



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Three Essays on Modelling and Testing the Conditional Risk Premium

Trinh Anh Khoa Hoang

Bachelor of Commerce (Honours)

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland on 14th October 2015

UQ Business School

Abstract

This thesis, a collection of three essays, focuses primarily on modelling conditional expectations and investigates the central research question: is the ex-ante market risk premium always positive?

Many asset pricing studies have focused on testing linear restrictions imposed by asset pricing models and largely ignore another important restriction: the positivity of the market risk premium. Aiming to enhance our understanding in this area, the first essay applies a novel market-based measure of conditional expected return, namely, the implied cost of capital, and examines whether or not the ex-ante risk premium is always positive. Employing economically meaningful information variables and a multiple inequality constraints framework, I find evidence that this positive condition is violated not only in the US, but also in other major international markets including Japan, Italy, and Germany. In stark contrast, when the realised return is used to proxy for the expected return, there is insufficient evidence against the hypothesis due to a high degree of noise embedded in the realised return proxy. Accordingly, I argue for the use of the implied cost of capital in modelling the time varying expected return.

In the second essay, I further examine the positive risk premium hypothesis. Specifically, I introduce a new two-stage method, involving Principal Component Analysis and Boosted Regression Tree to model conditional expected return. With these techniques, I address potential pitfalls associated with existing methods of capturing the true identity of investors' information sets, and how investors use the information in forming expectations. Consistent with the first essay's finding, the positivity condition of the risk premium is violated in the US. Collectively, the evidence suggests the rejection of the Conditional Capital Asset Pricing Model.

The implied cost of capital is appealing in its own right, yet its validity as a proxy for conditional expected return is open to debate. Returning to this measure, the third study investigates when and to what extent this estimate deviates from true expectations. In a simulation study allowing for time varying discount rates, I find that due to the constant term structure assumption embedded in the research method, the variation of the implied cost of capital is significantly lower than that of the true expected return. This feature reveals that in a standard regression, the economic significance interpretation of the coefficient is no longer appropriate. Additionally, I show analytically that when analyst forecasts are biased and/or unable to capture the full information of cash flow expectations incorporated in the market price, the derived implied cost of capital is contaminated. Moreover, there exists a real issue regarding spurious regressions involving the implied cost of capital.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person, except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature, and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer-reviewed paper(s):

- 1) Hoang, K. T., R. Faff, and M. Haq, 2014, Market discipline and bank risk taking, *Australian Journal of Management* 39, 327-350.

Publications included in this thesis

No publications included.

Contributions by others to the thesis

No contributions by others

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

My life has been so fortunate in that I always meet the right people at the right time. Without tremendous help and continuous support from my mentors, friends, and family, my PhD journey would not have found its way here.

First and foremost, I would like to express sincere gratitude to Professor Tom Smith, my PhD reader. I don't think I can thank you well enough for what you have done for me. Thank you for being a great mentor. Thank you for being a great friend. I thank you for everything, Tom.

I would like to thank my supervisors, Professor Robert Faff and Dr Mamiza Haq. I thank you for giving me the courage to pursue what I am passionate about. Without you, I wouldn't be here in the first place. Without your patience and understanding, I wouldn't be able to chase my goals. Such encouragement and support mean a great deal to me. Robert, your "roller coaster" story has always been a source of inspiration in every part of my life.

To my parents whom I love the most, your unconditional love and support are the main forces that have pushed me forward since I set foot in Australia. I will never be able to repay this debt. I also thank my sister who always tells me to keep challenging myself.

To my wife Chau, thank you for being supportive and taking care of me through this long journey. We will start a new chapter of our life and I promise that I will always love you.

To Sharif and Vu, I am humbled to share this journey with you. This journey would not have been as enjoyable without your advice and laughter. Cheers.

Keywords

positive risk premium, boosted regression tree, implied cost of capital

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC Code: 150201, Finance 50%

ANZSRC Code: 150202, Financial Econometrics 35%

ANZSRC Code: 150205, Investment and Risk Management 15%

Fields of Research (FoR) Classification

1502 Banking, Finance and Investment 100%

Table of Contents

| | |
|---|-----------|
| Publications included in this thesis | iv |
| Statement of parts of the thesis submitted to qualify for the award of another degree | v |
| Australian and New Zealand Standard Research Classifications (ANZSRC) | vii |
| Fields of Research (FoR) Classification..... | vii |
| Chapter 1 | 1 |
| 1.1 Motivation and Background | 1 |
| 1.2 Research Objectives | 2 |
| 1.3 Summary of Findings | 3 |
| 1.4 Contribution..... | 4 |
| Chapter 2 | 6 |
| 2.1 Introduction | 6 |
| 2.2 Ex-ante Risk Premium Restriction | 10 |
| 2.2.1 Ex-ante Risk Premium and Asset Pricing..... | 10 |
| 2.2.2 Multiple Conditional Inequalities Methodology | 12 |
| 2.2.3 Limitations of Realised Returns | 13 |
| 2.3 Implied Cost of Capital | 15 |
| 2.3.1 Analytical Framework | 15 |
| 2.3.2 ICC Construction..... | 16 |
| 2.4 Data and Sample Selection | 18 |
| 2.4.1 Data Sources | 18 |
| 2.4.2 Earnings Forecasts | 18 |
| 2.4.3 Plowback Ratio | 20 |
| 2.4.4 Putting It All Together..... | 20 |
| 2.5 Empirical Results..... | 22 |
| 2.5.1 Descriptive Statistics | 22 |
| 2.5.2 Test of Non-negative Market Risk Premium Hypothesis..... | 31 |
| 2.5.2.1 Instrumental Variables..... | 31 |
| 2.5.2.2 The US Market Evidence..... | 32 |
| 2.5.2.3 The International Markets Evidence..... | 35 |
| 2.6 Concluding Remarks | 37 |
| Chapter 3 | 39 |
| 3.1 Introduction | 39 |
| 3.2 Research Method | 44 |
| 3.2.1 Principal Component Analysis | 44 |
| 3.2.2 Regression Trees..... | 45 |
| 3.2.3 Boosting..... | 47 |
| 3.2.4 Estimation Period..... | 48 |
| 3.2.5 Model Assessment | 48 |

| | | |
|------------------------|---|-----|
| 3.2.5.1 | Relative Influence Measure | 49 |
| 3.2.5.2 | Partial Dependence Plots | 49 |
| 3.2.5.3 | Out-of-Sample Prediction | 50 |
| 3.2.6 | Inequality Constraint Method | 51 |
| 3.3 | Data Description | 53 |
| 3.4 | Empirical Results | 53 |
| 3.4.1 | Summary Statistics | 53 |
| 3.4.2 | Relative Influence Measure | 56 |
| 3.4.3 | Partial Dependence Plots | 56 |
| 3.4.4 | Out-of-Sample R^2_{OOS} | 58 |
| 3.4.4.1 | Comparison with Linear Models | 58 |
| 3.4.4.2 | Further Diagnostic Specifications | 59 |
| 3.4.5 | Inequality Test | 62 |
| 3.4.5.1 | Instrumental Variables | 62 |
| 3.4.5.2 | Main Inequality Results | 63 |
| 3.5 | Concluding Remarks | 66 |
| Chapter 4 68 | | |
| 4.1 | Introduction | 68 |
| 4.2 | Theoretical FrameWork | 72 |
| 4.2.1 | General Framework | 72 |
| 4.2.2 | Constant Term Structure | 76 |
| 4.2.3 | Spurious Regressions | 76 |
| 4.3 | Simulation Procedure | 77 |
| 4.3.1 | Mean and Variance of R_e of a Single Firm | 77 |
| 4.3.2 | Spurious Regression with Panel Data | 79 |
| 4.4 | Simulation Results | 81 |
| 4.4.1 | Constant Term Structure | 81 |
| 4.4.2 | Spurious Regression Effects | 90 |
| 4.5 | Concluding Remarks | 93 |
| Chapter 5 94 | | |
| 5.1 | Thesis Review | 94 |
| 5.2 | Review of Findings and Contributions | 94 |
| 5.3 | Future Research | 96 |
| References 97 | | |
| Appendices 102 | | |
| Appendix A 102 | | |
| A.1 | Correlation Matrix (Chapter 2) | 102 |
| A.2 | Least Angle Regression (LAR) in Chapter 3 | 103 |

| | |
|---|-----|
| A.3 Proof of Equations in Chapter 4 | 103 |
| A.3.1 Proof of Equation (4.7) | 103 |
| A.3.2 Proof of Equation (4.8) | 104 |
| A.3.3 Proof of Equation (4.12) | 104 |
| A.3.4 Proof of Equation (4.24) | 105 |

List of Figures and Tables

| | |
|--|-----|
| Table 2.1: Summary Statistics | 23 |
| Table 2.2: Correlation Matrix | 26 |
| Figure 2.1: Risk Premium of the CRSP Portfolio | 27 |
| Figure 2.2: Risk Premium of the S&P500 Portfolio | 28 |
| Figure 2.3: Risk Premium of the International Markets | 29 |
| Table 2.3: Inequality Tests on the Positivity Restriction of the Market Risk Premium | 34 |
| Table 2.4: Inequality Tests on the Positivity Restriction of the Implied Market Risk Premium: International Markets | 36 |
| Table 2.5: Inequality Tests on the Positivity Restriction of the Realised Market Risk Premium: International Markets | 37 |
| Figure 3.1: A Visual Demonstration of the Regression Tree Methodology | 42 |
| Table 3.1: Summary Statistics | 55 |
| Figure 3.2: Marginal Effect of the Information Variables on the Expected Return | 57 |
| Table 3.2: Out-of-Sample Prediction Comparison | 59 |
| Table 3.3: Out-of-Sample Prediction with Different Specifications | 61 |
| Table 3.4: Inequality Tests on the Positivity Restriction of the Market Risk Premium | 64 |
| Table 4.1: Comparison between the Mean and Standard Deviation of the ICC as opposed to that of the true Expected Return | 83 |
| Table 4.2: Difference between the Coefficients estimated from the Regression of the ICC on a Variable of Interest, and those from the Regression of the true Expected Returns on the same Variable of Interest | 84 |
| Table 4.3: Difference between the Coefficients estimated from the Regression of the Future Realised Returns on the ICC, as opposed to those from the Regressions of Future Realised Returns on the true Expected Returns | 88 |
| Table 4.4: Spurious Regression when Analyst Forecasts are unable to capture Full Cash Flow Expectations | 91 |
| Table 4.5: Spurious Regression when Analyst Cash Flow Forecasts are systematically biased | 92 |
| Table A.1.1: The Correlation Matrix between the Implied Risk Premium and the Realised Risk Premium for Non-US Countries | 102 |

List of Abbreviations used in the thesis

| | |
|------|------------------------------|
| BRT | Boosted Regression Tree |
| CAPM | Capital Asset Pricing Model |
| COV | Covariance |
| ICC | Implied Cost of Capital |
| PCA | Principal Component Analysis |
| VAR | Variance |

This thesis, a collection of three essays, focuses on modelling the conditional expected return and testing a central research question:

“Is the ex-ante risk premium always positive?”

1.1 MOTIVATION AND BACKGROUND

Understanding the risk-return trade-off is fundamental in finance. Intuitively, risk-averse and rational investors demand higher expected returns to compensate for taking on additional risk, leading to the ex-ante risk premium of aggregate wealth (i.e. the market portfolio) always being greater than, or equal to, zero. The positive risk premium ensures the mean-variance efficiency of the market portfolio; therefore it is a necessary condition for the Conditional Capital Asset Pricing Model (Conditional CAPM). Merton (1980) argues that this restriction should be explicitly included as a necessary condition for capital market equilibrium. However, under the general stochastic discount framework (Harrison and Kreps (1979)), a negative risk premium can theoretically exist when the marginal rate of substitution is positively correlated with market returns. The equity market, by providing infinite claims in the future, can help investors hedge against adverse changes in the investment opportunity set. As this hedging component is desirable, investors demand lower expected returns, especially in states associated with high probability of a regime shift (Whitelaw (2000)). Despite the important nature of the positive risk premium in theoretical modelling, much less effort has been devoted to testing whether or not the condition is violated in the data. My thesis aims to fill this gap.

As in any conditional asset pricing tests, operationalising this research question requires a tight and reliable proxy for the conditional expected return. However, modelling the conditional mean return is difficult (Merton (1980)), and existing methods have been subject to a number of criticisms.

The first common approach estimates the conditional mean return by linearly projecting future returns onto a set of predetermined variables (see, for example, Campbell and Thompson (2008); Rapach, Strauss and Zhou (2010); and Welch and Goyal (2008)). However, since the identity of the investors' information set is unobservable, various biases embedded in this methodology are likely to yield misleading inferences (see Ang and Bekaert (2007) for omitted variable biases; Foster, Smith and Whaley (1997) and Ferson, Sarkissian and Simin (2008) for data snooping bias). Additionally, a simple linear model is unlikely to capture a constantly changing

return data generating process, due to numerous information shocks from policy shifts, technology advancement, and institutional changes (Rapach, et al. (2010)).

Realised return is another widely used estimate of expected return, motivated by the belief that in the long run, if positive and negative information surprises in asset returns cancel each other out, the realised return is an unbiased proxy for the expected return. However, the noise in asset returns is large so that it might have significant impact on average realised return, even after a long time (Elton (1999)). Lundblad (2007) confirms this observation by showing in simulation analysis that it requires a very long sample period, well beyond 100 years, for the detection of a meaningful risk-return trade-off.

These criticisms highlight serious doubts about the construct validity of conditional asset pricing evidence. Elton (1999) emphasises in his American Finance Association presidential address that the quest for better measures of expected return and alternative ways of examining asset pricing models will yield a much better payoff than development of additional statistical tests that continue to rely on the realised return. My thesis follows this research agenda.

1.2 RESEARCH OBJECTIVES

In the first essay, I employ the Implied Cost of Capital (ICC) method to model the expected return, and examine the positivity condition of the risk premium. The ICC is the internal rate of return that equates the firm's stock price and present value of its future cash flows. Because the main inputs to compute the ICC are analyst earnings forecasts and stock prices, the ICC is forward-looking and does not assume that the information set observed by the economic agents is the same as that of the econometricians in forming expectations. Due to this appealing feature, the ICC has found increasing applications in both the accounting and the finance literature (see, for example, Chava and Purnanandam (2010) on default risk; Pastor, Sinha and Swaminathan (2008) on risk-return trade-off; Hwang, Lee, Lim and Park (2013) on probability of informed trading). I model the conditional expected return of the market portfolio alternatively as the value-weighted and equal-weighted averages of individual firms' ICCs. The market implied risk premium is obtained by subtracting long term government bond yields from the aggregate market ICC. To formally test the positivity condition of the risk premium, the implied measure is brought into the multiple inequality constraints testing framework developed by Boudoukh, Richardson and Smith (1993). The risk premium hypothesis is investigated in the context of G7 countries (the US, Canada, France, Italy, Germany, the UK, and Japan).

In the second essay, I propose a new two-stage procedure, taking into account a large amount of information and a complex return data generating process, to model the conditional mean

return. Principal Component Analysis (PCA) is performed in the first stage to find a set of common factors representing 160 financial variables from Ludvigson and Ng (2007) and Welch and Goyal (2008). By summarising a rich source of information into a small set of factors, I aim to span the investors' information set and avoid the curse of dimensionality (Ludvigson and Ng (2007)). The set of predictors is incorporated in the second stage, in which I employ the regression tree technique developed in the machine learning literature (see Hastie, Friedman and Tibshirani (2009) for literature review) to capture the complex relation between the conditional risk premium and the information variables. Specifically, without imposing strong modelling assumptions, the unknown function is approximated by carving out the predictor space through a sequence of piece-wise constant models. Additionally, a final stable model is constructed by using additive expansions of the simple regression trees, a supervised learning process known as 'boosting', similar to that discussed in the forecast combination literature (Rapach, et al. (2010)). Finally, the risk premium estimated from the two-stage model is incorporated in the multiple inequality constraints framework, to investigate the positive risk premium hypothesis in the US market from 1970 to 2012.

Although the ICC is intuitively attractive, its validity is subject to open debate. Coming back to this estimate in the third essay, I examine its validity as a proxy for the conditional expected return. In particular, a simulation study is conducted to investigate two aspects of measurement errors in the ICC. First, the primary focus is to understand how the constant term structure assumption embedded in the ICC methodology leads the mean and the variance of the ICC to deviate from those of the true expected return. The consequences of the deviation in the regression context, particularly related to the economic significance interpretation of the regression coefficients, are highlighted. Second, I extend the framework to a panel of firms and years to examine how measurement errors in cash flow forecasts, a critical input in calculating the ICC, can result in spurious regressions that involve the ICC as a proxy for the expected return.

1.3 SUMMARY OF FINDINGS

In the first essay, when the ICC is used as a proxy for the expected return, the positivity restriction of the risk premium is violated in the US (S&P 500 market portfolio), Japanese, German, and Italian markets. In contrast, there is insufficient evidence to conclude that the realised market returns are less than the risk free rate, with the exception of Germany, even though the realised risk premium has multiple instances of negativity in the sample period. Noise in the realised measure seems large, so it might obscure the true violation of the condition (Ostdiek (1998)).

In the second essay, I continue to find evidence against the positive risk premium hypothesis in the US market. Additionally, I show that the validity of the two-stage method in modelling the conditional expected return is justified through numerous specification tests. The superior performance, both statistically and economically, results from the technique ability in capturing the investors' information identity and the complicated return data generating process.

In the third essay, the simulation results suggest that the mean of the ICC and that of the true expected return are not significantly different from each other. Yet, the ICC variation seems to be significantly smoother than the true expected return variation. The lack of variation, resulting from the constant term structure assumption in the ICC method, can cause bias in the regression coefficient estimates, and thus leads researchers to invalidly interpret the economic significance of the relationship they examine. Further, when researchers aim to draw conclusions about the relationship between the expected return and the variables of interest from the regressions of the ICC on those variables, simulation evidence suggests that the analyst forecast errors can confound the inferences in such regressions. The confounding effects come from analyst forecasts being systematically biased and/or unable to capture the full cash flow expectations incorporated in the market price.

1.4 CONTRIBUTION

The first essay contributes by connecting important, yet unrelated, strands of the literature. The first strand focuses on the notion of risk-return trade-off. While a considerable segment of empirical work investigates the sign (see, for example, Pastor, et al. (2008); Ludvigson and Ng (2007); Ghysels, Santa-Clara and Valkanov (2005); or Rossi and Timmermann (2010)), the magnitude of the risk premium remains under-explored. This study, among a few, puts forward this restriction into direct empirical test (for example, Boudoukh, et al. (1993); Boudoukh, Richardson and Whitelaw (1997); Ostdiek (1998); and Walsh (2014)). Unlike previous studies, the positivity restriction is tested using a novel forward-looking measure of the expected return, the ICC. The second strand of literature centres on using the ICC as a measure of the expected return, yet mainly examines the cross-sectional properties of firms'/portfolios' ICC. The study, among a few, examines the aggregate ICC (Pastor, et al. (2008); and Li, Ng and Swaminathan (2013)). I find evidence that the positive risk premium is violated not only in the US, but also in other large economies, which implies the rejection of the Conditional CAPM in those markets, consistent with findings documented in Lewellen and Nagel (2006).

The contribution of the second essay lies in the new two-stage methodology in modelling the conditional expected return. The importance of addressing the identity of the investors'

information set, along with the complex return data generating process, in estimating the first moment of return is emphasised by the superior predictive performance of the proposed model. To the best of my knowledge, this is the first study to apply the boosted regression tree, a state-of-the-art methodology in the machine learning literature, in testing the positivity of risk premium hypothesis. The most similar paper to this essay is Rossi and Timmermann (2010), who also use this technique; however, they focus on testing the time varying risk-return trade-off.

The proposed contributions in the final essay are threefold. First, I provide a convenient analytical framework, allowing for time varying structure in both cash flow expectations and discount rates. This lays a useful foundation for assessing and comparing the properties of the ICC and those of the true conditional expected return. Second, simulation conducted under this framework reveals that the ICC variation is significantly lower compared to that of the true expected return. This feature warns researchers that they can no longer validly interpret the economic significance from the regression coefficient that involves the ICC. Even though assessing statistical significance plays a major role in empirical research, judging economic significance is equally important (McCloskey and Ziliak (1996)). This is the first study to show such evidence. Finally, I go further than existing works in the ICC measurement error literature (for example, see Wang (2015)) by demonstrating analytically and through simulation that the measurement errors in analyst forecast errors can be transferred to the ICC estimates which, in turn, induces the threat of spurious regressions.¹

The thesis is structured as follows. Chapter 2, Chapter 3, and Chapter 4 present Essay 1, Essay 2, and Essay 3 respectively. Chapter 5 offers a conclusion of the thesis.

¹ Existing studies simply describe or argue, without showing in analytical forms, that measurement errors in cash flow expectations can be transferred to the discount rate.

IS THE EX-ANTE RISK PREMIUM ALWAYS POSITIVE? A MARKET EXPECTATION PERSPECTIVE²

2.1 INTRODUCTION

Risk premium serves as a central theme in any asset pricing models. The idea that risk matters in determining expected return is intuitive and well-founded. If the marginal investors are strictly risk-averse and expected utility maximisers, they demand higher returns for investments that have a higher level of risk. As a result, the ex-ante market return should always exceed the risk free rate, leading to the positive risk premium. The positivity restriction is embedded as a necessary condition for the Conditional Capital Asset Pricing Model (Conditional CAPM) because it ensures the mean-variance efficiency of the market portfolio. This implies that violation of the restriction leads to the rejection of the Conditional CAPM. Furthermore, Merton (1980) supports that for equilibrium of the capital market, this condition should be explicitly incorporated as a necessary condition.

However, this is not as definitive as it sounds. Theoretically, in more general equilibrium asset pricing models (Harrison and Kreps (1979); and Lucas (1978)), the existence of negative risk premium is possible when the market excess return is positively correlated with the marginal rate of substitution. Because the market, generally equity, has indefinite claims in the future, it can provide investors hedging ability against adverse shocks to the investment opportunity set, induced by high probable regime shifting states of the economy (Whitelaw (2000)). The investors, facing such desirable hedge, might require a lower rate of return even though the risk in those states is high. Adding to the complication, the vast quantity of empirical literature on the true intertemporal risk-return relation is far from reaching a consensus. While some studies find a positive mean-variance trade-off (see, for example, Ludvigson and Ng (2007)), others find a negative relation (Guo and Whitelaw (2006)), and yet others document a highly non-linear structure (Glosten, Jagannathan and

² Earlier versions of this essay were presented at the 7th International Accounting & Finance Doctoral Symposium, Trondheim Norway, June 2014; SIRCA Inaugural Young Researcher Workshop, Sydney Australia, July 2014; and UQ Workshop with Professor Murillo Campello (Cornell University), Brisbane Australia, August 2014.

Runkle (1993). In light of the inconclusive evidence, it is surprising that much less effort has been devoted to testing whether or not the sign restriction of the risk premium is violated in the data. There are a few exceptions including Boudoukh, et al. (1993), Ostdiek (1998), and Walsh (2014). In this essay, without relying on the realised return as in previous studies, I re-examine this restriction with a novel market-based measure of the conditional expected return, the Implied Cost of Capital (ICC).

Asset pricing tests are replete with the use of the realised return as an empirical proxy for the expected return. Such a pragmatic approach is motivated by the belief that the realised return is an unbiased estimate of the expected return; if, in the long run, positive and negative information surprises in asset returns cancel each other out, the realised return closely matches the expected return. Nevertheless, there are strong reasons to believe that this might be a poor proxy. First, noise in asset returns tends to be large, which could significantly reduce the power of empirical analyses (Sharpe (1978); and Black (1986)). Additionally, information surprises are either large by themselves or highly correlated so that the aggregate effect is large, which might have significant impacts on the average realised return, even after a long period of time (Elton (1999)). So how long does it take for the realised return to “catch up” with the expected return? Lundblad (2007) studies this question in the context of risk-return trade-off and finds that the required data span could be well beyond 100 years. Second, the realised return contains both information about changes in expected cash flows and changes in expected return. Chen, Da and Zhao (2013) show that cash flow news has significant and dominant impacts on driving both aggregate and firm level average stock returns, relative to expected return news. As a result, the average ex-post return might be contaminated by these cash flow shocks. Third, the inferences about the expected return may be counterintuitive for equity. For example, there was a long period, commencing in the early 1990s, during which the Japanese stock market realised returns are lower than the risk-free rate. Would anyone conclude that the Japanese stock market became less risky during this period, just because its expected return, as proxied by the realised return, was low? Intuitively, the stock market should be more risky because of its prolonged poor performance.

All of these criticisms lead to one critical enquiry: if this proxy is so poor, the construct validity of existing empirical results is surely violated. Therefore, do these findings continue to hold under alternative proxies of the expected return? In his American Finance Association (AFA) presidential address, Elton takes the stand that looking for superior measures of the expected return, and different ways of testing asset pricing theories without using the realised return, will yield a much better outcome than developing further statistical tests that continue to rely on this noisy proxy.

Responding to these types of criticism, the ICC literature has developed. The ICC of a firm is the internal rate of return that equates to the firm's market value and present value of its expected future cash flows. This is an attractive alternative proxy for the expected return on several counts. In contrast to the ex-post realised return, the ICC, directly derived from a firm's stock price and cash flow forecasts, is forward-looking. As these primary inputs are determined by market participants, the ICC is the market-based measure of the time varying expected return and does not assume that the information set observed by the economic agents is the same as that of the econometricians in modelling the conditional expected return. Additionally, if the expected cash flows in the model truly capture the market expectation of the firm's future cash flows, the remaining price variation is attributed to the change in the expected return. As a result, the ICC measure might be less, if not, contaminated by the cash flow news (Cochrane (2011); and Chen, et al. (2013)).

Due to these appealing features, the ICC has recently gained attention from finance. Brav, Lehavy and Michaely (2005), and Lee, Ng and Swaminathan (2009), revisit traditional asset pricing tests in the US and the international contexts, respectively. Pastor, et al. (2008) examine the risk-return trade-off implication of asset pricing models; while Chava and Purnanandam (2010) investigate the relationship between the default risk and the stock expected returns. Chen, et al. (2013) examine the sources of stock price movements. Li, et al. (2013) find that the market ICC is a good predictor for future realised returns. Following the theme, this essay applies the ICC as a proxy for the conditional expected return, and tests the positivity restriction of the ex-ante market risk premium in the international G7 markets (the US, Canada, the UK, Germany, Italy, France, and Japan).

Specifically, I model the conditional mean return of the market portfolio as the value-weighted and equal-weighted averages of individual firms' ICC. Subsequently, long term government bond yield is subtracted from the aggregate market ICC to obtain the market implied risk premium. The implied measure is then brought into the multiple inequality constraints testing framework, which formally deals with the inequality nature of the positive risk premium hypothesis.

Given the critical assumption in the ICC methodology, that explicit cash flow forecasts need to capture the market expectation of future cash flows, I employ two cash flow estimates: analyst earnings forecasts from the I/B/E/S database and the forecasts from the regression model developed by Hou, Van Dijk and Zhang (2012). Consequently, the two alternative ICC estimates are labelled as the analyst-based and fundamental-based ICC, respectively. By employing the latter proxy, the test of the implied risk premium on the US sample can be extended to a longer time period (from 1962 to 2012), covers a larger number of firms, and overcomes biases pertaining to analyst earnings

forecasts (e.g. overoptimism, conflict of interest, sample selection, and survivorship biases). Alas, it is not possible to apply the fundamental-based ICC to other G-7 countries because the sample period is too short.³ To summarise, with respect to the analyst-based ICC, the US sample spans from 1977 to 2012; whereas for Canada, France, Germany, Italy, the United Kingdom, and Japan, sample spans are from 1990 to 2012. The fundamental-based ICC is only used for US firms and the sample period coverage starts from 1962 to 2012.

I find evidence that the positivity restriction of the market risk premium is violated for the US (S&P 500 market portfolio), Japanese, German, and Italian markets in some states of the economy, particularly associated with periods of downward-sloping term structure, high T-bill rates, and lagged negative risk premium. Motivated by empirical evidence, these periods signal the shift in regimes in which, predicted by the theory, the negative risk premium can exist (Whitelaw (2000)). For example, Harvey (1988) finds that the inverted term structure typically happens at the peak of the business cycle and therefore indicates that the economy moves towards a different regime, i.e. a contractionary state. In contrast, when the ex-post realised return is used to proxy for the expected return as in Boudoukh, et al. (1993) and Ostdiek (1998), there is insufficient evidence that the ex-post realised market return is less than the risk free rate, the exception being in the German market. Although the risk premium constructed by subtracting realised returns from risk free rate have multiple occasions of being negative in the study periods, noise in the realised measure tends to be large, so the test power might be significantly impaired (Ostdiek (1998)). Given the above evidence, I conclude that the implied risk premium can be served as a tight proxy for the conditional risk premium.

I aim to contribute by connecting three important, yet unrelated, areas of the literature. First, despite a vast empirical effort devoted to investigating the true nature of the time varying risk-return relation, it is surprising that there is significantly less research in testing whether and when the negative risk premium can occur. Although the positive risk premium is fundamental and has an important role in theoretical modelling, it remains an open and under-researched empirical question. This study, among a few, demonstrates such testing. Unlike previous studies using the realised return, I employ a novel proxy (ICC) for the conditional expected return to examine the hypothesis. Furthermore, because I do not employ predictive regressions in modelling the expected return in the

³ Fundamental-based ICC is computed at yearly frequency. Thus, if I compute fundamental-based ICC for G-7 countries other than the US, there are only 22 yearly time series data points between 1990 and 2012.

conditional inequalities framework, I therefore avoid model misspecification resulting from assuming functional forms between the risk premium and predetermined information variables (Boudoukh, et al. (1993)). Second, the revealing evidence regarding the negative-risk premium on the market portfolio implies the rejection of one of the most respected models, the Conditional CAPM (Jagannathan and Wang (1996)). While most of the existing evidence on the Conditional CAPM is conducted in the US and focuses on the linear restriction imposed by the model (Lewellen and Nagel (2006)), I find that the Conditional CAPM is rejected, not only in the US, but also in other international markets including Germany, Italy, and Japan. Third, unlike most of the accounting literature that examines firm level ICC, this study is among a few to study the ICC on the aggregate market level (Pastor, et al. (2008); and Li, et al. (2013)). The main advantage lies under aggregation at the market level ICC, which might help reduce estimation errors presented in the ICC calculation at firm level (Lee, So and Wang (2014)).

The chapter proceeds as follows. Section 2.2 provides formal representation of the sign restriction of the ex-ante market risk premium and its application in asset pricing tests. I then show how to test the restriction using the multiple conditional inequalities testing framework. Section 2.3 demonstrates the construction of the ICC. Section 2.4 describes the data for empirical analysis. Section 2.5 discusses the main findings. I offer brief concluding remarks in Section 2.6.

2.2 EX-ANTE RISK PREMIUM RESTRICTION

2.2.1 Ex-ante Risk Premium and Asset Pricing

The restriction that the conditional expected return exceeds the conditional risk free rate can be represented as:

$$E_t(R_{m,t+1}) - R_{ft} \geq 0 \tag{2.1}$$

where $R_{m,t+1}$ is the market expected return from time t to $t + 1$ and $R_{f,t}$ is the risk free rate from t to $t + 1$.

The idea is that all risk-averse investors should demand higher returns for risky assets than for the risk free asset. The non-negative ex-ante risk premium in all states is implied in the conditional version of CAPM, because it ensures that the conditional mean-variance efficiency of the market portfolio is satisfied. However, more general equilibrium models do not impose this restriction. Under the stochastic discount framework, that is, in the absence of arbitrage, there exists a stochastic discount factor, or pricing kernel, such that all asset returns follow the equation (Harrison and Kreps (1979)):

$$E_t(m_{t+1}R_{i,t+1}) = 1 \forall i, \forall t \quad (2.2)$$

where m_{t+1} is the marginal rate of substitution or the pricing kernel at time $t + 1$,
 $R_{i,t+1}$ is the gross return on asset i at time $t+1$.

Equation (2.2) holds for all assets, thus it holds for the market portfolio R_m and the risk free rate R_f . I express the excess market return within the stochastic discount framework by applying equation (2.2) to the market portfolio and the risk free asset, and subsequently taking the difference:

$$E_t(m_{t+1}R_{m,t+1}^e) = 0 \quad (2.3)$$

where $R_m^e = R_m - R_f$ is the market excess return.

Applying the Covariance formula $Cov(X, Y) = E(XY) - E(X)E(Y)$ and rearranging the equation (2.3), the conditional excess return of the market portfolio can be expressed separately as:

$$E_t(R_{m,t+1}^e) = -\frac{Cov_t(m_{t+1}, R_{m,t+1}^e)}{R_{f,t}} \quad (2.4)$$

In light of equation (2.4), the negative market excess return can only be possible if the return of the market portfolio positively covaries with the marginal rate of substitution. Whitelaw (2000) demonstrates that in the standard power utility with a regime-switching consumption process incorporated in the marginal rate of substitution, the negative risk premium can be obtained in states when a regime shift is most likely to occur. In such a state, the volatility of equity return is high, yet expected return might be low because the equity market can induce hedging demand from the investors due to its infinite claims in the future, as opposed to the marginal utility which depends only on the next state's consumption.

Testing the positive risk premium hypothesis provides a possible solution to the well-known critique, emphasised by Roll (1977), with respect to the identity of the market portfolio in any CAPM empirical tests. Roll's insight points to the idea that because the linear restriction hold as a mathematical identity, the important economic content in any appropriate standard CAPM tests lies on the ability of the portfolio proxy being able to represent the true, yet unobservable market portfolio. Such identification is daunting; and attempts to avoid it are largely unsuccessful (see Wheatley (1989)). Fortunately, testing the positive risk premium requires only a weak restriction that is to find a market portfolio proxy that is positively correlated with the true market portfolio,

⁴ Assuming R_f is constant, R_f is expressed in the context of (2.2): $E_t(m_{t+1})R_f = 1 \rightarrow R_f = \frac{1}{E_t(m_{t+1})}$

whereas the magnitude of the correlation is not crucial. To see the logic under the conditional CAPM, the return on the market portfolio proxy $R_{p,t}$ should follow the relation:

$$E_t(R_{p,t+1} - R_{f,t}) = \frac{Cov_t(R_{p,t+1}, R_{m,t+1})}{var_t(R_{m,t+1})} E_t(R_{m,t+1} - R_{f,t}) \quad (2.5)$$

As can be seen from equation (2.5), so long as the covariance between the market portfolio proxy and the market portfolio remain positive i.e. the first term on the right hand side, their risk premiums will have the same sign. Therefore, any evidence against the positive risk premium on the market portfolio proxy indicates the violation of the positive risk premium on the true market, which in turn implies the rejection of the Conditional CAPM. The positive co-movement between the proxy, a well-diversified portfolio containing a large number of assets, and the market portfolio is a weak condition to be satisfied.

2.2.2 Multiple Conditional Inequalities Methodology

Define μ_t as the ex-ante risk premium from (2.1):

$$E_t(R_{m,t+1} - R_{f,t}) = \mu_t \geq 0 \quad (2.6)$$

As the econometricians cannot observe the information set of the economic agents, they face the positivity restriction conditioning on their own information set. This implies multiple inequality restrictions. In particular, suppose $z_{i,t}^+$ is the strictly positive information set that is available to the econometricians,⁵ then the above inequality (2.6) implies multiple restrictions:

$$E_t(R_{m,t+1} - R_{f,t})z_{1,t}^+ = \mu_t \times z_{1,t}^+ \geq 0$$

$$E_t(R_{m,t+1} - R_{f,t})z_{2,t}^+ = \mu_t \times z_{2,t}^+ \geq 0$$

.....

$$E_t(R_{m,t+1} - R_{f,t})z_{i,t}^+ = \mu_t \times z_{i,t}^+ \geq 0$$

Rearranging the preceding set of inequalities and applying the law of iterated expectations, a system similar to the Generalized Method of Moments (GMM) (Hansen and Singleton (1982)) arises:

$$E \left((R_{m,t+1} - R_{f,t}) \otimes z_t^+ - \theta_{\mu z^+} \right) = 0 \quad (2.7)$$

where $\theta_{\mu z^+} = E(\mu_t \otimes z_t) \geq 0$.

⁵ $z_{i,t}^+ \geq 0$.

Different from the GMM, the parameter θ in this system is subject to a set of positivity constraints. The above restriction can be written as a system of N-moment conditions:

$$E[(R_{mt+1} - R_{ft})z_{1t}^+] = \theta_{\mu z_1^+}$$

.....

$$E[(R_{mt+1} - R_{ft})z_{Nt}^+] = \theta_{\mu z_N^+}$$

Boudoukh, et al. (1993) develop a formal framework that takes into account autocorrelation and cross-correlation of the conditional estimates subject to inequality constraints. First, the sample means of the product of the observable variables are estimated. In particular,

$$\hat{\theta}_{\mu z_i}^+ = \frac{1}{T} \sum_{t=1}^T [(R_{mt+1} - R_{ft})z_{i,t}^+] \quad \forall i = 1..N \quad (2.8)$$

$\hat{\theta}_{\mu z_i}^+$ are referred to as the unconstrained estimates because there is no sign restriction imposed on the parameters. They may be negative either because the null is false i.e. the violation of the positive risk premium or simply due to sampling errors.

Next, I calculate sample means under the inequality restriction in the null $\hat{\theta}_{\mu z_i}^R$ by minimising deviations from the unrestricted model under the quadratic form:

$$\min(\hat{\theta}_{\mu z^+}^+ - \theta_{\mu z^+})' \hat{\Omega} (-1)(\hat{\theta}_{\mu z^+}^+ - \theta_{\mu z^+}) \quad (2.9)$$

subject to $\theta_{\mu z^+} \geq 0$.

where $\hat{\Omega}$ is the consistent variance-covariance matrix of the moments. I employ the Bartlett kernel (Newey and West (1987)) to estimate $\hat{\Omega}$. Alternative consistent covariance matrix estimates proposed by Andrews (1991) could also be applied.

The test statistic is:

$$W \equiv T(\hat{\theta}_{\mu z}^R - \hat{\theta}_{\mu z}^+)' \hat{\Omega}^{-1} (\hat{\theta}_{\mu z}^R - \hat{\theta}_{\mu z}^+) \quad (2.10)$$

The idea of the test statistic W is to measure how close the parameters of the restricted model $\hat{\theta}_{\mu z}^R$ are to those of the unrestricted model $\hat{\theta}_{\mu z}^+$. Under the null, the difference should be small. Wolak (1989) shows that W is distributed as a weighted sum of χ^2 with different degrees of freedom $\sum_{k=0}^N \Pr[\chi_k^2 \geq c] w(N, N - k, \frac{\hat{\Omega}}{T})$ where c is the critical value for a given size test, and the weighting function $w(N, N - k, \frac{\hat{\Omega}}{T})$ has exactly $N - K$ positive elements.

2.2.3 Limitations of Realised Returns

Using annual realised return data spanning from 1802-1990, Boudoukh, et al. (1993) find that the ex-ante risk premium is negative in some states of the world, related to periods of high

expected inflation and a downward-sloping term structure. Ostdiek (1998) applies this methodology to test the world market portfolio.

Without modelling the conditional market risk premium directly, the previous authors iterate the conditional moments down to the unconditional expectation as in equation (2.7). However, iterating down to the unconditional expectation involves loss of the test power. This procedure becomes more severe with the use of ex post realised return because it implicitly assumes the convergence of the realised return toward the expected return. There are numerous reasons to believe that this is not the case. First, information surprises in the realised return are either large or highly correlated so that their cumulative effects are large and they might have a permanent impact on the mean realised return (Lundblad (2007)). Thus, it requires a time period spanning over a century for the realised return to approximate the expected return. Hoping for convergence towards the expected return over a realistic study period, if not in the long run, might not be a good idea (Elton (1999)). Second, stock price movements contain information about cash flows and discount rate shocks (Chen and Zhao (2009)). The heated debate on which sources are more important in driving stock prices: the change in cash flow expectations (Bansal and Yaron (2004)) or the change in expected returns (Campbell and Cochrane (1999)) is ongoing. Therefore, the realised return contains more information than that incorporated in the expected return. Third, the time varying expected return makes the convergence to the expected return counterintuitive for equity markets (Campello, Chen and Zhang (2008)). For example, the market realised risk premium, is often negative during bad times. Would anyone infer that stock markets are less risky in these periods? By contrast, it is more intuitive to think that rational investors would demand a higher risk premium in the corresponding times.

An alternative way to overcome this issue is to model the conditional expected return directly by employing predictive regressions using information variables. The idea is to regress future realised returns on a set of economic information variables under assumed functional forms (typically linear). Subsequently, the expected return can be proxied by the predicted value of the model. It can be formally represented in the inequalities framework as such:

$$E \begin{pmatrix} [(R_{m,t+1} - R_{f,t}) - \beta' z_t] \otimes z_t^+ \\ (\beta' z_t) \otimes z_t^+ - \theta_{\mu z} \end{pmatrix} = 0 \quad (2.11)$$

where $\theta_{\mu z} = E(\mu_t \otimes z_t^+) \geq 0$,

z_t is the set of information variables, and

β is the coefficients vector.

The above method implicitly assumes that the information variables in the predicting model can well capture the information set of the economic agents. However, the econometricians' information set is much smaller than that of the representative agents. Additionally, because there is

no consensus on which variables should be used, the method is subject to a data snooping criticism (Foster, et al. (1997); and Ferson, et al. (2008)). Furthermore, much less is known about how these information variables enter in forming expectations. Ghysels (1998) and Harvey (2001) highlight the problem of misspecification of the beta dynamics on inference and estimation, due to the linear functional form. Finally, as pointed out by Boudoukh, et al. (1993), because β needs to be estimated and not specified in the null inequality restriction, deriving global valid test statistics, which are valid for all values of the parameters can be problematic.

2.3 IMPLIED COST OF CAPITAL

2.3.1 Analytical Framework

In this section, I show analytically how the implied cost of capital (ICC) can be an excellent proxy for the time varying expected return. By definition, the ICC is the internal rate of return that equates market price with the present value of streams of future cash flows (earnings or dividends):

$$P_t = \sum_{k=1}^{\infty} \frac{E_t(D_{t+k})}{(1+r_e)^k} \quad (2.12)$$

where P_t is the firm stock price,

D is the stream of dividends,

r_e is the implied cost of capital.

To provide tractability, the present value formula can be represented in a log-linear approximation (Campbell and Shiller (1988)):

$$p_t = \frac{k}{1-\rho} + (1-\rho) \sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j}) - \sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}) \quad (2.13)$$

where $p_t = \log(P_t)$

$d_t = \log(D_t)$,

$$\rho = \frac{1}{1+\exp(\overline{d-p})},$$

$\overline{d-p}$ is the average dividend yield,

$$k = -\log(\rho) - (1-\rho)\log\left(\frac{1}{\rho-1}\right)$$

Analogous to equation (2.13), I define r_e that solves:

$$p_t = \frac{k}{1-\rho} + (1-\rho) \sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j}) - r_{e,t} \sum_{j=0}^{\infty} \rho^j \quad (2.14)$$

Equations (2.13) or (2.14) show that as an accounting identity, an increase (decrease) in the market price can be attributed to any increase (decrease) in expectation of the future cash flows

and/or decrease (increase) in the expected future returns. Under a further assumption that the log conditional expected return μ_t follows an AR(1) process:

$$\mu_{t+1} = \alpha + \gamma\mu_t + \epsilon, 0 < \gamma < 1 \quad (2.15)$$

Pastor, et al. (2008) demonstrate that:

$$r_{e,t} = \frac{\alpha}{1-\gamma} + \left(\mu_t - \frac{\alpha}{1-\gamma}\right) \frac{1-\rho}{1-\rho\gamma} \quad (2.16)$$

Indicated by equation (2.16), $r_{e,t}$ is perfectly correlated with the conditional expected return μ_t , and thus an excellent proxy for conditional expected return.

In stark contrast to the ex-post realised return, the ICC is forward-looking. As these primary inputs are determined by market participants, this market-based measure of the time varying expected return does not impose the assumption that the information set observed by the economic agents is the same as that of the econometricians in modelling the conditional expected return. Additionally, if the expected cash flows in the model truly capture the market expectation of the firm's future cash flows, the remaining price variation is attributed to changes in the expected return. As a result, the ICC measure might be less, if not, contaminated by the cash flow news.

2.3.2 ICC Construction

To calculate the ICC empirically, I follow Pastor, et al. (2008), Chava and Purnanandam (2010), and Li, et al. (2013). The model is a version of the residual income model of Gebhardt, Lee and Swaminathan (2001). At each time period the ICC of a firm is computed by equating the firm's stock price and its expected cash flows to equity and solving for the internal rate of return:

$$P_t = \sum_k^{\infty} \frac{E_t(FCFE_{t+k})}{(1+r_e)^k} \quad (2.17)$$

where P_t is the stock price at year t ,

$FCFE_{t+k}$ is the free cash flow to equity at year $t+k$,

r_e is the ICC at year t that will be solved numerically.

Next, the expected cash flow to equity is defined as:

$$E_t(FCFE_{t+k}) = FE_{t+k}(1-b_{t+k})$$

where FE_{t+k} is the earning forecast of year $t+k$ at year t ,

b_{t+k} is the plowback ratio at year $t+k$, thus $1-b_{t+k}$ is the payout ratio.

Because it is impractical to forecast earnings to infinity, equation (2.17) is decomposed into the explicit forecasts component spanning 15 years and the terminal value component starting from year 16. Specifically, equation (2.17) becomes:

$$P_t = \sum_{k=1}^{15} \frac{FE_{t+k}(1 - b_{t+k})}{(1 + r_e)^k} + \frac{FE_{t+16}}{r_e(1 + r_e)^{15}} \quad (2.18)$$

The explicit earnings forecast component is then divided into two sub-periods. The earnings forecasts for the first three years (FE_{t+1} , FE_{t+2} , and FE_{t+3}) are obtained directly. Subsequently, to compute FE_{t+4} to FE_{t+16} the long term growth forecast at year $t + 3$ is assumed to exponentially mean-revert to the steady-state value equal to the nominal GDP growth rate g :

$$FE_{t+k} = FE_{t+k-1} \times (1 + g_{t+k}) \quad (2.19)$$

$$g_{t+k} = g_{t+k-1} \times \exp\left(\frac{\log\left(\frac{g}{g_{t+3}}\right)}{13}\right), 4 \leq k \leq 16 \quad (2.20)$$

where g_{t+k} is the earnings forecast growth rate between year $t + k - 1$ to $t + k$,

g_{t+3} is long term growth forecast at year $t + 3$,

g is the steady-state value which is the nominal GDP growth rate at year t .

In terms of plowback ratios, b_{t+1} and b_{t+2} are obtained as the firm's most recent net payout ratio. From year $t + 3$ to $t + 16$, the plowback rates are assumed to follow a linear mean-reverting process to the steady-state value:

$$b_{t+k} = b_{t+k-1} - \frac{b_{t+2} - b}{15}, 3 \leq k \leq 16 \quad (2.21)$$

where $b = \frac{g}{r_e}$ is the steady-state value of plowback ratio,⁶

b_{t+k} is previously defined.

Plugging equations (2.19), (2.20), and (2.21) into equation (2.18) and solving numerically for r_e , the firm level ICC is obtained at each time period.

Note that the performance of the ICC in capturing the time varying expected return depends crucially on how well the earnings forecasts explain cash flow expectation variation. I employ a fundamental-based earnings forecast model recently developed by Hou, et al. (2012), in addition to the analyst earnings forecasts, to proxy for cash flow expectations. They show various advantages of the earnings produced by this method over the analyst forecasts in capturing the market expectations of cash flows. Biases pertaining to analyst forecasts such as optimism, survivorship, or

⁶ The steady-state plowback ratio $b = \frac{g}{r_e}$ implies that return on new investment is equal to the cost of equity r_e , meaning that after explicit forecast periods there is no additional economic value associated with new investments.

conflicts of interest can be overcome by the fundamental-based substitute. Unfortunately, due to the short data period, I do not apply the fundamental-based method for the international markets. The construction of the earnings forecast is discussed in more detail in the next section.

2.4 DATA AND SAMPLE SELECTION

2.4.1 Data Sources

For the US, I obtain data of firms listed on the NYSE, AMEX and NASDAQ exchanges (CRSP item EXCHCD with value of 1, 2, and 3), and exclude ADRs, close-end funds, and REITS (CRSP item SHRCDD not having value 10 and 11). Annual accounting information and stock market information are collected from the COMPUSTAT and the CRSP databases, respectively. Consensus analyst forecast information is from the I/B/E/S summary files. GDP growth rates are acquired from the Bureau of Economic Analysis. Government bond yields are from Federal Reserve Bank Reports. The sample period for the US starts from 1962 (1977) for the fundamental-based forecasts (analyst-based forecasts) and ends in 2012.

For non-US firms, I obtain firm level stock price, shares outstanding and earnings forecasts from the I/B/E/S database. Accounting and exchange rate data are collected from the GLOBAL COMPUSTAT vendor. Bond yields and the Morgan Stanley Capital International (MSCI) market return indices are from DATASTREAM, while GDP growth rates are provided by the World Bank. The sample for non-U.S. firms spans the course of 1990-2012. Data filtering and merging criteria between databases are discussed in the following sections.

2.4.2 Earnings Forecasts

For US firms, explicit earnings forecasts are acquired in two ways: analyst forecasts and fundamental-based forecasts (Hou et al. 2012). With respect to the analyst-based approach, I obtain the mean consensus earnings forecasts of each firm for the fiscal years FE_{t+1} , FE_{t+2} , and the long term growth forecast g_{t+3} (i.e. I/B/E/S item FPI with value 1, 2, and 0, respectively). I then calculate $FE_{t+3} = FE_{t+2}(1 + g_{t+3})$. If the long term growth forecast is missing, I replace it with the implicit figure as $g_{t+3} = \frac{FE_{t+2}}{FE_{t+1}} - 1$. Non-missing and non-zero data of FE_{t+1} and FE_{t+2} are required. Long term growth forecasts which are less (more) than 2% (100%), are set to 2% (100%),

except for firms with negative FE_{t+2} .⁷ The steady-state value of earning growth is the nominal GDP growth rate estimated by the historical average GDP growth rate, using an expanding window to capture the implicit inflation rate.

With respect to the fundamental-based approach, I use the methodology of Hou, et al. (2012) to forecast earnings for the first 3 years. Specifically, rolling pooled cross-sectional regressions using the previous 6 years' data are run for each year:

$$E_{i,t+\tau} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DivDum_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegEDum_{i,t} + \alpha_6 AC_{i,t} + \varepsilon_{t+\tau} \quad (2.22)$$

where $E_{i,t+\tau}$ is the earnings before extraordinary items of firm i in year $t + \tau$ ($\tau = 1$ to 3) (COMPUSTAT item IB),

$A_{i,t}$ is the total asset of firm i at year t (COMPUSTAT item AT),

$D_{i,t}$ is the dividend payment of firm i at year t (COMPUSTAT item DVC),

$DivDum_{i,t}$ is the dummy equal to 1 if $D_{i,t} > 0$ year t , and 0 otherwise,

$E_{i,t}$ is the earning before extraordinary items of firm i at year t ,

$NegEDum_{i,t}$ is the dummy variable equal to 1 if $E_{i,t} < 0$, and 0 otherwise,

$AC_{i,t}$ is the total accrual of firm i at year t , calculated as the change in total current assets (COMPUSTAT item ACT) minus the change in cash and short-term investment (COMPUSTAT item CHE) minus the change in total current liabilities (COMPUSTAT item LCT) plus the change in total debt in current liabilities (COMPUSTAT item DLC).

The first three year-ahead earnings (FE_{t+1} , FE_{t+2} , and FE_{t+3}) are forecasted, out-of-sample, based on the estimated coefficients. I require data availability for the level variables in the regressions. The earnings and the level variables are winsorised each year at the 1st and 99th percentiles to mitigate the effect of extreme observations. The long term growth forecast is calculated as $g_{t+3} = \frac{FE_{t+3}}{FE_{t+2}} - 1$. Other filtering criteria are similar to the analyst-based approach.

For non-US firms, I do not apply the fundamental-based methodology because at yearly frequency, my sample has only 22 data points for each country. The analyst-based earnings forecasts are constructed similar to US firms.

⁷ If $FE_{t+2} < 0$, I keep observations with long term growth forecasts greater than 100% because it is the only way the remaining years' forecasts do not end up with negative numbers. Dropping these observations does not affect the main result.

2.4.3 Plowback Ratio

With respect to US firms, for the first two years the plowback ratio is equal to one minus the current net payout ratio of the firm. The net payout ratio is calculated as

$$NPR_{i,t} = \frac{D_{i,t} + REP_{i,t} - NE_{i,t}}{NI_{i,t}} \quad (2.23)$$

where $NPR_{i,t}$ is the net payout ratio of firm i at year t ,

$D_{i,t}$ is the dividend payment of firm i at year t (COMPUSTAT item DVC),

$REP_{i,t}$ is the amount of common and preferred stock repurchased by firm i at year t (COMPUSTAT item PRSTKC),

$NE_{i,t}$ is the amount of common and preferred stock sold by firm i at year t (COMPUSTAT item SSTK),

$NI_{i,t}$ is the net income of firm i at year t (COMPUSTAT item NI).

Non-missing values for $D_{i,t}$ and $NI_{i,t}$ are required. If $NI_{i,t}$ is negative, it is then set to be 6% of total assets (Chen, et al. (2013)). If either $REP_{i,t}$ or $NE_{i,t}$ is missing, I replace with a value of 0.⁸ If $NPR_{i,t}$ is greater (less) than 1 (0) then a value of 1 (0) is assigned.

For other G-7 countries, due to the data limitation I use the dividend payout ratio as dividends divided by earnings. If firm earnings (GLOBAL COMPUSTAT item NICON) are negative, they are again first set to 6% of total assets; and subsequently I calculate the dividend payout ratio. If firm earnings are missing, I calculate them as the sum of income before extraordinary items (GLOBAL COMPUSTAT item IB), extraordinary items (GLOBAL COMPUSTAT item XI), and discontinued items (GLOBAL COMPUSTAT item XIDO).⁹ If the dividend payout ratio is greater (less) than 1 (0) then a value of 1 (0) is assigned.

2.4.4 Putting It All Together

With respect to the analyst-based forecasts approach in the US sample, the final sample is the intersection between the COMPUSTAT, the CRSP, and the I/B/E/S databases at monthly

⁸ A stricter requirement of dropping observations with missing values of $REP_{i,t}$ or $NE_{i,t}$ does not affect the main result.

⁹ NICON are missing partly due to missing values of XI and XIDO. By setting XI and XIDO to 0 when their values are missing, I implicitly assume Net Income (NICON) is equal to income before extraordinary items (IB).

frequency. To ensure the public availability of accounting information, i.e. to avoid the look-ahead bias, firms' fiscal year-end (COMPUSTAT) is required to be at least three months prior to the month when the ICC is calculated. The analyst forecasts are matched with the firms' stock price at the end of the month. Stock prices are adjusted for stock splits using an adjustment factor (CRSP item CFACPR). Earnings forecasts are already adjusted in the I/B/E/S database.

For non-US firms, the merged sample is from the GLOBAL COMPUSTAT and the I/B/E/S databases. Accounting information is required to be at least six months prior to the period of ICC computation. As earnings forecasts are often submitted a few days after mid-month, when the ICC is computed I merge firm earnings forecasts with firm stock prices on the closest trading day after the 15th of the month. Stock prices and the number of outstanding shares are already adjusted in the I/B/E/S database. To maintain consistency, I calculate realised returns from the first trading day after the 15th of the previous month to the first trading day after the 15th of the current month. The extreme values of the ICC in each month are winsorised at the 1st and 99th percentiles. The market-wide ICC is then computed, alternatively, as the value-weighted and equal-weighted averages of individual ICCs to represent the market portfolio's expected return. This process produces the analyst-based ICC.

With regard to the fundamental-based earnings forecast approach, only the COMPUSTAT and the CRSP databases are required at yearly frequency. I estimate the ICC for each firm at the end of June. To make sure the accounting information is publicly available at the time the ICC is calculated, a reporting lag of three months is again required. For example, the accounting information of firms having fiscal year-end between April last year and March this year is used to estimate the ICC in June this year. The ICC sample is winsorised at the 1st and 99th percentiles at the end of June each year. Unlike the analyst-based market-wide ICC, the model-based market-wide ICC is just an equal-weighted average of every firms' ICC each year end. Value-weighting is not employed in this instance since the earnings forecast regression in equation (2.22) has implicitly imposed the weight in earnings forecasts through the fixed coefficient on the $A_{i,t}$ variable; thus subsequently weighting the individual firms' ICC. To avoid double-weighting, an equal-weighted scheme is more appropriate than the value-weighted counterpart in calculating the market-wide ICC.

Finally, the implied risk premium is equal to the market-wide ICC minus long term government bonds, which are 10-year (7-year) government bond yields for the US, the UK, Japan, Germany, Canada, and France (Italy). To compute the realised risk premium, I subtract 1-year continuously compounded value-weighted market realised return from 1-year T-bill (1-year interbank offer rate from British Bankers Association BBA) for US and Canada (Germany, France,

Italy, UK, and Japan). For the US, I use two market portfolios CRSP index and S&P 500 Index. For the other G-7 countries, local MSCI return indices are employed as market portfolios.

2.5 EMPIRICAL RESULTS

2.5.1 Descriptive Statistics

Panels A and B of Table 2.1 report summary statistics for the annual risk premium, computed by the ICC method (fundamental-based and analyst-based) and ex-post realised returns, on two “market” proxies CRSP-AMEX/NYSE/NASDAQ (Panel A) and S&P 500 (Panel B). With respect to the CRSP portfolio, the average value-weighted and equal-weighted implied risk premium computed from analyst forecasts are 4.9% p.a. and 5.4% p.a., respectively. These figures are close to the 4.5% computed from the realised return. The equal-weighted implied annual risk premium derived from the fundamental-based earnings forecasts is quite high at 8%, partly reflecting the inclusion of a larger number of firms (2666 firms) that are mostly smaller than the analyst-based counterpart (2416 firms). The standard deviation of the fundamental-based implied risk premium, 4.5%, is considerably larger than that of the analyst-based value-weighted (equal-weighted) risk premium being 2.03% (1.8%). However, the standard deviation of the fundamental-based implied risk premium is substantially lower than that of the realised premium, 17.07%, which highlights the important point stressed earlier that the realised return is a noisy estimate for the conditional expected return.

With regard to the S&P 500 market portfolio in Panel B, the statistics are very similar except for the fundamental-based risk premium. In this case, the average annual risk premium undergoes a major reduction to 2% (Column 3). The average number of firms in the fundamental-based approach is now less than that of the analyst-based counterpart (391 versus 448 firms), as opposed to the CRSP portfolio (2666 versus 2416 firms). The difference is due to the data filtering criteria of the two earnings forecast methods.

Table 2.1: Summary Statistics

This table presents the summary statistics of risk premium (%) computed by the ICC methodology and ex-post realised returns of the CRSP (Panel A) and the S&P 500 (Panel B) portfolios. Summary statistics for other G7 countries are reported in Panel C (equal-weighted analyst-based risk premium), Panel D (value-weighted analyst-based risk premium), and Panel E (realised risk premium of local MSCI indices). At firm level, the implied cost of capital is derived from analyst earnings forecasts or fundamental-based earnings forecast regressions. I term these ‘analyst-based ICC’ and ‘fundamental-based ICC’. The market-wide ICC is the equal-weighted or value-weighted average of individual firms’ analyst-based ICC at monthly frequency. For fundamental-based ICC, the market-wide ICC is the equal-weighted average of individual firms’ ICC at yearly interval. The ex-post market realised returns are value-weighted returns including dividends at monthly frequency. The implied risk premium is computed by subtracting the yield of 10-year or 7-year Treasury Bond from the market-wide ICC. I compute realised risk premium for US and Canada market portfolios by subtracting the 1-year T-bill rate from market realised returns. For non-US and non-Canada firms, 1-year T-bill rates are replaced by 1-year interbank offer rates by British Banker Association BBA. The study period is from 1977-2012 for analyst-based ICC, while for fundamental-based ICC and realised-returns the study period spans 1962 to 2012. The sample period for other G-7 countries is from 1990 to 2012.

| Panel A: Risk Premium CRSP | | | | |
|-----------------------------------|----------------|----------------|----------------|------------------|
| | Analyst | | Fundamental | Realised Returns |
| | Value-weighted | Equal-Weighted | Equal-Weighted | Value-Weighted |
| Mean | 4.89 | 5.40 | 8.00 | 4.50 |
| Std Deviation | 2.03 | 1.80 | 4.45 | 17.07 |
| Minimum | 1.15 | 1.33 | 0.41 | -47.89 |
| Q1 | 3.46 | 4.36 | 4.63 | -6.01 |
| Median | 4.40 | 4.97 | 7.92 | 6.83 |
| Q3 | 5.92 | 6.44 | 11.55 | 15.53 |
| Maximum | 14.29 | 11.29 | 18.21 | 54.75 |
| Number of firms | 2416 | 2416 | 2666 | 5650 |

| Panel B: Risk Premium S&P 500 | | | | |
|--|----------------|----------------|----------------|------------------|
| | Analyst | | Fundamental | Realised Returns |
| | Value-Weighted | Equal-Weighted | Equal-Weighted | Value-Weighted |
| Mean | 5.16 | 4.72 | 1.98 | 4.49 |
| Std Deviation | 2.13 | 1.87 | 2.89 | 16.54 |
| Minimum | 1.04 | 0.81 | -3.95 | -46.05 |
| Q1 | 3.72 | 3.41 | -0.67 | -4.96 |
| Median | 4.71 | 4.17 | 2.50 | 6.26 |
| Q3 | 6.32 | 5.68 | 4.61 | 15.82 |
| Maximum | 16.74 | 11.26 | 8.02 | 50.00 |
| Number of firms | 448 | 448 | 391 | 500 |

Table 2.1 (continued)

| | CAN | JPN | GER | UK | ITA | FRA |
|---|--------------|---------------|---------------|--------------|---------------|---------------|
| Panel C: Equal-Weighted Analyst-based Risk Premium | | | | | | |
| Mean | 9.21 | 6.72 | 6.28 | 7.07 | 4.86 | 5.00 |
| Std Deviation | 3.08 | 4.71 | 4.25 | 3.13 | 4.10 | 2.66 |
| Min | 1.87 | -3.93 | -2.24 | 0.73 | -6.19 | -3.28 |
| Q1 | 7.20 | 3.22 | 3.20 | 4.19 | 2.28 | 3.36 |
| Median | 8.73 | 8.34 | 6.93 | 7.14 | 5.86 | 5.13 |
| Q3 | 10.76 | 9.90 | 9.59 | 9.35 | 7.89 | 6.99 |
| Max | 19.56 | 14.03 | 14.20 | 15.28 | 11.80 | 10.15 |
| Panel D: Value-Weighted Analyst-based Risk Premium | | | | | | |
| Mean | 5.38 | 5.01 | 3.27 | 3.12 | 2.10 | 2.53 |
| Std Deviation | 2.03 | 3.91 | 2.86 | 2.11 | 3.21 | 2.02 |
| Min | 0.13 | -3.29 | -2.20 | -1.23 | -6.94 | -3.28 |
| Q1 | 4.10 | 2.78 | 1.37 | 1.60 | -0.28 | 1.04 |
| Median | 4.93 | 5.08 | 2.89 | 2.71 | 2.54 | 2.09 |
| Q3 | 6.48 | 7.37 | 5.17 | 4.29 | 4.62 | 3.97 |
| Max | 10.79 | 13.44 | 10.36 | 7.92 | 8.68 | 7.16 |
| Number of firms | 241 | 839 | 242 | 591 | 87 | 242 |
| Panel E: Realised Risk Premium MSCI Index | | | | | | |
| Mean | 4.07 | -2.76 | 2.64 | 2.19 | -0.68 | 2.34 |
| Std Deviation | 17.85 | 23.65 | 24.58 | 15.50 | 24.36 | 22.35 |
| Min | -58.08 | -58.14 | -78.76 | -51.52 | -72.61 | -56.05 |
| Q1 | -7.42 | -19.04 | -13.51 | -4.51 | -17.49 | -13.42 |
| Median | 5.84 | -3.47 | 9.84 | 6.29 | 3.20 | 7.59 |
| Q3 | 15.06 | 13.63 | 19.77 | 12.25 | 16.63 | 19.49 |
| Max | 53.00 | 54.30 | 51.82 | 40.22 | 62.53 | 41.74 |

Panel C, D and E of Table 2.1 report the summary statistics for the equal-weighted, the value-weighted implied risk premium, and the realised risk premium of the local MSCI indices, respectively, for the international markets. These numbers are close to those reported in Pastor et al. (2008) (Table II, page 2875). With respect to the implied risk premium, the unconditional means of the equal-weighted measure are always greater than their value-weighted counterparts, indicating that the small-firm effect exists in the international markets. Across the implied measures, Canada has the highest risk premium (9.21% for equal-weighted and 5.38% for value-weighted); whereas the least risky stock market is Italy (4.86% for equal-weighted and 2.1% for value-weighted). Notably, the Japanese equity market which experiences a prolonged poor performance from the early 1990s, is the second (third) most risky market among G-7 countries according to equal-weighted (value-weighted) implied risk premium. Turning to the realised risk premium (Panel E of Table 2.1), consistent with the implied measures, Canada is still the most risky stock market among non-US countries (4.07%). However, Japan provides contrasting evidence, as its equity market is

now considered the least risky, with the mean (median) realised risk premium of -2.76% (-3.84%). It is quite counterintuitive to rationalise the inference, because rational investors should see the Japanese stock market as considerably risky. Additionally, there are multiple periods where the realised risk premium are negative, reflected by the fact that the first quartile statistic (Q1) is negative in all of the countries; whereas there are only a few negative observations in the implied samples. Finally, the standard deviation in the realised risk premium is considerably larger than its implied counterparts emphasising the noise embedded in the realised proxy.

Table 2.2 presents the correlation among measures of the market risk premium for the US sample.¹⁰ The correlation of the ex-post realised risk premium of the CRSP and S&P 500 portfolios is 98.5% and significantly positive at the 1% level of significance. In contrast, the ex-post realised measures are negatively correlated with the analyst-based and fundamental-based counterpart, yet not statistically significant. Finally, the implied risk premium are all positively correlated with each other (value-weighted versus equal-weighted, analyst-based versus fundamental-based). This is not surprising considering that the implied risk premium is derived from the same residual income model.

Figure 2.1 plots the time series of the risk premium on the CRSP portfolios. Panel A presents the implied risk premium for the analyst-based and the fundamental-based approaches. The analyst-based value-weighted and the equal-weighted plots tend to move in tandem, consistent with the 74.4% correlation shown in Table 2.2. For instance, there is an increase in the implied risk premium from 2007 to 2009, which intuitively matches the recent financial crisis. This pattern is also followed by the fundamental-based risk premium. Divergence between the analyst-based equal-weighted and the value-weighted expected returns occurs, particularly around the late 1990s and the early 2000s, potentially reflecting the period of technology stock bubbles.

Panel B of Figure 2.1 displays the time series plots of the implied risk premium with the ex-post realised risk premium. Consistent with Panel A of Table 2.1, the ex-post realised risk premium is very noisy compared to the implied risk premium series. There is no particular pattern that can be observed with realised returns. In fact, some of the trends are missed. For example, while my implied risk premium reveals that equity is risky in this period, the ex post realised premium presents a significantly negative risk premium.

¹⁰ For brevity, I do not report the correlation matrix for non-US countries here because the correlation between the implied and the realised risk premium exerts the same patterns. The correlation matrix is reported in Table A.1.1 in Appendix A.1.

Table 2.2: Correlation Matrix

This table reports the correlation matrix among measures of risk premium, computed by the implied cost of capital methodology and ex-post realised returns of the CRSP and S&P 500 portfolios. At firm level, the implied cost of capital is derived from analyst earnings forecasts or fundamental-based earnings forecast regressions. I term these as analyst-based ICC and fundamental-based ICC. The market-wide ICC is the equal-weighted or value-weighted average of individual firms' analyst-based ICC at monthly frequency. For fundamental-based ICC, the market-wide ICC is the equal-weighted average of individual firms' ICC at yearly interval. The ex-post market realised returns are value-weighted returns including dividends at monthly frequency. The risk premium is computed by subtracting the yield of 10-year Treasury Bond from the market-wide ICC or the market ex-post realised returns. The study period is from 1976-2012 for analyst-based ICC, while for fundamental-based ICC and realised-returns the study period spans from 1962 to 2012. Superscripts *, **, *** indicate statistically significant levels at 10%, 5%, and 1%, respectively.

| | <i>Realised_{CRSP}</i> | <i>Value Analyst_{CRSP}</i> | <i>Equal Analyst_{CRSP}</i> | <i>Fundamental_{CRSP}</i> | <i>Realised_{S&P}</i> | <i>Value Analyst_{S&P}</i> | <i>Equal Analyst_{S&P}</i> | <i>Fundamental_{S&P}</i> |
|--|--------------------------------|-------------------------------------|-------------------------------------|-----------------------------------|-----------------------------------|--|--|--------------------------------------|
| <i>Realised_{CRSP}</i> | 1 | | | | | | | |
| <i>Value Analyst_{CRSP}</i> | -0.0263 | 1 | | | | | | |
| <i>Equal Analyst_{CRSP}</i> | -0.144 | 0.744*** | 1 | | | | | |
| <i>Fundamental_{CRSP}</i> | -0.0446 | 0.560*** | 0.635*** | 1 | | | | |
| <i>Realised_{S&P}</i> | 0.985*** | -0.108 | -0.212 | -0.107 | 1 | | | |
| <i>Value Analyst_{S&P}</i> | -0.0269 | 0.991*** | 0.669*** | 0.543*** | -0.105 | 1 | | |
| <i>Equal Analyst_{S&P}</i> | -0.168 | 0.905*** | 0.885*** | 0.595*** | -0.25 | 0.877*** | 1 | |
| <i>Fundamental_{S&P}</i> | -0.0225 | 0.743*** | 0.659*** | 0.630*** | -0.107 | 0.732*** | 0.778*** | 1 |

One important thing to note in this graph is that the conditional implied risk premium is always greater than zero. The minimum values of the analyst equal-weighted, value-weighted and the fundamental premium are 1.3%, 1.1%, and 0.4%, respectively (Panel A, Table 2.1). A direct implication of these observations is that the inequality constraint hypothesis cannot be rejected (with the associated p-value equal to 1) because there is no difference between estimates of the restricted and unrestricted models. Therefore, this restriction is not violated, if 1) the market proxy is the CRSP portfolio and 2) expected return is proxied by the ICC. In contrast to the implied premium, the ex-post realised premium has many cases below zero. This could be due to sampling error, or a reflection that the positivity restriction of the market risk premium is truly violated.

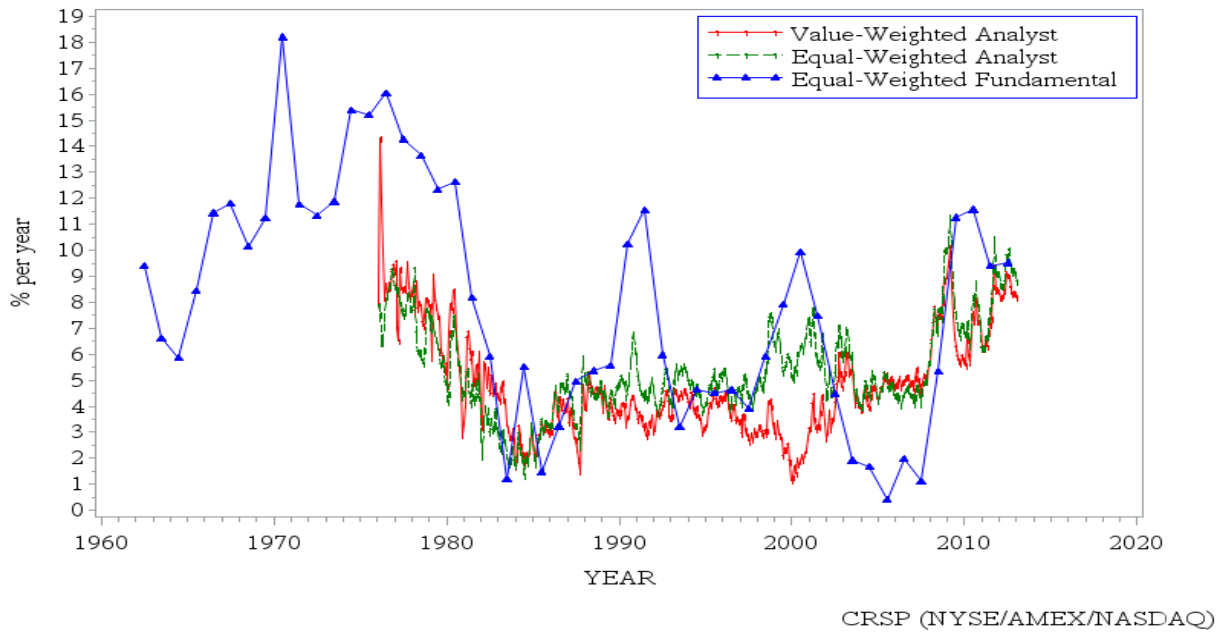
Panel A of Figure 2.2 plots the time series of the implied risk premium of the S&P 500 “market” portfolio. Similar to Figure 2.1, the analyst-based implied risk premium never becomes negative, which implies that the positivity restriction on risk premium cannot be rejected with a p-value equal to 1. On the other hand, the fundamental-based approach produces negative risk premium in some instances. Specifically, the negative risk premium occurs in the late 1980s and 1990s corresponding to periods with relatively low risk premium produced by the analyst-based approach. Once again, a negative risk premium frequently occurs when the ex-post realised returns are used.

Figure 2.3 plots the time series of the ICC (left-hand-side of each panel) and the implied risk premium (right-hand-side of each panel) for the international markets. With the exception of Japan, there is a strikingly similar mean-reverse pattern for the ICC. The market-wide ICC is high in the early 1990 and starts to decrease to the lowest level around 2000. This pattern is quite interesting because it corresponds to the period leading up to the European Monetary Union (EMU) formation. Since 2000, the market-wide ICC exhibits an increasing trend, with a typical spike around the Global Financial Crisis period. After added back the risk free rates into the ICC plots, there is a clear upward trend of the implied risk premium across all international markets partly associated with declining risk free rates. With the exception of Canada, negative risk premium is observed in non-US samples. As a consequence, I only apply the multiple conditional inequalities test to these samples because I know for certain that the positivity of the Canadian implied market risk premium cannot be rejected.

Figure 2.1: Risk Premium of the CRSP Portfolio

Panel A plots the time series of the value-weighted and equal-weighted analyst-based, and the equal-weighted fundamental-based implied risk premium. Panel B plots the value-weighted analyst-based, equal-weighted fundamental-based, and value-weighted ex-post realised risk premium of the CRSP market portfolio.

Panel A: Implied Risk Premium



Panel B: Risk Premium

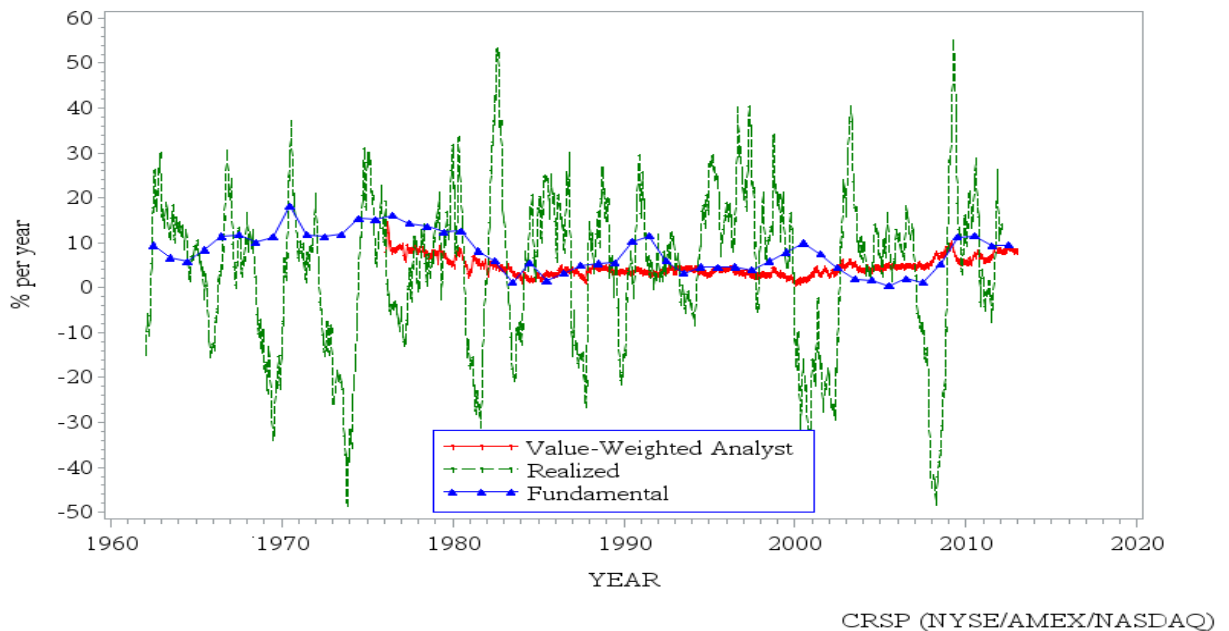
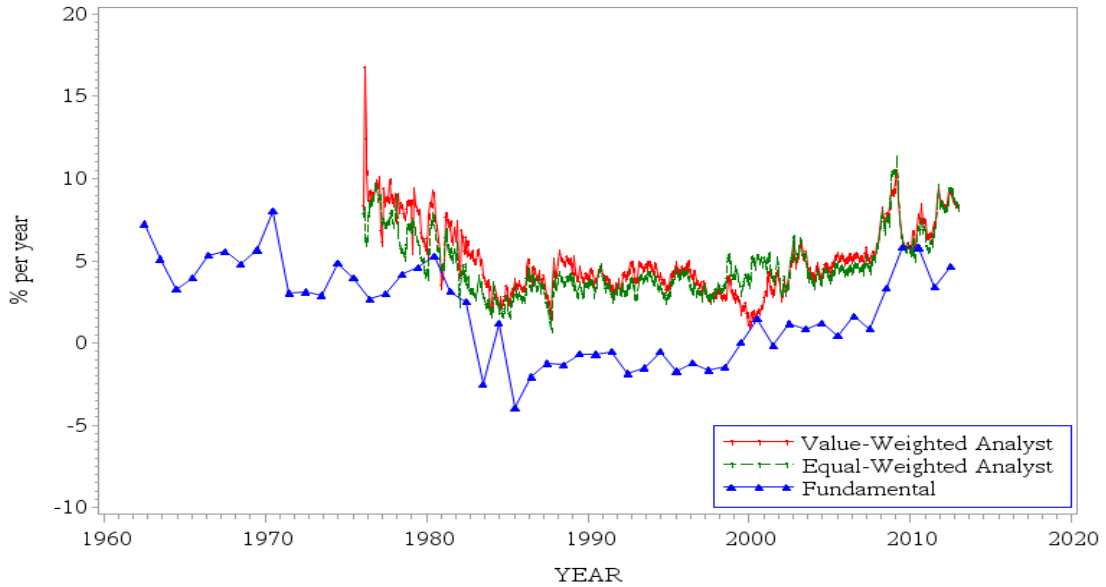


Figure 2.2: Risk Premium of the S&P500 Portfolio

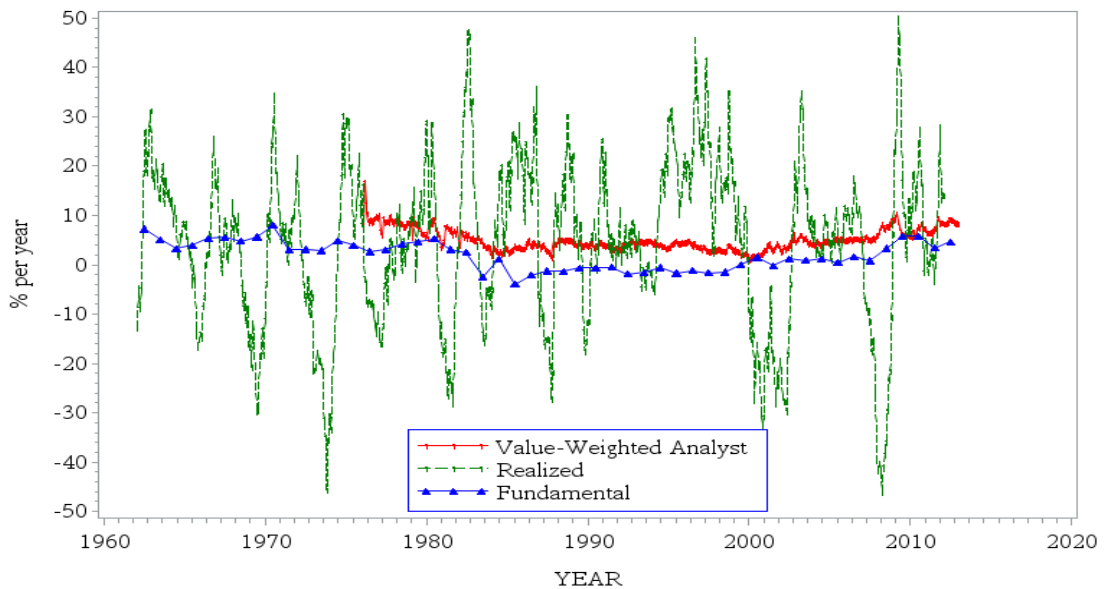
Panel A plots the time series of the value-weighted and equal-weighted analyst-based, and the equal-weighted fundamental-based implied risk premium. Panel B plots the value-weighted analyst-based, equal-weighted fundamental-based, and value-weighted ex-post realised risk premium of the S&P 500 market portfolio.

Panel A: Implied Risk Premium



SP500

Panel B: Risk Premium

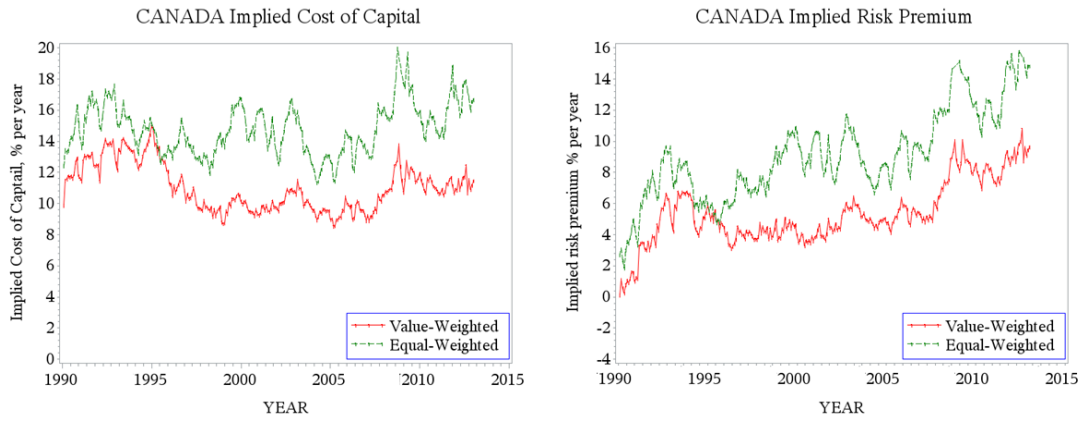


SP500

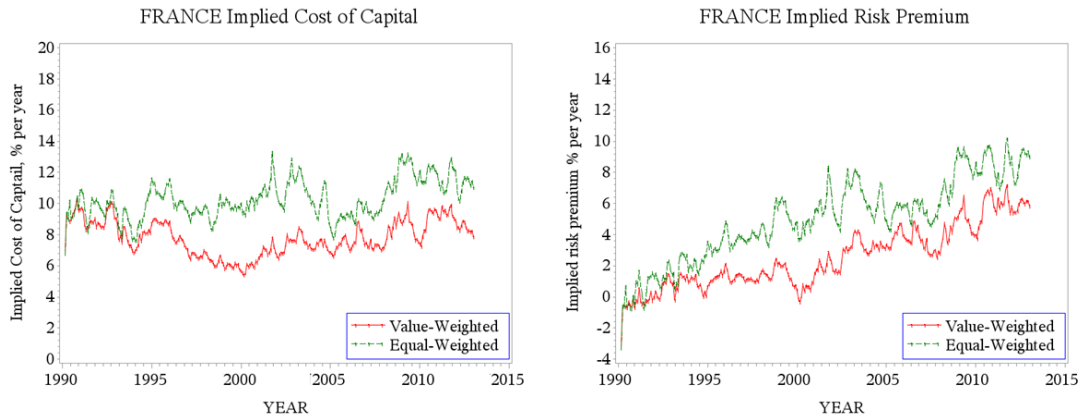
Figure 2.3: Risk Premium of the International Markets

The time series of the analyst-based implied cost of capital (left-hand-side) and the implied risk premium (right-hand-side) are plotted for non-US countries.

Panel A



Panel B



Panel C

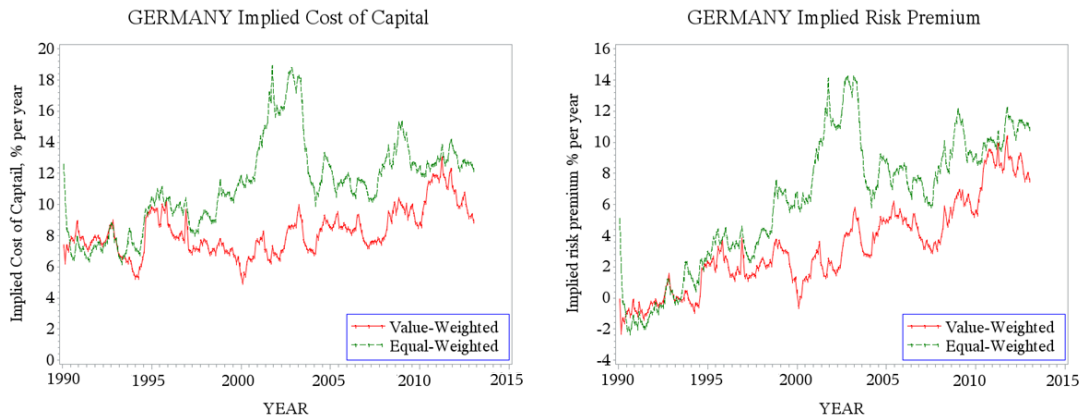
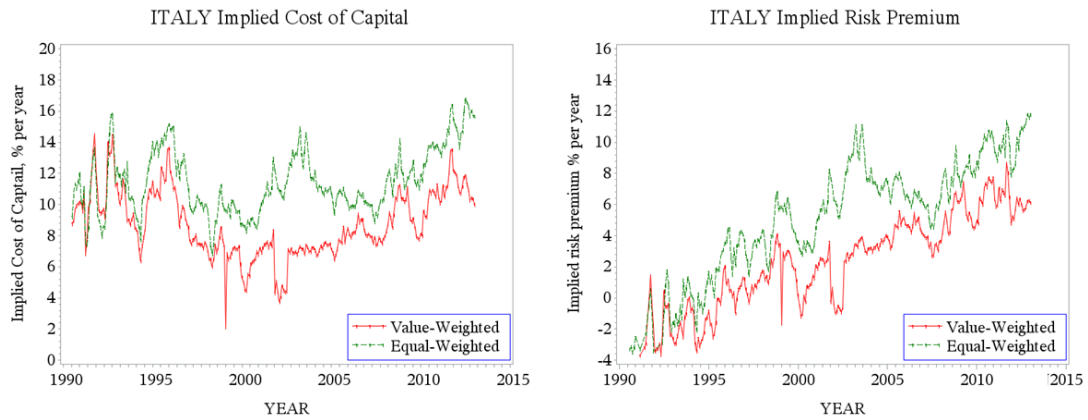
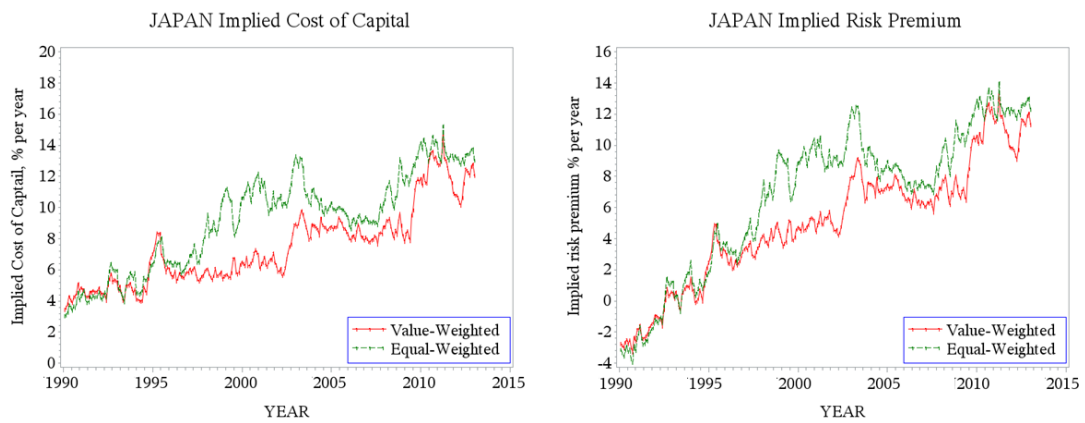


Figure 2.3 (Continued)

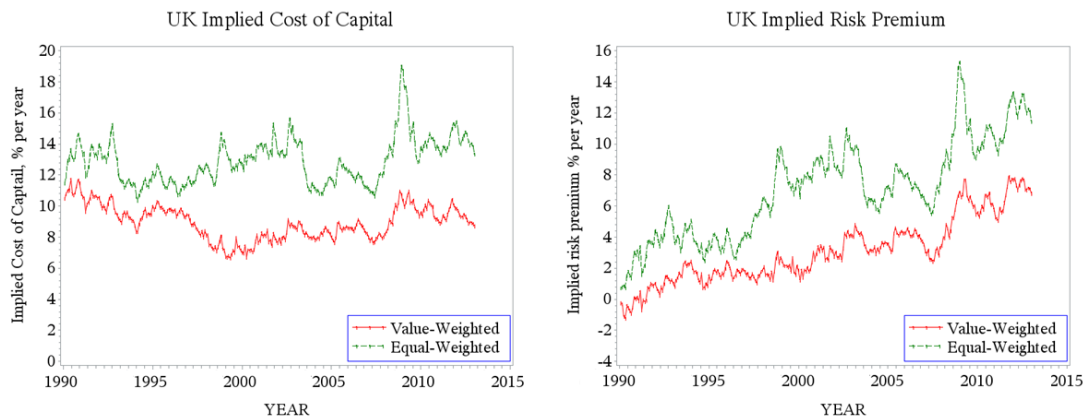
Panel D



Panel E



Panel F



2.5.2 Test of Non-negative Market Risk Premium Hypothesis

2.5.2.1 Instrumental Variables

In the inequality constraint framework, I employ three instrumental variables: 1-year nominal Treasury bill or interbank offer yields R_{bill} ; term structure [with a difference between 5-

year (10-year) Treasury bond R_{bond} for the US (non-US) and 1-year T-bill (1-year interbank offer rate) nominal yields for the US (non-US); and lag negative risk premium $R_{m,t-1} - R_{f,t-1}$. The choice of these instruments is based on existing empirical evidence. In particular, it has been documented that high T-bill rates and downward-sloping term structures (see, for example, Fama and Schwert (1977) for T-bill rates; Campbell (1987) for term structures) tend to be associated with periods of low stock returns. The lag negative risk premium can be thought of as an equity counterpart for downward-sloping term structure. Additionally, these information variables tend to signal regime switching states, in which theories predict the negative risk premium can occur. As such, these economically motivated instruments should have the power to reflect periods associated with low risk premium/negative risk premium.

The variables are constructed to be strictly positive so that I can preserve the inequality restriction (see equations (2.6) and (2.7)). High T-bill rate is defined when it lies above the unconditional mean $z_{1,t}^+ = \max(0, R_{bill,t} - E[R_{bill}])$. This definition not only maintains the strict positivity of T-bill rate, but also takes into account the magnitude of those high rates. To provide an economic interpretation, I normalise this variable as $z_{1,t}^* = \frac{z_{1,t}^+}{E(z_{1,t}^+)}$ where $E(z_{1,t}^+)$ is the sample mean of $z_{1,t}^+$ in the context of equation (2.8) $\hat{\theta}_{\mu z_i}^+ = \frac{1}{T} \sum_{t=1}^T [(R_{m,t+1} - R_{f,t}) z_{1,t}^+]$. Therefore, $\hat{\theta}_{\mu z_1}^+$ is the conditional sample mean of risk premium, weighted by the magnitude of high T-bill rates. In a similar manner, I define $z_{2,t}^+ = \max(0, -(R_{bond,t} - R_{bill,t}))$ as the downward-sloping term structure, and $z_{3,t}^+ = \max(0, -(R_{m,t-1} - R_{f,t-1}))$ as the lag negative risk premium. Their normalised versions are $z_{2,t}^* = \frac{z_{2,t}^+}{E(z_{2,t}^+)}$ and $z_{3,t}^* = \frac{z_{3,t}^+}{E(z_{3,t}^+)}$, respectively.

2.5.2.2 The US Market Evidence

With respect to the US market, I apply the conditional inequality constraints test to the three cases where the negative risk premium occurs: the S&P 500 fundamental-based implied risk premium, and the S&P 500 and the CRSP ex-post realised risk premium, since the other implied risk premium cases imply that the non-negativity restriction cannot be rejected. Empirical results are reported in Table 2.3. For the S&P 500 fundamental-based implied risk premium (Column 1), the probabilities of high T-bill rate, inverted yield curve, and lag negative risk premium occurring are 22%, 47% and 28%, respectively. The high T-bill rate and the downward-sloping term structure do not seem to have power in detecting negative implied risk premium periods. Conditioning on these two instruments, the conditional means of the implied risk premium weighted by the magnitude of these states are positive, ranging from 2.07% (high T-bill rate) to 3.79% (downward-

sloping term structure). In contrast, the magnitude and sign of the risk premium is successfully captured by the negative risk premium of the preceding period. The corresponding conditional implied risk premium is -1.01%, weighted by the absolute value of the instrument. This result highlights the desirable persistent property of the expected return (Cochrane (2011)). Turning to the ex-post realised risk premium cases of the CRSP (column 2) and the S&P 500 (column 3) market portfolios, the power of the instruments in predicting realised negative risk premium is somewhat different. A downward sloping term structure state is highly associated with low risk premium states, resulting in the weighted conditional mean realised risk premium of -4.81% (-5.17%) for the CRSP (S&P500) market portfolio. Conditioning on the remaining instruments, the mean risk premium is positive (2.56% for high T-bill rate and 8.95% for lag negative risk premium).

However, these univariate results that detect the negative risk premium in some conditional states are only suggestive of the violation of positivity restriction. To formally test the null hypothesis of whether or not market risk premium is always positive, it is necessary to take into account the correlation across individual conditional estimates $\hat{\theta}_{\mu, z_t^+}$. The high autocorrelation structure of the instrumental variables results in noisy estimates of the conditional means in small samples.¹¹ I apply the Newey and West (1987) consistent covariance matrix method to adjust for the embed correlation structure, and provide the resulting W test statistic. The W test statistics are 3.47 (associated p-value is 0.08), 0.98 (p-value=0.35), and 1.36 (p-value=0.29) for the S&P 500 implied risk premium, the CRSP realised risk premium, and the S&P 500 realised risk premium, respectively. As a result, while I find evidence against the positivity of the implied market risk premium at the significance level of 10%, I cannot reject the hypothesis in the realised risk premium counterpart.

¹¹ The autocorrelation of the instrument variables is quite high, ranging from 0.86 to 0.95 for lag negative implied risk premium and for T-bill rate, respectively. I do not report this in a table for the sake of brevity.

Table 2.3: Inequality Tests on the Positivity Restriction of the Market Risk Premium

This table reports the statistic to test the hypothesis of whether the ex-ante market risk premium is always positive $E_t[(R_{mt+1} - R_{ft}) \otimes z_t^+] - \theta_{\mu z}^+ = 0$ where $\theta_{\mu z}^+ = E[\mu_t \otimes z_t^+] \geq 0$. The S&P 500 fundamental-based implied risk premium, and the CRSP and S&P 500 ex-post realised risk premium are tested. W is a joint test of multiple inequality restrictions associated with the high T-bill rate, downward-sloping yield curve, and lag negative risk premium periods. $\hat{\theta}_{\mu z_i^+}$ represents the conditional mean of the risk premium, weighted on the magnitudes of these states. Particularly, the high T-bill rate is defined as $z_{1,t}^+ = \max(0, R_{bill,t} - E[R_{bill}])$; the downward-sloping yield curve is defined as $z_{2,t}^+ = \max(0, -(R_{bond,t} - R_{bill,t}))$ where R_{bond} and R_{bill} are the yields of 5-year and 1-year treasury bonds (bills), respectively; lag negative risk premium is $z_{3,t}^+ = \max(0, -(R_{m,t-1} - R_{f,t}))$. I also report the standard errors of these conditional means and the probability that these states occur. All of the estimates are corrected for conditional heteroskedasticity and autocorrelation using Newey and West (1987) method.

| | S&P 500 Implied Risk Premium | CRSP Realised Returns | S&P 500 Realised Returns |
|--|------------------------------------|--------------------------|--------------------------------|
| High T-bill Rate | | | |
| Probability of State | 22.25 | 27.24 | 22.85 |
| Conditional mean $\hat{\theta}_{\mu z_1^+}$ | 2.07 | 2.56 | 2.88 |
| (Standard Error) | (1.25) | (3.36) | (3.12) |
| Downward-Sloping Term Structure | | | |
| Probability of State | 46.87 | 45.07 | 44.85 |
| Conditional mean $\hat{\theta}_{\mu z_2^+}$ | 3.79 | -4.81 | -5.17 |
| (Standard Error) | (1.57) | (4.87) | (4.44) |
| Lag Negative Risk Premium | | | |
| Probability of State | 27.75 | 22.76 | 22.15 |
| Conditional mean $\hat{\theta}_{\mu z_3^+}$ | -1.01 | 8.95 | 8.18 |
| (Standard Error) | (0.54) | (4.48) | (4.23) |
| Multiple Inequality Restriction Statistic W | | | |
| | 3.4680 | 0.9786 | 1.3586 |
| p-value | 0.0788 | 0.3521 | 0.2873 |

2.5.2.3 The International Markets Evidence

Turning to the non-US samples, Table 2.4 and Table 2.5 report the empirical results for the value-weighted implied risk premium and realised risk premium, respectively. Focusing on Table 2.4, the most successful instrument in capturing the negative risk premium is the lag negative risk premium. Conditioning on the lag negative risk premium, the means of risk premium, taking into account the absolute magnitude of the risk premium are all negative across international markets - 2.13% (Japan) to -0.4% (France).

The conditional means of the implied risk premium, weighted by the magnitude of high short-term rate (1-year interbank offer rate), are negative for Japan (-1.44%) and Italy (-1.76%) only, demonstrating that the high T-bill rate has some success in predicting low market risk premium. The downward sloping term structure does not fare well in capturing the negative implied risk premium. Conditioning on the states where the long rate is less than the short rate, the means weighted by the absolute value of the term structure premium are mostly positive for non-US stock markets, with the exception of the Japan sample (-2.13%). Taking into account the correlation structure of the multiple conditional estimates, I can reject the positivity market risk premium hypothesis for Japan ($W=5.22$) at the 5% level, Italy ($W=8.336$) at the 1% level, and Germany ($W=4.94$) at the 5% level.

Turning to Table 2.5, the downward-sloping term structure is the strongest predictor for the negative realised risk premium with its weighted conditional means negative across countries ranging from -1.8% (Canada) to -18.55% (Italy). The high T-bill rate and the lag negative risk premium show some suggestive evidence of the negative risk premium for Japan, Italy, and Germany. The p-values indicate that the magnitudes of these conditional realised means are much larger than those of the implied means. However, I do not find any evidence to support the violation of the positivity restriction on the market, except marginally for Germany at the 10% level of significance ($p\text{-value}=0.0987$). In light of the empirical results, I suggest that the noise in realised returns is large that it significantly reduces the power of the test.

Table 2.4: Inequality Tests on the Positivity Restriction of the Implied Market Risk Premium: International Markets

This table reports the statistic to test the hypothesis of whether the value-weighted analyst-based implied ex-ante market risk premium is always positive $E_t[(R_{mt+1} - R_{ft}) \otimes z_t^+] - \theta_{\mu z}^+ = 0$ where $\theta_{\mu z}^+ = E[\mu_t \otimes z_t^+] \geq 0$ for non-US countries. W is a joint test of multiple inequality restrictions associated with the high T-bill rate, downward-sloping yield curve, and lag negative risk premium. $\hat{\theta}_{\mu z_i^+}$ represents the conditional mean of the risk premium, weighted on the magnitudes of these states. Particularly, the high T-bill rate is defined as $z_{1,t}^+ = \max(0, R_{bill,t} - E[R_{bill}])$; the downward-sloping yield curve is defined as $z_{2,t}^+ = \max(0, -(R_{bond,t} - R_{bill,t}))$ where R_{bond} and R_{bill} are the yields of 10-year (7-year for Italy) treasury bonds and 1-year interbank offer rates from the British Banker Association, respectively. Lag negative risk premium is $z_{3,t}^+ = \max(0, -(R_{m,t-1} - R_{f,t-1}))$. I also report the standard errors of these conditional means and the probability that these states occur. All of the estimates are corrected for conditional heteroskedasticity and autocorrelation using the Newey and West (1987) method.

| | JPN | ITA | UK | FRA | GER |
|--|---------------|---------------|--------|--------|---------------|
| High T-bill Rate | | | | | |
| Probability of State | 5.36 | 26.11 | 25.45 | 27.84 | 23.46 |
| Conditional mean $\hat{\theta}_{\mu z_1^+}$ | -1.44 | -1.76 | 0.54 | 0.36 | 0.05 |
| (Standard Error) | (0.7) | (0.61) | (0.2) | (0.17) | (0.26) |
| Downward-Sloping Term Structure | | | | | |
| Probability of State | 49.85 | 47.62 | 45.81 | 44.8 | 47.41 |
| Conditional mean $\hat{\theta}_{\mu z_2^+}$ | -2.23 | 3.39 | 1.75 | 1.00 | 1.20 |
| (Standard Error) | (1.12) | (1.98) | (0.39) | (0.47) | (0.76) |
| Lag Negative Risk Premium | | | | | |
| Probability of State | 44.64 | 23.89 | 24.55 | 22.16 | 26.54 |
| Conditional mean $\hat{\theta}_{\mu z_3^+}$ | -2.13 | -2.85 | -0.55 | -0.4 | -0.77 |
| (Standard Error) | (0.93) | (1.07) | (0.41) | (0.26) | (0.35) |
| Multiple Inequality Restriction Statistic W | | | | | |
| | 5.2222 | 8.336 | 1.7914 | 2.3371 | 4.9437 |
| p-value | 0.0153 | 0.0068 | 0.2126 | 0.1694 | 0.0368 |

Table 2.5: Inequality Tests on the Positivity Restriction of the Realised Market Risk Premium: International Markets

This table presents statistics that examine whether or not This table reports the statistic to test the hypothesis of the realised market risk premium is always positive $E_t[(R_{mt+1} - R_{ft}) \otimes z_t^+] - \theta_{\mu z}^+ = 0$ where $\theta_{\mu z}^+ = E[\mu_t \otimes z_t^+] \geq 0$ for non-US countries. I use local MSCI indices to proxy for local market portfolio. W is a joint test of multiple inequality restrictions associated with the high T-bill rate, downward-sloping yield curve, and lag negative risk premium. $\hat{\theta}_{\mu z_t^+}$ represents the conditional mean of the risk premium, weighted on the magnitudes of these states. Particularly, the high T-bill rate is defined as $z_{1,t}^+ = \max(0, R_{bill,t} - E[R_{bill}])$; the downward-sloping yield curve is defined as $z_{2,t}^+ = \max(0, -(R_{bond,t} - R_{bill,t}))$ where R_{bond} and R_{bill} are the yields of 10-year (7-yaer for Italy) treasury bonds and 1-year interbank offer rates from the British Banker Association, respectively. Lag negative risk premium is $z_{3,t}^+ = \max(0, -(R_{m,t-1} - R_{f,t-1}))$. I also report the standard errors of these conditional means and the probability that these states occur. All of the estimates are corrected for conditional heteroskedasticity and autocorrelation using the Newey and West (1987) method.

| | JPN | ITA | UK | FRA | GER | CAN |
|--|--------|--------|--------|--------|---------------|--------|
| High T-bill Rate | | | | | | |
| Probability of State | 22.69 | 26.05 | 30.71 | 27.5 | 27.94 | 31.55 |
| Conditional mean $\hat{\theta}_{\mu z_1^+}$ | -10.46 | 3.32 | 2.13 | 0.27 | -4.02 | 0.97 |
| (Standard Error) | 7.18 | 5.47 | 1.52 | 2.21 | 3.39 | 2.07 |
| Downward-Sloping Term Structure | | | | | | |
| Probability of State | 47.2 | 45.38 | 42.96 | 45.23 | 45.09 | 42.93 |
| Conditional mean $\hat{\theta}_{\mu z_2^+}$ | -10.95 | -18.55 | -3.96 | -3.00 | -11.41 | -1.87 |
| (Standard Error) | 8.17 | 21.28 | 2.64 | 4.72 | 6.22 | 1.32 |
| Lag Negative Risk Premium | | | | | | |
| Probability of State | 27.31 | 23.95 | 19.29 | 22.5 | 22.06 | 18.45 |
| Conditional mean $\hat{\theta}_{\mu z_3^+}$ | -0.87 | -1.08 | 4.26 | 0.88 | 3.4 | 8.84 |
| (Standard Error) | 4.6 | 5.74 | 6.33 | 6.04 | 7.18 | 5.66 |
| Multiple Inequality Restriction Statistic W | | | | | | |
| Restriction Statistic W | 2.1601 | 0.7598 | 2.2437 | 0.399 | 3.3635 | 2.0129 |
| p-value | 0.159 | 0.3918 | 0.1945 | 0.5088 | 0.0987 | 0.2225 |

2.6 CONCLUDING REMARKS

In this essay, I employ an appealing market-based measure of the conditional expected return, the Implied Cost of Capital, in testing the positivity restriction on the market risk premium. Numerous asset pricing studies focus on studying the linear restrictions implied by asset pricing

models, but they largely ignore the importance of the positive market risk restriction in testing framework. With economically meaningful instruments that can signal the regime shift in the business cycles (high T-bill rates, downward sloping term structure, and lag negative risk premium) I detect the violation of the positive implied market risk premium, which necessarily implies the rejection of the conditional CAPM. Interestingly, it is in these states that the theory predicts the existence of negative risk premium (Whitelaw (2000)). The “market portfolios” that exhibit this violation include S&P 500, Japan, Italy, and Germany. These results on the implied risk premium are in stark contrast to the realised risk premium counterpart. In six out of seven market portfolios, I do not find any convincing evidence against the positivity condition of the market realised risk premium. In light of the noisy realised return’s numerous limitations in proxying for the expected return, I attribute the success in detecting the violation of the non-negativity restriction to the ICC and therefore conclude that the ICC can serve as a tight proxy for the time varying expected return.

*IS THE EX-ANTE RISK PREMIUM ALWAYS POSITIVE? EVIDENCE FROM A NEW CONDITIONAL EXPECTATIONS MODEL*¹²

3.1 INTRODUCTION

Under rational expectations, if marginal investors are strictly risk-averse and expected utility maximisers, they demand higher returns for bearing additional risk. This leads the ex-ante market risk premium to be greater than or equal to zero. The positivity of market risk premium is necessary for the conditional mean-variance efficiency of the market portfolio; therefore it is a necessary condition of the Conditional Capital Asset Pricing Model (Conditional CAPM). Violation of this restriction implies a rejection of the Conditional CAPM. Merton (1980) further suggests that the positive risk premium should be included as a necessary condition for capital market equilibrium, although more general equilibrium asset pricing theories do not impose this restriction (Lucas (1978)). The belief in the positive risk premium is so strong that the recent empirical literature has gone so far as to impose this constraint directly in the predictive stock returns models (for example, see Campbell and Thompson (2008); and Pettenuzzo, Timmermann and Valkanov (2013)). Despite its debatable nature in theoretical modelling (Whitelaw (2000)), the positive risk premium hypothesis remains an open empirical question. Yet, little effort has been devoted to conducting this testing. In this paper, I continue to examine this hypothesis with a new conditional expectation model.

Specifically, I model the conditional risk premium for the US market (CRSP index) from 1960 to 2011 through a two-step procedure, in which the Principal Component Analysis and the Boosted Regression Trees (BRT) techniques (Friedman (2001); and Breiman, Friedman, Stone and Olshen (1984)) are employed. The technique aims to circumvent two major problems on how

¹² Earlier versions of this essay were presented at the UQ Workshop with Professor Allen Kleidon (Cornerstone Research), Brisbane Australia, November 2014; 6th Financial Markets and Corporate Governance Conference, Perth Australia, April 2015; and Accounting and Finance Association of Australia and New Zealand Conference, Hobart Australia, July 2015.

researchers capture the identity of the investors' information set and their use of the information in forming expectations. I highlight the problems below to motivate my method.

Conditional expectations (mean, variance, and higher-order moments) are unobservable, thus testing the asset pricing theories requires reliable empirical proxies. For the expected return, a common approach is to linearly project ex-post realised excess returns on a small number of available information variables space. However, since the identity of the investors' information set is unknown, omitted variable bias surrounding the linear regression is likely to yield misleading inferences (Ang and Bekaert (2007) for the empirical perspective; and Hansen and Richard (1987) for the theoretical). The inclusion of as many variables as possible to span the true information set does not solve the problem, but points to another statistical issue in which the degrees of freedom quickly exhaust when the number of predictors approaches the number of observations (Ludvigson and Ng (2007)). Adding to this complication, data snooping bias is prevalent and the statistical results are sensitive to the choice of conditioning variables (Foster, et al. (1997); and Harvey (2001)).

Second and more importantly, the question on how investors use the information to generate the expected return is highly debatable. A highly complex and uncertain environment, where investors face numerous information shocks as a result of institutional change, evolution of information technology, or policy shift, leads to a constantly changing return data generating process. As a result, it is quite unlikely that a simple linear predictive model can capture such complexity. A linear model, at best, provides an approximation, and is more likely to be misspecified (Harvey (2001)). Not surprisingly, Welch and Goyal (2008), after a careful assessment of the literature, reach a disappointing conclusion that most existing economic variables do not outperform a naïve historical mean in predicting future stock returns. If returns are not predictable, testing conditional asset pricing models becomes an elusive task (Chen and Zhao (2009)).

To address the above critiques, I propose a two-stage procedure, including principal component analysis and a state-of-the-art supervised learning technique, known as Boosted Regression Trees (BRT), in modelling the conditional risk premium (see Hastie, et al. (2009) for a discussion of the literature). Specifically in the first stage, the principal component analysis (PCA) is performed to capture the common information underlying 160 financial variables from Ludvigson and Ng (2007) and Welch and Goyal (2008). With this step, I aim to span the identity of the investors' information set, without relying on a small subset of arbitrary conditioning variables, by compressing a much richer source of information into a small set of factors.

My new set of predictors, including a small number of principal components and the well-known predicting variables in Welch and Goyal (2008), enters the second stage. In this stage, to address the inherent, highly complex and non-linear relations between the expected excess return

and the information set, I use the BRT technique that is developed in the machine learning. Unlike any parametric models imposing strong modelling assumptions, the non-parametric regression trees estimate the unknown function by breaking up the predictor space through a sequence of piecewise constant models. To understand the method, it is easier to express the conditional expectation estimation through the concept of basis functions. Generally, the conditional expectation of a variable Y , the regression problem seeks to approximate the unknown function from basis function expansions of input X (information variables)

$$f(x) = E(Y|X = x) = \sum_{i=1}^p \beta_i h(x; \theta) \quad (3.1)$$

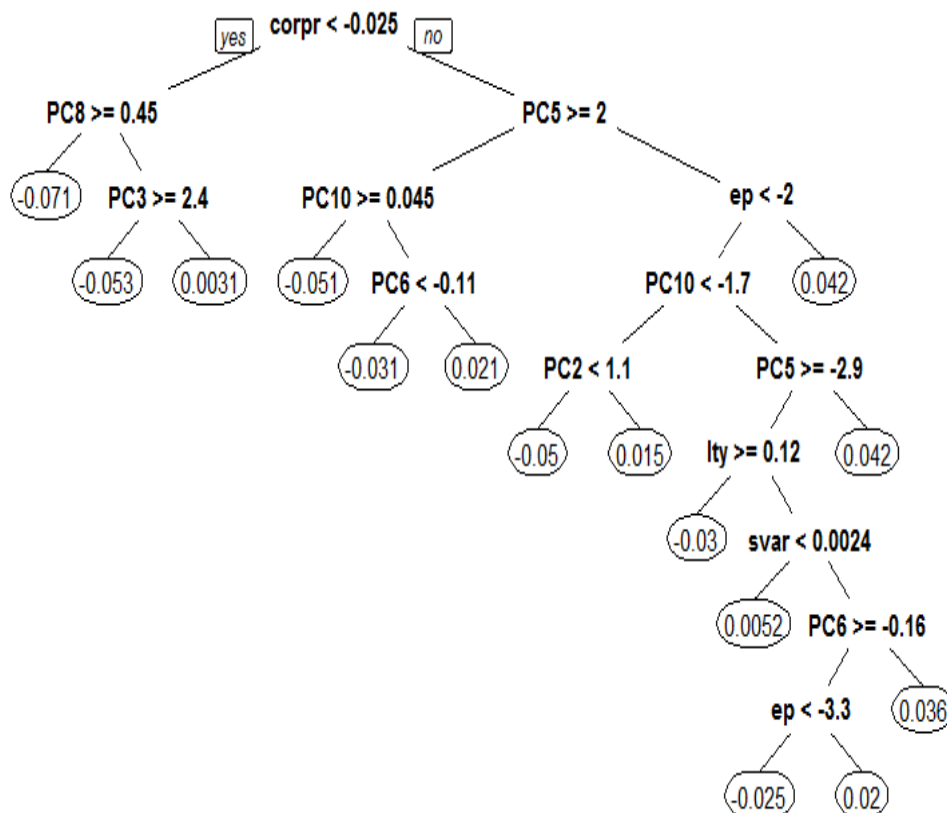
where β is the expansion coefficients and h is the basis functions of the input X , parameterized by θ . Traditionally, the linear regression is convenient because it approximates $f(x)$ by first-order Taylor expansion and $h(x, \theta) = x$. However, linearity is unlikely in modelling the conditional expected return because the data relates in a complicated and non-linear nature (Brandt and Wang (2007)). Instead of assuming a linear functional form, the regression trees approximate the unknown function by recursively partitioning the predictor space X into disjoint sub-regions; and ideally until the information is “tamed”, simple constant models are fit into these regions. The regression tree method forms the basis functions of X through parameter θ which decides the partitioning variables, partitioning values, and the prediction rule within regions.

Figure 3.1 provides a visual explanation of how the regression tree works. The variables and their associated values are the splitting parameters. Numbers within circles are constant values within each partitioned region, which are used to model the expected risk premium. For example, corporate bond return is first chosen to split the sample at -2.5%. In the region where corporate bond return ($corpr$) is less than -2.5%, the 8th principal component ($PC8$) is a subsequent split at the value of 0.45. If the $PC8 \geq 0.45$, the constant -7.1% is a fitted value of the equity premium, whereas a further split is needed for the region in which $PC8 < 0.45$. The procedure continues until some stopping criteria are reached. The model can then form the prediction by summing up constant values in corresponding bins.

Furthermore, I employ a type of ensemble methods, known as boosting, to improve the model’s fit. The idea of the ensemble scheme, similar to those discussed in the forecast combination literature (Rapach, et al. (2010)) is to aggregate the predictions from models which do not perform well individually into one with considerably improved properties. In this context, the simple individual models are the trees. The algorithm builds a sequence of small trees (typically after 1 split), in which subsequent trees seek to minimise the residuals weighted by previous trees’ errors. As a result, the final model is the sum of the forecasts from all of the small trees. The BRT methodology,

perhaps one of the most powerful ideas recently developed in machine learning literature, not only helps to extract unknown functional form, but also has the ability to handle high dimensional data (Hastie, et al. (2009)). Yet, this method has only found a few applications in the finance literature (see Rossi and Timmermann (2010) on risk-return trade-off; and Ng (2014) on business cycles).

Figure 3.1: A Visual Demonstration of the Regression Tree Methodology



To avoid overfitting and maintain generalisation of the model, it is important to build the model in a training set and form predictions on a test set, i.e. out-of-sample.¹³ In this study the risk

¹³ Ideally, a sample is divided into 3 parts: namely, a training period, a validation period, and an evaluation period, to estimate and assess a model. The training period is used to fit the model; the validation period is used to select the best set of model parameters by minimising the prediction error criteria; once the final model with the best set of parameters is built, the test period is used to evaluate the true test errors of the resulting model. If the fitting model is adapted to the training data by minimising an error criterion, it is apparent that the training error rate underestimates the true error rate. As a result, it is necessary to assess the generalisation error of the model by test set. I omit the validation step because the data is insufficient (Hastie, et al. (2009), page 241).

premium is estimated out-of-sample with recursive training periods. Following Ludvigson and Ng (2007), I use the first ten years from 1960:01 to 1969:12 to fit the PCA and the BRTs, and in turn estimate the first predicted risk premium in 1970:01. The forecast for 1970:02 is produced by fitting the two-stage procedure in an expanding training window from 1960:1 to 1970:01. The predicted values in the out-of-sample test set serve as the empirical proxy in the main hypothesis test of whether or not the risk premium is always positive.

The findings can be summarised as follows. First, the out-of-sample R^2 supports the validity of my two-stage methodology in forming the conditional risk premium. The predictive power of future returns consistently beats that of the historical mean returns and their linear counterparts (kitchen-sink OLS and Least Angle Regression of Efron, Hastie, Johnstone and Tibshirani (2004)). I attribute this superior performance to the ability of the technique in capturing the investors' information identity and its underlying complicated data generating process. Second, I find strong evidence to support the existence of negative risk premium in some states of the world, which is related to periods in which corporate bond returns and long term government bond rates of return are low, along with the preceding period experiencing a negative risk premium, and a downward-sloping term structure of interest rate. My result raises a question mark over a recent practice that imposes the positive risk premium constraint in predictive models (Pettenuzzo, et al. (2013)).

I relate and contribute to several strands of literature. Empirically, the positive risk premium restriction remains an under-researched question despite its importance in theoretical modelling. I am among a few who have conducted a direct empirical test with this restriction (for example, Boudoukh, et al. (1997); Boudoukh, et al. (1993); and Walsh (2014)). Unlike previous studies using the realised return as a proxy for the expected return, I explicitly model conditional expected return by adopting the superior conditional expectations model. I am the first to apply the BRT in testing this empirical question. Rossi and Timmermann (2010) also use this technique and focus on the time varying risk-return trade-off, while Ng (2014) aims to detect the peaks and troughs of business cycles. I highlight the importance to address the non-linear nature of the information set and expectations. Finally, my PCA is in a similar spirit, yet simpler than that of Ludvigson and Ng (2007) in addressing the dimensionality problem in the standard linear regression as well as the identity of the investors' information set.¹⁴

¹⁴ Ludvigson and Ng (2007) adopt dynamic factor analysis to extract latent factors underlying the original set of predictors. However, they adopt linear regression to form expectations.

This chapter is structured as follows. Section 3.2 illustrates the research method, followed by a brief data description in Section 3.3. Empirical results are discussed in Section 3.4. The chapter conclusion is offered in Section 3.5.

3.2 RESEARCH METHOD

3.2.1 Principal Component Analysis

Suppose there are a $T \times K$ matrix X and a $T \times 1$ vector Y representing K information variables and a dependent variable rp_{t+1} (ex-post stock excess returns), respectively, over T periods. The traditional linear regression faces a practical degrees-of-freedom problem when the number of potential predictors K approaches the number of observations T ; and when $K > T$, the estimation cannot be performed. The principal component analysis (PCA) circumvents this issue by linearly combining the available predictors into several orthogonalised common factors that collectively help to explain most variations among the original set. In particular, the first principal component Z_1 has the form:

$$Z_1 = Xv_1 \quad (3.2)$$

where v_1 is the $K \times 1$ loading vector

X is the $T \times K$ matrix of the information variables.

Z_1 is the $T \times 1$ principal component.

To capture a large fraction of variation in X , the loadings vector v_1 is the solution of the optimisation problem that ensures the sample variance of Z_1 is the largest:

$$\max_{v_1} v_1^T \Sigma v_1 \text{ subject to } v_1^T v_1 = 1 \quad (3.3)$$

where $\Sigma = X^T X / (NT - 1)$ is the sample covariance matrix of X with elements being centred.

The normalisation constraint $v_1^T v_1 = 1$ limits the value of v_1 so that it is not arbitrarily large, otherwise the variance of the component could be arbitrarily large. Differentiating (3.3) with respect to v_1 , imposing the normalisation constraint, I obtain

$$\Sigma v_1 = \lambda_1 v_1 \quad (3.4)$$

where λ is the Lagrange multiplier associated with the constraint.

It is easily seen that (3.4) is the characteristic equation where v_1 is the eigenvector with the associated eigenvalue λ_1 . Therefore, the largest eigenvalue of Σ should be chosen because the objective function (3.3) is maximised.

The second component is constructed in a similar fashion, with an additional constraint that it is uncorrelated with the first component. The objective function can be expressed as:

$$\max_{v_2} v_2^T \Sigma v_2$$

subject to:

$$v_2^T v_2 = 1$$

$$Cov(Xv_2, Xv_1) = v_2^T \Sigma v_1 = v_2^T \lambda_1 v_1 = 0$$

Differentiating (3.5) w.r.t v_2 and set the first order condition (FOC) to be 0:

$$\Sigma v_2 - \lambda_2 v_2 - \phi \lambda_1 v_1 = 0 \quad (3.6)$$

where λ_2 and ϕ are the Lagrange multipliers associated with two constraints in (3.5).

Pre-multiply v_1 on both side of (3.6), and the FOC becomes:

$$v_1^T \Sigma v_2 - \lambda_2 v_1^T v_2 - \phi \lambda_1 v_1^T v_1 = 0$$

Since $cov(Xv_1, Xv_2) = 0$, the first two terms are 0 and $v_1^T v_1 = 1$; therefore ϕ is necessarily equal to 0. The quantity reduces to

$$\Sigma v_2 = \lambda_2 v_2 \quad (3.7)$$

Again, equation (3.7) is the characteristic equation and eigenvector v_2 is associated with the second largest eigenvalue λ_2 of Σ chosen. Continuing the strategy until k^{th} step, I first obtain k loading vectors, which are the eigenvector of Σ corresponding to the k largest eigenvalues. These k components summarise a large amount of information inherent in the original set of predictors.

The principal components obtained in this stage will enter the second stage as predictors in addition to well-known variables documented in previous literature. In particular, I choose the first 10 components, in conjunction with the 13 predictors documented in Welch and Goyal (2008) to serve as independent variables in my regression tree stage, which is discussed below.

3.2.2 Regression Trees

I provide the intuition how tree-based regression works; a more detailed material discussion can be found in Hastie, et al. (2009). Because the functional forms between the dependent and the independent variables are unknown, the regression tree seeks to estimate the unknown functional form through elementary basis functions. In particular, it divides the predictor space into non-overlapping regions and simply models the response variable as a constant within each region. The splitting variables, splitting points and constant fit in each region form the nature of basis functions. Formally, in the first step the tree algorithm searches through the entire predictor space; and choose an independent variable j and a splitting point s so that the two defined planes are given by:

$$R_1(j, s) = \{X | X_j \leq s\} \text{ and } R_2(j, s) = \{X | X_j > s\} \quad (3.8)$$

which minimises the sum of squared residuals criterion in the resulting regions. Specifically, the optimal parameters set (j, s) is the solution of:

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (rp_{t+1} - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (rp_{t+1} - c_2)^2 \right] \quad (3.9)$$

For a given pair (j, s) , (\hat{c}_1, \hat{c}_2) is the solution of the inner minimisation, which is just the average of the response variable in each region.¹⁵

$$\begin{aligned} \hat{c}_1 &= \frac{1}{\sum_t^T I(x_t \in R_1(j, s))} \sum_t^T rp_{t+1} I(x_t \in R_1(j, s)) \\ \hat{c}_2 &= \frac{1}{\sum_t^T I(x_t \in R_2(j, s))} \sum_t^T rp_{t+1} I(x_t \in R_2(j, s)) \end{aligned} \quad (3.10)$$

where $I(x \in R)$ is an indicator variable equal to 1 if $x \in R$ and is 0 otherwise,

In a similar manner, the subsequent optimal split (a splitting point and a predictor) is decided by minimising the sum of squared residuals in the resulting regions from previous splits, instead of searching on the entire sample space. The process continues until some stopping criteria are satisfied. Suppose there are M regions R_1, R_2, \dots, R_m at the final step, then the fitted value of the regression tree is:

$$\hat{f}(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (3.11)$$

where $I(x \in R_m)$ is an indicator variable equal to 1 if $x \in R_m$ and is 0 otherwise, c_m is the average of the response variable in region R_m .

Regression tree analysis is a flexible method that can capture complex features of the predictors and the response variable. However, one problem associated with the size of the tree is the number of splits that should be performed. A large tree associated with more splitting points will result in fewer observations in each region, and therefore increases the risk of overfitting data; that is, the method might perform well in-sample, but the out-of-sample predicting accuracy is poor. On the other hand, a small tree might not be sufficient to capture important features in the data. Moreover, the above method of building a regression tree is a “greedy”, top down algorithm in which a subsequent split is performed dependent upon the result of the previous split. Therefore, there is no guarantee that this approach can obtain a global optimal solution, because what seems to

¹⁵ That is what it means by modelling the response variable as constant within each region.

be an inferior split initially might become a valuable step in getting a good split in the next iteration. I discuss in the next section the boosting algorithm that can overcome these potential issues.

3.2.3 Boosting

Instead of fitting the data into a single large tree with the greedy approach, boosting builds a number of simple trees in a sequential way and seeks an efficient way to combine them. This is a type of ensemble method. It initiates the idea that the collection of poor individual models can have considerably improved properties.

More generally, boosting is a supervised learning algorithm that can fit additive expansions of elementary basis functions of the input, which are the simple regression trees, to approximate the unknown function. The aim is to build B simple trees in B iterations. There are three steps in one iteration. First, a small tree is built, typically after a small number of splits by fitting the residuals $\varepsilon = rp_{t+1} - \hat{f}(x_t)$, rather than the original response variable in each region.¹⁶ The reason behind fitting residuals is that it aims to grow a new tree in the region where the current fitted model does not perform well, which results in a high sum of squared residuals. That is the essential idea of a supervised learning technique, that it can iteratively “learn” the true functional form in response to differences between the original and generated output, $rp_{t+1} - \hat{f}(x_t)$. Second, the additional information of the new tree is “slowly” updated to the current predicting model. Third, a new set of residuals is calculated with the new model to prepare for the next iteration. The process continues after B iterations and the final model is just the sum of B fitted trees. The algorithm can be summarised as follows:

Step 1: Initialise $\hat{f}(x_t) = 0$ and $\varepsilon_{t+1} = rp_{t+1}$

Step 2: For $b = 1..B$ iterations

Build a simple tree \hat{f}^b with one split, fitting the residuals ε_{t+1}

“Slowly” update \hat{f}^{b-1} with the information from \hat{f}^b , using shrinkage parameter λ

$$\hat{f}(x_t) \leftarrow \hat{f}(x_t) + \lambda \hat{f}^b(x_t)$$

Update the residuals:

$$\varepsilon_{t+1} \leftarrow \varepsilon_{t+1} - \lambda \hat{f}^b(x_t)$$

Step 3: Final predicted model

¹⁶ I use one split, which effectively focuses on the main effect of individual trees, on approximation of the functional form.

$$\hat{f}(x_t) = \sum_{b=1}^B \lambda \hat{f}^b(x_t) \quad (3.12)$$

The superior forecasting performance of using boosting compared to the greedy top down approach can be attributed to three factors. First, building small trees can reduce the risk of overfitting in the regression tree. Second, and more critical to the boosting algorithm, is the use of a small λ shrinkage parameter that controls for the learning rate of new information in the current model. Empirically, a small λ , i.e. a slow updating rate, tends to produce better out-of-sample forecast accuracy because it avoids fitting the data too intensely Friedman (2001). Finally, boosting provides a type of model averaging by summing up all of the small trees, which enhances the stability of the forecasts (Rapach, et al. (2010)).

Three main intertwined parameters are required for the empirical implementation: 1) the number of splits to use in building a tree in each iteration, 2) the number of trees to be built, i.e. the number of boosting iterations, 3) the value of the learning rate λ . I follow common practice in the machine learning literature and adopt: 1) one split, 2) 1000 boosting iterations, and 3) $\lambda = 0.001$ (Hastie, et al. (2009)).

3.2.4 Estimation Period

It is critical to build and test the model in two different data sets because the method tends to understate the error rate in-sample (Hastie, et al. (2009), page 219). As a result, I follow Ludvigson and Ng (2007) and adopt a recursive window, with the first training period being 120 months, to build my model; that is, I construct my components and initial regression trees using data up to 1969:12 and form the prediction in 1970:01, which serves as the first estimate for expected risk premium. The second estimate in 1970:02 is calculated from the PCA and the regression tree fitted during the training period from 1960:01 to 1970:01. As a result, the time series risk premium calculated out-of-sample includes 504 observations. These observations are fed into the multiple inequality constraints framework to test the null hypothesis of the positive risk premium.

3.2.5 Model Assessment

I discuss measures to assess the performance of the predictors and the model in this subsection. Relative Influence and Partial Dependence statistics measure the importance of the predictors in the model, while the out-of-sample R^2 assesses the performance of the method in modelling the conditional risk premium.

3.2.5.1 Relative Influence Measure

Breiman, et al. (1984) propose the influence measure to evaluate the contribution of the predictors in a single regression tree as below:

$$\tau_k^2 = \sum_{j=2}^J \hat{v}_j^2 I(v_j = k) \quad (3.13)$$

where $\hat{v}^2 = \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_{t,j-1}^2 - \hat{\varepsilon}_{t,j}^2)$ is the reduction in the squared forecast error at the j^{th} node,

$I(v_j = k)$ is the indicator equal to 1 if variable k is chosen at j^{th} split.

The idea of this measure is straightforward. For example, at the j^{th} step, a predictor and a splitting point are chosen to partition an identified region from previous steps into two sub-regions. The predictor is one that satisfies the objective function (3.9), i.e. minimises the sum of squared errors in partitioned regions. Therefore, if a variable is chosen multiple times, its importance in the model can be measured as the sum of the reduction in squared errors \hat{v}^2 across the regions where it is chosen as the splitting variable. I then average τ_k^2 across boosting iterations to obtain the measure of influence of variable k , $\widehat{\tau}_k^2$. Finally, I obtain the relative influence measure, RI_k , of the variable k by dividing the influence measure $\widehat{\tau}_k^2$ by the total influence of all variables across the boosting iterations:

$$RI_k = \frac{\widehat{\tau}_k^2}{\sum_{i=1}^K \widehat{\tau}_i^2} \quad (3.14)$$

where $\widehat{\tau}_i^2 = \frac{1}{B} \sum_{b=1}^B \tau_k^2$ is the influence measure of variable i across B iterations.

3.2.5.2 Partial Dependence Plots

While the relative influence shows the importance of predictors in fitting future stock returns, partial dependence plots inform on the marginal effect of individual variables on the conditional expected return. The marginal effect of variable X_k on rp , $f_k(X_k)$, is defined below:

$$f_k(X_k) = E_{X_C} f(X_k, X_C) \quad (3.15)$$

where X_C is the information set that excludes variable X_k .

The idea of this measure is that it averages the effect of all variables in X_C for each value of X_k , and thus tracks the effect of X_k on the predicted value of the response variable. The sample estimate of $f_k(X_k)$ can be calculated as:

$$\hat{f}_k(X_k) = \frac{1}{T} \sum_{t=1}^T f(X_k, x_{C,t}) \quad (3.16)$$

3.2.5.3 Out-of-Sample Prediction

Although the above measures offer useful information about the validity of my two-stage method, they do not directly indicate that my model is in fact superior in modelling conditional risk premium. To assess the performance of my two-stage method in modelling the conditional risk premium, I adopt the out-of-sample R^2 statistic that is commonly used in the literature (see, for example, Campbell and Thompson (2008); and Welch and Goyal (2008)):

$$R_{OOS}^2 = 1 - \frac{\sum_{t=1}^T (rp_{t+1} - \hat{f}(x))^2}{\sum_{t=1}^T (rp_{t+1} - \bar{rp}_{t+1})^2} \quad (3.17)$$

where $\hat{f}(x)$ is my two-stage method's predicted value of expected risk premium, \bar{rp}_t is the historical average of stock excess returns until time $t - 1$.

Following Rapach, et al. (2010), I assess the statistical significance of the R_{OOS}^2 using the Clark and West (2007) adjusted statistic:

$$f_{t+1} = (rp_{t+1} - \bar{rp}_{t+1})^2 - ((rp_{t+1} - \widehat{rp}_{t+1})^2 - (\bar{rp}_{t+1} - \widehat{rp}_{t+1})^2) \quad (3.18)$$

Regressing the out-of-sample value f_t on a constant, and calculating the p-value of the one-sided test associated with the constant, I obtain the p-value for the R_{OOS}^2 .

To gauge the economic significance of the R_{OOS}^2 , I calculate the utility gains for mean-variance investors with relative risk aversion γ , who form monthly portfolios between stocks and bonds using forecasts from the predictive models as opposed to forecasts based on the historical mean (Li, et al. (2013)). Specifically, the allocation to stocks in the next period based on historical forecasts of expected return and variance is illustrated below:

$$w_{0,t} = \left(\frac{1}{\gamma}\right) \left(\frac{\bar{rp}_{t+1}}{\hat{\sigma}_{t+1}^2}\right) \quad (3.19)$$

Similarly, $w_{1,t}$ represents the allocation to stocks in the next period if the investors employ a predictive model of returns:

$$w_{1,t} = \left(\frac{1}{\gamma}\right) \left(\frac{\widehat{rp}_{t+1}}{\hat{\sigma}_{t+1}^2}\right) \quad (3.20)$$

In both investment decisions, I forecast variance of stock returns $\hat{\sigma}_{t+1}^2$ using a 10-year rolling window of monthly returns (Li, et al. (2013)). The investors' average utility, based on the allocation using the historical mean and the predictive model over the out-of-sample period, are illustrated below:

$$U_0 = \mu_0 - \frac{1}{2}\gamma \hat{\sigma}_0^2 \quad (3.21)$$

and

$$U_1 = \mu_1 - \frac{1}{2}\gamma \hat{\sigma}_1^2 \quad (3.22)$$

where $(\mu_0, \hat{\sigma}_0^2)$ and $(\mu_1, \hat{\sigma}_1^2)$ are the means and variances of the portfolios' returns based on the historical mean and the predictive model forecasts, respectively.

The utility gain is the difference between U_1 and U_0 . To report the annualised percentage return, I multiply $(U_1 - U_0)$ by 1200. The risk aversion coefficients $\gamma = 1, 3,$ and 5 are chosen for the main results.

In addition to R_{OOS}^2 , I conduct the Mincer-Zarnowitz regression test of unbiased forecasts by simply regressing rp_{t+1} on the forecasts \hat{rp}_{t+1} , out-of-sample, and jointly testing the null that the intercept is equal to 0 and the coefficient is equal to 1.

3.2.6 Inequality Constraint Method

Defining μ_t as the true ex-ante risk premium, the positive risk premium condition can be expressed as below:

$$\mu_t \geq 0 \quad (3.23)$$

As μ_t is unobservable, I replace the μ_t in the inequality (3.23) with its proxy, $\hat{f}(x_t)$, estimated by the two-stage procedure. If the positive condition is true, it should hold across different environments, which in turn implies multiple inequality restrictions. In particular, suppose $z_{i,t}^+$ is the strictly positive information set¹⁷ that is available to econometricians, then the above inequality implies multiple restrictions:

$$E_t(\hat{f}(x_t) \times z_{1,t}^+) = \mu_t \times z_{1,t}^+ \geq 0$$

$$E_t(\hat{f}(x_t) \times z_{2,t}^+) = \mu_t \times z_{2,t}^+ \geq 0$$

.....

$$E_t(\hat{f}(x_t) \times z_{i,t}^+) = \mu_t \times z_{i,t}^+ \geq 0$$

Furthermore, if I arrange the preceding set of inequalities and apply the law of iterated expectations:

¹⁷ $z_{i,t}^+ \geq 0$

$$E(\hat{f}(x_t) \otimes z_t^+ - \theta_{\mu z^+}) = 0 \quad (3.24)$$

where $\theta_{\mu z^+} = E(\mu_t \otimes z_t) \geq 0$.

Equation (3.24) is now similar to the Generalized Method of Moments (Hansen and Singleton (1982)), but the parameters in this system are subject to a set of positivity constraints.

The above restriction can be written as a system of N-moment conditions:

$$\begin{aligned} E[\hat{f}(x_t) z_{1t}^+] &= \theta_{\mu z_1^+} \\ &\dots\dots\dots \\ E[\hat{f}(x_t) z_{Nt}^+] &= \theta_{\mu z_N^+} \end{aligned}$$

Given the null,

$$H_0: \theta_{\mu z_i}^+ \geq 0 \quad \forall i = 1 \dots N$$

versus

$$H_A: \theta_{\mu z_i}^+ \in R^N$$

To test this hypothesis, Boudoukh, et al. (1993) develop a formal framework that takes into account multiple autocorrelation and cross-correlation of the conditional estimates, subject to inequality constraints. First, I estimate the sample means of the observable variables' product. Specifically,

$$\hat{\theta}_{\mu z_i}^+ = \frac{1}{T} \sum_{t=1}^T [\hat{f}(x_t) z_{i,t}^+] \quad \forall i = 1..N \quad (3.25)$$

I refer to this as the unconstrained model because there is no restriction on the sign of these estimates. They can be negative either because the null is false or due to sampling errors.

Next, I calculate the sample means under the inequality restriction in the null $\hat{\theta}_{\mu z_i}^R$ by minimising deviations from the unrestricted model under the quadratic form:

$$\begin{aligned} \min(\hat{\theta}_{\mu z^+}^+ - \theta_{\mu z^+})' \hat{\Omega}^{-1} (\hat{\theta}_{\mu z^+}^+ - \theta_{\mu z^+}) \\ \text{subject to } \theta_{\mu z^+} \geq 0. \end{aligned} \quad (3.26)$$

where $\hat{\Omega}$ is the consistent variance-covariance matrix of the moments. I employ the Quadratic Spectral kernel with asymptotically optimal lag length to estimate $\hat{\Omega}$ (Andrews (1991)). The test statistic is illustrated below:

$$W \equiv T(\hat{\theta}_{\mu z}^R - \hat{\theta}_{\mu z}^+)' \hat{\Omega}^{-1} (\hat{\theta}_{\mu z}^R - \hat{\theta}_{\mu z}^+) \quad (3.27)$$

The idea of the test statistic W is to measure how close the parameters of the restricted model $\hat{\theta}_{\mu z}^R$ are to those of the unrestricted model $\hat{\theta}_{\mu z}^+$. Under the null, the difference should be small. Wolak (1989) shows that the W statistic is distributed as a weighted sum of χ^2 with different degrees of freedom:

$$\sum_{k=0}^N \Pr[\chi_k^2 \geq c] w(N, N - k, \frac{\hat{\Omega}}{T})$$

where c is the critical value for a given size test, and the weighting function $w(N, N - k, \frac{\hat{\Omega}}{T})$ has exactly $N - K$ positive elements.

3.3 DATA DESCRIPTION

I consider 160 financial variables that have been used in the return predictability literature. In particular, the information set includes 147 financial variables documented in Ludvigson and Ng (2007). The other 13 predictors constructed by Welch and Goyal (2008) include the following:

Log dividend price ratio (dp),

Earning price ratio (ep),

Default yield spread calculated as the difference between BAA and AAA corporate bond yields (def),

Term structure (tms),

Long term government bond yield (lty),

Long term bond returns (ltr),

Stock variance measured as squared daily returns ($svar$),

Three-month T-bill rate (tbl),

Inflation rate ($infl$),

Lag excess returns ($lret$),

Net equity expansion is measured as the sum of 12 months net issue for NYSE stocks divided by the market capitalisation of the stocks ($ntis$) (Boudoukh, Michaely, Richardson and Roberts (2007)),

Book to market ratio (bm) Kelly and Pruitt (2013), and

Corporate bond returns ($corpr$).

The stock market excess return (rp) is calculated as monthly CRSP market portfolio return including dividend minus the 1-month T-bill rate ($Rfree$) (continuously compounded). The sample period spans 1960 to 2011.

3.4 EMPIRICAL RESULTS

3.4.1 Summary Statistics

Panel A of Table 3.1 reports the summary statistics of the predictors from Welch and Goyal (2008) and the conditional risk premium generated by the two-stage method. The statistics for

independent variables are similar to those that have been reported elsewhere (see, for example, Pettenuzzo, et al. (2013)). The monthly mean risk premium $\hat{f}(x)$ is 0.28% with the associated standard deviation of 0.47%. It is interesting to see that there are negative observations in my risk premium prediction, indicating that it is meaningful to detect whether the positivity of the equity risk premium states is violated, or whether the negative fitted values are just an artefact of my sampling error. This is where I deviate from the prior literature, by not restricting my forecasts above 0 to improve the predictive performance (Campbell and Thompson (2008)), but rather testing whether or not these negative states provide evidence against the positivity condition of the risk premium.

For brevity, Panel B presents the statistics of the first ten principal components estimated in the first stage of the procedure. The first component (*PC1*) explains the largest proportion of variation, 62%, in the original set of 160 financial variables. Orthogonalised to the first component, the second component (*PC2*) explains 4.48% of the variation in the predictor space. The contribution explained by subsequent components is steadily declining, with the 10th component (*PC10*) capturing only 0.75% of the variation in my information set. I show later in Table 3.3 that the predictive performance of my model improves significantly once the first component is introduced, and remains stable as the number of components increases.

Table 3.1: Summary Statistics

This table presents the summary statistics of the monthly risk premium on the CRSP US market from 1970:1 to 2011:12. These are estimated by the two-stage method described in Section 3.2.1 and 3.2.2, and those of 13 information variables documented in Welch and Goyal (2008) (Panel A) from 1960:1 to 2011:12. Panel B reports the relative and cumulative importance of the first 10 common components, R^2 and Cum, respectively. R^2 is reflected by the fraction of total variance in the original 160 information variables collected from Ludvigson and Ng (2007) and Welch and Goyal (2008). Panel C reports the relative influence of the 10 principal components and 13 information variables in Welch and Goyal (2008) entering the boosted regression tree stage described in Sections 3.2.5. The cumulative relative influence of the Top 3, 5, and 10 are also reported in Panel C.

| Panel A: Instrumental Variables | | | | | | | |
|--|---------------|----------------|---------------|---------------|---------------|---------------|---------------|
| | Mean | Min | Q1 | Median | Q3 | Max | Std. Dev. |
| $r\hat{p}$ | 0.0028 | -0.0162 | 0.0013 | 0.0040 | 0.0053 | 0.0168 | 0.0047 |
| <i>b.m</i> | 0.5189 | 0.1205 | 0.2949 | 0.4943 | 0.6872 | 1.2065 | 0.2633 |
| <i>lty</i> | 0.0688 | 0.0248 | 0.0481 | 0.0641 | 0.0821 | 0.1482 | 0.0250 |
| <i>ntis</i> | 0.0131 | -0.0576 | 0.0055 | 0.0155 | 0.0258 | 0.0512 | 0.0189 |
| <i>Rfre</i> | 0.0043 | 0.0000 | 0.0027 | 0.0041 | 0.0054 | 0.0136 | 0.0025 |
| <i>infl</i> | 0.0033 | -0.0192 | 0.0012 | 0.0029 | 0.0051 | 0.0179 | 0.0035 |
| <i>ltr</i> | 0.0065 | -0.1124 | -0.0095 | 0.0045 | 0.0229 | 0.1523 | 0.0292 |
| <i>corp</i> | 0.0065 | -0.0949 | -0.0070 | 0.0051 | 0.0190 | 0.1560 | 0.0261 |
| <i>svar</i> | 0.0022 | 0.0001 | 0.0006 | 0.0011 | 0.0022 | 0.0709 | 0.0045 |
| <i>ep</i> | -2.8275 | -4.8365 | -3.0330 | -2.8581 | -2.5646 | -1.8987 | 0.4440 |
| <i>dp</i> | -3.5587 | -4.5240 | -3.8794 | -3.4974 | -3.3076 | -2.7533 | 0.4004 |
| <i>def</i> | 0.0102 | 0.0032 | 0.0072 | 0.0089 | 0.0121 | 0.0338 | 0.0046 |
| <i>tms</i> | 0.0175 | -0.0365 | 0.0064 | 0.0168 | 0.0301 | 0.0455 | 0.0150 |
| <i>lret</i> | 0.0033 | -0.2605 | -0.0223 | 0.0080 | 0.0341 | 0.1470 | 0.0455 |

| Panel B: Principal Components (%) | | | | | | | | | | |
|--|------|------|------|------|------|------|------|------|------|------|
| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 |
| R^2 | 62.8 | 4.5 | 3.5 | 3.2 | 1.7 | 1.3 | 1.2 | 1.1 | 1.0 | 0.8 |
| Cum. R^2 | 62.8 | 67.3 | 70.8 | 74.0 | 75.6 | 77.0 | 78.2 | 79.3 | 80.3 | 81.0 |

| Panel C: Relative Influence (%) | | | | | | | | | | | | |
|--|-------|-------|--------|------|------|------|------|------|------|------|------|------|
| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| Variable | corpr | PC5 | PC10 | svar | b.m | dp | PC9 | ep | ltr | lret | def | |
| RI | 32.02 | 26.36 | 15.78 | 5.48 | 4.13 | 3.42 | 2.10 | 1.80 | 1.60 | 1.32 | 1.19 | |
| Rank | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| Variable | infl | ntis | PC1 | tms | PC4 | tbl | PC8 | PC3 | PC6 | PC7 | PC2 | lty |
| RI | 1.18 | 0.89 | 0.77 | 0.56 | 0.35 | 0.31 | 0.22 | 0.19 | 0.15 | 0.12 | 0.06 | 0.00 |
| | Top 3 | Top 5 | Top 10 | | | | | | | | | |
| Cum.RI | 74.16 | 83.77 | 94.01 | | | | | | | | | |

3.4.2 Relative Influence Measure

Table 3.1 Panel C presents the relatively influence statistics which measure the contribution of individual predictors in estimating the equity premium. Corporate bond returns appear to be the most important variable which obtains the relative influence weighting of 32%. Interestingly, although *PC5* and *PC10* only capture a small variation in the original information set (1.7% and 0.8%, Table 3.1 Panel B), they are the second and third most important factors in modelling the risk premium, which attract the relative influence scores of 26.3% and 15.78%, respectively.¹⁸ The well-known dividend price ratio is at sixth rank with the relative influence weighting of just 3.42%. The top three predictors dominate the sample and aggregate to a combined weighting of close to three-quarters. This evidence suggests the importance of some hidden factors that help to span investors' information set in forming expectations. Furthermore, as I show later in Panel A of Table 3.3, the predictive performance of my model improves significantly once the first component is introduced and remains stable as the number of components increases.

3.4.3 Partial Dependence Plots

Turning to Figure 3.2, which presents the partial dependence plots of 5 predictors, including corporate bond returns, 5th principal component (*PC5*), stock variance, term structure, and dividend price ratio in modelling the equity risk premium, I gain some insight into the existence of non-linearity in the relationship between the equity premium and information variables.¹⁹ With respect to the corporate bond returns, the relation is highly non-linear. In particular, when returns fall in the range of -10% to -5%, the risk premium does not change and remains around -1%. In contrast, when the bond returns increase from -4% to -2%, a sharp increase in the risk premium is observed. The relation again becomes flat when the predictor passes into its positive domain. Turning to the 5th principal component (*PC5*) plot, there is a flat structure in the negative and extreme positive levels. Yet, in the intermediate range, a higher value of the component is associated with a lower risk premium reaching its minimum at -0.6%. The non-linear pattern persists in the term structure of the interest rate. Although there is a general positive relation

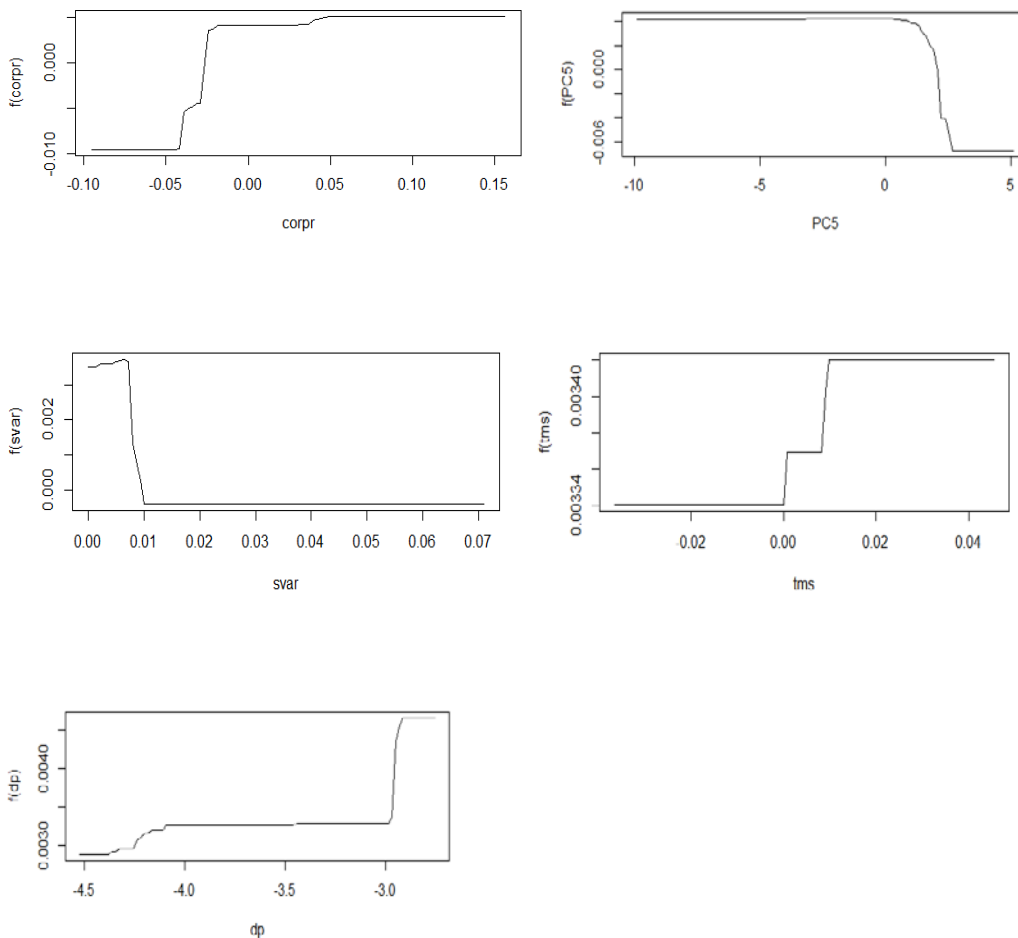
¹⁸ Note that the relative influence measure provides the ranking of 23 variables in the predictor space (13 variables from Welch and Goyal (2008) and 10 components from step 1. The ranking of the three most important variables remains robust to the number of components entering the BRT stage.

¹⁹ For brevity, I present the plots for 5 predictors. Other predictors also exhibit a complicated relationship with the expected-risk premium.

between the term structure and the equity risk premium, the slope is particularly steep across the mid-range observations. Finally, the dividend price ratio shows the strongest positive relation across the high value range, whereas it remains constant mostly across lower values. Clearly, the partial dependence plots raise a concern over the assumption of using a linear functional form in modelling the conditional expected return in the prior literature. As the non-linearities are evident across all predictors, accounting for these effects might help capture the true conditional risk premium. I turn next to the out-of-sample evaluation to provide formal evidence in relation to this claim.

Figure 3.2: Marginal Effect of the Information Variables on the Expected Return

These figures present partial dependence plots for the mean risk premium, based on the five predictors during the full sample period 1960:1-2011:12: namely, corporate bond returns (*corpr*), 5th principal component (*PC5*), stock variance (*svar*), and term structure (*tms*). The horizontal axis presents the sample values of the predictors. The vertical axis illustrates the conditional risk premium as a function of the individual predictor.



3.4.4 Out-of-Sample R_{OOS}^2

3.4.4.1 Comparison with Linear Models

Panel A of Table 3.2 directly compares the out-of-sample R_{OOS}^2 and the associated p-values of my method, as opposed to those of pooled OLS (kitchen sink) and Least Angle Regression (LAR).²⁰ LAR can be thought of as a linear counterpart to my regression trees. In this setting, I build a LAR model and form prediction of stock returns with the top 3 predictors.²¹ I also report the economic significance of these R_{OOS}^2 through the gain in utility for mean-variance investors, with three different degrees of relative risk aversion γ ($\gamma = 1, 3, \text{ and } 5$).

Not surprisingly, the kitchen sink OLS performs the worst in predicting future stock excess returns, with R_{OOS}^2 being -10.41%. This is high in absolute value, so the historical average outperforms prediction generated by the OLS model. The low R_{OOS}^2 of the OLS might be due to the linear assumption and degrees-of-freedom issues, which tend to cause the model highly unstable. LAR addresses the latter problem and in fact, shows a large improvement compared to the OLS, with the associated R_{OOS}^2 equal to 0.66% and statistically significant at the 10% significance level. However, the LAR's forecasts yield disappointing utility losses of 0.13% and 0.52% annually, for investors with risk aversion coefficients equal to 3 and 5, respectively. It turns out that the boosted tree estimation provides superior predictive performance. The R_{OOS}^2 , 1.58%, with the associated p-value of 1%, is twice as large as that of LAR. Encouragingly, the utility gains based on these superior forecasts are consistently positive and economically sizable, ranging from around 1.34% ($\gamma = 5$) to nearly 6% ($\gamma = 1$). These utility gains are all larger than those of the kitchen sink OLS and LAR, across different degrees of risk aversion.

Panel B of Table 3.2 further reports the Mincer-Zarnowitz test discussed in Section 3.2.5.3. The joint test that the intercept is equal to 0 and the coefficient is equal to 1 offers a similar conclusion. The p-value near 0% indicates that the traditional OLS is highly misspecified, whereas the BRT model and LAR are not (with associated p-values of 91% and 87%, respectively).

²⁰ LAR, recently proposed in the machine learning literature, is an efficient model selection mechanism that allows researchers to choose the best linear subset among potential conditioning variables in predicting, in this case, stock excess returns. I provide a brief description of LAR in Appendix A.2, while a more detailed discussion can be found in Efron, et al. (2004).

²¹ For brevity, I choose the 3 best predictors in LAR. In an untabulated table, I find that the predictive performance of BRT outperforms that of LAR, regardless of the number of predictors chosen in LAR.

Table 3.2: Out-of-Sample Prediction Comparison

This table compares the out-of-sample predictive performance of the two-stage method described in Section 3.2.5.3, as opposed to those of kitchen sink OLS and Least Angle Regression (Efron, et al. (2004)), from 1970:1 to 2011:12. The benchmark boosted regression tree uses 1000 boosting iterations, 10 principal components as well as 13 economic variables documented in Welch and Goyal (2008). The Least Angle Regression forms prediction based on the best 3 variables. Panel A reports the out-of-sample R_{OOS}^2 and their Clark and West (2007) associated p-value. The economic significance of the R_{OOS}^2 , measured by the average utility gain for mean-variance investors with three different relative risk aversion coefficients $\gamma = 1, 3, \text{ and } 5$, are presented. Panel B illustrates the Mincer-Zarnowitz misspecification test discussed in Section 3.2.5.3.

| Panel A: Out-of-Sample R_{OOS}^2 | | | | | |
|--|------------------|--------------------|----------------|--------------|--------------|
| | R_{OOS}^2 | <i>p-value</i> | $\gamma = 1$ | $\gamma = 3$ | $\gamma = 5$ |
| Boosted Regression Tree | 1.58% | 0.01 | 5.99% | 2.15% | 1.34% |
| Kitchen Sink OLS | 10.41% | 0.04 | 4.84% | 0.87% | -1.18% |
| Least Angle Regression | 0.66% | 0.08 | 2.80% | -0.13% | -0.52% |
| Panel B: Mincer-Zarnowitz Test | | | | | |
| | <i>Intercept</i> | <i>Coefficient</i> | <i>p-value</i> | | |
| Boosted Regression Tree | 0.0003 | 1.1425 | 0.91 | | |
| Kitchen Sink OLS | 0.0029 | 0.2361 | 0.00 | | |
| Least Angle Regression | 0.0008 | 0.7118 | 0.87 | | |

3.4.4.2 Further Diagnostic Specifications

In Table 3.3, I provide a more detailed picture of how my method fares in different specifications. I first show the importance of identifying the true investors' information set by presenting the model's predictive performance associated with varying numbers of principal components entering stage two. "0 component" is when I use the raw input (160 variables) directly in step two, without performing step one (PCA). I then increase the number of components entering the second step to 1, 2, 5, 10, 50, 100, and 120.

Panel A of Table 3.3 shows that without step 1, PCA (0 component), the BRT performs quite poorly in predicting future excess returns. The R_{OOS}^2 stays at 0.45% and is not statistically significant. However, once the first principal component is placed in the second stage with the 13 predictors of Welch and Goyal (2008), predictive performance improves significantly to 1.27%, significant at the 5% level. It also generates sizable average utility gain of 1.26%, twice as large as that of no component. As the number of components increases, my method continues to beat the historical average, and reaches the best performance at 10 components. It is consistent with the evidence in Panel C of Table 3.1, in which PC5 and PC10 are among the top three most important predictors in forming the expected risk premium. The combining evidence highlights the

importance of identifying the true investors' information set, which is essentially what my first stage tries to address.

Turning attention to Panel B of Table 3.3, I present the model performance when I vary the number of boosting iterations to 2000, 3000, 4000, and 5000 boosting trees. The overall observation is that the performance is not sensitive to this choice. All of the R_{OOS}^2 are positive and statistically significant at the 1% level. Although there is a potential threat of overfitting once I increase the number of trees to 5000 as the R_{OOS}^2 declines to 0.34%, the forecasts generated by this specification yield the largest utility gain of 3.28% for investors with a relative risk aversion γ of 3.

Next, I examine the robustness of the predictive performance as I vary the lengths of estimation and evaluation periods. Specifically, Panel C of Table 3.3 presents the out-of-sample performance when I change initial training periods from 10 years to 20 years, and then to 30 years, and thus my evaluation periods are effectively 1970:1-2011:12, 1980:1-2011:12, and 1990:1-2011:12. Additionally, following Rapach, et al. (2010), other evaluation samples include Post Oil Shock (1976:1-2011:12), Technology Bubble (1:2000-2011:12), and the recent Great Recession (1970:1-2008:1). Overall, across different scenarios, I consistently find the out-of-sample R_{OOS}^2 positive, ranging from 1.29% to 1.58%. Most of them are statistically significant at least at the 10% level of significance, with the exception of the period between 2000:1 and 2011:12. This seems to be the strictest period, because within the course of only 11 years there are two major crises. Furthermore, the predictive performance seems to decrease after the Great Recession period. For example, the R_{OOS}^2 declines from 1.49% to 1.45% when comparing period (2000:1-2007:12) as opposed to (2000:1-2011:12). On the other hand, the utility gains are economically large, ranging from 1.03% in the period spanning from the Post Oil Shock (1976:1) to the Pre-Recession (12:2007), to 2.78% in the period covering the Technology Bubble (1-2001) up until the Post-Recession (2011:12).

Table 3.3: Out-of-Sample Prediction with Different Specifications

This table reports the predictive performance of the two-stage method, discussed in Section 3.2.5 under various specifications. Panel A presents the R_{OOS}^2 , the Clark and West (2007) p-value, and the average utility gain for mean-variance investors with a relative risk aversion coefficient equal to 3, when 0, 1, 3, 5, 10, 20, 50, 100, and 120 components are included with 13 economic variables from Welch and Goyal (2008) in the boosted regression tree stage. Panel B reports similar statistics when the number of boosting iterations ranges from 1000 to 5000. Panel C shows the predictive performance of the method in different estimation and evaluation periods. The periods 1970:1-2011:12, 1980:1-2011:12 and 1990:1-2011:12 reflect that the initial training period are 120, 240, and 360, respectively. Pre-Recession 1970:1-2007:12, Oil Shock 1976:1-2011:12, Oil Shock and Pre-Recession 1976:1-2007:12, Technology Bubble 2000:1-2011:12, and Technology Bubble Pre-Recession 2000:1-2007:12 periods are also under examination.

| Panel A: Number of Principal Components | | | |
|---|-------------|----------------|--------------|
| | R_{OOS}^2 | <i>p-value</i> | $\gamma = 3$ |
| 0 Component | 0.45% | 0.14 | 0.63% |
| 1 Component | 1.27% | 0.03 | 1.26% |
| 3 Components | 1.25% | 0.03 | 0.98% |
| 5 Components | 1.25% | 0.02 | 1.15% |
| 10 Components | 1.58% | 0.01 | 2.15% |
| 20 Components | 1.15% | 0.02 | 1.88% |
| 50 Components | 0.66% | 0.08 | 1.54% |
| 100 Components | 1.03% | 0.03 | 2.13% |
| 120 Components | 0.95% | 0.04 | 2.10% |
| Panel B: Number of Boosting Iterations | | | |
| | R_{OOS}^2 | <i>p-value</i> | $\gamma = 3$ |
| 1000 Iterations | 1.58% | 0.01 | 2.15% |
| 2000 Iterations | 1.65% | 0.01 | 2.29% |
| 3000 Iterations | 1.44% | 0.01 | 2.89% |
| 4000 Iterations | 0.89% | 0.01 | 3.01% |
| 5000 Iterations | 0.34% | 0.00 | 3.28% |
| Panel C: Estimation and Evaluation Periods | | | |
| | R_{OOS}^2 | <i>p-value</i> | $\gamma = 3$ |
| 120 Periods 1970:1-2011:12 | 1.58% | 0.01 | 2.15% |
| 240 Periods 1980:1-2011:12 | 1.47% | 0.02 | 1.36% |
| 360 Periods 1990:1-2011:12 | 1.29% | 0.08 | 1.85% |
| Pre-Recession 1970:1-2007:12 | 1.58% | 0.01 | 1.98% |
| Oil Shock 1976:1-2011:12 | 1.38% | 0.02 | 1.39% |
| Oil Shock Pre-Recession 1976:1-2007:12 | 1.32% | 0.01 | 1.03% |
| Technology Bubble 2000:1-2011:12 | 1.45% | 0.13 | 2.78% |
| Technology Bubble Pre-Recession 2000:1-2007:12 | 1.49% | 0.09 | 2.39% |

3.4.5 Inequality Test

The evidence above supports the validity of the two-stage method. I now apply the forecasts generated by the model to the multiple inequalities framework and test my main hypothesis: the positivity of the risk premium. In addition to testing the full sample (1970:1-2011:12), I consider a sub-period 1970:1-2007:12 which excludes the post Great Recession period based on previous evidence that my predictive performance deteriorates after the 2008 recession (see Table 3.3 Panel C). For the purpose of comparison, I also conduct the same analysis yet using the realised risk premium as a proxy for the expected risk premium.

3.4.5.1 Instrumental Variables

I construct the first set (Set A) of instrumental variables that are used in existing literature, including corporate bond returns (*corpr*), long term government bond returns (*ltr*), and lagged risk premium in the inequality testing framework. Additionally, I consider a second set of instrumental variables (Set B) in which I replace the long term government bond returns with the slope of the term structure of interest rates, based on evidence that the term structure contains useful information about regime shifting states in business cycles (Campbell (1987); and Boudoukh, et al. (1997)). These states, predicted by theory, are likely to be associated with the negative risk premium periods (Whitelaw (2000)). I choose the corporate bond returns because the variable seems to be the best predictor in my predicting model (see relative influence measure in Panel C of Table 3.1). The choice of long term bond returns is the government counterpart to corporate bond returns. Finally, I choose the lagged risk premium following Ostdiek (1998).

The variables should be constructed to be non-negative, so that the inequality restrictions in the testing framework are preserved. In this regard, I employ two transformation methods that correspond a priori to periods of low implied premium (i.e. low corporate and government bond returns, lagged negative risk premium, and downward-sloping term structure). In transformation 1, the low corporate bond return state is defined when it lies below the unconditional mean, and takes the value of $z_{1,t}^+ = 1$, and 0 otherwise. Similarly, $z_{2,t}^+ = 1$ indicates the state when long term government bond returns are below its long run mean, and 0 otherwise. $z_{3,t}^+ = 1$ is when the previous period has a negative risk premium. Finally, the term structure in set B $z_{3B,t}^+ = 1$ is when it is downward sloping.

For transformation 2, I aim to improve the power of the test by taking into account the economic magnitude of states corresponding to low risk premium periods. More specifically, I define the low corporate bond return state as $z_{1,t}^+ = |\min(0, corpr_t - E[corpr_t])|$, where $E[corpr_t]$ is the long run mean of corporate bond returns. In a similar manner, the low long term

government rate of returns state is $z_{1,t}^+ = |\min(0, ltr_t - E[ltr_t])|$, where $E[ltr_t]$ is the long run mean of corporate bond returns. I define $z_{3,t}^+ = \max(0, -r(\widehat{p}_t))$ as the lagged negative risk premium, and $z_{3B,t}^+ = \max(0, -tms_t)$ as downward-sloping term structure. To provide an economic interpretation, I normalise these variable as $z_{i,t}^* = \frac{z_{i,t}^+}{E(z_{i,t}^+)}$ where $i = 1, 2, 3, 3B$ and $E(z_{i,t}^+)$ is the sample mean of $z_{i,t}^+$ in the context $\widehat{\theta}_{\mu z_i}^+ = \frac{1}{T} \sum_{t=1}^T [\widehat{f}(x) z_{i,t}^+]$.

3.4.5.2 Main Inequality Results

Table 3.4 reports the empirical results for the positivity risk premium test. First, I present the results that employ the risk premium proxy generated by the two-stage method, in both the full period (1970:1-2011:12) and the pre-recession period (1970:1-2007:12), using the instrumental variables in set A (Panel A). Across all sample periods and transformation methods, the lagged negative risk premium has the strongest predictive power in detecting the negative risk premium periods. For example, with respect to the transformation 2 conditioning on this instrument, the conditional means of the risk premium for the full and pre-recession samples, weighted by the magnitude of this state, is -0.21% and -0.16%, respectively. Focusing on the dummy transformation 1, conditional on the lagged negative risk premium, the means of the risk premium are slightly lower at -0.13% and -0.11%, respectively. For the full period using transformation 2, the low corporate and long term bond returns are associated with the negative equity premium states. In these, the corresponding weighted means of the risk premium conditioning on the low returns states are -0.13% and -0.08%, respectively. In contrast, the negative risk premium is not successfully captured by these two instrumental variables if I transform them using dummy method 1. One notable observation is that these univariate results seem to be weaker in the pre-recession period (1970:1-2007:12), relative to the full period (1970:1-2011:12).

Turning to Panel B of Table 3.4, when the ex-post realised return proxy for the expected return, the lag negative realised risk premium continues to be a predictor in detecting negative realised risk premium. Conditioning on this variable, the sample means of the realised risk premium range from -0.05% to -0.42%. Interestingly, the corporate and government bond returns are now more strongly associated with low negative realised risk premium states. In particular, regardless of the samples and transformation methods conditioning on these two events, the sample averages of the risk premium vary from -0.07% (1970:1-2011:12; low long term government bond returns; and transformation 1) to 1.3% (1970:1-2011:12; low corporate bond returns; and transformation 2).

Table 3.4: Inequality Tests on the Positivity Restriction of the Market Risk Premium

This table reports the statistics of the hypothesis test: whether or not the ex-ante equity risk premium is always positive for two sample periods, 1970:1-2011:12 and 1970:1-2007:12. W is a joint test of multiple inequality restrictions. The inequality testing framework is discussed in Section 3.2.6. To apply information variables to the multiple inequalities framework, I employ dummy transformation 1, and informative transformation 2. Detailed discussion of the transformation methods is presented in Section 3.4.5.1. Multivariate results are demonstrated for both transformations. Panel A reports the results where the predicted values from the two-stage method serve as proxy for expected risk premium. Panel B shows the analysis where realised measure proxies for expected risk premium. Panel A and Panel B use set A of instrumental variables, including low corporate bond returns, long term government bond returns and lagged negative risk premium. Panel C illustrates the results where set B instrumental variables are used. These variables include downward-sloping term structure, lagged negative risk premium, and low corporate bond returns. For brevity, only critical value, W statistic, and p-value are shown in Panel C. The predicted risk premium indicates that the risk premium generated by the two-stage method proxies for expected risk premium. Realised risk premium uses realised measure to proxy for expected risk premium. I also report the standard errors of these conditional means and the probability that these states occur. All of the estimates are corrected for conditional cross-correlation and autocorrelation, using Quadratic Spectral kernel with asymptotically optimal lag length (Andrews (1991)).

| Panel A: Predicted Risk Premium | | | | |
|--|-----------------------|------------------|-----------------------|------------------|
| | 1970:1-2011:12 | | 1970:1-2007:12 | |
| | Transformation 1 | Transformation 2 | Transformation 1 | Transformation 2 |
| Lag risk premium | | | | |
| Probability of State | 18.08% | 21.46% | 18.82% | 22.55% |
| Conditional Mean | -0.13% | -0.21% | -0.11% | -0.16% |
| Standard Error | -0.0007 | 0.0012 | 0.0007 | 0.0011 |
| Long term Government Returns | | | | |
| Probability of State | 48.48% | 48.32% | 48.41% | 48.26% |
| Conditional Mean | 0.09% | -0.08% | 0.09% | -0.06% |
| (Standard Error) | 0.0003 | 0.0006 | 0.0003 | 0.0005 |
| Corporate Bond Returns | | | | |
| Probability of State | 31.92% | 28.54% | 31.18% | 27.45% |
| Conditional Mean | 0.07% | -0.13% | 0.08% | -0.08% |
| (Standard Error) | 0.0003 | 0.0007 | 0.0003 | 0.0006 |
| Critical Value 5% Level | 4.02 | 4.19 | 4.06 | 4.25 |
| W Statistic | 3.79 | 4.95 | 2.68 | 3.16 |
| p-value | 0.06 | 0.03 | 0.11 | 0.09 |

Table 3.4 (Continued)

| Panel B: Realised Risk Premium | | | | |
|--|------------------|------------------|------------------|------------------|
| | 1970:1-2011:12 | | 1970:1-2007:12 | |
| | Transformation 1 | Transformation 2 | Transformation 1 | Transformation 2 |
| Lag risk premium | | | | |
| Probability of State | 17.53% | 20.78% | 16.68% | 19.09% |
| Conditional Mean | -0.23% | -0.42% | -0.21% | -0.05% |
| Standard Error | 0.0040 | 0.0068 | 0.0040 | 0.0056 |
| Long term Government Returns | | | | |
| Probability of State | 48.20% | 47.49% | 48.41% | 48.33% |
| Conditional mean | -0.07% | -0.40% | -0.13% | -0.27% |
| (Standard Error) | 0.0029 | 0.0047 | 0.0030 | 0.0044 |
| Corporate Bond Returns | | | | |
| Probability of State | 32.47% | 29.22% | 33.32% | 30.91% |
| Conditional mean | -0.2% | -1.3% | -0.2% | -0.7% |
| (Standard Error) | 0.0031 | 0.0071 | 0.0030 | 0.0053 |
| Critical Value 5% Level | 4.01 | 4.23 | 3.95 | 4.08 |
| W Statistic | 0.37 | 3.18 | 2.68 | 1.54 |
| p-value | 0.43 | 0.09 | 0.42 | 0.20 |
| Panel C: Set B Instrumental Variables | | | | |
| | 1970:1-2011:12 | | 1970:1-2007:12 | |
| | Transformation 1 | Transformation 2 | Transformation 1 | Transformation 2 |
| Predicted Risk Premium | | | | |
| Critical Value | 4.53 | 4.99 | 4.50 | 4.99 |
| W Statistic | 4.40 | 4.95 | 3.52 | 3.16 |
| <i>p – value</i> | 0.05 | 0.05 | 0.08 | 0.12 |
| Realised Risk Premium | | | | |
| Critical Value | 4.49 | 4.75 | 4.43 | 4.84 |
| W Statistic | 2.71 | 3.29 | 2.72 | 1.65 |
| p-value | 0.13 | 0.11 | 0.12 | 0.26 |

These univariate results that detect the negative risk premium in some conditional states are only suggestive of the violation of the positivity restriction. To formally test the null hypothesis, whether or not the equity risk premium is always positive, I need to take into account the correlation across individual conditional estimates $\hat{\theta}_{\mu, z_t^+}$. The autocorrelation structure of the instrumental variables results in noisy estimates of the conditional means in small samples, because the instrumental variables might be picking up similar sampling errors. Applying the inequality

procedure described in Section 3.2.6, I now turn to the formal evidence reflected by the W statistic. Looking back to Panel A of Table 3.4, I find strong evidence against the null of positive risk premium in the full sample period (1970:1-2011:12). After adjusting for the autocorrelation and cross-correlation structures of the information variables, the W statistic is equal to 3.79 (4.95) in the dummy transformation 1 (informative transformation 2). These statistics imply that I can reject the null of the positive risk premium at the 5% and 10% levels, for the informative transformation 2 and the dummy transformation 1, respectively. There is a slight sensitivity when I test in the pre-recession period. I cannot reject the null of negative risk premium if I use transformation 1. However, with the more powerful transformation 2, I continue to reject the null of positive risk premium at the 10% level.

Focusing on Panel B of Table 3.4, the multivariate tests provide a stark contrast to the univariate test in the case of realised risk premium. Although I find strong suggestive univariate evidence that the information variables are associated with realised risk premium, after taking into account the autocorrelation and cross-correlation structures of these instruments, the multivariate test fails to reject the null of positive risk premium in three out of four scenarios. I only find evidence against the null for the full sample period and informative transformation 2. This stresses a disadvantage of the ex-post realised returns being a noisy estimate of the expected return.

Applying set B instruments, results reported in Panel C of Table 3.4 offer a similar picture.²² Focusing on the risk premium estimated by the two-stage method, in the full sample period the positivity of risk premium is violated at the 10% level in most cases. There is a minor sensitivity, and surprising evidence for the pre-recession period. Dummy transformation 1 in fact rejects the null ($W=3.52$), whereas the more powerful informative transformation 2 fails to do so ($W=3.16$). Turning to the realised risk premium counterpart, no evidence against the negative realised risk premium is detected. The W statistic values range from 1.65 to 3.29 and are not statistically significant at the 10% level.

3.5 CONCLUDING REMARKS

A large number of studies focus on testing the linear restrictions imposed by theoretical models, yet mostly ignore the positivity condition of the risk premium. In this chapter, I empirically

²² I only report critical value, W statistic, and p-value on Panel C for brevity. The univariate results are qualitatively similar.

test this important restriction. The study's contribution lies in the manner of modelling the conditional risk premium, in which I address two issues surrounding the traditional linear instrumental approach. Specifically, I adopt the principal component analysis and BRT techniques to alleviate criticisms pertaining to the identity of the investors' information set, and the non-linear structure of the return data generating process.

The empirical evidence suggests that my two-stage methodology is superior to other linear methods (OLS and LAR) in modelling the conditional risk premium. More importantly, I document that, in the US market, the positive risk premium condition is violated in some states of the economy, such as low corporate and government bond returns, downward-sloping term structure, and lag negative risk premium periods. The result implies a rejection of the conditional CAPM. Finally, I raise a concern over the current practice of imposing directly the positive risk premium constraint in a predictive model.

THE IMPLIED COST OF CAPITAL: A CLOSER LOOK AT MEASUREMENT ERRORS

4.1 INTRODUCTION

The implied cost of capital (ICC) is the internal rate of return that equates the firm's market value and present value of expected future cash flows. Due to its forward-looking nature, the ICC is an attractive proxy for the expected return, and has increasingly gained attention in the accounting and the finance literature.²³ Much of the existing research aims to assess the relation between the ICC and firm characteristics or regulatory events, and draws inferences of those variables on the expected return (see, for example, Naiker, Navissi and Truong (2013) on option trading; Chen, Chen, Lobo and Wang (2011) on audit quality; Hwang, et al. (2013) on probability of informed trading PIN; and Chava and Purnanandam (2010) on default risk).

However, the construct validity of the ICC as an expected return proxy is open to serious debate. Botosan, Plumlee and Wen (2011) report that after controlling for discount rate news and cash flow news, some ICC estimates are related to firm risk proxies, but others are not. Easton and Monahan (2005) conclude that the ICC, at either the firm or portfolio levels, is not a good predictor for future realised returns. In contrast, realising that the noise in computing the ICC at the firm level Li, et al. (2013) find that the aggregate ICC is quite successful in predicting future returns. The inconclusive evidence urges deeper enquiry, and has recently sparked a new research avenue focusing on why the ICC deviates from the true expected return, i.e. measurement errors. In his ICC survey, Easton (2009) emphasises that measurement errors should be one of the focuses of future research on these estimates.

The current chapter follows this research agenda. In particular, I build an analytical framework allowing cash flow expectations and discount rates to be time varying. I then conduct

²³ Unlike factor-based models or realised returns which rely on noisy ex-post data, the ICC is forward-looking and directly derived from the firm's current stock price and expectation of future firm's fundamentals, e.g. analyst earnings forecasts. As the primary inputs to compute the ICC are analyst earnings forecasts and stock prices, which are determined by market participants, the ICC does not assume that the information set observed by the economic agents is the same as that of the econometricians in modelling conditional expected returns.

simulation and study the two aspects of measurement errors in the ICC. First, in the time series dimension, I seek to understand the extent to which the mean and variance of the ICC deviate from those of the true expected return, due to the constant term structure assumption. I highlight the consequences of the deviation in the regression context, particularly related to the interpretation of the regression coefficients' economic significance. Second, I extend the framework to a panel of firms and years, and examine how measurement errors in cash flow forecasts, a critical input in calculating the ICC, can result in spurious regressions.

Lambert (2009) suggests that among all the issues pertaining to the ICC, the lack of constancy is the most under-researched area. Despite a number of ICC models being developed, they all employ a common assumption, namely, the flat term structure of the discount rate. This assumption is in stark contrast to extensive finance research documenting the dynamic of the discount rate and a large proportion of cost of equity variation resulting from the intertemporal changes in the equity premium (Cochrane (2011); and Croce, Lettau and Ludvigson (2015)). Furthermore, recent evidence suggests that the term structure of equity is not flat (Ang and Liu (2004); and Callen and Lyle (2014)). If the expected return is time varying, the constant term structure ICC is the weighted average of the discount rates over time. The constant ICC will overstate/understate the short rates/long rates, depending on the shape of the term structure (Cready (2001)).

A subsequent inquiry is to understand how the violation of the constant term structure affects established results, using the ICC as a measure of the expected return. It could be the case that even though the equity term structure is not flat, its effects are of second-order importance. For example, Botosan, et al. (2011) and Li, et al. (2013) still find that some static ICC models can track future realised returns quite well. However, Callen and Lyle (2014) show that the non-flat term structure ICC can have better predictive power of future realised returns than its static counterpart, and thus is a better proxy for the expected return. Little effort has been devoted to examining this area.

In this regard, I examine the extent to which the constant term structure leads to the mean and variance of the ICC deviating from the true time varying expected return. I find that, if conditional expected return is time varying, the variation of the ICC is significantly smoother than that of the true expected return. The lack of variation in the ICC can severely bias the regression coefficient estimates, and thus lead researchers to wrongly interpret the economic significance of

the relationship they examine. Specifically, I find that an estimated coefficient is biased downward when the ICC serves as a dependent variable.²⁴ Because an interpretation of economic significance depends on the magnitude of the coefficient and the distribution of the independent variable, given that the distribution of the independent variable is unchanged, researchers will understate the economic impact of the independent variables on the true expected return.²⁵

The second type of measurement error in the ICC stems from measurement errors in analyst cash flow forecasts, the most critical input into the ICC estimation. I show analytically and in simulations that this type of measurement error confounds the researchers' inferences, i.e. it creates spurious regressions. The measurement errors stem from analyst forecasts that are 1) not sufficiently capturing the true market expectations of future cash flows, or 2) being systematically biased towards certain types of firms.

To understand the first source of spurious regressions, note that price is a function of the future cash flow expectation and the expected return, thus any changes in price reflect shocks to the expected cash flows and/or shocks to the discount rate in an indefinite period. However, analysts only update their cash flow forecasts in definite short-term periods, typically one, two, or three years ahead, and then assume a constant growth rate into the distant future. As a result, they potentially miss a fraction of price variation resulting from cash flow expectation shocks of more distant periods (4,5,6,... ∞). If the market price and these limited cash flow forecasts are used to solve for the ICC, the missing variation is transferred to the variation of the estimated ICC. If the cash flow expectations are truly related to firm characteristics, the missing variation is related to firm characteristics and transferred to the discount rate. The ICC is then spuriously related to firm

²⁴ Researchers who seek to examine the relationship between expected return and variable of interest will draw inference from regressions of the ICC on the variable proxy. For example, Naiker, et al. (2013) study the effect of option listing on expected return on equity, by running regression of the ICC on firms with/without listed options. The inference is drawn from the coefficient estimate.

²⁵ Consider that a regression of the ICC serves as a dependent variable on a variable of interest X . The coefficient on X , β_{ICC} , is a function of the correlation between X and the ICC , the variance of X , and the variance of ICC : $\beta_{ICC} = \frac{Cov(X,ICC)}{Var(X)} = Corr(X,ICC) * \frac{\sigma_{ICC}}{\sigma_X}$, where Cov , Var , and $Corr$ are Covariance, Variance, and Correlation operators, respectively. Compare β_{ICC} with the true coefficient, $\beta_{\mu} = \frac{Cov(X,\mu)*\sigma_{\mu}}{\sigma_X}$ when regressing the true expected returns on X , the difference between the coefficients comes from the difference between the standard deviations of the ICC and the true expected returns. As $\sigma_{ICC} < \sigma_{\mu}$, therefore $\beta_{ICC} < \beta_{\mu}$.

characteristics, even though in the theoretical framework the true expected returns are only related to systematic risk.

The second source of measurement error comes from analyst cash flows being biased. These systematic biases are manifested in different forms and well documented in the literature. For example, Guay, Kothari and Shu (2011) find that consensus analyst earnings forecasts tend to be over-optimistic for growth firms. In addition, the forecast errors can result from sample selection bias in which, for example, analysts tend to cover large and financially sound firms (Diether, Malloy and Scherbina (2002)). Irrespective of the forms of bias, the cash flow measurement errors are correlated with firm characteristic variables. If these mismeasured cash flows are used to back out the ICC, the resulting ICC will be correlated with firm characteristic variables.

To better understand the issue, consider that the analyst forecasts are biased, whereas the price is efficient and captures the true expected cash flows and the expected return. If solving for the ICC from “contaminated” cash flow forecasts and true price, the ICC will deviate from the true expected return by the amount of measurement error in the cash flow forecasts. And if the measurement error is related to some types of firms, the spurious regression problem arises when one regresses the ICC on the firm characteristics and finds significant results. The measurement error gives the false impression that the expected return is related to firm characteristics.

I contribute to the existing literature in several ways. First, I provide a convenient analytical framework, allowing for time varying cash flow expectation and expected return, which can be useful in assessing and comparing the ICC and the true conditional expected return. Although the ICC tracks the true time varying expected return quite well, the ICC variation is significantly lower than that of the true expected return, due to the constant term structure assumption. This lays the foundation for my second contribution, in which I show that researchers can no longer correctly interpret the economic significance from the regressions that involve the ICC. Albeit statistical significance is critical in empirical research, correctly assessing the economic significance is equally important (McCloskey and Ziliak (1996)). The ICC fails in this regard. The third contribution surrounds the notion of spurious regressions resulting from the mismeasured analyst cash flow forecasts. With respect to the biased forecasts, i.e. analyst forecast being biased towards certain types of firms (see, for example, Ecker, Francis, Olsson and Schipper (2013)), for the source of biases, I go further than Wang (2015) by showing analytically how the bias is transferred to the discount rate, which in turn results in spurious regressions. Finally, I show that spurious regression can manifest under the inability of analysts to capture the entire cash flow expectation. Given that firm cash flows are related to firm characteristics, the missing variation of the cash flows not captured by analyst forecasts transfers to the discount rate, creating a false impression that the ICC

is related to firm characteristics. To the best of my knowledge, this is the first study to document this phenomenon.

This chapter is structured as follows. Section 4.2 outlines the analytical framework which helps derive the ICC. Section 4.3 shows the simulation procedures for both time series and panel data. Section 4.4 reports and discusses results. Section 4.5 concludes.

4.2 THEORETICAL FRAMEWORK

4.2.1 General Framework

The implied cost of capital is defined as the internal rate of return that equates the present value of future cash flows (dividends) to the firm's current stock price:²⁶

$$P_t = \sum_{i=1}^{\infty} \frac{E_t(D_{t+i})}{(1+r_e)^i}$$

where P_t is the firm's stock price,

$E_t(D_{t+i})$ is the expected dividend at time $t+i$ conditioning on the information at time t .

For tractability, I follow Pastor, et al. (2008) to approximate the present value formula:

$$p_t = \frac{k}{1-\rho} + (1-\rho) \sum_{j=0}^{\infty} E_t(d_{t+1+j}) - \sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}) \quad (4.1)$$

where p_t is the log price,

d_t is the log dividend,

$$\rho = \frac{1}{1+\exp(\overline{d-p})},$$

$$k = -\log(\rho) - (1-\rho)\log\left(\frac{1}{\rho} - 1\right),$$

$\overline{d-p}$ is the average log dividend price ratio.

I define the ICC $r_{e,t}$ in the context of equation (4.1):

$$p_t = \frac{k}{1-\rho} + (1-\rho) \sum_{j=0}^{\infty} E_t(d_{t+1+j}) - r_{e,t} \sum_{j=0}^{\infty} \rho^j \quad (4.2)$$

²⁶ I define dividend generally as the true cash flows that are distributed to the shareholders. Various ICC models can further impose the clean surplus accounting assumption to link dividends with earnings forecasts (Gebhardt, et al. (2001); and Hou, et al. (2012)). Regardless of the definition, the implications remain the same.

I assume that the log dividend growth $g_{t+1} = d_{t+1} - d_t$ is time varying and follows an AR(1) process:

$$g_{t+1} = \alpha_g + \phi_g g_t + e_{g,t+1}, 0 < \phi_g < 1, e_g \sim N(0, \sigma_g^2) \quad (4.3)$$

In addition, I assume that the log expected excess return μ_t is specified as the Conditional Capital Asset Pricing Model (Jagannathan and Wang (1996)).²⁷

$$\mu_t = \beta_t \lambda_t \quad (4.4)$$

where β_t is the firm beta,

λ_t is the market risk premium.

To allow μ_t to be time varying, I specify the dynamics of β and λ as mean reverting:

$$\beta_{t+1} = \alpha_\beta + \phi_\beta \beta_t + e_{\beta,t+1} \quad (4.5)$$

$$\lambda_{t+1} = \alpha_\lambda + \phi_\lambda \lambda_t + e_{\lambda,t+1} \quad (4.6)$$

where $0 < \phi_\beta, \phi_\lambda < 1$ and $(e_\beta, e_\lambda) \sim N\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \sigma_\beta^2 & 0 \\ 0 & \sigma_\lambda^2 \end{bmatrix}\right)$.²⁸

These specifications are convenient for us to extend to panel data later, by modelling each firm by its individual β (see the simulation procedure in Section 4.3.2) (Ferson, et al. (2008)). Given the dynamics of the key parameters, Appendix A.3.1 and A.3.2 show that:

$$\sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j}) = \frac{d_t}{1-\rho} + \frac{\alpha_g}{(1-\rho)^2(1-\rho\phi_g)} + \frac{\phi_g g_t(1-\rho)}{(1-\rho)(1-\rho\phi_g)} \quad (4.7)$$

$$\begin{aligned} \sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}) &= \frac{\beta_t \lambda_t}{1-\rho\phi_\beta\phi_\lambda} + \left(\frac{\rho\alpha_\lambda\phi_\beta}{(1-\rho\phi_\beta)(1-\rho\phi_\beta\phi_\lambda)} \right) \beta_t \\ &+ \left(\frac{\rho\alpha_\beta\phi_\lambda}{(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \right) \lambda_t \\ &+ \frac{\rho\alpha_b\alpha_\lambda(1-\rho^2\phi_b\phi_\lambda)}{(1-\rho)(1-\rho\phi_\beta)(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \end{aligned} \quad (4.8)$$

Substituting (4.7) and (4.8) into (4.1), true price can be expressed as a function of the cash flow expectation and the discount rate:

²⁷ Instead of modelling μ_t following AR(1) process as in Pastor, et al. (2008), I model μ_t as a function of β and risk premium λ under the Conditional CAPM.

²⁸ The dynamics of β and λ are well documented in the literature (for time varying risk premium, see Cochrane (2011) for a review; for evidence on the time varying beta, see Faff, Hillier and Hillier (2000); for modelling the conditional expected return using the Conditional CAPM with time varying beta and risk premium, see Ang and Liu (2004)).

$$\begin{aligned}
p_t = & \frac{k}{1-\rho} + d_t + \frac{\alpha_g}{(1-\rho)(1-\rho\phi_g)} + \frac{\phi_g g_t}{(1-\rho\phi_g)} \\
& - \left(\frac{\beta_t \lambda_t}{1-\rho\phi_\beta\phi_\lambda} + \left(\frac{\rho\alpha_\lambda\phi_\beta}{(1-\rho\phi_\beta)(1-\rho\phi_\beta\phi_\lambda)} \right) \beta_t \right. \\
& \left. + \left(\frac{\rho\alpha_\beta\phi_\lambda}{(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \right) \lambda_t \right. \\
& \left. + \frac{\rho\alpha_b\alpha_\lambda(1-\rho^2\phi_b\phi_\lambda)}{(1-\rho)(1-\rho\phi_\beta)(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \right)
\end{aligned} \tag{4.9}$$

Once the true price is known, I plug equation (4.7) into (4.2) and rearrange to obtain $r_{e,t}$ in which analysts have a full knowledge of expected future dividends:

$$r_{e,t} = k + (d_t - p_t)(1 - \rho) + \frac{\alpha_g}{1 - \rho\phi_g} + \frac{\phi_g g_t(1 - \rho)}{(1 - \rho\phi_g)} \tag{4.10}$$

Realistically, it is not possible to have full horizon information about expected future cash flows. Thus, I relax this assumption and compute several variations of $r_{e,t}$ in which analysts can only forecast cash flows accurately for finite T periods, and subsequently assume cash flows will grow at the constant long term rate $\bar{g} = \frac{\alpha_g}{1-\phi_g}$ at the terminal period. Specifically, the forecasts comprise explicit forecasting and terminal periods:

$$\sum_{j=0}^{\infty} E_t(d_{t+1+j}) = \sum_{j=0}^{T-1} \rho^j (d_t + \sum_{i=1}^{j+1} g_{t+i}) + \sum_{j=1}^{\infty} \rho^{T+j-1} (d_{t+T} + j\bar{g}) \tag{4.11}$$

Under this specification, Appendix A.3.3 shows that

$$\begin{aligned}
\sum_{j=0}^{\infty} E_t(d_{t+1+j}) = & \frac{d_t}{1-\rho} + \left(\frac{\phi_g(1-\rho^T\phi_g^T)}{(1-\rho)(1-\rho\phi_g)} \right) g_t \\
& + \left(\frac{(1-\phi_g + (1-\rho)\rho^T\phi_g^{T+1})}{(1-\rho)^2(1-\phi_g)(1-\rho\phi_g)} \right) \alpha_g
\end{aligned} \tag{4.12}$$

Substituting (4.12) into (4.2), I solve for $r_{e,t}^T$:

$$\begin{aligned}
r_{e,t}^T = & k + (d_t - p_t)(1 - \rho) + \left(\frac{\phi_g(1 - \rho^T\phi_g^T)(1 - \rho)}{(1 - \rho\phi_g)} \right) g_t \\
& + \left(\frac{(1 - \phi_g + (1 - \rho)\rho^T\phi_g^{T+1})}{(1 - \phi_g)(1 - \rho\phi_g)} \right) \alpha_g
\end{aligned} \tag{4.13}$$

where $r_{e,t}^T$ is the ICC derived from the cash flows explicitly forecasted until time T, and then is assumed to grow at a constant rate.

To incorporate measurement errors in analyst cash flow forecasts, I first specify the measurement error w_t following an AR(1) process:²⁹

$$w_{t+1} = \alpha_w + \phi_w w_t + e_{w,t+1} \quad (4.14)$$

where $0 < \phi_w < 1$, $e_w \sim N(0, \sigma_w^2)$.

To address the existing issue pertaining to biases inherent in analyst forecasts, I impose measurement errors in the cash flow forecasts:

$$\overline{d_{t+1}} = d_t + g_{t+1} + w_{t+1} \quad (4.15)$$

where $\overline{d_{t+1}}$ is the cash flow with measurement error.

Thus the ICC with measurement errors in cash flow forecasts r_{ew} can be expressed as:

$$\begin{aligned} r_{ew,t} = & k + (d_t - p_t)(1 - \rho) + \frac{\alpha_g}{1 - \rho\phi_g} + \frac{\phi_g g_t (1 - \rho)}{(1 - \rho\phi_g)} \\ & + \frac{\alpha_w}{1 - \rho\phi_w} + \frac{\phi_w w_t (1 - \rho)}{(1 - \rho\phi_w)} \end{aligned} \quad (4.16)$$

I assume that the long term measurement error reverts to 0; thus $\alpha_w = 0$. Although the long term measurement error is 0, as can be seen in equation (4.16) the ICC $r_{ew,t}$ derived from the biased analyst forecast is a function of the cash flow growth g_t , the dividend price ratio $d_t - p_t$, and the measurement error of the cash flow forecasts w_t .

I compute some variations of $r_{ew,t}$ where analysts only have limited information about expected future dividends. Yet, in this case the analyst explicit forecasts are contaminated with errors.

$$\begin{aligned} r_{ew,t}^T = & k - (d_t - p_t)(1 - \rho) + \left(\frac{\phi_g (1 - \rho^T \phi_g^T) (1 - \rho)}{(1 - \rho\phi_g)} \right) g_t \\ & + \left(\frac{(1 - \phi_g + (1 - \rho)\rho^T \phi_g^{T+1})}{(1 - \phi_g)(1 - \rho\phi_g)} \right) \alpha_g \\ & + \left(\frac{\phi_w (1 - \rho^T \phi_w^T) (1 - \rho)}{(1 - \rho\phi_w)} \right) w_t \end{aligned} \quad (4.17)$$

²⁹ This specification is consistent with the evidence from Wang (2015) documenting that measurement errors are persistent.

4.2.2 Constant Term Structure

The framework conveniently allows us to compare different variations of the ICC and true expected return μ . In the context of a single firm, I first compare how the mean and variance of r_e differ from those of μ without the threat of spurious regression. To show how the smoothing variance affects regression coefficients, I consider two regressions where the expected return proxies serve as dependent as well as independent variables. First, I impose the condition that the variable x_t is actually related to the true expected return μ_t :

$$\mu_t = a_{\mu x} + b_{\mu x} * x_{\mu x,t} + e_{\mu x,t} \quad (4.18)$$

Second, I consider the predictive regression of the ex-post realised returns on the expected returns proxies:

$$r_{t+1} = a_{\mu r} + b_{\mu r} \mu_t + e_{\mu,r,t} \quad (4.19)$$

In both cases, I compare $b_{\mu x}$ and $b_{\mu r}$ with the coefficients obtained from the regressions involving the ICC. In particular, I run the regression where the expected return proxies are the dependent variables:

$$r_{e,t}^* = a_x + b_{r_{ex}} x_{\mu x,t} + \epsilon_t \quad (4.20)$$

where $r_{e,t}^*$ stands for any expected return proxies.

I run the predictive regression where the expected return proxies are the predictors.

$$r_{t+1} = a_r + b_{r_{er}} r_{e,t}^* + \epsilon_t \quad (4.21)$$

The aim is to study the difference between two pairs: 1) $b_{\mu x}$ in (4.18) and $b_{r_{ex}}$ in (4.20); and 2) $b_{\mu r}$ in (4.19) and $b_{r_{er}}$ in (4.21). As projected, the smoothing variation of the r_e^* induces $b_{r_{ex}}$ to be significantly lower than $b_{\mu x}$, and $b_{r_{er}}$ to be significantly higher than $b_{\mu r}$.

4.2.3 Spurious Regressions

The literature typically regresses the ICC on the variable of interest, and draws inferences on the relationship between the variable and the expected returns. I address the threat of confounding inferences in such regressions. There are two scenarios in which spurious regressions can arise. The first arises because the analysts have limited ability to capture the cash flow expectations related to the variable of interest, x . I relate the true cash flow forecasts with the variable x through g_t :

$$g_t = a_{gx} + b_{gx} x_{gx,t} + e_{gx,t} \quad (4.22)$$

To understand how spurious regressions can occur in this scenario, it should be noted that analysts only update their cash flow forecasts for a finite horizon, typically one, two, or three years ahead, and then assume a constant growth rate into the distant future. Yet, if the market price changes reflect all of the changes in the discount rates and/or the cash flow expectation, analyst

forecasts may potentially miss a fraction of price variation resulting from the cash flow shocks of more distant periods (rather than the explicit forecast periods: for example, periods 4, 5, ... ∞). If the market price and these limited cash flow forecasts are used to solve for the ICC, the missing variation is transferred to the variation of the ICC. Because the cash flow expectations are truly related to the interest variable (equation (4.22)) the missing variation is also related to that variable. As a result, the ICC inherits the missing variation and is thus spuriously related to the variable of interest.

Second, spurious regressions can arise when analyst cash flow forecasts are systematically biased. Empirical evidence on these biases, manifested in different forms, is well documented in the literature. For example, Easton and Sommers (2007) show that analyst earnings forecasts are typically over-optimistic; and the level of optimism varies across firm size. Guay, et al. (2011) document that analysts tend to provide more optimistic forecasts for growth firms. Additionally, biased cash flow forecasts can result from sample selection bias manifested under analyst coverage. For instance, small or distressed firms are under-presented in the I/B/E/S database, particularly in the early period (Diether, et al. (2002)). Regardless of the form of the biases, the measurement errors in analyst forecasts are related to firm variables. I relate the firm characteristic $x_{w,t}$ to w_t :

$$w_t = a_{wx} + b_{wx}x_{wx,t} + e_{wx,t} \quad (4.23)$$

Equations (4.16) and (4.17) show that the ICC resulting from the “contaminated” cash flow forecasts and the true price, is a function of the measurement errors in the cash flows forecasts. As the measurement errors are related to the firm characteristics, the ICC is related to the firm characteristics.

4.3 SIMULATION PROCEDURE

4.3.1 Mean and Variance of R_e of a Single Firm

I simulate in the context of a single firm across time because this is where the flat term structure becomes an issue. I describe the simulation procedure to generate the time series of μ_t , $r_{e,t}^T$, and $r_{ew,t}^T$. I choose values of key parameters as follows. To obtain the yearly unconditional mean of the market risk premium 6%, I set $\alpha_\lambda = 0.1\%$ per month and $\phi_\lambda = 0.8$. Next, $\phi_\beta = 0.8$, $\alpha_\beta = 0.2$, and $\sigma_\beta = 0.6$ imply the unconditional mean and variance of β equal to 1. Therefore, the

unconditional mean of μ is 6% per year. With respect to the dividend growth, I obtain the unconditional mean of g at 2.4% per year by setting $\alpha_g = 0.04\%$ per month and $\phi_g = 0.8$. The unconditional means of g and μ result in $\rho = 0.9734$ and $k = 0.1227$.³⁰

Appendix A.3.4 shows that the return variance can be explained by the variance of the expected return, determined by σ_λ , and σ_β , and the variance of dividend growth rate σ_g :

$$\sigma_t^2 = \frac{\sigma_g^2}{(1 - \rho\phi_g)^2} + \rho^2 \left(\frac{\sigma_\lambda^2 \sigma_\beta^2}{(1 - \rho\phi_\beta\phi_\lambda)^2} + \frac{\sigma_\lambda^2 (\alpha_\beta + \beta_t \phi_\beta (1 - \rho\phi_\lambda))^2}{(1 - \rho\phi_\lambda)^2 (1 - \rho\phi_\beta\phi_\lambda)^2} + \frac{\sigma_\beta^2 (\alpha_\lambda + \lambda_t (1 - \rho\phi_\beta)\phi_\lambda)^2}{(1 - \rho\phi_\beta)^2 (1 - \rho\phi_\beta\phi_\lambda)^2} \right) \quad (4.24)$$

I set the unconditional mean of return variance σ_t^2 at $(16\%)^2$ per year, to match the S&P500 variance data. I choose $\sigma_g = (0.559\%, 0.722\%, 0.855\%)$ so that the fraction of return variance that is explained by dividend growth $\phi = \frac{\sigma_g^2}{(1 - \rho\phi_g)^2 \sigma_t^2}$ is $(0.3, 0.5, 0.7)$.³¹ Setting $\sigma_\beta = 0.6$ and $\beta_t = 1$, the fraction of return variance is explained by the variance of β being 0.082. The remaining $(0.618, 0.418, 0.218)$ explanatory power goes to the variance of λ resulting in σ_λ^2 equal to $(0.19\%, 0.156\%, 0.113\%)^2$ per month. Finally, I set $a_{\mu x} = 0$, $b_{\mu x} = 1$, and $\sigma_{\mu x} = \sigma_\lambda$.

The initial values g_0 , β_0 , and λ_0 are set at their unconditional means. d_0 is initialised at 0. The initial price p_0 is calculated from equation (4.9). The simulation below is repeated in each period t , $t = 1 \dots T_{obs}$, given the information at time $t - 1$:

- 1) Draw g_t from (4.3).
- 2) Compute $d_t = d_{t-1} + g_t$.
- 3) Draw β_t and λ_t from (4.5) and (4.6).
- 4) Calculate μ_t from (4.4).
- 5) Compute true price p_t from (4.9).

³⁰ The parameters are chosen to match the empirical data and existing evidence (see Cochrane (2011) and Fama and French (2002) for the equity premium and cash flow growth; see Faff, et al. (2000) for time varying β).

³¹ The quantity of the stock variance driven by expected returns and expected cash flow shocks, is still open to debate (see, for example, Chen, et al. (2013)). I choose the range of the parameter to address this open research question.

- 6) Compute $r_{e,t}$, and $r_{e,t}^T$ from (4.10) and (4.13). The explicit forecast periods T are 0, 3, 5, and 10.
- 7) Generate $x_{\mu x}$ from (4.18).

The process above allows us to generate time series for all variables that I use to study the properties of the ICC in comparison to true expected returns. I consider the difference in mean and variance of μ_t and those of the ICC proxies. I obtain the mean difference $\Delta Mean$ by subtracting the average of the ICC proxies from the average of true expected returns in each simulation. Similarly, the difference in standard deviation, ΔStd , is calculated as standard deviation of the true expected return less standard deviation of the ICC proxies in each simulation. Next, I highlight the implication of the difference in the context of regression, when expected return proxies serve as dependent or independent variables. To do so, I calculate $\Delta Coef$ which are differences between two coefficient pairs 1) $b_{\mu x}$ in (4.18) and $b_{r_{ex}}$ in (4.20); and 2) $b_{\mu r}$ in (4.19) and $b_{r_{er}}$ in (4.21).

Three alternative sample sizes, $T=360$, 480, and 600 months, are considered. I run 1000 repetitions of the simulation for each sample size. The true values of the statistics (coefficients, difference in mean, difference in standard deviation, difference in coefficients) are their average values across simulations. The “t-statistic” is calculated as the average values of the statistics divided by their standard deviations across simulations.³²

4.3.2 Spurious Regression with Panel Data

In this section, I describe a simulation procedure that generates a panel of 100 firms and 40 years of $\mu_{i,t}, r_{i,e,t}^T$ and $r_{i,ew,t}^T$, and study how measurement errors in the cash flows can confound the inferences of regressions involving the ICC. As discussed above, the sources of measurement errors in the cash flows come from 1) the inability of analysts to capture full cash flow expectation and 2) analyst forecasts being systematically biased.

I choose values of key parameters as follows. To obtain the yearly unconditional mean of the market risk premium 6%, I choose $\alpha_\lambda = 0.012\%$ per year and $\phi_\lambda = 0.8$. The cross section of firms features the cross-sectional variation in the conditional β and dividend growth rates. In particular, I generate long term β for each firm from a normal distribution with mean of 1 and

³² As a robustness check, I calculate the t-statistics corrected for Newey and West (1994) in regressions (4.18), (4.19), (4.20), and (4.21). The results are qualitatively similar.

standard deviation of 0.5, which consequently determines $\alpha_{\beta,i}$ in equation (4.5). Similarly, I generate long term dividend growth from a normal distribution with the mean of 2.4% and standard deviation of 2%. I describe the evolution of a firm's β and g_t by setting each firm's $\phi_\beta = 0.8$ and $\phi_g = 0.8$, respectively. I compute numerically for $\rho = 0.9734$ and $k = 0.1227$.³³

I set the conditional mean of return variance σ_t^2 at $(16\%)^2$ per year. I choose $\sigma_g = (1.939\%, 2.503\%, 2.962\%)$ so that the fraction of return variance explained by dividend growth $\frac{\sigma_g^2}{(1-\rho\phi_g)^2\sigma_t^2}$ is (0.3, 0.5, 0.7). The remaining variations of return are explained by σ_β and σ_λ . Specifically, I compute three pairs of $(\sigma_\beta; \sigma_\lambda)$ as (0.237; 1.424%), (0.305; 1.833%), and (0.360; 2.161%) so that variation of the conditional expected return captures 30%, 50%, and 70% of return variance, respectively.³⁴ I set $a_{gx} = a_{wx} = 0$, $b_{gx} = b_{wx} = 1$, and $\sigma_{gx} = \sigma_{wx} = \sigma_g$, in equations (4.22) and (4.23).

For each firm, the initial values g_0 , β_0 , and λ_0 are set at their unconditional means. d_0 is initialised at 0. The initial price p_0 is calculated from equation (4.9). For each firm, the simulation below is repeated in each period t , $t = 1 \dots T_{obs}$, given the information at time $t - 1$:

- 1) Draw g_t from (4.3).
- 2) Compute $d_t = d_{t-1} + g_t$.
- 3) Draw β_t and λ_t from (4.5) and (4.6).
- 4) Calculate μ_t from (4.4).
- 5) Compute true price p_t from (4.9).
- 6) Compute $r_{e,t}$, and $r_{e,t}^T$ from (4.10) and (4.13). The explicit forecast periods T are 0, 3, 5,

³³ Ferson, et al. (2008) simulate cross-section of firms by modelling the β .

³⁴ I obtain $(\sigma_\beta; \sigma_\lambda)$ simultaneously by solving for:

$$\frac{\rho^2}{\sigma_t^2} \left(\frac{\sigma_\lambda^2 \sigma_\beta^2}{(1-\rho\phi_\beta\phi_\lambda)^2} + \frac{\sigma_\lambda^2 (\alpha_\beta + \beta_t \phi_\beta (1-\rho\phi_\lambda))^2}{(1-\rho\phi_\lambda)^2 (1-\rho\phi_\beta\phi_\lambda)^2} + \frac{\sigma_\beta^2 (\alpha_\lambda + \lambda_t (1-\rho\phi_\beta)\phi_\lambda)^2}{(1-\rho\phi_\beta)^2 (1-\rho\phi_\beta\phi_\lambda)^2} \right) = 1 - \frac{\sigma_g^2}{\sigma_t^2 (1-\rho\phi_g)^2}$$

$$\frac{\sigma_\lambda^2 (\alpha_\beta + \beta_t \phi_\beta (1-\rho\phi_\lambda))^2}{(1-\rho\phi_\lambda)^2 (1-\rho\phi_\beta\phi_\lambda)^2} - \frac{\sigma_\beta^2 (\alpha_\lambda + \lambda_t (1-\rho\phi_\beta)\phi_\lambda)^2}{(1-\rho\phi_\beta)^2 (1-\rho\phi_\beta\phi_\lambda)^2} = 0$$

The first equation is to make sure that the remaining fraction of return variance is explained by $(\sigma_\beta; \sigma_\lambda)$, while the second equation assumes that σ_β and σ_λ explain the same amount of return variation. I set initial values of β_t and λ_t as their unconditional values of 1 and 6%, respectively.

and 10.

- 7) Draw w_t from (4.14).
- 8) Compute $r_{ew,t}$ and $r_{ew,t}^T$ from (4.16) and (4.17), respectively. Again, T explicit dividend forecast periods are 3, 5, and 10.
- 9) Generate x_{gx} and x_{wx} from (4.22) and (4.23).

The process repeats for the cross section of firms, and allows us to generate a panel of all variables that I use to study measurement errors leading to spurious regression.

With spurious regressions, I consider two cases highlighted below:

$$r_{e,t}^* = a_{r_{egx}} + b_{r_{egx}}x_{gx,t} + \epsilon_{r_{egx},t} \quad (4.25)$$

$$r_{e,t}^* = a_{r_{ewx}} + b_{r_{ewx}}x_{wx,t} + \epsilon_{r_{ewx},t} \quad (4.26)$$

I examine whether or not the coefficients $b_{r_{egx}}$ and $b_{r_{ewx}} \neq 0$ through the t-statistic and its 95th confidence interval in the regressions (4.25) and (4.26). In each regression I control for the firm β . I run 1000 simulations and obtain 2-way clustered standard error t-statistics in each simulation (Petersen (2009)). The value of the t-statistic is its average value across simulations. To obtain the 95% confidence interval band for the t-statistic, I rank the t-statistics obtained from 1000 simulations, and the two empirical cut-off points are observations at the 2.5th and 97.5th percentiles.

4.4 SIMULATION RESULTS

4.4.1 Constant Term Structure

In this section, I report the simulation results for the differences in mean and standard deviation of the true expected returns μ and those of ICC ($r_{e,0}$, $r_{e,3}$, $r_{e,5}$, $r_{e,10}$, and r_e) (Table 4.1). These in turn affect the economic significance interpretation of regressions involving the ICC (Table 4.2 and Table 4.3).

In Table 4.1, I observe that the ICC proxies do not deviate significantly from the true expected returns in magnitude. For instance, with T=360 and $\phi=0.3$, the mean difference ($\Delta Mean$) between μ and r_e is very close to zero, with the associated t-statistic being 0.0087. Moreover, the differences in mean are robustly negligible across all scenarios when I vary T and ϕ . In contrast, the variation of the ICC is significantly smoother than that of the true expected returns. The difference ΔStd ranges from 0.41% to 0.59% per month. As I decrease the relative importance of cash flow in explaining return variance ϕ , the differences increase. For instance, focusing on r_e and T=600, ΔStd increases from 0.49% to 0.59% when ϕ is from 0.7 to 0.3. Most importantly, these

differences are all statistically significant at the 1% level. The t-statistics are large, ranging from 7 to 12, and they increase as analysts become better able to capture cash flow expectations in more distant periods.

Table 4.2 shows the direct consequence of smoothing variance on the regressions involving the ICC as the dependent variable. In this case, the coefficients obtained from the regressions of the ICC are downwardly biased (equation (4.20)). I compare the coefficients from (4.18) to those from equation (4.20), measured by $\Delta Coef$ and its associated t-statistic. As can be seen from Table 4.2, the ICC is quite powerful in tracking the true relation of a variable and expected return, reflected by the t-statistic of the coefficient, t_{Coef} ; however, all of the coefficients, $Coef$, from the ICC regressions are significantly lower than the true coefficient of 1. The coefficient values, $Coef$, range from 0.1006 ($r_{e,0}, \phi = 0.3$, and $T = 480$) to 0.1113 ($r_{e,0}, \phi = 0.7$, and $T = 600$). The differences between the coefficients from the ICC regression and those from the true expected return regression, $\Delta Coef$, are statistically significant at the 1% level, indicated by $t_{\Delta Coef}$. The result is in line with the evidence in Table 4.1 that the ICC's standard deviation is significantly lower than that of the true expected return.

Table 4.1: Comparison between the Mean and Standard Deviation of the ICC as opposed to that of the true Expected Return

This table reports differences in mean and standard deviation between the ICC proxies ($r_{e,0}, r_{e,3}, r_{e,5}, r_{e,10}$, and r_e) and the true expected returns μ . $\Delta Mean$ is measured by averaging the mean difference between μ and the ICC proxies across 1000 simulations. The associated t-statistic $t_{\Delta Mean}$ is calculated as the average value of the statistic divided by its standard deviation across simulations. ΔStd is measured by averaging the standard deviation difference between μ and the ICC proxies across 1000 simulations. The associated $t_{\Delta Std}$ is calculated as the average value of the statistic divided by its standard deviation across simulations. I vary the number of observations T (360,480, and 600) and the fraction of return variance explained by cash flow news ϕ (0.3, 0.5, and 0.7). A detailed simulation procedure is discussed in Section 4.3.

| | T=360 | | | | T=480 | | | | T=600 | | | | |
|------------------------------|---------------------------------|-------------------------------------|--------------------------------|------------------------------------|---------------------------------|-------------------------------------|--------------------------------|------------------------------------|---------------------------------|-------------------------------------|--------------------------------|------------------------------------|---------|
| Proxies | $\Delta Mean$ | $t_{\Delta Mean}$ | ΔStd | $t_{\Delta Std}$ | $\Delta Mean$ | $t_{\Delta Mean}$ | ΔStd | $t_{\Delta Std}$ | $\Delta Mean$ | $t_{\Delta Mean}$ | ΔStd | $t_{\Delta Std}$ | |
| $\phi=0.3$ | $r_{e,0}$ | 2.70E-06 | 0.0030 | 0.0054 | 7.1551 | -7.65E-06 | -0.0100 | 0.0055 | 8.5089 | 3.15E-05 | 0.0451 | 0.0055 | 9.3860 |
| | $r_{e,3}$ | 5.37E-06 | 0.0060 | 0.0057 | 7.7615 | -8.31E-06 | -0.0110 | 0.0058 | 9.2446 | 3.18E-05 | 0.0459 | 0.0058 | 10.1957 |
| | $r_{e,5}$ | 6.32E-06 | 0.0071 | 0.0058 | 7.9227 | -8.55E-06 | -0.0113 | 0.0058 | 9.4349 | 3.19E-05 | 0.0461 | 0.0059 | 10.4126 |
| | $r_{e,10}$ | 7.35E-06 | 0.0082 | 0.0059 | 8.0234 | -8.80E-06 | -0.0117 | 0.0059 | 9.5466 | 3.21E-05 | 0.0463 | 0.0059 | 10.5505 |
| | r_e | 7.77E-06 | 0.0087 | 0.0059 | 8.0341 | -8.91E-06 | -0.0118 | 0.0059 | 9.5542 | 3.21E-05 | 0.0464 | 0.0059 | 10.5665 |
| $\phi=0.5$ | $r_{e,0}$ | 9.70E-06 | 0.0118 | 0.0048 | 7.3404 | -7.62E-06 | -0.0106 | 0.0048 | 8.3208 | 2.49E-07 | 0.0004 | 0.0048 | 9.2467 |
| | $r_{e,3}$ | 8.73E-06 | 0.0108 | 0.0052 | 8.3385 | -1.05E-05 | -0.0150 | 0.0052 | 9.4588 | 2.25E-06 | 0.0034 | 0.0053 | 10.5050 |
| | $r_{e,5}$ | 8.39E-06 | 0.0104 | 0.0053 | 8.6374 | -1.15E-05 | -0.0165 | 0.0054 | 9.7982 | 2.95E-06 | 0.0045 | 0.0054 | 10.8801 |
| | $r_{e,10}$ | 8.01E-06 | 0.0100 | 0.0054 | 8.8394 | -1.26E-05 | -0.0182 | 0.0054 | 10.0232 | 3.72E-06 | 0.0057 | 0.0055 | 11.1323 |
| | r_e | 7.86E-06 | 0.0098 | 0.0054 | 8.8603 | -1.30E-05 | -0.0188 | 0.0054 | 10.0438 | 4.03E-06 | 0.0062 | 0.0055 | 11.1579 |
| $\phi=0.7$ | $r_{e,0}$ | -1.88E-05 | -0.0241 | 0.0041 | 7.1353 | 4.11E-06 | 0.0059 | 0.0041 | 8.3278 | 3.37E-06 | 0.0055 | 0.0041 | 9.3084 |
| | $r_{e,3}$ | -1.83E-05 | -0.0246 | 0.0046 | 8.4980 | 6.92E-06 | 0.0102 | 0.0047 | 10.0198 | -2.17E-06 | -0.0036 | 0.0047 | 11.2202 |
| | $r_{e,5}$ | -1.82E-05 | -0.0246 | 0.0048 | 8.9289 | 7.91E-06 | 0.0118 | 0.0048 | 10.5745 | -4.11E-06 | -0.0068 | 0.0048 | 11.8450 |
| | $r_{e,10}$ | -1.80E-05 | -0.0246 | 0.0049 | 9.2292 | 9.00E-06 | 0.0135 | 0.0049 | 10.9856 | -6.26E-06 | -0.0104 | 0.0049 | 12.2986 |
| | r_e | -1.79E-05 | -0.0245 | 0.0049 | 9.2514 | 9.44E-06 | 0.0141 | 0.0049 | 11.0307 | -7.12E-06 | -0.0118 | 0.0049 | 12.3424 |

Table 4.2: Difference between the Coefficients estimated from the Regression of the ICC on a Variable of Interest, and those from the Regression of the true Expected Returns on the same Variable of Interest

This table compares two sets of the coefficients: those in which the ICC proxies ($r_{e,0}, r_{e,3}, r_{e,5}, r_{e,10}$, and r_e), and ex-post realised returns r_1 serve as the dependent variables (equation (4.20)); and those in which the true expected return μ is used as the dependent variable (equation (4.18)). I report the coefficients $Coef$ and their associated statistics t_{Coef} . $\Delta Coef$ is the difference between the coefficients obtained from equation (4.18) and those from equation (4.20). $t_{\Delta Coef}$ is the associated t-statistic of $\Delta Coef$. The statistic is calculated by averaging the estimates across 1000 simulations. The t-statistic is calculated by dividing the average statistic by the standard deviation of the statistics across 1000 simulations. A detailed discussion of the simulation procedure is in Section 4.3.

| | Proxies | $\phi = 0.3$ | | | | $\phi = 0.5$ | | | | $\phi = 0.7$ | | | |
|--------------|------------|--------------|------------|---------------|-------------------|--------------|------------|---------------|-------------------|--------------|------------|---------------|-------------------|
| | | $Coef$ | t_{Coef} | $\Delta Coef$ | $t_{\Delta Coef}$ | $Coef$ | t_{Coef} | $\Delta Coef$ | $t_{\Delta Coef}$ | $Coef$ | t_{Coef} | $\Delta Coef$ | $t_{\Delta Coef}$ |
| T=360 | μ | 0.9207 | 39.9799 | - | - | 0.9369 | 52.3331 | - | - | 0.9591 | 76.7328 | - | - |
| | $r_{e,0}$ | 0.1011 | 7.1152 | 0.8196 | 30.9629 | 0.1053 | 5.4295 | 0.8316 | 32.2628 | 0.1105 | 4.3303 | 0.8485 | 30.9614 |
| | $r_{e,3}$ | 0.1009 | 14.1871 | 0.8198 | 35.2158 | 0.1050 | 11.0515 | 0.8319 | 42.4813 | 0.1105 | 9.0540 | 0.8486 | 51.4094 |
| | $r_{e,5}$ | 0.1009 | 21.0200 | 0.8198 | 36.2432 | 0.1049 | 17.0565 | 0.8320 | 45.6018 | 0.1104 | 14.4784 | 0.8486 | 61.5132 |
| | $r_{e,10}$ | 0.1008 | 35.3106 | 0.8199 | 36.8921 | 0.1047 | 34.5587 | 0.8322 | 47.5466 | 0.1104 | 34.1911 | 0.8487 | 68.8459 |
| | r_e | 0.1008 | 38.8481 | 0.8199 | 36.9899 | 0.1047 | 41.8480 | 0.8322 | 47.7145 | 0.1104 | 43.1048 | 0.8487 | 68.8698 |
| | r_1 | 0.9192 | 3.6761 | 0.0015 | 0.0059 | 0.9393 | 2.9764 | -0.0024 | -0.0076 | 0.9546 | 2.2920 | 0.0045 | 0.0108 |
| T=480 | μ | 0.9218 | 50.0365 | - | - | 0.9373 | 61.5055 | - | - | 0.9601 | 85.9031 | - | - |
| | $r_{e,0}$ | 0.1006 | 8.1733 | 0.8211 | 39.3805 | 0.1045 | 6.0081 | 0.8328 | 37.3651 | 0.1110 | 4.8786 | 0.8491 | 33.8242 |
| | $r_{e,3}$ | 0.1008 | 16.2622 | 0.8210 | 44.8094 | 0.1047 | 12.3350 | 0.8327 | 49.9836 | 0.1106 | 10.1333 | 0.8495 | 55.6963 |
| | $r_{e,5}$ | 0.1008 | 24.0459 | 0.8209 | 45.9038 | 0.1047 | 19.1601 | 0.8326 | 53.8078 | 0.1105 | 16.1643 | 0.8496 | 67.0680 |
| | $r_{e,10}$ | 0.1009 | 40.7153 | 0.8209 | 46.3364 | 0.1047 | 39.3962 | 0.8326 | 55.9902 | 0.1103 | 39.3315 | 0.8497 | 76.9500 |
| | r_e | 0.1009 | 45.4949 | 0.8209 | 46.2638 | 0.1048 | 47.5616 | 0.8325 | 56.0429 | 0.1103 | 53.2044 | 0.8498 | 78.0664 |
| | r_1 | 0.9114 | 4.1590 | 0.0104 | 0.0459 | 0.9249 | 3.3139 | 0.0124 | 0.0438 | 0.9570 | 2.7527 | 0.0030 | 0.0086 |

Table 4.2 (Continued)

| Proxies | $\phi = 0.3$ | | | | $\phi = 0.5$ | | | | $\phi = 0.7$ | | | |
|------------------------|--------------|-------------------------|----------------------|--------------------------|--------------|-------------------------|----------------------|--------------------------|--------------|-------------------------|----------------------|--------------------------|
| | <i>Coef</i> | <i>t_{Coef}</i> | Δ <i>Coef</i> | <i>t_{ΔCoef}</i> | <i>Coef</i> | <i>t_{Coef}</i> | Δ <i>Coef</i> | <i>t_{ΔCoef}</i> | <i>Coef</i> | <i>t_{Coef}</i> | Δ <i>Coef</i> | <i>t_{ΔCoef}</i> |
| μ | 0.9233 | 56.2518 | - | - | 0.9381 | 68.5402 | - | - | 0.9601 | 97.7309 | - | - |
| $r_{e,0}$ | 0.1017 | 9.7055 | 0.8216 | 42.3919 | 0.1054 | 6.9592 | 0.8327 | 40.6261 | 0.1113 | 5.5682 | 0.8488 | 38.4623 |
| T=600 $r_{e,3}$ | 0.1013 | 19.4570 | 0.8220 | 48.6721 | 0.1051 | 14.3605 | 0.8330 | 54.1155 | 0.1108 | 11.5857 | 0.8493 | 63.5251 |
| $r_{e,5}$ | 0.1012 | 29.0047 | 0.8222 | 50.2594 | 0.1050 | 22.4434 | 0.8331 | 58.5097 | 0.1106 | 18.4965 | 0.8495 | 76.5967 |
| $r_{e,10}$ | 0.1010 | 48.9004 | 0.8223 | 51.3210 | 0.1048 | 45.7219 | 0.8333 | 61.5343 | 0.1104 | 44.0465 | 0.8497 | 87.8964 |
| r_e | 0.1009 | 53.1242 | 0.8224 | 51.5103 | 0.1048 | 52.7055 | 0.8333 | 61.9474 | 0.1104 | 56.5879 | 0.8498 | 89.1080 |
| r_1 | 0.9021 | 4.6444 | 0.0213 | 0.1062 | 0.9155 | 3.6590 | 0.0226 | 0.0891 | 0.9151 | 2.9073 | 0.0450 | 0.1424 |

Given that the distribution of the independent variables does not change, the low coefficients will result in understating the true economic significance of the variable on expected returns. To see the issue analytically, consider that the regression of the ICC serves as a dependent variable on the variable X . The coefficient on β_{ICC} can be expressed as below:

$$\beta_{ICC} = \frac{Cov(X, ICC)}{Var(X)} = \frac{Corr(X, ICC) * \sigma_X * \sigma_{ICC}}{\sigma_X^2} \quad (4.27)$$

where Cov , Var , and $Corr$ are Covariance, Variance, and Correlation operators, respectively. Comparing β_{ICC} with the true coefficient, $\beta_{\mu} = \frac{Corr(X, \mu) * \sigma_X * \sigma_{\mu}}{\sigma_X^2}$ when I regress the true expected returns on X , the difference between the standard deviations of the ICC and the true expected returns produce the difference between the coefficients. As $\sigma_{ICC} < \sigma_{\mu}$, therefore $\beta_{ICC} < \beta_{\mu}$.

In contrast, despite being less powerful in tracking the time varying expected returns, which are reflected by the t-statistics being always less than those of the ICC, the magnitude of the coefficients from the realised return regression r_1 are similar to the counterpart case involving the true expected returns μ . For example, with $\phi = 0.5$ and $T = 600$, the coefficient of μ regression is 0.9381, which is quite close to the value of 0.9155 of the coefficient of r_1 regression. The difference between the two coefficients (0.0226) is not statistically significant (t-statistic=0.0891).

Table 4.3 presents the results with respect to the predictive regression, using the ICC proxies as predicting variables. I report the coefficients from regressing ex-post realised returns on the true expected return and the ICC proxies, and the difference between these coefficients ($\Delta Coef$). Statistically, the ICC is powerful in predicting future realised returns, reflected by the large t-statistics associated with the estimated coefficients. It is interesting to see that the larger the cash flow shocks are affecting stock price movement (larger ϕ), the weaker the predictive power of the ICC (t-statistics associated with the coefficients decrease as I move from left to right of the Table). This observation is consistent with Botosan, et al. (2011) who highlight the importance of controlling for cash flow news in predictive regressions. Similar to Table 4.2, I can see that the coefficients on the ICC proxies deviate from the true coefficient of 1. The negative sign on the $\Delta Coef$ implies the coefficients on ICC are larger than those on μ . In particular, most of the

coefficients on the ICC are statistically larger than 1, with the exception of $r_{e,0}$ ($r_{e,3}$) where ϕ is 0.5 and 0.7.³⁵

Interestingly, the more successful analysts are at capturing cash flow expectations into the distant future, the larger the difference between the coefficients. For example, focusing on $T=480$ and $\phi = 0.5$, the coefficient on r_e (perfect cash flow information) is 8.55 (t-statistic=3.4) while that on $r_{e,3}$ (accurate cash flow forecasts up to 3 years) is 5.21 (t-statistic=2.53). The differences $\Delta Coef$ between these coefficients and those on the true expected returns μ (0.99) are statistically significant (-4.23 with t-statistic=-2.25 for $r_{e,3}$; and -7.39 with t-statistic=-3.37 for r_e). This evidence is consistent with the observation in Table 4.1, in which the more accurately the analyst captures future cash flow, the smoother the ICC's variation becomes due to constant term structure assumption. Due to the significant difference between the coefficients reflected by $\Delta Coef$, the economic significance interpretation of the coefficients, from regressions where the ICC is the independent variable, is no longer meaningful. In this case, researchers overstate the true economic impact of the ICC on the variable of interest.

³⁵ Given that the $\sigma_{r_e} < \sigma_\mu$ the coefficient $\beta_{r_e} = \frac{Cov(r_e, r)}{Var(r_e)} = \frac{\rho_{r_e} \sigma_r}{\sigma_{r_e}} > \frac{\rho_\mu \sigma_r}{\sigma_\mu} = \beta_\mu$.

Table 4.3: Difference between the Coefficients estimated from the Regression of the Future Realised Returns on the ICC, as opposed to those from the Regressions of Future Realised Returns on the true Expected Returns

This table compares two sets of coefficients: those in the predictive regression of ex-post realised returns where the ICC proxies ($r_{e,0}, r_{e,3}, r_{e,5}, r_{e,10}$, and r_e serve as the independent variables (equation (4.21)); and those where the true expected return μ is used as the predicting variable (equation (4.19)). I report the coefficients $Coef$ and their associated statistics t_{Coef} . $\Delta Coef$ is the difference between the coefficients obtained from equation (4.19) and those from equation (4.21). $t_{\Delta Coef}$ is the associated t-statistic of $\Delta Coef$. The statistic is calculated by averaging the estimates across 1000 simulations. The t-statistic is calculated by dividing the average statistic by the standard deviation of the statistics across 1000 simulations. A detailed discussion of the simulation procedure is in Section 4.3.

| Proxies | $\phi=0.3$ | | | | $\phi=0.5$ | | | | $\phi=0.7$ | | | | |
|--------------|------------|------------|---------------|-------------------|------------|------------|---------------|-------------------|------------|------------|---------------|-------------------|---------|
| | $Coef$ | t_{Coef} | $\Delta Coef$ | $t_{\Delta Coef}$ | $Coef$ | t_{Coef} | $\Delta Coef$ | $t_{\Delta Coef}$ | $Coef$ | t_{Coef} | $\Delta Coef$ | $t_{\Delta Coef}$ | |
| T=360 | μ | 0.9971 | 3.6214 | - | - | 0.9981 | 3.0549 | - | - | 0.9956 | 2.3377 | - | - |
| | $r_{e,0}$ | 3.5535 | 2.2084 | -2.5564 | -1.6849 | 2.2570 | 1.4962 | -1.2589 | -0.8686 | 1.5765 | 0.9842 | -0.5809 | -0.3787 |
| | $r_{e,3}$ | 6.5749 | 2.9718 | -5.5778 | -2.7667 | 5.2900 | 2.2701 | -4.2919 | -2.0138 | 4.2775 | 1.5761 | -3.2819 | -1.3247 |
| | $r_{e,5}$ | 7.7850 | 3.2783 | -6.7880 | -3.1733 | 6.9947 | 2.6243 | -5.9965 | -2.4974 | 6.2209 | 1.9044 | -5.2252 | -1.7784 |
| | $r_{e,10}$ | 8.6426 | 3.5753 | -7.6455 | -3.5494 | 8.4701 | 2.9560 | -7.4720 | -2.9280 | 8.2220 | 2.2582 | -7.2264 | -2.2373 |
| | r_e | 8.7322 | 3.6686 | -7.7351 | -3.6598 | 8.6485 | 3.0426 | -7.6504 | -3.0314 | 8.4646 | 2.3436 | -7.4689 | -2.3401 |
| T=480 | μ | 0.9912 | 4.2061 | - | - | 0.9859 | 3.3819 | - | - | 0.9992 | 2.7470 | - | - |
| | $r_{e,0}$ | 3.5008 | 2.5431 | -2.5096 | -1.9265 | 2.2004 | 1.6667 | -1.2145 | -0.9648 | 1.4131 | 1.0214 | -0.4140 | -0.3099 |
| | $r_{e,3}$ | 6.5211 | 3.5012 | -5.5299 | -3.2644 | 5.2134 | 2.5267 | -4.2275 | -2.2520 | 4.0739 | 1.7340 | -3.0748 | -1.4266 |
| | $r_{e,5}$ | 7.7350 | 3.8605 | -6.7437 | -3.7450 | 6.9092 | 2.9077 | -5.9233 | -2.7742 | 6.0519 | 2.1433 | -5.0527 | -1.9825 |
| | $r_{e,10}$ | 8.5972 | 4.1940 | -7.6059 | -4.1696 | 8.3731 | 3.2754 | -7.3872 | -3.2464 | 8.1959 | 2.6239 | -7.1967 | -2.5947 |
| | r_e | 8.6899 | 4.2975 | -7.6986 | -4.2921 | 8.5509 | 3.3806 | -7.5650 | -3.3707 | 8.5240 | 2.7623 | -7.5249 | -2.7595 |

Table 4.3 (Continued)

| | | <u>$\phi = 0.3$</u> | | | | <u>$\phi = 0.5$</u> | | | | <u>$\phi = 0.7$</u> | | | |
|---------------------|------------|--------------------------------|-------------------------|----------------------|--|--------------------------------|-------------------------|----------------------|--|--------------------------------|-------------------------|----------------------|--|
| | | <i>Coef</i> | <i>t_{Coef}</i> | Δ <i>Coef</i> | <i>t_{ΔCoef}</i> | <i>Coef</i> | <i>t_{Coef}</i> | Δ <i>Coef</i> | <i>t_{ΔCoef}</i> | <i>Coef</i> | <i>t_{Coef}</i> | Δ <i>Coef</i> | <i>t_{ΔCoef}</i> |
| | μ | 0.9787 | 4.6919 | - | - | 0.9778 | 3.7316 | - | - | 0.9541 | 2.9382 | - | - |
| | $r_{e,0}$ | 3.3175 | 2.6392 | -2.3388 | -1.9689 | 2.0607 | 1.6367 | -1.0829 | -1.0829 | 1.3671 | 1.1289 | -0.4130 | -0.3585 |
| <u>T=600</u> | $r_{e,3}$ | 6.2996 | 3.7547 | -5.3209 | -3.4757 | 5.0147 | 2.5913 | -4.0370 | -4.0370 | 3.9130 | 1.8838 | -2.9588 | -1.5639 |
| | $r_{e,5}$ | 7.5339 | 4.2299 | -6.5552 | -4.0892 | 6.7129 | 3.0814 | -5.7352 | -5.7352 | 5.7975 | 2.3180 | -4.8434 | -2.1543 |
| | $r_{e,10}$ | 8.4637 | 4.6808 | -7.4850 | -4.6544 | 8.2465 | 3.6133 | -7.2688 | -7.2688 | 7.8363 | 2.8102 | -6.8821 | -2.7815 |
| | r_e | 8.5942 | 4.8168 | -7.6155 | -4.8149 | 8.4740 | 3.7770 | -7.4962 | -7.4962 | 8.1458 | 2.9384 | -7.1917 | -2.9337 |

4.4.2 Spurious Regression Effects

In this section, I present simulation results on spurious regressions that are caused by measurement errors in analyst cash flow forecasts. The sources of measurement error are from 1) the limited ability of analysts to capture full cash flow information and 2) the systematic bias of analyst forecasts. Evidence of spurious regressions is reflected by the 95% confidence interval band for the t-statistics being different from the standard range (-1.98, 1.98).

Table 4.4 shows how limited information in the cash flow forecasts can give rise to a mechanical relation between variable of interest and the ICC (equation (4.25)). Focusing on the true expected returns μ_t and the ICC r_e in which analysts have full information about true cash flows, there is no evidence regarding the existence of spurious results. In Panel A, the t-statistics for r_e and the true expected return μ are -0.02, and 0.04, respectively. Moving across the panels as cash flows become more important in explaining return variance, the t-statistics remain stable around -0.01 for r_e and about 0.4 for μ . The confidence intervals for the t-statistic for r_e and μ are (-2.09, 2.11) and (-2.23, 2.09), respectively, which are close to the standard 5% interval (-1.98, 1.98).

However, moving to the ICC proxies backed out from stock price and limited cash flow forecasts $r_{e,0}$, $r_{e,3}$, $r_{e,5}$, and $r_{e,10}$, there is clear evidence that the missing variation in the cash flows is transferred to the discount rate, which then induces the spurious regression effect. In Panel A, when the analyst has no information about cash flows and cash flow news captures 30% of the return variation, the ICC $r_{e,0}$ is negatively related to the variable of interest x (t-statistic = -2.39; 2.5% cut-off = -4.49; 97.5% cut-off = -0.26) even though the true expected return μ is only related to β by construction. Note that any increase in price results from either an increase in cash flow expectation and/or decrease in discount rate. As I specify that the variable x_{gx} is positively related to cash flow growth, the missing variation is incorporated into the discount rate and leads to a false impression that the variable of interest reduces expected returns.

The more important the cash flows in driving stock returns variance, the more serious the spurious regression problem becomes. In Panel C, when 70% of price movement is driven by cash flow news ($\phi = 0.7$, Panel C), even though analysts can capture accurately up to 5 year cash flows, the implied cost of capital $r_{e,5}$ is still negatively correlated to the variable of interest at the 10% level (t-statistic=-1.76; 2.5% cut-off=-3.84; 97.5% cut-off=0.34). The t-statistic monotonically decreases with the amount of cash flow information analysts can capture, thus the spurious regression becomes less of a problem. In particular, when the cash flows explain 70% of return variance and analysts can forecast accurately up to 10 years (t-statistic=-0.51), there is less spurious threat. However, the left tailed t-statistic of -2.53 falls outside the traditional 5% of critical values -1.98.

Table 4.4: Spurious Regression when Analyst Forecasts are unable to capture Full Cash Flow Expectations

This table reports average value, 2.5%, and 97.5% empirical cut-off points of a 2-way clustered t-statistic in the regression of different ICC proxies on firm characteristics (equation (4.25)). 1000 simulations are run. In each simulation, a balanced panel data including 100 firms and 40 years is generated. Panels A, B, and C vary the results in accordance with changes in the amount of return variance explained by cash flow news from 30%, 50%, to 70%, respectively.

| Panel A: $\phi=0.3$ | | | |
|---------------------------------------|---------------|-------------|--------------|
| Proxies | t-stat | 2.5% | 97.5% |
| r_e | -0.02 | -2.09 | 2.11 |
| μ | 0.04 | -2.23 | 2.09 |
| $r_{e,0}$ | -2.39 | -4.49 | -0.26 |
| $r_{e,3}$ | -1.14 | -3.16 | 1.01 |
| $r_{e,5}$ | -0.69 | -2.77 | 1.43 |
| $r_{e,10}$ | -0.21 | -2.29 | 1.91 |
| Panel B: $\phi=0.5$ | | | |
| Proxies | t-stat | 2.5% | 97.5% |
| r_e | -0.01 | -1.95 | 2.10 |
| μ | 0.04 | -2.07 | 2.03 |
| $r_{e,0}$ | -3.93 | -6.19 | -1.77 |
| $r_{e,3}$ | -1.87 | -3.95 | 0.20 |
| $r_{e,5}$ | -1.14 | -3.15 | 0.94 |
| $r_{e,10}$ | -0.33 | -2.31 | 1.77 |
| Panel C: $\phi=0.7$ | | | |
| Proxies | t-stat | 2.5% | 97.5% |
| r_e | -0.01 | -2.02 | 2.13 |
| μ | 0.04 | -2.09 | 2.02 |
| $r_{e,0}$ | -6.02 | -8.65 | -3.80 |
| $r_{e,3}$ | -2.88 | -5.12 | -0.83 |
| $r_{e,5}$ | -1.76 | -3.84 | 0.34 |
| $r_{e,10}$ | -0.51 | -2.53 | 1.64 |

Table 4.5: Spurious Regression when Analyst Cash Flow Forecasts are systematically biased

This table reports average value, 2.5%, and 97.5% empirical cut-off points of a 2-way clustered t-statistic in the regression of different implied cost of capital proxies on firm characteristics (equation (4.26)). 1000 simulations are run. In each simulation, balanced panel data, including 100 firms and 40 years, is generated. Panels A, B, and C vary the results in accordance with changes in the amount of return variance is explained by cash flows news from 30%, 50%, and 70%, respectively.

| Panel A: $\phi=0.3$ | | | |
|---------------------------------------|---------------|-------------|--------------|
| Proxies | t-stat | 2.5% | 97.5% |
| μ | 0.05 | -1.98 | 1.98 |
| $r_{ew,3}$ | 2.27 | -1.70 | 6.53 |
| $r_{ew,5}$ | 3.05 | -0.99 | 7.57 |
| $r_{ew,10}$ | 3.87 | -0.23 | 8.52 |
| r_{ew} | 4.19 | 0.07 | 4.74 |
| Panel B: $\phi=0.5$ | | | |
| Proxies | t-stat | 2.5% | 97.5% |
| μ | 0.02 | -1.89 | 2.02 |
| $r_{ew,3}$ | 2.80 | 0.75 | 5.05 |
| $r_{ew,5}$ | 3.80 | 1.67 | 6.20 |
| $r_{ew,10}$ | 4.89 | 2.71 | 7.37 |
| r_{ew} | 5.33 | 19.84 | 9.47 |
| Panel C: $\phi=0.7$ | | | |
| Proxies | t-stat | 2.5% | 97.5% |
| μ | 0.05 | -1.96 | 1.98 |
| $r_{ew,3}$ | 4.22 | 0.54 | 8.34 |
| $r_{ew,5}$ | 5.55 | 1.69 | 9.99 |
| $r_{ew,10}$ | 6.87 | 2.81 | 11.98 |
| r_{ew} | 7.36 | 3.08 | 12.71 |

In Table 4.5, I present evidence that shows if analyst forecasts are biased towards certain firm characteristics, the expected return proxy spuriously relates to that firm characteristic, even though the true expected return is generated purely from the conditional CAPM (equation (4.26)). Similar to Table 4.4, I examine how much of the threat increases as I raise the importance of cash flow expectation shocks in driving stock price. First, I start with the benchmark true expected returns μ . Regardless of the importance of the cash flows in driving stock movements, spurious regression is not a major concern. The t-statistic remains around 0.05 when I move across the panels. Additionally, the range of the t-statistic stays around (-1.89, 2.02) which is very close to the standard critical values for the two-tailed 5% level of significance (-1.98, 1.98).

Although the true expected return is not related to the firm characteristic, the ICC $r_{ew,3}$, $r_{ew,5}$, $r_{ew,10}$ and r_{ew} backed out from “contaminated” cash flow forecasts induce a serious spurious issue. In particular, consider the case where cash flows drive 30% of return variance ($\phi = 0.3$) and analysts capture three years of forecasts, with some bias towards certain firms ($r_{ew,3}$), the implied cost of capital $r_{ew,3}$ is significantly related to firm characteristics, reflected by the t-statistic of 2.27. Similar to the prior analysis, the threat is more pronounced where cash flow news becomes increasingly important in driving price variation. Looking at Panels B and C, the $r_{ew,3}$'s t-statistic increases to 2.80 and 4.22, as cash flows explain 50% and 70% of return variance.

If analyst forecast errors extend into the more distant future, the problem gets worse. In particular, if analysts try to capture up to five years of future cash flows, and are biased, the t-statistics obtained from regression of $r_{ew,5}$ on firm characteristics are all higher than those of three-year ahead forecasts (3.05 Panel A; 3.80 Panel B; and 5.55 Panel C). With respect to $r_{ew,10}$, the average t-statistics are 3.87 (Panel A), 4.89 (Panel B), and 6.87 (Panel C).

4.5 CONCLUDING REMARKS

In this paper, I derive an analytical framework, allowing for time varying expected returns and time varying cash flow expectation, and conduct simulation to study in which circumstances the ICC deviates from true expected returns. First, I find that the ICC does not deviate from the true expected return in terms of magnitude, meaning that it does not overstate or understate the mean of the true expected returns. However, due to the constant term structure assumption, the variation of the ICC is significantly less than that of the true expected returns. It has a direct consequence for regressions involving the ICC. In particular, the coefficient on the regression of the ICC and any particular variable significantly understates the true relation. As a result, the economic significance cannot be interpreted based on the resulting coefficient. One remedy is to interpret the economic significance from a realised returns proxy. Although the realised return proxy is noisy, which makes it difficult to detect the statistical relationship, its resulting coefficient is unbiased. The smoothing variance is also a problem in predictive regression. The coefficient on the ICC is greater than the true coefficients. My findings lend support to the recent movement in studying the term structure of equity (Lambert (2009); and Callen and Lyle (2014)).

In addition, I study the threats of spurious regression in ICC studies. The source of the spurious regression comes from the critical input of the ICC, analyst forecasts. I find that if analysts 1) are biased in their cash flow forecasts toward certain types of firms, and 2) have limited cash flow forecasts, these measurement errors are translated to the discount rate and give us a false impression that the expected return is related to certain firm characteristics.

5.1 THESIS REVIEW

This thesis is a trilogy of essays focused on modelling conditional expected return and testing a central research question: “Is the ex-ante market risk premium always positive?”

The notion that risk matters in determining returns is intuitive. To compensate for bearing additional risk, rational and risk-averse investors should demand higher returns for risky assets than for the risk free rate. As a result, the risk premium of the market portfolio always exceeds zero. Nevertheless, it is not, as it seems, so straightforward. An intense debate surrounding this positivity restriction continues to challenge asset pricing theorists. One of the most well respected asset pricing models, the Conditional CAPM imposes the positive risk premium as a necessary condition, whereas more general theories show that the negative risk premium is possible when market returns positively covary with the marginal rate of substitution (Whitelaw (2000); and Harrison and Kreps (1979)). Moreover, an overwhelming amount of research that documents the complex nature of the risk-return trade-off remains heated and unsettled (Pastor, et al. (2008); or Rossi and Timmermann (2010)). Given these observations, it is surprising that not much empirical research endeavours to test the positive risk premium hypothesis.

The studies contained in this thesis examine the question with novel measures of conditional expected return. Through these estimates, I further highlight the critical role of valid empirical proxies for unobservable conditional expectations, used for testing conditional asset pricing models. The thesis seeks to contribute by connecting different and important strands of research.

5.2 REVIEW OF FINDINGS AND CONTRIBUTIONS

The first essay uses the implied cost of capital (ICC), a forward-looking proxy for the conditional expected return (Pastor, et al. (2008)). The ICC is defined as the internal rate of return that equates the firm’s value with the present value of its expected cash flows. Because the ICC uses earnings forecasts and stock prices, it is strictly ex-ante and thus overcomes numerous pitfalls associated with the noisy realised return being a proxy for the expected return (Elton (1999); Campello, et al. (2008); and Chen, et al. (2013)). Applying this proxy in the multiple inequality constraints framework, I find evidence that the positive risk premium is violated in the US, German, Japanese, and Italian markets. Such evidence implies that the Condition CAPM does not hold in these markets (Lewellen and Nagel (2006)).

These findings contribute to the literature by bridging unrelated areas of the accounting and finance research. First, while empirical asset pricing has devoted considerable effort to investigating the linear restrictions implied by models, it has devoted little attention to an important condition: positivity of the risk premium. This first study, among a few, directly examines whether or not this condition is violated in the international context. Second, unlike most accounting literature studying the cross-sectional properties of the ICC, I focus on the time series characteristic of the aggregate ICC.

The second study introduces a new two-stage procedure to model the conditional mean return. Using Principal Component Analysis (PCA) in the first stage, I seek to capture the investors' information set by summarising the variation underlying 160 financial variables from Ludvigson and Ng (2007) and Welch and Goyal (2008) into a few common factors. Next, a set of principal components are included in the second stage where a regression tree technique is employed. This technique approximates the unknown function of information variables in producing conditional risk premium, by breaking up the predictor space sequentially through piece-wise constant models. Additionally, an ensemble method, known as boosting, aims to combine the individual trees into a final model with much improved properties (Rapach, et al. (2010); and Ng (2014)). The risk premium, which is estimated out-of-sample from the two-stage method, serves as a conditional risk premium proxy in the multiple inequality constraints framework in order to investigate the positive risk premium hypothesis in the US market from 1970 to 2012. I again find evidence that the positive risk premium condition is violated in the US market. Also, the two-stage method survives through various validity tests, indicating that it is a good proxy for the conditional risk premium.

With the introduction of the superior two-stage procedure, the second essay contributes by highlighting the importance of capturing the investors' information set along with the complex return data generating process. This is the first study to apply the boosting regression tree, a state-of-the-art method in the machine learning literature, in testing the positivity of the risk premium hypothesis.

While appealing in its own right, the ICC's validity as a proxy for the conditional expected return is open to debate. Accordingly, the third study investigates when and to what extent this estimate deviates from true expectations. In a simulation study allowing for a time varying discount rate, I find that due to the constant term structure assumption embedded in the ICC method, the variation of the ICC is significantly lower than that of the true expected return. This feature reveals that in a standard regression, the economic significance interpretation of the coefficient is no longer appropriate. The coefficient is downward (upward) biased when the ICC is an independent (dependent) variable. Next, I extend the work from Wang (2015) by showing analytically that the measurement errors in analyst forecasts can result in the estimated ICC being contaminated. In the

simulation, the spurious regressions become a real issue when analyst forecasts are biased and/or unable to capture the full information of the cash flow expectation incorporated in market price.

5.3 FUTURE RESEARCH

This thesis opens several future research avenues. As the studies considered here only focus on a set of major international markets, i.e. the G7 countries, it would be interesting to see whether or not the negative risk premium continues to exist in other countries as well as in the global context. Such testing would provide evidence with respect to the external validity of the established results.

Next, because the aggregate ICC is an excellent proxy for tracking the market time varying expected return, future work can find other applications utilising this estimate. For example, paying attention to Figure 2.3, it can be seen that the risk premiums in major European countries tend to mirror each other, especially since the formation of the European Monetary Union. Moreover, the risk premium in these markets has increased ever since the event. Investigating how such policy affects the underlying risk premium at the individual country level, as well as the implications for market integration, is worthy of future study.

It would be beneficial to expand our understanding with respect to the time varying ICC at the firm level. However, caution is warranted. As highlighted in the third study, spurious regressions due to measurement errors in analyst forecasts can become a problem. Accordingly, searching for a solution to these measurement errors might prove rewarding, given the attractiveness of the forward-looking ICC.

Finally, as researchers are exposed to increasingly large amounts of data, further applications of the PCA and BRT are a promising research avenue. Due to the superior ability in handling high dimensional data and complex data generating structures, approximating high moments of expectation and examining implications for ex-ante asset pricing models with these techniques is a potential solution to the growing dissatisfaction with the traditional instrumental method. In so doing, the construct validity of established results in the literature can be reassessed.

REFERENCES

- Andrews, D. W., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica: Journal of the Econometric Society* 817-858.
- Ang, A., and G. Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651-707.
- Ang, A., and J. U. N. Liu, 2004, How to discount cashflows with time-varying expected returns, *The Journal of Finance* 59, 2745-2783.
- Bansal, R., and A. Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *The Journal of Finance* 59, 1481-1509.
- Black, F., 1986, Noise, *The Journal of Finance* 41, 529-543.
- Botosan, C. A., M. A. Plumlee, and H. E. Wen, 2011, The relation between expected returns, realized returns, and firm risk characteristics*, *Contemporary Accounting Research* 28, 1085-1122.
- Boudoukh, J., R. Michaely, M. Richardson, and M. R. Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *The Journal of Finance* 62, 877-915.
- Boudoukh, J., M. Richardson, and T. Smith, 1993, Is the ex ante risk premium always positive?: A new approach to testing conditional asset pricing models, *Journal of Financial Economics* 34, 387-408.
- Boudoukh, J., M. Richardson, and R. F. Whitelaw, 1997, Nonlinearities in the relation between the equity risk premium and the term structure, *Management Science* 43, 371-385.
- Brandt, M., and L. Wang, 2007, Measuring the time-varying risk-return relation from the cross-section of equity returns, *Unpublished Working Paper, Duke University*.
- Brav, A., R. Lehavy, and R. Michaely, 2005, Using expectations to test asset pricing models, *Financial Management* 34, 31-64.
- Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen, 1984. *Classification and regression trees* (CRC press).
- Callen, J. L., and M. R. Lyle, 2014, The term structure of implied costs of equity capital, *Rotman School of Management Working Paper*.
- Campbell, J. Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-399.
- Campbell, J. Y., and J. H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *The Journal of Political Economy* 107, 205-251.

- Campbell, J. Y., and R. J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195-228.
- Campbell, J. Y., and S. B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509-1531.
- Campello, M., L. Chen, and L. Zhang, 2008, Expected returns, yield spreads, and asset pricing tests, *Review of Financial Studies* 21, 1297-1338.
- Chava, S., and A. Purnanandam, 2010, Is default risk negatively related to stock returns?, *Review of Financial Studies* 23, 2523-2559.
- Chen, H., J. Z. Chen, G. J. Lobo, and Y. Wang, 2011, Effects of audit quality on earnings management and cost of equity capital: Evidence from china*, *Contemporary Accounting Research* 28, 892-925.
- Chen, L., Z. Da, and X. Zhao, 2013, What drives stock price movements?, *Review of Financial Studies* 26, 841-876.
- Chen, L., and X. Zhao, 2009, Return decomposition, *Review of Financial Studies* 22, 5213-5249.
- Clark, T. E., and K. D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291-311.
- Cochrane, J. H., 2011, Presidential address: Discount rates, *The Journal of Finance* 66, 1047-1108.
- Cready, W. M., 2001, The separation theorem, investor “myopia”, and market prices: A discussion of “do institutional investors prefer near-term earnings over long-run value?”*, *Contemporary Accounting Research* 18, 247-256.
- Croce, M. M., M. Lettau, and S. C. Ludvigson, 2015, Investor information, long-run risk, and the term structure of equity, *Review of Financial Studies* 28, 706-742.
- Diether, K. B., C. J. Malloy, and A. Scherbina, 2002, Differences of opinion and the cross section of stock returns, *The Journal of Finance* 57, 2113-2141.
- Easton, P., 2009. *Estimating the cost of capital implied by market prices and accounting data* (Now Publishers Inc).
- Easton, P. D., and S. J. Monahan, 2005, An evaluation of accounting-based measures of expected returns, *The Accounting Review* 80, 501-538.
- Easton, P. D., and G. A. Sommers, 2007, Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts, *Journal of Accounting Research* 45, 983-1015.
- Ecker, F., J. Francis, P. Olsson, and K. Schipper, 2013, Associations between realized returns and risk proxies using non-random samples, (Working Paper).
- Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004, Least angle regression, 407-499.
- Elton, E. J., 1999, Presidential address: Expected return, realized return, and asset pricing tests, *The Journal of Finance* 54, 1199-1220.

- Faff, R. W., D. Hillier, and J. Hillier, 2000, Time varying beta risk: An analysis of alternative modelling techniques, *Journal of Business Finance & Accounting* 27, 523-554.
- Fama, E. F., and K. R. French, 2002, The equity premium, *The Journal of Finance* 57, 637-659.
- Fama, E. F., and G. W. Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115-146.
- Ferson, W. E., S. Sarkissian, and T. Simin, 2008, Asset pricing models with conditional betas and alphas: The effects of data snooping and spurious regression, *Journal of Financial and Quantitative Analysis* 43, 331-353.
- Foster, F. D., T. O. M. Smith, and R. E. Whaley, 1997, Assessing goodness-of-fit of asset pricing models: The distribution of the maximal r^2 , *The Journal of Finance* 52, 591-607.
- Friedman, J. H., 2001, Greedy function approximation: A gradient boosting machine, *The Annals of Statistics* 1189-1232.
- Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan, 2001, Toward an implied cost of capital, *Journal of Accounting Research* 39, 135-176.
- Ghysels, E., 1998, On stable factor structures in the pricing of risk: Do time-varying betas help or hurt?, *The Journal of Finance* 53, 549-573.
- Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, There is a risk-return trade-off after all, *Journal of Financial Economics* 76, 509-548.
- Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance* 48, 1779-1801.
- Guay, W., S. Kothari, and S. Shu, 2011, Properties of implied cost of capital using analysts' forecasts, *Australian Journal of Management* 36, 125-149.
- Guo, H. U. I., and R. F. Whitelaw, 2006, Uncovering the risk–return relation in the stock market, *The Journal of Finance* 61, 1433-1463.
- Hansen, L. P., and S. F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587-613.
- Hansen, L. P., and K. J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269-1286.
- Harrison, J. M., and M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381-408.
- Harvey, C. R., 1988, The real term structure and consumption growth, *Journal of Financial Economics* 22, 305-333.
- Harvey, C. R., 2001, The specification of conditional expectations, *Journal of Empirical Finance* 8, 573-637.

- Hastie, T., J. Friedman, and R. Tibshirani, 2009. *The elements of statistical learning* (Springer New York).
- Hou, K., M. A. Van Dijk, and Y. Zhang, 2012, The implied cost of capital: A new approach, *Journal of Accounting and Economics* 53, 504-526.
- Hwang, L.-S., W.-J. Lee, S.-Y. Lim, and K.-H. Park, 2013, Does information risk affect the implied cost of equity capital? An analysis of pin and adjusted pin, *Journal of Accounting and Economics* 55, 148-167.
- Jagannathan, R., and Z. Wang, 1996, The conditional capm and the cross-section of expected returns, *The Journal of Finance* 51, 3-53.
- Kelly, B., and S. Pruitt, 2013, Market expectations in the cross-section of present values, *The Journal of Finance* 68, 1721-1756.
- Lambert, R. A., 2009, Discussion of “on the relation between expected returns and implied cost of capital”, *Review of Accounting Studies* 14, 260-268.
- Lee, C., D. Ng, and B. Swaminathan, 2009, Testing international asset pricing models using implied costs of capital, *Journal of Financial and Quantitative Analysis* 44, 307.
- Lee, C., E. C. So, and C. C. Wang, 2014, Evaluating firm-level expected-return proxies, *Harvard Business School Accounting & Management Unit Working Paper*.
- Lewellen, J., and S. Nagel, 2006, The conditional capm does not explain asset-pricing anomalies, *Journal of Financial Economics* 82, 289-314.
- Li, Y., D. T. Ng, and B. Swaminathan, 2013, Predicting market returns using aggregate implied cost of capital, *Journal of Financial Economics* 110, 419-436.
- Lucas, R. E., Jr., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429-1445.
- Ludvigson, S. C., and S. Ng, 2007, The empirical risk–return relation: A factor analysis approach, *Journal of Financial Economics* 83, 171-222.
- Lundblad, C., 2007, The risk return tradeoff in the long run: 1836–2003, *Journal of Financial Economics* 85, 123-150.
- McCloskey, D. N., and S. T. Ziliak, 1996, The standard error of regressions, *Journal of Economic Literature* 34, 97-114.
- Merton, R. C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323-361.
- Naiker, V., F. Navissi, and C. Truong, 2013, Options trading and the cost of equity capital, *The Accounting Review* 88, 261-295.
- Newey, W. K., and K. D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

- Newey, W. K., and K. D. West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631-653.
- Ng, S., 2014, Viewpoint: Boosting recessions, *Canadian Journal of Economics/Revue canadienne d'économique* 47, 1-34.
- Ostdiek, B., 1998, The world ex ante risk premium: An empirical investigation, *Journal of International Money and Finance* 17, 967-999.
- Pastor, L., M. Sinha, and B. Swaminathan, 2008, Estimating the intertemporal risk–return tradeoff using the implied cost of capital, *The Journal of Finance* 63, 2859-2897.
- Petersen, M. A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435-480.
- Pettenuzzo, D., A. G. Timmermann, and R. I. Valkanov, 2013, Forecasting stock returns under economic constraints.
- Rapach, D. E., J. K. Strauss, and G. Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821-862.
- Roll, R., 1977, A critique of the asset pricing theory's tests part i: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129-176.
- Rossi, A., and A. Timmermann, 2010, What is the shape of the risk-return relation?, *In AFA 2010 Atlanta Meetings Paper*.
- Sharpe, W. F., 1978, Discussion, *The Journal of Finance* 33, 917-920.
- Walsh, K., 2014, The investment horizon and asset pricing models, *Australian Journal of Management* 40, 277-294.
- Wang, C. C., 2015, Measurement errors of expected returns proxies and the implied cost of capital, *Harvard Business School Accounting & Management Unit Working Paper*.
- Welch, I., and A. Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455-1508.
- Wheatley, S. M., 1989, A critique of latent variable tests of asset pricing models, *Journal of Financial Economics* 23, 325-338.
- Whitelaw, R., 2000, Stock market risk and return: An equilibrium approach, *Review of Financial Studies* 13, 521-547.
- Wolak, F. A., 1989, Testing inequality constraints in linear econometric models, *Journal of Econometrics* 41, 205-235.

Appendix A

A.1 Correlation Matrix (Chapter 2)

Table A.1.1: The Correlation Matrix between the Implied Risk Premium and the Realised Risk Premium for Non-US Countries

This table reports the correlation matrix among measures of risk premium computed by the Implied Cost of Capital methodology, denoted as IRP and ex-post realised returns, denoted as RRP, of the non-US countries (UK, Canada, Italy, France, Germany, and Japan) from 1990 to 2012. At firm level, the implied cost of capital is derived from analyst earnings forecasts. The market-wide ICC is the value-weighted average of individual firms at monthly frequency. The ex-post market realised returns are value-weighted returns including dividends at monthly frequency. The risk premium is computed by subtracting the yield of 10-year Treasury Bonds (7-year for Italy) from the market-wide ICC or the market ex-post realised returns. Superscripts *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

| | IRP_{Canada} | $IRP_{Germany}$ | IRP_{France} | IRP_{UK} | IRP_{Japan} | IRP_{Italy} | RRP_{Canada} | $RRP_{Germany}$ | RRP_{France} | RRP_{UK} | RRP_{Japan} | RRP_{Italy} |
|-----------------|----------------|-----------------|----------------|------------|---------------|---------------|----------------|-----------------|----------------|------------|---------------|---------------|
| IRP_{Canada} | 1.00 | | | | | | | | | | | |
| $IRP_{Germany}$ | 0.77*** | 1.00 | | | | | | | | | | |
| IRP_{France} | 0.82*** | 0.96*** | 1.00 | | | | | | | | | |
| IRP_{UK} | 0.86*** | 0.93*** | 0.95*** | 1.00 | | | | | | | | |
| IRP_{Japan} | 0.69*** | 0.92*** | 0.90*** | 0.88*** | 1.00 | | | | | | | |
| IRP_{Italy} | 0.70*** | 0.90*** | 0.91*** | 0.88*** | 0.90*** | 1.00 | | | | | | |
| RRP_{Canada} | 0.07 | 0.12** | 0.13** | 0.10* | 0.05 | 0.12** | 1.00 | | | | | |
| $RRP_{Germany}$ | 0.13** | 0.20*** | 0.18*** | 0.15** | 0.09 | 0.16*** | 0.84*** | 1.00 | | | | |
| RRP_{France} | 0.04 | 0.13** | 0.10 | 0.08 | 0.02 | 0.10* | 0.85*** | 0.95*** | 1.00 | | | |
| RRP_{UK} | 0.19*** | 0.21*** | 0.18*** | 0.17*** | 0.06 | 0.09 | 0.75*** | 0.87*** | 0.86*** | 1.00 | | |
| RRP_{Japan} | 0.28*** | 0.35*** | 0.33*** | 0.31*** | 0.28*** | 0.26*** | 0.67*** | 0.67*** | 0.68*** | 0.66*** | 1.00 | |
| RRP_{Italy} | 0.04 | 0.01 | 0.02 | 0.02 | -0.03 | 0.02 | 0.74*** | 0.83*** | 0.86*** | 0.78*** | 0.63*** | 1.00 |

A.2 Least Angle Regression (LAR) in Chapter 3

LAR, proposed by Efron, et al. (2004), is an efficient model selection mechanism that allows researchers to choose the best linear subset among potential conditioning variables in predicting, in this case, stock returns. Traditionally, researchers can self-select and report the best subset and disregard the effort of mining the data, so data snooping bias becomes an issue. Additionally, the computation quickly becomes infeasible when the number of predictors becomes large.³⁶ Instead, LAR algorithmically builds the model through a sequence of steps, starting with a zero active set. At the first step, it adds the variable that mostly correlates with the dependent variable to the active set; and then it moves the coefficient on that predictor towards its least squares' coefficient, until another predictor has as much correlation with the current residual as the first predictor does. This is when the second predictor is added to the active set. The procedure continues in the same fashion, but now the coefficients on the active predictors move in their joint least squares' direction, until the one independent variable has as much correlation with the current residual. The process can be stopped until the pre-specified number of predictors enters the active set, or the model eventually includes all predictors in which the solution is the ordinary least squares.³⁷

A.3 Proof of Equations in Chapter 4

A.3.1 Proof of Equation (4.7)

Note that by iterating equation (4.3), I obtain $E_t(g_{t+k}) = \phi_g^k g_t + \alpha_g \frac{1-\phi_g^k}{1-\phi_g}$. In addition, by definition $d_{t+k} = d_t + \sum_{i=1}^k g_{t+i}$, the LHS of equation (4.7) becomes:

³⁶ Suppose the number of potential predictors is p , the possible subsets are 2^p . Therefore, if $p = 15$ then researchers have 32768 models to choose from.

³⁷ Note that the coefficients on the predictors at the intermediate steps are not equal to the least square solution coefficients.

$$\begin{aligned}
\sum_{j=0}^{\infty} \rho^j E_t(d_{t+1+j}) &= \sum_{j=0}^{\infty} \rho_j \left(d_t + \sum_{i=1}^{j+1} g_{t+i} \right) \\
&= \sum_{j=0}^{\infty} \rho_j \left(d_t + \sum_{i=1}^{j+1} \left(\phi_g^k g_t + \alpha_g \frac{1 - \phi_g^k}{1 - \phi_g} \right) \right)
\end{aligned} \tag{A.3.1}$$

At this step, I use MATHEMATICA ® functions Sum and Coefficient to expand and to collect terms for d_t , α_g , and g_t respectively.³⁸ It is then simple to obtain the expression as in equation (4.7).

A.3.2 Proof of Equation (4.8)

Iterating (4.4), (4.5), and (4.6), I express the conditional expected return:

$$\begin{aligned}
\mu_{t+j} &= E_t(r_{t+1+j}) \\
&= \sum_{i=0}^{j-1} \phi_\beta^i \left(\sum_{l=0}^{j-1} \alpha_\lambda \phi_\lambda^l + \phi_\lambda^j \lambda_t \right) \alpha_\beta + \left(\sum_{i=0}^{j-1} \phi_\lambda^i \phi_\beta^j \beta_t \right) \alpha_\lambda \\
&\quad + \phi_\lambda^j \phi_\beta^j \beta_t \lambda_t
\end{aligned} \tag{A.3.2}$$

Substitute into the LHS of equation (4.8):

$$\begin{aligned}
&\sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}) \\
&= \sum_{j=0}^{\infty} \rho^j \left(\sum_{i=0}^{j-1} \phi_\beta^i \left(\sum_{l=0}^{j-1} \alpha_\lambda \phi_\lambda^l + \phi_\lambda^j \lambda_t \right) \alpha_\beta \right. \\
&\quad \left. + \left(\sum_{i=0}^{j-1} \phi_\lambda^i \phi_\beta^j \beta_t \right) \alpha_\lambda + \phi_\lambda^j \phi_\beta^j \beta_t \lambda_t \right)
\end{aligned} \tag{A.3.3}$$

Similarly, Sum and Coefficient functions are used to expand and collect terms for $\beta_t \lambda_t$, β_t , λ_t , α_λ , and α_β , respectively. The closed form expression as in equation (4.8) follows.

A.3.3 Proof of Equation (4.12)

To prove equation (4.12), I substitute $\bar{g} = \frac{\alpha_g}{1 - \phi_g}$ into equation (4.11):

³⁸ The codes are available upon request.

$$\begin{aligned} \sum_{j=0}^{\infty} E_t(d_{t+1+j}) &= \sum_{j=0}^{T-1} \rho^j (d_t + \sum_{i=1}^{j+1} g_{t+i}) \\ &+ \sum_{j=1}^{\infty} \rho^{T+j-1} \left(d_{t+T} + j \frac{\alpha_g}{1-\phi_g} \right) \end{aligned} \quad (\text{A.3.4})$$

Similarly, it is straightforward to expand and collect terms for d_t , α_g , and g_t . Additionally, to simplify the expressions in each term, I use the FullSimplify function. Equation (4.12) is then obtained.

A.3.4 Proof of Equation (4.24)

Campbell and Shiller (1988) provide an approximation for $r_{t+1} = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right)$:

$$r_{t+1} \approx k + \rho p_{t+1} + (1 - \rho)d_{t+1} - p_t \quad (\text{A.3.5})$$

Substituting the true price in equation (4.9) into p_{t+1} in (A.3.5) and gathering all the terms known at time t as a constant k_t , I can express the log future returns as

$$\begin{aligned} r_{t+1} \approx & k_t + d_{t+1} + \frac{\rho\phi_g g_{t+1}}{1-\rho\phi_g} \\ & - \rho \left(\frac{\beta_t \lambda_t}{1-\rho\phi_\beta\phi_\lambda} + \left(\frac{\rho\alpha_\lambda\phi_\beta}{(1-\rho\phi_\beta)(1-\rho\phi_\beta\phi_\lambda)} \right) \beta_t \right. \\ & + \left(\frac{\rho\alpha_\beta\phi_\lambda}{(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \right) \lambda_t \\ & \left. + \frac{\rho\alpha_b\alpha_\lambda(1-\rho^2\phi_b\phi_\lambda)}{(1-\rho)(1-\rho\phi_\beta)(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \right) \end{aligned} \quad (\text{A.3.6})$$

Taking variance of (A.3.6) and assuming the covariance between cash flows g_t and expected returns μ_t are 0, I approximate the return variance as:

$$\begin{aligned} \text{Var}_t(r_{t+1}) &= \sigma_{g,t+1}^2 + \frac{\rho^2\phi_g^2\sigma_{g,t+1}^2}{(1-\rho\phi_g)} \\ &+ \text{Var}_t \left[-\rho \left(\frac{\beta_t \lambda_t}{1-\rho\phi_\beta\phi_\lambda} + \left(\frac{\rho\alpha_\lambda\phi_\beta}{(1-\rho\phi_\beta)(1-\rho\phi_\beta\phi_\lambda)} \right) \beta_t \right. \right. \\ &+ \left. \left(\frac{\rho\alpha_\beta\phi_\lambda}{(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \right) \lambda_t \right. \\ &+ \left. \left. \frac{\rho\alpha_b\alpha_\lambda(1-\rho^2\phi_b\phi_\lambda)}{(1-\rho)(1-\rho\phi_\beta)(1-\rho\phi_\lambda)(1-\rho\phi_\beta\phi_\lambda)} \right) \right] \\ &+ \frac{2\rho\phi_g}{(1-\rho\phi_g)} \sigma_{g,t+1}^2 \end{aligned} \quad (\text{A.3.7})$$

Note that in $\text{Var}_t(\beta_{t+1}) = \sigma_\lambda^2$, $\text{Var}_t(\lambda_{t+1}) = \sigma_\lambda^2$, $\text{Var}_t(\beta_{t+1}\lambda_{t+1}) = (\alpha_\lambda + \phi_\lambda\lambda_t)^2\sigma_\beta^2 + (\alpha_\beta + \phi_\beta\beta_t)^2\sigma_\lambda^2 + \sigma_\beta^2\sigma_\lambda^2$, $\text{Cov}_t(\beta_{t+1}, \lambda_{t+1}) = 0$, $\text{Cov}_t(\beta_{t+1}, \beta_{t+1}\lambda_{t+1}) = (\alpha_\lambda + \phi_\lambda\lambda_t)\sigma_\beta^2$, and

$Cov_t(\lambda_{t+1}, \beta_{t+1} \lambda_{t+1}) = (\alpha_\beta + \phi_\beta \beta_t) \sigma_\lambda^2$, I expand and collect terms for $\sigma_\lambda^2 \sigma_b^2$, σ_λ^2 , and σ_b^2 in the third term of (A.3.7). Subsequently, I simplify (A.3.7) to get equation (4.24).