estimates in view of the constraints. In contrast, the Bayesian methodology which we propose in this paper incorporates this information in the parameter estimates. Table 2 shows that there are clearly considerable gains from adopting our methodology. For example, under the Normal-Gamma prior with $\psi = 0.1$, the average $R^2$—value from the univariate forecasting models more than doubles from 0.15 under the constrained OLS forecasts to 0.33. This increase reflects improved forecasting performance under the Bayesian method for 11 of the 16 models.

Similar results are obtained under the Jeffreys priors. Imposing the constraints lead to an improvement in out-of-sample forecasting performance for 13 or 14 out of 16 models, depending on whether $\sigma_\beta = 0.02$ or $\sigma_\beta = 0.2$. This reflects an increase in the average $R^2$—value of 0.10 when $\sigma_\beta = 0.02$ and an increase in this statistic close to 0.40 when $\sigma_\beta = 0.2$. Clearly the constraints matter significantly to forecasting performance. Moreover, under the Jeffreys priors on average the constrained Bayesian models continue to outperform the truncated OLS forecasts.

We conclude from these results that imposing constraints on the equity premium leads to substantial improvements in the out-of-sample forecasting performance of the majority of univariate forecasting models with only four of 16 vari-
ables being unable to predict returns. Imposing such constraints cuts the number of cases with negative $R^2$—values roughly in half from 27 to 15 cases across all priors and univariate forecasting models (i.e. out of a total of 64 cases) and raises the $R^2$—value for twelve of the sixteen variables. In addition, we find significant improvements in the precision of the forecasts by exploiting these constraints to inform the posterior distribution of the coefficients by using our Bayesian methodology as opposed to simply truncating the forecasts at zero.

3.5 Stambaugh Bias

Many of the predictor variables used in the return forecasting literature are known to be highly persistent. Moreover, innovations to such variables and innovations to returns are in some cases strongly correlated. Such conditions are known to introduce a significant bias in the coefficient estimates and a resulting bias in the forecast. This phenomenon, which is generally known as the Stambaugh bias (Stambaugh (1986, 1999)) is particularly important for valuation ratios such as the dividend yield or the price-earnings ratio. These variables are highly persistent with innovations that are strongly negatively correlated with returns.

A large literature has attempted to address the small-sample bias problem by developing refined econometric estimators and bias-reduction methods. Examples include Ang and Bekaert (2007), Campbell and Yogo (2006), Cavanagh, Elliott and Stock (1995), Jansson and Moreira (2006), Lewellen (2004), Polk, Thompson

Our approach provides an effective way to reduce the Stambaugh bias and, as a result, improve on the precision of the forecasts. A feature of our approach which distinguishes it from the more statistically-based methods for dealing with the problem is that it uses an estimator that is economically motivated by a constraint that holds as a consequence of simple equilibrium arguments. Imposing this constraint on the forecast in every period is particularly appropriate for the simple forecasting models which, as pointed out by Ang and Bekaert (2007), can reasonably be viewed as representing approximations rather than an exact relationship.

Along with most studies in the literature, we consider the following simple model for returns \( r_{t+1} \) and the predictor variable \( x_{t+1} \):

\[
\begin{align*}
  r_{t+1} &= \mu_r + \beta_r x_t + u_{t+1} \\
  x_{t+1} &= \mu_x + \beta_x x_t + v_{t+1}.
\end{align*}
\] (3.21)

Returns are predictable by means of past values of \( x \) whenever \( \beta_r \neq 0 \). Moreover, the predictor variable follows a first-order autoregressive process whose persistence is measured by \( \beta_x \in [0, 1] \). Another key parameter is the correlation between shocks to returns and shocks to the state variable, \( \rho_{uv} = \sigma_{uv}/(\sigma_v \sigma_u) \). As noted by Stam-
baugh (1999), OLS estimates of $\beta_r$, denoted by $\hat{\beta}_r$, are biased by approximately

$$E[\hat{\beta}_r - \beta_r] \approx -\frac{\sigma_{uv}}{\sigma_v^2} \left( \frac{1 + 3\beta_x}{T} \right),$$ (3.22)

where $T$ is the sample size, $\sigma_{uv}$ is the covariance between $u$ and $v$ and $\beta_x$ is the persistence parameter for the predictor variable. Valuation ratios such as the earnings-price ratio or the dividend yield are highly persistent with estimates of $\beta_x$ close to one and a correlation between $u$ and $v$ that is large and negative. This introduces a large upwards bias in the estimate of the slope coefficient $\beta_r$ in (3.21) given $\sigma_{uv} < 0$. This has the dual effect of making inference on $\hat{\beta}_r$ very difficult and also leads to worse forecasting performance since the return forecasts become too sensitive to the value of $x$.

Figure 3 shows how the equity premium constraint works when $x \geq 0$ (top window) or $x < 0$ (bottom window). For a given value of $x$, $\mu + \beta x \geq 0$ is a joint constraint on both $\mu$ and $\beta$. When plotted against $\mu$ and $\beta$, the boundary of this constraint always goes through zero and has a slope $\mu/\beta = -x$ (for $\beta \neq 0$) that reflects the value of the predictor variable, $x$: The intercept-slope ratio $\mu/\beta$ must lie between $-\max(x)$ and $\min(x)$. Together with the objective of obtaining the "best fit" by trading off the intercept and the slope coefficient, the equity premium constraint tends to drag $\beta$ towards zero.
3.5.1 Simulation Results

To see how the bias in the slope coefficient and the forecasting performance is affected by the equilibrium constraints, we conduct a series of simulation experiments. We simulate returns under the model (3.21) fitted to stock returns and the dividend yield, using OLS estimates of the coefficients \( \{\hat{\mu}_r, \hat{\beta}_r, \hat{\sigma}_r, \hat{\mu}_x, \hat{\beta}_x, \hat{\sigma}_x, \hat{\sigma}_{rx}\} \) in place of their unknown values. In particular, we generate 5,000 random draws of time series of returns using sample sizes similar to those studied by Stambaugh (1999), namely \( T = 200, 500, 800 \).

For each draw of returns and the predictor variable, we next obtain estimates of the slope coefficient, \( \beta_r \), using OLS estimation as well as Bayesian estimates based on the unconstrained model (3.1), the model that constrains the equity premium (3.3) and the double-constrained model that restricts both the conditional equity premium and the sign (3.4).

Table 3 reports the outcome of these simulations. The true parameters for the simulation are set as:

\(^{10}\)To simulate returns, we use the Cholesky decomposition

\[
\begin{align*}
    u_{t+1} &= \hat{\sigma}_r \varepsilon_{rt+1} \\
    v_{t+1} &= \left( \frac{\hat{\sigma}_r^2 \hat{\sigma}_x^2 - \hat{\sigma}_{xr}^2}{\hat{\sigma}_r^2} \right)^{1/2} \varepsilon_{xt+1} + \frac{\hat{\sigma}_{xr}}{\hat{\sigma}_r} \varepsilon_{rt+1},
\end{align*}
\]

where \( \varepsilon_{rt+1}, \varepsilon_{xt+1} \) are IID and mutually uncorrelated random variables.
\[ r_{t+1} = 0.033 + 0.008x_t + u_{t+1} \]
\[ x_{t+1} = -0.014 + 0.996x_t + v_{t+1} \]
\[ \text{var}(u_{t+1}) = 0.0018, \text{var}(v_{t+1}) = 0.0019, \text{cov}(u_{t+1}, v_{t+1}) = 0.0018 \]

The table shows the mean, skewness and kurtosis of the coefficient estimates as well as the root mean squared error (RMSE) of the one-step-ahead forecast. In common with other researchers we find that the OLS estimates are heavily biased. In the smallest sample, \( T = 200 \), the bias amounts to more than 100% of the true parameter value (0.8) assumed in the simulations. This bias is shared also with the unconstrained Bayesian estimates, which is perhaps unsurprising since we are using uninformative priors in the analysis. As the sample size increases to 500 observations, as expected the bias declines but is still quite substantial. Even in the largest sample with 800 observations some bias remains.

Very different results emerge for the constrained models. Imposing the constraints cuts the bias by more than half in the smallest sample, \( T = 200 \), and largely eliminates the bias in the two largest samples with 500 or 800 observations. Moreover, both the skewness and the kurtosis of the distribution of the coefficient estimates are significantly reduced under the constrained models.

Turning to the forecasting performance listed in the final column of Table 3, this is also significantly improved under the constrained models. In the smallest
sample with 200 observations, the root mean squared forecast error gets reduced from 1.6% (using OLS) or 1.9 (using unconstrained Bayesian estimates) to 1.2 under the equity premium constraint and only 1.0 under the sign- and equity premium constrained model. Even in the largest samples with 500 or 800 observations, the constraints on the forecasting model continue to lead to a large reduction in RMSE-values both in relative and absolute terms.

3.5.2 Forecasts Based on Valuation Ratios: Results from Longe r Samples

As mentioned previously, the Stambaugh bias is a particular concern for the valuation ratios which are highly persistent with innovations that are strongly correlated with returns. For three of these, namely the dividend price ratio, the dividend yield and the earnings-price ratio we have particularly long data samples spanning the period 1871-2005:12, while for a fourth, the smoothed earnings price ratio, the data sample is 1881-2005:12. Recent results that have questioned that these variables can predict stock returns ex-ante (e.g. Goyal and Welch (2003, 2007)) have reported results going back to 1927. It is therefore of interest to consider the forecasting performance of these variables in the longer sample.

To this end, Table 4 reports the out-of-sample forecasting performance for these four predictor variables using the longest-available sample to estimate the parameters of the forecasting models. The out-of-sample forecast period begins
in 1927 and the parameters are updated recursively through time using only data that was available prior to the date of the forecast. For each of the four priors under consideration we report the RMSE and out-of-sample $R^2$-values for the unconstrained and constrained models.

The forecasting performance of the unconstrained models is quite similar under the Normal-Gamma priors as well as under the Jeffreys prior with $\sigma_\beta = 0.2$. Under these priors the out-of-sample $R^2$ is negative for the forecasts based on the dividend-price ratio or the dividend yield, while it is around 0.25 and 0.08 under the earnings-price ratio and smoothed earnings-price ratio models. Forecasting performance improves for the dividend-based predictor variables under the Jeffreys prior with $\sigma_\beta = 0.02$, but it worsens for the earnings-price ratio predictor variable.

Turning to the results under the constrained forecasting models, the forecast precision improves significantly. The out-of-sample $R^2$-value is now always positive and it improves in 15 of 16 cases, in many cases more than doubling the $R^2$-value.

### 3.5.3 Parameter Estimation Error

Our results so far suggest that, for the valuation ratios, the improvement in forecasting performance under the equilibrium restriction on the return forecasting models can be understood in terms of its ability to reduce the finite-sample (Stambaugh) bias. For variables that are either less persistent or whose innovations are less correlated with returns—including managerial decision variables such as the
dividend payout ratio and net equity expansion (Baker et al (2006))—this bias is less likely to be the explanation for the improved forecasting performance of the constrained models. This is also an issue for variables such as the T-bill rate or the inflation rate for which this bias is not a great concern.

To understand why the constraints work for these models, we turn to another explanation, namely the effect of parameter estimation error. It is widely known that it is difficult to estimate the slope coefficients of the return forecasting models with much precision. As a result, the forecasting performance is likely to be improved by estimation methods that reduce parameter uncertainty. Imposing the constraint achieves this objective. As shown in Table 3, the standard deviation—computed across the 5,000 Monte Carlo simulations—is much lower for the constrained than for the unconstrained forecasting models. Hence lower parameter estimation error is another way to understand why the constrained return forecasting models work better than the unconstrained ones.

3.6 Conclusion

We presented a new approach to forecasting stock returns out-of-sample that optimally incorporates information embedded in theoretical restrictions on the conditional equity premium and on the sign of the coefficient of the predictor variable. When implemented empirically, this approach was found to be highly successful at improving the precision of out-of-sample forecasts of stock returns.
3.7 Acknowledgement

Chapter 3, in full, is a reprint of the material as it appears in the Working Paper by Davide Pettenuzzo, Allan Timmermann, Rossen Valkanov, and Rosalin Wu (2008). The dissertation author was one of the primary authors of this paper. I would like to thank my co-authors for kindly allowing me to include the paper as part of my dissertation.
3.8 Tables and Figures:
Table 3.1: Full-sample estimates of slope coefficients for the individual return forecasting models: Part 1

<table>
<thead>
<tr>
<th>Part I</th>
<th>OLS</th>
<th>t-stat</th>
<th>Unconstrained Model</th>
<th>N-G Priors</th>
<th>Jeffreys Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(\Psi = 0.1)</td>
<td>(\Psi = 1)</td>
</tr>
<tr>
<td>Dividend Price ratio</td>
<td>0.008</td>
<td>2.35</td>
<td>0.0080</td>
<td>0.0081</td>
<td>0.0019</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>0.008</td>
<td>2.37</td>
<td>0.0081</td>
<td>0.0081</td>
<td>0.0020</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>0.008</td>
<td>2.17</td>
<td>0.0079</td>
<td>0.0080</td>
<td>0.0019</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>0.011</td>
<td>2.40</td>
<td>0.0105</td>
<td>0.0104</td>
<td>0.0025</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.008</td>
<td>1.35</td>
<td>0.0086</td>
<td>0.0082</td>
<td>0.0019</td>
</tr>
<tr>
<td>T-Bill rate</td>
<td>-0.117</td>
<td>-2.42</td>
<td>-0.1107</td>
<td>-0.1168</td>
<td>-0.0280</td>
</tr>
<tr>
<td>Long term yield</td>
<td>-0.082</td>
<td>-1.61</td>
<td>-0.0785</td>
<td>-0.0813</td>
<td>-0.0200</td>
</tr>
<tr>
<td>Term spread</td>
<td>0.256</td>
<td>2.16</td>
<td>0.2066</td>
<td>0.2488</td>
<td>0.0619</td>
</tr>
<tr>
<td>Default yield spread</td>
<td>-0.304</td>
<td>-0.85</td>
<td>-0.0984</td>
<td>-0.2362</td>
<td>-0.0767</td>
</tr>
<tr>
<td>Default return spread</td>
<td>0.091</td>
<td>0.65</td>
<td>0.0700</td>
<td>0.0879</td>
<td>0.0223</td>
</tr>
<tr>
<td>CS premium</td>
<td>2.153</td>
<td>3.12</td>
<td>0.2221</td>
<td>1.1578</td>
<td>0.4793</td>
</tr>
<tr>
<td>Long term return</td>
<td>0.146</td>
<td>2.35</td>
<td>0.1382</td>
<td>0.1444</td>
<td>0.0345</td>
</tr>
<tr>
<td>Stock variance</td>
<td>-0.391</td>
<td>-0.80</td>
<td>-0.0735</td>
<td>-0.2729</td>
<td>-0.0962</td>
</tr>
<tr>
<td>Dividend Payout ratio</td>
<td>0.006</td>
<td>0.75</td>
<td>0.0056</td>
<td>0.0058</td>
<td>0.0013</td>
</tr>
<tr>
<td>Net equity expansion</td>
<td>-0.197</td>
<td>-1.97</td>
<td>-0.1700</td>
<td>-0.1901</td>
<td>-0.0474</td>
</tr>
<tr>
<td>Inflation</td>
<td>-1.311</td>
<td>-3.57</td>
<td>-0.3957</td>
<td>-1.0761</td>
<td>-0.3218</td>
</tr>
</tbody>
</table>

Note: Please see the notes at the end of Part II
Table 3.2: Full-sample estimates of slope coefficients for the individual return forecasting models: Part 2

<table>
<thead>
<tr>
<th></th>
<th>Part II</th>
<th>OLS</th>
<th>t-stat</th>
<th>Constrained Model</th>
<th>N-G Priors</th>
<th>Jeffreys Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\Psi = 0.1$</td>
<td>$\Psi = 1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\sigma_\beta = 0.02$</td>
<td>$\sigma_\beta = 0.2$</td>
</tr>
<tr>
<td>Dividend Price ratio</td>
<td>0.008</td>
<td>2.35</td>
<td>0.0039</td>
<td>0.0038</td>
<td>0.0022</td>
<td>0.0042</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>0.008</td>
<td>2.37</td>
<td>0.0039</td>
<td>0.0039</td>
<td>0.0022</td>
<td>0.0041</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>0.008</td>
<td>2.17</td>
<td>0.0038</td>
<td>0.0038</td>
<td>0.0023</td>
<td>0.0040</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>0.011</td>
<td>2.40</td>
<td>0.0047</td>
<td>0.0048</td>
<td>0.0028</td>
<td>0.0051</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.008</td>
<td>1.35</td>
<td>0.0074</td>
<td>0.0075</td>
<td>0.0032</td>
<td>0.0074</td>
</tr>
<tr>
<td>T-Bill rate</td>
<td>-0.117</td>
<td>-2.42</td>
<td>-0.0375</td>
<td>-0.0371</td>
<td>-0.0274</td>
<td>-0.0402</td>
</tr>
<tr>
<td>Long term yield</td>
<td>-0.082</td>
<td>-1.61</td>
<td>-0.0447</td>
<td>-0.0457</td>
<td>-0.0273</td>
<td>-0.0464</td>
</tr>
<tr>
<td>Term spread</td>
<td>0.256</td>
<td>2.16</td>
<td>0.0823</td>
<td>0.0837</td>
<td>0.0626</td>
<td>0.0875</td>
</tr>
<tr>
<td>Default yield spread</td>
<td>-0.304</td>
<td>-0.85</td>
<td>-0.2442</td>
<td>-0.3993</td>
<td>-0.1707</td>
<td>-0.3957</td>
</tr>
<tr>
<td>Default return spread</td>
<td>0.091</td>
<td>0.65</td>
<td>0.0695</td>
<td>0.0693</td>
<td>0.0523</td>
<td>0.0687</td>
</tr>
<tr>
<td>CS premium</td>
<td>2.153</td>
<td>3.12</td>
<td>0.3324</td>
<td>0.9424</td>
<td>0.5280</td>
<td>1.3250</td>
</tr>
<tr>
<td>Long term return</td>
<td>0.146</td>
<td>2.35</td>
<td>0.0434</td>
<td>0.0435</td>
<td>0.0341</td>
<td>0.0457</td>
</tr>
<tr>
<td>Stock variance</td>
<td>-0.391</td>
<td>-0.80</td>
<td>-0.0514</td>
<td>-0.0527</td>
<td>-0.0513</td>
<td>-0.0505</td>
</tr>
<tr>
<td>Dividend Payout ratio</td>
<td>0.006</td>
<td>0.75</td>
<td>0.0068</td>
<td>0.0067</td>
<td>0.0036</td>
<td>0.0064</td>
</tr>
<tr>
<td>Net equity expansion</td>
<td>-0.197</td>
<td>-1.97</td>
<td>-0.1146</td>
<td>-0.1191</td>
<td>-0.0606</td>
<td>-0.1243</td>
</tr>
<tr>
<td>Inflation</td>
<td>-1.311</td>
<td>-3.57</td>
<td>-0.2094</td>
<td>-0.2382</td>
<td>-0.2065</td>
<td>-0.2610</td>
</tr>
</tbody>
</table>

Note: This table presents estimates of the slope coefficients of the univariate forecasting models using monthly stock returns in excess of a 1-month T-bill rate over the period 1940:1 – 2005:12. The first and second columns show the ordinary least squares (OLS) estimates and their t-statistics. The next four columns in Part I show the posterior means of the slope coefficients for unconstrained models estimated using Bayesian methods under different priors. Those in Part II show the posterior means of the slope coefficients obtained using Bayesian methods that constrain the equity premium to be non-negative and also constrain the sign of the slope coefficient. "CS premium" stands for cross-sectional premium.
<table>
<thead>
<tr>
<th></th>
<th>R2 in-sample</th>
<th>OLS</th>
<th>N-G Priors (Ψ=0.1)</th>
<th>ψ=1</th>
<th>Jeffreys Priors (σ₂=0.02)</th>
<th>σ₂=0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dividend Price ratio</td>
<td>0.569</td>
<td>0.208</td>
<td>0.234</td>
<td>0.104</td>
<td>0.200</td>
<td>0.258</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>0.582</td>
<td>0.227</td>
<td>0.250</td>
<td>0.021</td>
<td>0.202</td>
<td>0.275</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>0.466</td>
<td>0.252</td>
<td>0.243</td>
<td>-0.005</td>
<td>0.135</td>
<td>0.275</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>0.602</td>
<td>-0.724</td>
<td>-0.519</td>
<td>-1.325</td>
<td>0.231</td>
<td>-0.499</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.106</td>
<td>-1.418</td>
<td>-1.326</td>
<td>-1.438</td>
<td>0.018</td>
<td>-1.179</td>
</tr>
<tr>
<td>T-Bill rate</td>
<td>0.610</td>
<td>-0.089</td>
<td>0.315</td>
<td>0.295</td>
<td>0.363</td>
<td>0.099</td>
</tr>
<tr>
<td>Long term yield</td>
<td>0.202</td>
<td>-1.289</td>
<td>-0.142</td>
<td>-0.876</td>
<td>0.202</td>
<td>-0.965</td>
</tr>
<tr>
<td>Term spread</td>
<td>0.461</td>
<td>0.109</td>
<td>0.283</td>
<td>0.041</td>
<td>0.308</td>
<td>0.200</td>
</tr>
<tr>
<td>Default yield spread</td>
<td>0.092</td>
<td>-0.183</td>
<td>-0.023</td>
<td>-0.128</td>
<td>-0.028</td>
<td>-0.158</td>
</tr>
<tr>
<td>Default return spread</td>
<td>0.053</td>
<td>-0.544</td>
<td>-0.142</td>
<td>-0.336</td>
<td>-0.046</td>
<td>-0.502</td>
</tr>
<tr>
<td>CS premium</td>
<td>1.142</td>
<td>-0.028</td>
<td>0.131</td>
<td>0.585</td>
<td>0.322</td>
<td>0.231</td>
</tr>
<tr>
<td>Long term return</td>
<td>0.570</td>
<td>0.036</td>
<td>0.298</td>
<td>-0.273</td>
<td>0.325</td>
<td>0.212</td>
</tr>
<tr>
<td>Stock variance</td>
<td>0.081</td>
<td>-3.175</td>
<td>-1.018</td>
<td>-1.329</td>
<td>-0.418</td>
<td>-2.960</td>
</tr>
<tr>
<td>Dividend Payout ratio</td>
<td>0.071</td>
<td>-0.293</td>
<td>-0.283</td>
<td>-0.357</td>
<td>-0.014</td>
<td>-0.275</td>
</tr>
<tr>
<td>Net equity expansion</td>
<td>0.365</td>
<td>0.032</td>
<td>0.193</td>
<td>-0.037</td>
<td>0.153</td>
<td>0.072</td>
</tr>
<tr>
<td>Inflation</td>
<td>1.468</td>
<td>1.140</td>
<td>0.385</td>
<td>1.259</td>
<td>0.282</td>
<td>1.125</td>
</tr>
</tbody>
</table>

Note: Please see the notes at the end of Part II.
Table 3.4: Forecasting performance of individual return models: Part 2.

<table>
<thead>
<tr>
<th>R2</th>
<th>in-sample</th>
<th>out-of-sample</th>
<th>Constrained Models</th>
<th>N-G Priors</th>
<th>Jeffreys Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td></td>
<td></td>
<td>(\Psi=0.1)</td>
<td>(\sigma_\beta=0.02)</td>
</tr>
<tr>
<td></td>
<td>(\Psi=1)</td>
<td>(\sigma_\beta=0.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dividend Price ratio</td>
<td>0.569</td>
<td>0.424</td>
<td>0.972</td>
<td>0.960</td>
<td>0.3585</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>0.582</td>
<td>0.427</td>
<td>1.004</td>
<td>1.009</td>
<td>0.3796</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>0.466</td>
<td>0.197</td>
<td>0.454</td>
<td>0.435</td>
<td>0.278</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>0.602</td>
<td>-0.108</td>
<td>0.689</td>
<td>0.702</td>
<td>0.3747</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.106</td>
<td>-1.231</td>
<td>-0.116</td>
<td>-0.115</td>
<td>0.1027</td>
</tr>
<tr>
<td>T-Bill rate</td>
<td>0.610</td>
<td>0.675</td>
<td>0.863</td>
<td>0.465</td>
<td>0.5512</td>
</tr>
<tr>
<td>Long term yield</td>
<td>0.202</td>
<td>0.557</td>
<td>0.777</td>
<td>0.498</td>
<td>0.3713</td>
</tr>
<tr>
<td>Term spread</td>
<td>0.461</td>
<td>0.262</td>
<td>0.357</td>
<td>0.151</td>
<td>0.3332</td>
</tr>
<tr>
<td>Default yield spread</td>
<td>0.092</td>
<td>-0.088</td>
<td>0.030</td>
<td>-0.262</td>
<td>-0.004</td>
</tr>
<tr>
<td>Default return spread</td>
<td>0.053</td>
<td>-0.279</td>
<td>-0.440</td>
<td>-0.664</td>
<td>-0.0981</td>
</tr>
<tr>
<td>CS premium</td>
<td>1.142</td>
<td>0.491</td>
<td>0.398</td>
<td>0.816</td>
<td>0.4618</td>
</tr>
<tr>
<td>Long term return</td>
<td>0.570</td>
<td>0.196</td>
<td>0.393</td>
<td>0.381</td>
<td>0.4677</td>
</tr>
<tr>
<td>Stock variance</td>
<td>0.081</td>
<td>-0.205</td>
<td>-0.253</td>
<td>-1.160</td>
<td>-0.6857</td>
</tr>
<tr>
<td>Dividend Payout ratio</td>
<td>0.071</td>
<td>-0.128</td>
<td>-0.158</td>
<td>-0.150</td>
<td>0.0584</td>
</tr>
<tr>
<td>Net equity expansion</td>
<td>0.365</td>
<td>0.068</td>
<td>0.179</td>
<td>0.124</td>
<td>0.2541</td>
</tr>
<tr>
<td>Inflation</td>
<td>1.468</td>
<td>1.223</td>
<td>0.192</td>
<td>0.024</td>
<td>0.2674</td>
</tr>
</tbody>
</table>

Note: This table presents forecasting results for monthly stock returns using univariate forecasting models. Column 1 shows in-sample R2-values while the other columns show out-of-sample R2-values computed over the period 1950:01 - 2005:12. The data goes back to 1940:01 and the first 10 years of observations are used to obtain initial estimates of the parameters. Subsequently an expanding window is used to estimate the models recursively over time. Positive R2-values show that a forecasting model produces more precise forecasts (lower root mean squared forecast errors) than the prevailing mean model. "CS premium" stands for cross-sectional premium.
Table 3.5: Estimates of slope coefficient and forecast performance for the dividend yield model, Part I

<table>
<thead>
<tr>
<th>Panel</th>
<th>T=200</th>
<th>T=500</th>
<th>T=800</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>skewness</td>
</tr>
<tr>
<td>OLS</td>
<td>2.092</td>
<td>1.659</td>
<td>1.979</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>1.839</td>
<td>1.838</td>
<td>1.676</td>
</tr>
<tr>
<td>Equity Premium Constrained</td>
<td>1.281</td>
<td>1.181</td>
<td>0.646</td>
</tr>
<tr>
<td>Equity Premium and Sign Constrained</td>
<td>1.461</td>
<td>0.989</td>
<td>1.233</td>
</tr>
</tbody>
</table>

Note: See notes at the end of the next table.
Table 3.6: Estimates of slope coefficient and forecast performance for the dividend yield model, Part II

<table>
<thead>
<tr>
<th>Panel</th>
<th>T=200</th>
<th>T=500</th>
<th>T=800</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>skewness</td>
</tr>
<tr>
<td><strong>OLS</strong></td>
<td>2.092</td>
<td>1.659</td>
<td>1.979</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>2.093</td>
<td>2.005</td>
<td>1.612</td>
</tr>
<tr>
<td>Equity Premium Constrained</td>
<td>1.348</td>
<td>1.189</td>
<td>0.512</td>
</tr>
<tr>
<td>Equity Premium &amp; Sign Constrained</td>
<td>1.538</td>
<td>1.043</td>
<td>1.133</td>
</tr>
<tr>
<td><strong>Panel B: T=500</strong></td>
<td>mean</td>
<td>s.d.</td>
<td>skewness</td>
</tr>
<tr>
<td>OLS</td>
<td>1.226</td>
<td>0.598</td>
<td>1.785</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>1.230</td>
<td>0.755</td>
<td>1.564</td>
</tr>
<tr>
<td>Equity Premium Constrained</td>
<td>0.875</td>
<td>0.464</td>
<td>0.668</td>
</tr>
<tr>
<td>Equity Premium &amp; Sign Constrained</td>
<td>0.893</td>
<td>0.445</td>
<td>0.868</td>
</tr>
<tr>
<td><strong>Panel C: T=800</strong></td>
<td>mean</td>
<td>s.d.</td>
<td>skewness</td>
</tr>
<tr>
<td>OLS</td>
<td>1.071</td>
<td>0.394</td>
<td>1.619</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>1.071</td>
<td>0.506</td>
<td>1.378</td>
</tr>
<tr>
<td>Equity Premium Constrained</td>
<td>0.734</td>
<td>0.302</td>
<td>0.720</td>
</tr>
<tr>
<td>Equity Premium &amp; Sign Constrained</td>
<td>0.736</td>
<td>0.298</td>
<td>0.831</td>
</tr>
</tbody>
</table>

Note: This table shows the result of 5,000 Monte Carlo simulations for the dividend yield forecasting model with parameters set to match the full-sample estimates for US stock returns: $r_t = \mu + \beta x_{t-1} + \varepsilon_t, x_t = g + \gamma x_{t-1} + \nu_t$. The (posterior) mean of the slope coefficient on the dividend yield in the return equation is shown in the first column followed by the skew and kurtosis of this coefficient. The fourth column shows the root mean squared forecast error based on forecasts of returns for the following period. In all cases an uninformative prior is used with $\psi = 0.1$. 
Table 3.7: Out-of-sample R2 of individual return forecasting models

<table>
<thead>
<tr>
<th>Panel A:</th>
<th>sample period</th>
<th>N-G Priors</th>
<th>Jeffreys Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconstrained Models</td>
<td></td>
<td>(\Psi=0.1)</td>
<td>(\Psi=1)</td>
</tr>
<tr>
<td>Dividend Price ratio</td>
<td>1871.2–2005.12</td>
<td>-0.332</td>
<td>-0.331</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>1871.2–2005.12</td>
<td>0.048</td>
<td>-0.246</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>1871.2–2005.12</td>
<td>0.151</td>
<td>0.254</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>1881.1–2005.12</td>
<td>0.169</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\sigma_\beta=0.02)</td>
<td>(\sigma_\beta=0.2)</td>
</tr>
<tr>
<td>Panel B:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constrained Models (I)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dividend Price ratio</td>
<td>1871.2–2005.12</td>
<td>0.034</td>
<td>0.138</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>1871.2–2005.12</td>
<td>0.031</td>
<td>0.176</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>1871.2–2005.12</td>
<td>0.031</td>
<td>0.166</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>1881.1–2005.12</td>
<td>0.049</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.241</td>
<td>-0.601</td>
</tr>
<tr>
<td>Panel C:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constrained Models (II)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dividend Price ratio</td>
<td>1871.2–2005.12</td>
<td>0.015</td>
<td>0.130</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>1871.2–2005.12</td>
<td>0.062</td>
<td>0.242</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>1871.2–2005.12</td>
<td>0.070</td>
<td>0.273</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>1881.1–2005.12</td>
<td>0.065</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.191</td>
<td>0.189</td>
</tr>
</tbody>
</table>

Note: This table presents forecasting results for monthly stock returns using univariate forecasting models. Column 1 shows sample periods for each individual model. Columns 2-5 presents the forecasting performance in terms of out-of-sample R2 using Bayesian estimation methods under different prior specifications. Panel A contains results of models without any constraint. Panel B contains results of models with non-negative return premium constraint. Panel C contains results of models with both non-negative return premium constraint and sign constraint on the slope of forecasting equation.
Table 3.8: Compare Model C and Model D

<table>
<thead>
<tr>
<th>State variable</th>
<th>OOS R2</th>
<th>Normal-Gamma Priors</th>
<th>Normal-Gamma Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\Psi = 1)</td>
<td>(\Psi = 0.1)</td>
<td>(\Psi = 0.1)</td>
</tr>
<tr>
<td>Dividend Price ratio</td>
<td>0.586</td>
<td>0.960   -0.374</td>
<td>0.557</td>
</tr>
<tr>
<td>Earnings Price ratio</td>
<td>0.104</td>
<td>0.435   -0.331</td>
<td>0.071</td>
</tr>
<tr>
<td>Smooth E/P ratio</td>
<td>0.419</td>
<td>0.702   -0.284</td>
<td>0.385</td>
</tr>
<tr>
<td>Dividend Payout ratio</td>
<td>-0.202</td>
<td>-0.150  -0.052</td>
<td>-0.241</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>-0.214</td>
<td>-0.115  -0.099</td>
<td>-0.256</td>
</tr>
<tr>
<td>T-Bill rate</td>
<td>0.123</td>
<td>0.465   -0.342</td>
<td>0.333</td>
</tr>
<tr>
<td>Long term yield</td>
<td>0.145</td>
<td>0.498   -0.353</td>
<td>0.292</td>
</tr>
<tr>
<td>Long term return</td>
<td>-0.049</td>
<td>0.381   -0.431</td>
<td>-0.048</td>
</tr>
<tr>
<td>Term spread</td>
<td>0.092</td>
<td>0.151   -0.059</td>
<td>0.074</td>
</tr>
<tr>
<td>Default yield spread</td>
<td>-0.017</td>
<td>-0.262  0.245</td>
<td>-0.011</td>
</tr>
<tr>
<td>Default return spread</td>
<td>-0.396</td>
<td>-0.664  0.267</td>
<td>-0.329</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.081</td>
<td>0.024   -0.105</td>
<td>-0.025</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>0.614</td>
<td>1.009   -0.395</td>
<td>0.575</td>
</tr>
<tr>
<td>Stock variance</td>
<td>-0.939</td>
<td>-1.160  0.221</td>
<td>-0.139</td>
</tr>
<tr>
<td>Net equity expansion</td>
<td>-0.104</td>
<td>0.124   -0.228</td>
<td>-0.056</td>
</tr>
<tr>
<td>CS premium</td>
<td>0.542</td>
<td>0.816   -0.274</td>
<td>0.138</td>
</tr>
</tbody>
</table>

Note:

M_C: Model C, conditional equity premium constraint

M_D: Model D, double constraints: conditional premium+sign
Figure 3.1: Posterior Distribution of Slope Coefficients, Part I
Figure 3.2: Posterior Distribution of Slope Coefficients, Part II
Figure 3.3: Posterior Distribution of Slope Coefficients, Part III
Figure 3.4: Posterior Distribution of Slope Coefficients, Part IV
Figure 3.5: Out-of-sample Forecasts at 1-month frequency
Part I: suppose $x>0$, shaded area stands for $\mu + \beta x \geq 0$

\[ \text{Slope} = -\frac{1}{\min(x)} \]

Part II: suppose $x<0$, shaded area stands for $\mu + \beta x \geq 0$

\[ \text{Slope} = -\frac{1}{\max(x)} \]

Figure 3.6: The Effect of Constraints on Regression Parameters
Figure 3.7: Out-of-sample Forecasts with different frequency
Bibliography


Appendix A

Log-linearized Present Value Model

This appendix provides a brief sketch of the derivation of variance decomposition as in equation (1.1). The derivation follows Campbell (1991).

Let $R_{t+1}$ denote the return on the stock between $t$ and $t+1$, $P_t$ denote price of the financial asset at time $t$, and $D_{t+1}$ the dividend at time $t + 1$. Define $R_{t+1}$ as:

$$1 + R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$$

Taking logs and letting $r_{t+1} \equiv \ln(1 + R_{t+1})$ yields:

$$r_{t+1} = p_{t+1} - p_t + \ln(1 + \exp(d_{t+1} - p_{t+1})) \quad (A.1)$$

where $p_{t+1}$, and $d_{t+1}$ denote the corresponding logarithmic values of $P_{t+1}$ and $D_{t+1}$. 

133
Approximating the final term in Equation (A.1), $\ln(1 + \exp(d_{t+1} - p_{t+1}))$, with a first order Taylor series expansion around the steady-state log-dividend-price ratio $\overline{d} - p$ we get:

$$\ln(1 + \exp(d_{t+1} - p_{t+1})) \approx h + (1 - \rho)(d_{t+1} - p_{t+1}) \quad (A.2)$$

where $\rho \equiv \frac{P}{D+P}$ represents the steady state ratio of stock price to price plus dividend. It is a number slightly smaller than one, and could be interpreted as the discount factor; $h \equiv -\ln(\rho) - (1 - \rho)\ln\left(\frac{1}{\rho} - 1\right)$ is a constant from the Taylor approximation. Substituting (A.2) back into (A.1) gives:

$$p_t = h + \rho p_{t+1} + (1 - \rho)d_{t+1} - r_{t+1} \quad (A.3)$$

Solving the above difference equation forward, in the infinite-horizon limit we get

$$p_t = \frac{h}{1 - \rho} + \sum_{i=0}^{\infty} \rho^i((1 - \rho)d_{t+1+i} - r_{t+1+i}) \quad (A.4)$$

The convergence of the sum is assured under the assumption that the dividend and return processes are cointegrated. Equation (A.4) also holds ex ante as a present-value model:

$$p_t = \frac{h}{1 - \rho} + \mathbb{E}_t\left[\sum_{i=0}^{\infty} \rho^i((1 - \rho)d_{t+1+i} - r_{t+1+i})\right] \quad (A.5)$$
with $E_t$ denoting an expectation formed at the end of period $t$. Substituting (A.3) back into (A.2) and collecting terms yields the formula for unexpected stock return as a function of forecast revisions in future dividend growth rate and discount rate:

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

(A.6)
Appendix B

Data Summary

I study a five-variable VAR system. The variables are all at monthly frequency. They are:

1. The market return ($Rm$): the value-weighted S&P 500 index returns, as reported by the Center for Research in Security Prices (CRSP) at the University of Chicago.

2. The dividend growth rate ($Gd$): the log difference of dividends. Dividends are twelve-month moving sums of dividends paid on the S&P 500 index. The data are from Amit Goyal’s website.

3. The smoothed price-earnings ratio ($PE_{10}$): the moving ten-year average of earnings divided by price.

4. The dividend yield ($YLD$): the difference between the log of dividends and the log of lagged prices.
5. Inflation ($INFL$): measured as the Consumer Price Index (All Urban Consumers) from the Bureau of Labor Statistics.