ESSAYS IN ASSET PRICING

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF HUMANITIES

2017

Liu Liu
Alliance Manchester Business School
# Contents

Abstract 10

Declaration 11

Copyright Statement 12

Acknowledgements 13

1 Introduction 15

1.1 Motivation ................................. 15

1.2 Thesis Structure ............................ 20

2 The Yen Risk Premiums: A Story of Regime Shifts in Bond Markets 24

2.1 Introduction ................................ 25

2.2 Related Literature ............................ 28

2.3 The Model ................................. 29

2.3.1 Regime switching dynamic term structure model (RS-DTSM) 30

2.3.2 Exchange rates and exchange rate risk premiums ............... 33

2.4 Model Estimation ............................ 36

2.4.1 Minimum-chi-square estimation of RS-DTSM ................. 36
2.4.2 Identification of RS-DTSM ........................................... 38
2.4.3 Exchange rate risk premium regressions ......................... 40
2.5 Empirical Analysis ....................................................... 42
  2.5.1 Bond yield and foreign exchange data .......................... 42
  2.5.2 Latent factors ....................................................... 44
  2.5.3 Regime probabilities .............................................. 46
  2.5.4 Exchange rate risk premium regressions ......................... 49
2.6 Concluding Remarks ................................................... 53

Bibliography ................................................................. 56

Appendices ...................................................................... 79
  2.A Bond Prices ............................................................. 79
  2.B Price of Factor Risk .................................................... 80
  2.C Reduced-Form Models ............................................... 80

3 Correcting Estimation Bias in Regime Switching Dynamic Term Structure Models ...................................................... 86
  3.1 Introduction ............................................................... 87
  3.2 Related Literature ...................................................... 89
  3.3 The Model ................................................................. 91
    3.3.1 Latent factors ...................................................... 92
    3.3.2 Bond yields and forward rates ................................. 93
    3.3.3 The pricing kernel ............................................... 95
  3.4 Model Estimation and Estimation Bias .............................. 96
4.4 Results: Conditional Moments ........................................ 155
  4.4.1 Distorted probability .............................................. 155
  4.4.2 Conditional moments .............................................. 156
  4.4.3 Comparative statics .............................................. 159
4.5 Simulation Results ..................................................... 161
  4.5.1 Simulated probability .............................................. 161
  4.5.2 Before and at the onset of recessions .......................... 162
4.6 Concluding Remarks ................................................... 166
Bibliography ................................................................. 168

Appendices .................................................................. 192
  4.A Value Function ....................................................... 192
  4.B P/D Ratio .............................................................. 195

5 Conclusions ............................................................... 197

Word count 32,163
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Summary statistics of bond yield data</td>
<td>60</td>
</tr>
<tr>
<td>2.2</td>
<td>Summary statistics of foreign exchange data</td>
<td>61</td>
</tr>
<tr>
<td>2.3</td>
<td>Parameter estimates in the US term structure model</td>
<td>62</td>
</tr>
<tr>
<td>2.4</td>
<td>Parameter estimates in the Japanese term structure model</td>
<td>63</td>
</tr>
<tr>
<td>2.5</td>
<td>Summary statistics of transition probabilities and latent factors</td>
<td>64</td>
</tr>
<tr>
<td>2.6</td>
<td>Summary statistics of model-implied risk premiums</td>
<td>65</td>
</tr>
<tr>
<td>2.7</td>
<td>Risk premium regressions: ordinary least squares</td>
<td>66</td>
</tr>
<tr>
<td>2.8</td>
<td>Risk premium regressions: GARCH(1,1)</td>
<td>67</td>
</tr>
<tr>
<td>2.9</td>
<td>Risk premium regressions: robustness checks</td>
<td>68</td>
</tr>
<tr>
<td>2.C.1</td>
<td>Estimates of the US reduced-form term structure model</td>
<td>84</td>
</tr>
<tr>
<td>2.C.2</td>
<td>Estimates of the Japanese reduced-form term structure model</td>
<td>85</td>
</tr>
<tr>
<td>3.1</td>
<td>Summary statistics of bond yields</td>
<td>121</td>
</tr>
<tr>
<td>3.2</td>
<td>Estimates of reduced-form parameters, before bias correction</td>
<td>122</td>
</tr>
<tr>
<td>3.3</td>
<td>Estimates of structural parameters, before bias correction</td>
<td>123</td>
</tr>
<tr>
<td>3.4</td>
<td>Estimates of reduced-form parameters, after bias correction</td>
<td>124</td>
</tr>
<tr>
<td>3.5</td>
<td>Estimates of structural parameters, after bias correction</td>
<td>125</td>
</tr>
<tr>
<td>3.6</td>
<td>Parameter estimates in the procedure of bootstrap bias correction</td>
<td>126</td>
</tr>
</tbody>
</table>
3.7 Volatility of forward rates before and after bias correction . . . . . . . . 127
3.8 Simulation results: 6-month forward rates . . . . . . . . . . . . . . . . 128
3.9 Simulation results: 2-year forward rates . . . . . . . . . . . . . . . . . . 129
3.10 Simulation results: 5-year forward rates . . . . . . . . . . . . . . . . . 130
3.11 Simulation results: 10-year forward rates . . . . . . . . . . . . . . . . . 131
4.1 Model and preference parameters . . . . . . . . . . . . . . . . . . . . . 171
4.2 Moments of simulated probability . . . . . . . . . . . . . . . . . . . . . 172
4.3 Before and at the onset of recessions: $p_0 = 0.5$ . . . . . . . . . . . . 173
4.4 Before and at the onset of recessions: $p_0 = 0.9$ . . . . . . . . . . . . 174
4.5 Before and at the onset of recessions: $p_0 = 0.1$ . . . . . . . . . . . . 175
4.6 Long-sample moments: $p_0 = 0.5$ . . . . . . . . . . . . . . . . . . . . . . . . 176
4.7 Long-sample moments: $p_0 = 0.9$ . . . . . . . . . . . . . . . . . . . . . . . . 177
4.8 Long-sample moments: $p_0 = 0.1$ . . . . . . . . . . . . . . . . . . . . . . . . 178
List of Figures

2.1 Bond yield data ......................................................... 69
2.2 Foreign exchange data .............................................. 70
2.3 Forward premium and interest rate differentials .................... 71
2.4 US factor loadings of bond yields .................................. 72
2.5 Japanese factor loadings of bond yields ............................ 73
2.6 Model-implied latent factors ......................................... 74
2.7 US filtered, smoothed and transition probabilities .............. 75
2.8 Japanese filtered, smoothed and transition probabilities ........ 76
2.9 US risk premiums ...................................................... 77
2.10 Japanese risk premiums .............................................. 78

3.1 Bond yield data ......................................................... 132
3.2 Filtered and smoothed probabilities .................................. 133
3.3 Factor loadings of bond yields ...................................... 134
3.4 6-month forward rates before and after bias correction .......... 135
3.5 2-year forward rates before and after bias correction .......... 136
3.6 5-year forward rates before and after bias correction .......... 137
3.7 10-year forward rates before and after bias correction .......... 138
4.1 Distorted probability of the long-recession parameter value . . . . . . 179
4.2 Effects of parameter learning on the pricing kernel . . . . . . . . . . 180
4.3 Effects of parameter learning on equity returns . . . . . . . . . . . . 181
4.4 Effects of parameter learning on volatilities . . . . . . . . . . . . . . 182
4.5 Effects of ambiguity aversion coefficient on the pricing kernel . . . . 183
4.6 Effects of ambiguity aversion coefficient on equity returns . . . . . . 184
4.7 Effects of ambiguity aversion coefficient on volatilities . . . . . . . . 185
4.8 Effects of risk aversion coefficient on the pricing kernel . . . . . . . . 186
4.9 Effects of risk aversion coefficient on equity returns . . . . . . . . . . 187
4.10 Effects of risk aversion coefficient on volatilities . . . . . . . . . . . . 188
4.11 Effects of ambiguity and risk aversion coefficients on the pricing kernel 189
4.12 Effects of ambiguity and risk aversion coefficients on equity returns . . 190
4.13 Effects of ambiguity and risk aversion coefficients on volatilities . . . . 191
This thesis improves our understanding of asset prices and returns as it documents a regime shift risk premium in currencies, corrects the estimation bias in the term premium of bond yields, and shows the impact of ambiguity aversion towards parameter uncertainty on equities. The thesis consists of three essays.

The first essay “The Yen Risk Premiums: A Story of Regime Shifts in Bond Markets” documents a new monetary mechanism, namely the shift of monetary policies, to account for the forward premium puzzle in the USD-JPY currency pair. The shift of monetary policy regimes is modelled by a regime switching dynamic term structure model where the risk of regime shifts is priced. Our model estimation characterises two policy regimes in the Japanese bond market—a conventional monetary policy regime and an unconventional policy regime of quantitative easing. Using foreign exchange data from 1985 to 2009, we find that the shift of monetary policies generates currency risk: the yen excess return is predicted by the Japanese regime shift premium, and the emergence of the yen carry trade in the mid 1990s is associated with the transition from the conventional to the unconventional monetary policy in Japan.

The second essay “Correcting Estimation Bias in Regime Switching Dynamic Term Structure Models” examines the small sample bias in the estimation of a regime switching dynamic term structure model. Using US data from 1971 to 2009, we document two regimes driven by the conditional volatility of bond yields and risk factors. In both regimes, the process of bond yields is highly persistent, which is the source of estimation bias when the sample size is small. After bias correction, the inference about expectations of future policy rates and long-maturity term premia changes dramatically in two high-volatility episodes: the 1979–1982 monetary experiment and the recent financial crisis. Empirical findings are supported by Monte Carlo simulation, which shows that correcting small sample bias leads to more accurate inference about expectations of future policy rates and term premia compared to before bias correction.

The third essay “Learning about the Persistence of Recessions under Ambiguity Aversion” incorporates ambiguity aversion into the process of parameter learning and assess the asset pricing implications of the model. Ambiguity is characterised by the unknown parameter that governs the persistence of recessions, and the representative investor learns about this parameter while being ambiguity averse towards parameter uncertainty. We examine model-implied conditional moments and simulated moments of asset prices and returns, and document an uncertainty effect that characterises the difference between learning under ambiguity aversion and learning under standard recursive utility. This uncertainty effect is asymmetric across economic expansions and recessions, and this asymmetry generates in simulation a sharp increase in the equity premium at the onset of recessions, as in the recent financial crisis.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s Policy on Presentation of Theses.
Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Stuart Hyde, Dr Sungjun Cho and Dr Hening Liu, for their continuous support of my PhD study and for their patience, motivation and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. Professor Stuart Hyde inspired my journey in asset pricing. He trained me hard in programming at the beginning of my PhD, from which I benefited indefinitely in the following years. He also takes great care of my personal development as a researcher—I still remember his talk about business cycles in conducting research, a fundamental concept in my research while not willingly accepted in life. Dr. Sungjun Cho usually spends hours discussing models with me. He asks questions and comments on all the necessary details. Several times he helped check my codes line by line—I was relieved that no errors were found, and I always benefited greatly from his suggestions on improving programming skills. Dr. Hening Liu guided me into the world of theoretical asset pricing. It is always interesting to talk to him about recent research, and I am inspired by his passion for doing research.

I would like to thank Professor Massimo Guidolin, who was my supervisor in the first two years of my PhD before leaving Manchester. During these two years, he raised the technical level of my research, and emailed me whenever he had any good idea that would improve my work.

Besides my supervisors, I would like to thank the panel members in my mid-year and annual reviews: Dr. Kevin Aretz, Professor Michael Bowe, Professor Ian Garrett, Dr. Olga Kolokolova, Dr. Alex Kostakis, Professor Ser-Huang Poon, Professor Richard Stapleton, Dr. Alex Taylor, for their insightful comments, encouragement and also the hard questions. I express my gratitude to my thesis examiners, Dr. Alex Kostakis and Professor Neil Kellard, for their insightful and constructive questions and comments.
I am deeply indebted to my father, Endong Liu, and my mother, Lijia Liu, for their unconditional love and for their faith in me. I would not be the person I am today without their constant patience, love and support. Being a scholar of ancient Chinese and English literature, my father shows to me what lifelong learning is before I know these words. I receive a gift for mathematics and science from my mother, who is also an optimist and helps me look at the positive side as I tend to be pessimistic.

I would like to thank Dr Maria Marchica, Dr. Jingya Wang, Ben Tang, Ryan White, Dr. Yomna Ahmed Yousif and my PhD colleagues at the Manchester Accounting and Finance Group for their encouragement and help.

Finally, I would like to acknowledge the financial support from the President’s Doctoral Scholar Award of the University of Manchester and the Alliance MBS PhD Hardship funding; and the technical support from the Business Data Service of the University of Manchester Library and High Throughput Computing using Condor at Manchester.
Chapter 1

Introduction

1.1 Motivation

Uncertainty is a fundamental theme in the field of asset pricing. As Campbell, Lo and MacKinlay put it:

What distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation. The starting point for every financial model is the uncertainty facing investors, and the substance of every financial model involves the impact of uncertainty on the behavior of investors and, ultimately, on market prices. . . . The random fluctuations that require the use of statistical theory to estimate and test financial models are intimately related to the uncertainty on which those models are based. (Campbell, Lo and MacKinlay, 1997, pp. 3)

Along the lines of their arguments, my thesis, containing three essays, separately examines three uncertainties that have an impact on the determination or the estimation of asset prices and returns, namely the uncertainty over monetary policy regime shifts, the statistical bias (uncertainty in its broadest sense) in parameter estimates, and the uncertainty over the persistence of recessions. In order to capture the uncertainty over monetary policy regime shifts in the first essay, I employ the regime switching dynamic term structure model of Dai, Singleton and Yang (2007), and examine its implications
CHAPTER 1. INTRODUCTION

for currency excess returns. In the empirical estimation of the model, concerns arise as it is well documented in the literature that the maximum likelihood estimates of affine term structure models suffer from small sample bias, while the magnitude of any estimation bias in regime switching term structure models is unclear. This concern is addressed in the second essay, where I study the estimation bias in a variant of Dai, Singleton and Yang (2007): instead of using their specification that the transition probability of regimes under the physical measure is time-varying, I assume that it is constant. With this simplified specification of regimes, I am able to focus on the estimation bias in the process of bond yields while also dealing with the estimation bias in the persistence of regimes. However, I did not correct the estimation bias in the first essay as the specification of time-varying transition probabilities imposes additional computational burden in the simulation-based bias correction procedure, and I leave it to future work. While the three essays work with various risk premiums, the first two essays are fundamentally different from the third one in the nature of uncertainties. The regime risk premium and the factor risk premium in the first essay and the term premium in the second essay compensate for the uncertainties over random outcomes where the distribution of random outcomes is known. In this case, uncertainty refers to pure risk. In contrast, the uncertainty over the persistence of recessions in the third essay describes a situation where the distribution of random outcomes is unknown, and in this situation, uncertainty refers to ambiguity rather than risk. While I frequently use the term “uncertainty” throughout the thesis, it refers to risk in the first two essays and refers to ambiguity in the third essay, and I also distinguish between risk and ambiguity in the third essay.

The first essay is an empirical study, showing that the uncertainty over monetary policy regime shifts implied from the international bond markets is priced in the foreign exchange market. The theoretical setup of this essay is based on the seminal work of Backus, Foresi and Telmer (2001), who derive, under the assumption of no-arbitrage, the relationship between exchange rates and the pricing kernel in term structure models for each currency. In the case of regime shifts where regimes are assumed to be independent across countries, this no-arbitrage relationship implies that any shocks to the regime in the bond markets are absorbed by exchange rates. As a consequence, it is predicted that investors require a positive risk premium on taking a short forward
position in a currency if, in its home country, there is greater uncertainty over monetary policy shifts.

I study the USD-JPY currency pair and employ the regime switching dynamic term structure model of Dai, Singleton and Yang (2007) where the uncertainty of regime shifts is priced. I estimate the term structure model for US and Japanese bond yields independently, and highlight the regimes identified from the Japanese bond market: a conventional monetary policy regime that prevails the sample period and an unconventional policy regime of quantitative easing from 2001 to 2006. Moreover, there is substantial regime uncertainty during the transition from the conventional to the unconventional policy regime from 1995 to 2000, implying substantial risk in Japanese yen, and it is in this period that the famous yen carry trade emerges. This observed link is supported by risk premium regressions, and is consistent with the prediction—investors require a compensation in the yen carry trade (equivalent to taking a short forward position in the low interest rate yen) for the substantial uncertainty over monetary policy shifts in Japan.

This essay contributes to the literature by documenting a new monetary mechanism, namely monetary policy regime shifts, to explain currency excess returns, as the extant literature concentrates on the monetary channels based on Taylor (1993) rules or traditional monetary models of exchange rates (see Brennan and Xia, 2004; Dong, 2006; Sarno, Schneider and Wagner, 2012). The finding that the uncertainty over monetary policy shifts is priced in the foreign exchange market is also of fundamental importance to monetary authorities as the interaction between monetary policies and currency excess returns is the starting point for the understanding of more complex systems, e.g., the interaction between monetary policies and the stabilising or destabilising effects of carry trade speculation as studied by Plantin and Shin (2011).

The second essay examines the small sample bias in the estimation of a regime switching dynamic term structure model. Many empirical studies document regime switching behaviour in interest rates (see Ang and Timmermann, 2012 for a survey). Landén (2000), Dai and Singleton (2003) and Evans (2003) propose regime switching dynamic term structure models that have closed-form solutions for zero-coupon bond prices. While this class of models offers fascinating insights into the interaction of bond
yields, risk factors and changes in business cycle conditions and monetary policies, problems arise as the maximum likelihood (ML) estimates of the model potentially suffer from two sources of bias: the bias in the estimation of vector autoregression models and the bias in the estimation of regimes. Both parallel the well-documented estimation bias in autoregression models where the ML estimates are biased towards a process that is less persistent than the true process, and the bias is particularly severe when the true process is highly persistent and the sample size is small. Unfortunately, in empirical studies, the sample size is not sufficiently big: the typical data sample dating back to the 1970s has only a few business cycles. Moreover, bond yields are highly persistent and the estimated regimes are usually very persistent. As a result, the persistence of bond yields and the persistence of regimes are likely to be underestimated. The downward bias in the estimated persistence of regimes distorts the inference about regimes, and the downward bias in the estimated persistence of bond yields distorts the inference about risk neutral rates and term premia.

Using US data from 1971 to 2009, I document two regimes driven by the conditional volatility of bond yields and risk factors. In both regimes, the process of bond yields is highly persistent, which is the source of estimation bias when the sample size is small. After bias correction, the inference about expectations of future policy rates and long-maturity term premia changes dramatically in two high-volatility episodes: the 1979–1982 monetary policy experiment and the recent financial crisis. It turns out that the exceptionally high long-term forward rates in the 1979–1982 episode and the exceptionally low rates in the recent financial crisis are to a great extent driven by the expectations of future policy rates. While the term premia are high in these two periods, they belong to the normal business-cycle variation of term premia in the full sample period, without being exceptionally high in the 1979–1982 episode or being counter-intuitively low in the recent financial crisis as in the case before bias correction. These findings are important to both researchers and policy makers, as one question central to the understanding of the dynamics of interest rates and risk factors is to what extent the variation in long-term bond yields is driven by the variation in expectations of future policy rates or the variation in term premia.
CHAPTER 1. INTRODUCTION

The third essay is a theoretical study, showing that learning about the uncertainty over the persistence of recessions under ambiguity aversion has important implications for the equity premium at the onset of recessions. In an economy with two regimes characterising business cycle fluctuations and labelled as “recession” and “expansion” respectively, uncertainty arises from the unknown parameter that governs the persistence of the recession regime. Learning about this parameter is difficult due to the less frequent nature of recessions. Collin-Dufresne, Johannes and Lochstoer (2016) show further that the uncertainty over the persistence of recessions leads to the biggest welfare loss and therefore has the largest asset pricing impact, compared to the uncertainty over the magnitude and the volatility of consumption shocks in recessions. In addition, it is well known that the equity premium is countercyclical (Campbell and Shiller, 1988a, 1988b; Fama and French, 1989). Recently, Martin (2016) estimates that in November 2008—the height of the recent financial crisis, the (monthly) annualised expected return on the US stock market peaked at 55.0%, and this is in sharp contrast to the Great Moderation period from 2004 to 2006 when the average expected return was 1.86%—the equity premium at the height of the crisis is 30 times as much as that during the Great Moderation at the one-month horizon and is 10 times as much at the one-year horizon. I show that incorporating ambiguity aversion into the process of learning about the persistence of recessions is able to reproduce in simulation the sharp increase in the equity premium at the onset of recessions.

In the benchmark model, the parameter that governs the persistence of recessions is unknown, and the representative investor learns about this parameter while being ambiguity averse towards parameter uncertainty. I assess the asset pricing implications of the model, and document an uncertainty effect that characterises the difference between learning under ambiguity aversion and learning under recursive preferences. Specifically, an ambiguity-averse investor requires higher compensation for parameter uncertainty compared to an ambiguity-neutral investor. More importantly, this uncertainty effect, driven by the distorted probability in the ambiguity model, is asymmetric across regimes—the size of the uncertainty effect is moderate in the expansion regime and is substantial in the recession regime. This asymmetry plays a key role in matching the sharp increase in the equity premium at the onset of recessions. Our simulation is designed such that after 100 years of learning—comparable length with the 1890–1994
sample of consumption data used in Cecchetti, Lam and Mark (2000)—the onset of
the next recession represents the height of the recent financial crisis and the preceding
period represents the Great Moderation period. At the annual frequency, our simula-
tion results show that at the onset of the next recession, ambiguity models are able to
generate 10–30 times the equity premium in the preceding period, and increasing the
magnitude of ambiguity aversion will further increase the ratio. By contrast, recur-
sive utility models generate 3–12 times the equity premium in the preceding period,
and increasing the magnitude of risk aversion does not help increase the ratio. In this
sense, ambiguity models offer desirable flexibility in matching the sharp increase in the
equity premium at the onset of recessions. This finding improves our understanding
of the countercyclical behaviour of equity premium, and proposes a new mechanism
that induces substantial increase in the equity premium at the onset of recessions, as
in the recent financial crisis.

1.2 Thesis Structure

The thesis structure follows the format accepted by the Manchester Accounting and
Finance Group, Alliance Manchester Business School at the University of Manchester.
It allows chapters to be incorporated into a format suitable for submission and pub-
lication in peer-reviewed academic journals. Therefore, this thesis is structured around
three essays containing original research in chapters 2, 3, and 4. The chapters are
self-contained, i.e., each chapter has a separate literature review, answers unique and
original questions, and employs distinct analysis. Page numbers, titles, and subtitles
have a sequential order throughout the thesis.

The remainder of the thesis proceeds as follows. Chapter 2 examines the effects
of the uncertainty over monetary policy regime shifts implied from the international
bond markets on the currency excess return in the foreign exchange market. Chapter 3
examines the small sample bias in the estimation of a regime switching dynamic term
structure model and the effects of bias correction on the inference about expectations of
future policy rates and term premia. Chapter 4 examines the implications of learning
about the uncertainty over the persistence of recessions under ambiguity aversion for
asset prices and returns, in particular, the equity premium at the onset of recessions. Chapter 5 concludes. In chapters 2–4, I use the first person plural (we, our) rather than the singular (I, my), as these chapters are in the form of working papers co-authored with my supervisors. Chapter 2 is co-authored with Dr. Sungjun Cho and Professor Stuart Hyde, Chapter 3 is co-authored with Dr. Sungjun Cho, and Chapter 4 is co-authored with Dr. Hening Liu.
Bibliography


Chapter 2

The Yen Risk Premiums: A Story of Regime Shifts in Bond Markets

Abstract

We document a new monetary mechanism, namely the shift of monetary policies, to account for the forward premium puzzle in the USD-JPY currency pair. The shift of monetary policy regimes is modelled by a regime switching dynamic term structure model where the risk of regime shifts is priced. Our model estimation characterises two policy regimes in the Japanese bond market—a conventional monetary policy regime and an unconventional policy regime of quantitative easing. Using foreign exchange data from 1985 to 2009, we find that the shift of monetary policies generates currency risk: the yen excess return is predicted by the Japanese regime shift premium, and the emergence of the yen carry trade in the mid 1990s is associated with the transition from the conventional to the unconventional monetary policy in Japan.
2.1 Introduction

Uncovered interest rate parity (UIP) postulates that the high interest rate currency is expected to depreciate in the next period by the same amount as the interest rate differential (or the forward premium due to covered interest rate parity). Equivalently, the excess return on taking a forward position, which is equal to the difference between the exchange rate change and the forward premium, is expected to be zero. Empirically, however, the high interest rate currency tends to appreciate, implying a positive carry trade return on taking a long forward position in the high interest rate currency (see Hansen and Hodrick, 1980; Bilson, 1981; Fama, 1984). The empirical violation of UIP is referred to as the forward premium puzzle.

A large theoretical and empirical literature has pointed out that the puzzle is able to be accounted for by time-varying risk premiums. For example, Lustig and Verdelhan (2007), Verdelhan (2010), Bansal and Shaliastovich (2013) and Zviadadze (2016) propose consumption-based explanations; Clarida, Davis and Pedersen (2009) and Menkhoff, Sarno, Schmeling and Schrimpf (2012) find that currency excess returns are related to volatilities; Brunnermeier, Nagel and Pedersen (2009), Burnside, Eichenbaum, Kleshchelski and Rebelo (2011), Dobrynskaya (2014) and Chernov, Graveline and Zviadzade (2016) provide empirical evidence that currency risk premiums compensate for jump, crash or downside risk; Dong (2006) and Backus, Gavazzoni, Telmer and Zin (2013) propose monetary policy explanations. In the recent literature, Lustig and Verdelhan (2007) sort currencies into portfolios based on interest rate differentials and document a cross sectional variation of portfolio returns. Lustig, Roussanov and Verdelhan (2011) identify two global (common) factors that explain currency portfolio returns: a level (dollar) factor where all currency portfolios have equal factor loadings, and a slope (carry) factor where factor loadings are asymmetric across portfolios. In this paper, we also account for the forward premium puzzle from a risk-based point of view, where currency risk premiums are implied from bond markets. Following the modelling strategy in the literature on the relationship between exchange rates and interest rates, we study a currency pair rather than using the portfolio approach of Lustig and Verdelhan (2007).
Specifically, we study the USD-JPY currency pair—the most heavily used currencies in carry trades—and account for the forward premium puzzle using the risk premium that compensates for the risk of regime shifts. Regime shifts are characterised by changes in business cycle conditions and/or monetary policies, and are estimated from US and Japanese bond markets. The theoretical setup of the link between bond yields and exchange rates is based in the seminal work of Backus, Foresi and Telmer (2001), who derive, under the assumption of no-arbitrage, the relationship between exchange rates and the pricing kernel in term structure models for each currency. In the case of regime shifts where regimes are assumed to be independent across countries, this no-arbitrage relationship implies that any shocks to the regime in the bond markets are absorbed by exchange rates. As a consequence, the regimes identified from the bond markets should help identify or explain the regimes in the foreign exchange market and therefore contain important information about exchange rate changes. Empirically, our research is motivated by the observation that the emergence of the yen carry trade (equivalent to taking a short forward position in the low interest rate yen) in the mid 1990s coincided with the end of a series of conventional monetary policies in Japan around the early 1990s and the start of the unconventional monetary policy of quantitative easing from 2001 to 2006, in support of a link between exchange rate excess returns and transitions in monetary policy regimes.

The postulate that the regimes in the bond markets contain important information about the variation in exchange rates is also guided by the empirical findings that regimes in interest rates are often linked to monetary policies (see Ang and Timmermann, 2012 for a survey), combined with the literature on the relationship between monetary policies and carry trade returns (see Dong, 2006; Plantin and Shin, 2011; Backus, Gavazzoni, Telmer and Zin, 2013). In the empirical studies that document regime switching behaviour in interest rates, regimes are mostly characterised by the difference in the volatility of interest rates, and are often linked to monetary policies: the Federal Reserve experiment from 1979 to 1982 (see, among many others, Hamilton, 1988; Gray, 1996; Ang and Bekaert, 2002; Dai, Singleton and Yang, 2007), the variables that drive monetary policy decisions—real rate, inflation, and output gap (see Garcia and Perron, 1996; Evans, 2003; Ang, Bekaert and Wei, 2008), and the response of monetary authorities (see Bikbov and Chernov, 2013; Ahn, Chib and
Kang, 2014). There are similar findings in Japanese data (see Clarida, Sarno, Taylor and Valente, 2003, 2006; Kimura and Nakajima, 2016). Through this monetary policy channel, interest rate regimes are linked to exchange rates by the literature on the relationship between monetary policies and carry trade returns. While not modelling regimes explicitly, Backus, Gavazzoni, Telmer and Zin (2013) point out asymmetries in the response of monetary authorities across countries to output gap and inflation—variables in Taylor (1993) rules. They postulate that a currency tends to appreciate if the monetary policy in this country responds more to output gap and less to inflation relative to its foreign counterpart.

Our research builds on the above arguments that interest rate regimes should contain important information about exchange rates through the monetary policy channel. We contribute to the literature by documenting a new monetary mechanism, namely the shift of monetary policies. Specifically, instead of utilising the Taylor-rule arguments, we postulate that the shift of policies (regimes) generates currency risk. To capture the risk of regime shifts, we employ the term structure model of Dai, Singleton and Yang (2007) where the risk of regime shifts is priced, and our work can be viewed as a direct extension of their model to the foreign exchange literature. We estimate the regime switching dynamic term structure model for the US and Japanese bond yields. Consistent with the extant literature, we find two regimes in the conditional volatility of the underlying latent factors—a low-volatility regime and a high-volatility regime. We highlight the volatility regime in the Japanese bond market and its implications for currency excess returns. In Japan, the high volatility regime that prevails the sample period is interpreted as a conventional monetary policy regime, and the low volatility regime from 2001 to 2006 is interpreted as an unconventional policy regime—consistent with the finding in Kimura and Nakajima (2016). Our model reveals substantial regime uncertainty during the transition from the conventional to the unconventional policy regime from 1995 to 2000, implying substantial risk in Japanese yen, and it is in this period that the yen carry trade emerges. In August 1995, the Japanese yen depreciated against the US dollar by 10.24%—the largest depreciation of yen during the sample period from 1985 to 2009. In the Japanese bond market, from June to July, the risk premium that compensates for the risk of regime shifts increased more than ten times—exhibiting a clear signal for the depreciation in the following
month. By contrast, the forward premium was silent about this depreciation—with no significant changes in the premium between June and August.

The remainder of the paper proceeds as follows. Section 2.2 reviews related literature with an emphasis on the relationship between interest rates and exchange rates, implied by single regime term structure models. Section 2.3 introduces the regime switching dynamic term structure model and its implications for the exchange rate risk premiums. Section 2.4 explains the estimation of the term structure model and the specification of risk premium regressions. Section 2.5 discusses the empirical results. Section 2.6 offers some concluding remarks.

2.2 Related Literature

Our research is guided by the literature on the relationship between interest rates in (single regime) term structure models and exchange rate risk premiums in terms of modelling strategy. The seminal work of Backus, Foresi and Telmer (2001) adapts modern no-arbitrage term structure models to a multi-currency setting, and derives, under the assumption of no-arbitrage, the relationship between exchange rates and the pricing kernel for each currency. Within this framework, the modelling strategy usually falls into two categories: a country-specific model that estimates the pricing kernel independently for each country and then computes model-implied foreign exchange risk premiums (see Brennan and Xia, 2006), or a global model that estimates the pricing kernel in each country and the foreign exchange risk premiums jointly (see Inci and Lu, 2004; Mosburger and Schneider, 2005; Dong 2006; Sarno, Schneider and Wagner, 2012). Sarno, Schneider and Wagner (2012) highlight an empirical trade-off in the two modelling strategies. The country-specific model fits bond yield data accurately while model-implied exchange rates are very different from actual rates. Conversely the global model matches foreign exchange data closely while fitting the bond yield data less accurately. We employ the country-specific model, taking care of the estimation of the regime switching term structure model for each country. We show that model-implied risk premiums do contain information about currency excess returns despite a lack of accuracy in matching the actual returns.
Our research is also guided by this literature in terms of the channels through which the exchange rate risk premium is linked to the risk premiums implied from the international bond markets. Brennan and Xia (2004) estimate a so-called “essentially” affine term structure model independently for five major currencies, and find that the currency excess return is explained by the risk premium that compensates for the risk in the investment opportunity set consisting of real interest rate, expected inflation and maximum Sharpe ratio. Using US and German data, Dong (2006) estimates a term structural model with macroeconomic variables, and finds that the output gap and inflation account for about 70% of the variation in the exchange rate risk premium. Sarno, Schneider and Wagner (2012) jointly estimate a global factor model of bond yields and exchange rates for six major currencies. Their model-implied risk premiums predict exchange rate excess returns, and are closely related to various economic variables, including the difference between money supply and output—the variable in traditional monetary models of exchange rates. In summary, the literature supports a monetary channel that the currency excess return is explained by the risk premiums implied from bond markets, which is the base for our proposed mechanism—the shift of monetary policies.

In a related work, Ahn, Chib and Kang (2014) estimate a two-country regime switching dynamic term structure model for US and Canadian bond yields using both bond yield and exchange rate data. However, while the focus of their study is the term structure of the cross-country correlation of bond returns, this paper studies the relationship between the risk premiums implied from bond markets and the exchange rate excess return. Further our regime switching model is fundamentally different in terms of the specification of regimes and that it allows regime risk to be priced. The regime risk premium is a key ingredient in explaining currency excess returns.

2.3 The Model

In the no-arbitrage relationship between the exchange rate, the domestic pricing kernel and the foreign pricing kernel, any one single element can be implied from the other two. In this section, we specify the domestic and the foreign pricing kernels through
a regime switching dynamic term structure model, and derive their implications for exchange rates.

2.3.1 Regime switching dynamic term structure model (RS-DTSM)

We employ the regime switching dynamic term structure model Dai, Singleton and Yang (2007). A novel feature of their model is that the risk of regime shifts is priced. The price of regime risk is specified as the difference between the time-varying transition probability of regimes under the physical measure and the constant transition probability under the risk neutral measure, and the associated risk premium compensates for the risk of regime shifts per se.

The process of latent state variables under the physical measure ($P$) and under the risk neutral measure ($Q$) is specified as

$$ F_{t+1} = F_t - \kappa^P(q_t) \cdot \left[ F_t - \theta^P(q_t) \right] + \Sigma(q_t) \cdot u_{t+1}^P $$  
(2.1)

$$ F_{t+1} = F_t - \kappa^Q \cdot \left[ F_t - \theta^Q(q_t) \right] + \Sigma(q_t) \cdot u_{t+1}^Q, $$  
(2.2)

where $F_t$ is an $N \times 1$ vector of latent state variables; $q_t$ follows an observable\(^1\) two regime Markov process; the autoregressive parameter under the $Q$-measure, $\kappa^Q$, is assumed to have a single regime; $u_{t}^{P,Q} \sim i.i.d. N(0, I_N)$; the conditional volatility, $\Sigma(q_t)$, is regime-dependent but does not depend on time. Let $\mu_t^P$ and $\mu_t^Q$ denote time $t$’s conditional mean of $F_{t+1}$ under the $P$-measure and under the $Q$-measure respectively. The conditional means are time-varying and regime-dependent, and are given by

$$ \mu_t^P(q_t) = F_t - \kappa^P(q_t) \cdot \left[ F_t - \theta^P(q_t) \right] $$  
(2.3)

$$ \mu_t^Q(q_t) = F_t - \kappa^Q \cdot \left[ F_t - \theta^Q(q_t) \right]. $$  
(2.4)

It is assumed that the transition probability of regimes under the $P$-measure, denoted by $\Pi^P$, is time-varying, and the transition probability under the $Q$-measure,\(^1\)The assumption that the regime is observable is also made by Dai, Singleton and Yang (2007). Specifically, the regime is observable to investors but is unobservable to econometricians.
denoted by $\Pi^Q$, is constant over time. As will be shown in section 2.5, the two regimes are driven by the conditional volatility of latent factors, characterising a low-volatility regime and a high-volatility regime, denoted by $q_t = j$, $j = L, H$. Using this label of regimes, the transition probability is written as

$$
\Pi^P = \begin{bmatrix}
\pi_{LL,t}^P & \pi_{LH,t}^P \\
\pi_{HL,t}^P & \pi_{HH,t}^P
\end{bmatrix}, \quad (2.5)
$$

and

$$
\Pi^Q = \begin{bmatrix}
\pi_{LL}^Q & \pi_{LH}^Q \\
\pi_{HL}^Q & \pi_{HH}^Q
\end{bmatrix}, \quad (2.6)
$$

where $\pi_{jk} = p(q_{t+1} = k|q_t = j)$, $\forall j, k = L, H$, is the probability of being in regime $k$ next period given that the current regime is regime $j$. The time-varying transition probability under the $P$-measure is a function of latent factors, i.e.,

$$
\pi_{jk,t}^P = p(q_{t+1} = k|q_t = j; F_t) = \frac{1}{1 + e^{(\eta_0^q + \eta_1^q F_t)}}, \quad j \neq k. \quad (2.7)
$$

Bond prices and bond yields are functions of the latent factors. Let $P_{n,t}$ denote the price of a zero-coupon bond with a maturity of $n$ months at time $t$, and $y_{n,t} = \frac{-\log P_{n,t}}{n} \times 12$ is the annualised yield on this zero-coupon bond. It is assumed that the yield on the one-month zero-coupon bond (not annualised), denoted by $r_t$, depends on the latent factors $F_t$ and the current regime $q_t$ according to the function

$$
r_t(q_t) = \delta_0(q_t) + \delta_1^t F_t, \quad (2.8)
$$

where $\delta_0(q_t)$ is a regime-dependent scalar, and the factor loading $\delta_1$ is an $N \times 1$ vector that does not depend on regimes. This functional form of $r_t$ gives closed-form solutions for the price of zero-coupon bonds. It can be shown that the price of an $n$-month maturity bond, denoted by $P_{n,t}$, satisfies

$$
P_{n,t}(q_t) = e^{-A_n(q_t) - B_n^t F_t}, \quad (2.9)
$$

where $A_n(q_t)$ is a scalar with two regimes, and the factor loading $B_n$ is an $N \times 1$ vector with a single regime. Given the regimes in the current and the next periods, i.e., $q_t = j$ and $q_{t+1} = k$, the scalar $A_n(q_t)$ and the factor loading $B_n$ are given by

$$
A_{n+1}^j = \delta_0^j - \log \left( \sum_{k=L,H} \pi_{jk}^Q \cdot e^{-A_k^j} \right) + B_n^t \kappa^Q \theta^Q j - \frac{1}{2} B_n^t \Sigma^j \Sigma^j B_n \quad (2.10)
$$

and

$$
B_{n+1} = \delta_1^j + B_n - \kappa^Q B_n, \quad (2.11)
$$
CHAPTER 2. THE YEN RISK PREMIUMS

with initial conditions $A_0^j = 0$ and $B_0^j = 0_{N \times 1}$. The derivation of equations (2.9)–(2.11) is given in Appendix 2.A. Following equation (2.9), the annualised bond yield on a zero-coupon bond with a maturity of $n$ months is given by

$$y_{n,t}(q_t) = \frac{A_n(q_t)}{n/12} + \frac{B_n'(q_t)}{n/12}F_t = a_n(q_t) + b_n'F_t,$$

where $a_n(q_t)$ is a scalar with two regimes, and the factor loading $b_n$ is an $N \times 1$ vector with a single regime.

The pricing kernel is regime-dependent and is specified as

$$\log M_{t,t+1}(q_t, q_{t+1}) = -r_t(q_t) - \Gamma_{t,t+1}(q_t, q_{t+1}) - \frac{1}{2} \lambda_t(q_t)'\lambda_t(q_t) - \lambda_t(q_t)'u_{t+1},$$

where $\Gamma_{t,t+1}(q_t, q_{t+1})$ is the price of regime risk, and $\lambda_t(q_t)$ is the price of factor risk. The price of regime risk is defined as the difference between the transition probability under the $P$-measure and under the $Q$-measure, i.e.,

$$\Gamma_{t,t+1}^{jk} = \Gamma_{t,t+1}(q_t = j, q_{t+1} = k) \equiv \log \left( \frac{\pi^P_{jk,t}}{\pi^Q_{jk}} \right).$$

To see the interpretation of $\Gamma_{t,t+1}^{jk}$, consider a security with payoff $1_{\{q_{t+1} = k\}}$, i.e., paying 1 if and only if $q_{t+1} = k$. Given the current regime $q_t = j$, the log expected excess return of this security is

$$\log \frac{E_t[P_{t+1}\mid q_t = j]}{P^j_t} - r^j_t = \log \frac{E_t[1_{\{q_{t+1} = k\}}\mid q_t = j]}{e^{-r^j_t}E^Q_t[1_{\{q_{t+1} = k\}}\mid q_t = j]} - r^j_t = \log \frac{\pi^P_{jk,t}}{\pi^Q_{jk}} - r^j_t = \log \frac{\pi^P_{jk,t}}{\pi^Q_{jk}} = \Gamma_{t,t+1}^{jk},$$

so that $\Gamma_{t,t+1}^{jk}$ is interpreted as the price of regime risk. The regime risk is priced if $\Gamma_{t,t+1}^{jk} \neq 0$, or equivalently, $\pi^P_{jk,t} \neq \pi^Q_{jk}$. The price of factor risk $\lambda_t(q_t)$ is specified as

$$\lambda_t(q_t) = \left( \Sigma(q_t) \right)^{-1} \left( \lambda_0(q_t) + \lambda_1(q_t)F_t \right).$$

It is shown in Appendix 2.B that $\lambda_t(q_t)$ satisfies

$$\mu^Q_t(q_t) = \mu^P_t(q_t) - \Sigma(q_t)\lambda_t(q_t),$$

where $\mu^Q_t$ and $\mu^P_t$ are the conditional means of latent factors under the physical and the risk neutral measure. It follows that given $q_t = j$,

$$\lambda_0^j = \kappa^Pj\theta^Pj - \kappa^Qj\theta^Qj$$

and

$$\lambda_1^j = \kappa^Qj - \kappa^Pj.$$
To see the interpretation of $\lambda_t$, consider a security with payoff $e^{b'F_{t+1}}$. Given the current regime $q_t = j$, the log expected excess return of this security is

$$\log E_t^P \left[ e^{b'F_{t+1}} / P_t^j \right] - r_t^j = \log E_t^P \left[ e^{b'F_{t+1}} / e^{-r_t^j E_t^Q \left[ e^{b'F_{t+1}} \right]} \right] - r_t^j = b' \left( \mu_t^P - \mu_t^Q \right) = (b' \Sigma^j) \lambda_t^j, \quad (2.20)$$

where $(b' \Sigma^j)$ is interpreted as the exposure of this security to factor risk, and $\lambda_t^j$ is interpreted as the price of factor risk.

### 2.3.2 Exchange rates and exchange rate risk premiums

In the two-country setting of the regime switching dynamic term structure model, latent factors and regimes in one country are assumed to be independent of those in the other country.\(^2\) Variables of the foreign country are labelled by a superscript *. The spot exchange rate, denoted by $S_t$, and the forward exchange rate, denoted by $F_t$, are expressed as US dollars per unit of foreign currency. Lower cases $s_t$ and $f_t$ denote taking the log of the respective variables. An increase in $S_t$ represents an appreciation of the foreign currency and a depreciation of the US dollar.

As shown in Backus, Foresi and Telmer (2001), the assumption of no-arbitrage imposes a relationship between the exchange rate and the pricing kernels in the two countries. Specifically, consider an asset denominated in US dollars with a gross return of $R_{t+1}$. For a US investor who invests one US dollar in this asset, the gross return satisfies

$$1 = E_t \left[ \mathcal{M}_{t,t+1}(q_t, q_{t+1}) R_{t+1} \right]. \quad (2.21)$$

For a foreign investor who invests one unit of foreign currency in this asset, she converts the foreign currency into US dollars, invests in the US asset, and one period later, the

\(^2\)These are two underlying assumptions of our country-specific model. The first assumption of independent latent factors might not capture the cross-country behaviour of bond yields very well, given that the correlation between US and Japanese yields is high at 0.6218–0.8903 in Table 2.1. Similarly, as shown in Table 2.5, the across-country correlation of model-implied latent factors is at 0.6124–0.8625, while Japanese factors are less correlated with the US curvature factor with correlations of 0.2925 and 0.3932. In this sense, the latent factors in the two countries may reduce to two common (possibly US) factors and a country-specific factor, and this does not contradict our empirical finding that the US factor risk premium explains the excess return on the USD-JPY currency pair. Note that even in the global model of Sarno, Schneider and Wagner (2012), they firstly estimate two global factors using US data only and then include these factors in the estimation of term structure models for each country. The second assumption of independent regimes is more supported by the data, in terms of both the identification of regimes and a relative low across-country correlation of transition probabilities at 0.0716–0.5026 in Table 2.5.
converts the payoff back into the foreign currency. Then the gross return satisfies

\[ 1 = E_t \left[ M_{t,t+1}(q^*_t, q^*_t q^*_{t+1}) S_t R_{t+1} / S_{t+1} \right]. \] (2.22)

To ensure arbitrage-free markets, the exchange rate must satisfy

\[ \frac{S_{t+1}(q^*_t, q^*_t q^*_{t+1})}{S_t(q^*_t, q^*_t)} = \frac{M_{t,t+1}(q^*_t, q^*_t q^*_{t+1})}{M_{t,t+1}(q^*_t, q^*_t q^*_{t+1})}, \] (2.23)

where time \( t \)'s exchange rate depends on time \( t \)'s domestic and foreign regimes \( \{q_t, q^*_t\} \). Given the specification of the pricing kernel in equation (2.13), the change in the log spot exchange rate is

\[
\Delta s_{t+1} = \ln S_{t+1} - \ln S_t = \ln M_{t,t+1}(q^*_t, q^*_t q^*_{t+1}) - \ln M_{t,t+1}(q_t, q^*_t q^*_{t+1}) \\
= -r^*_t(q^*_t) - \Gamma_{t,t+1}(q^*_t, q^*_{t+1}) - \frac{1}{2} \lambda_t^*(q^*_t) \lambda_t^*(q^*_t) - \lambda_t^*(q^*_t) u^*_{t+1} \\
- \left( -r_t(q_t) - \Gamma_{t,t+1}(q_t, q^*_t q^*_{t+1}) - \frac{1}{2} \lambda_t(q_t) \lambda_t(q_t) - \lambda_t(q_t) u^*_{t+1} \right) \\
= r_t(q_t) - r^*_t(q^*_t) + \left( \Gamma_{t,t+1}(q_t, q^*_t q^*_{t+1}) - \Gamma_{t,t+1}^*(q^*_t, q^*_{t+1}) \right) \\
+ \left( \frac{1}{2} \lambda_t(q_t) \lambda_t(q_t) - \frac{1}{2} \lambda_t^*(q^*_t) \lambda_t^*(q^*_t) \right) + \lambda_t(q_t) u^*_{t+1} - \lambda_t^*(q^*_t) u^*_{t+1}, \] (2.24)

where regimes \( \{q_t, q^*_t\} \) are independent of each other, and factor shocks \( \{u_t, u^*_t\} \) are also independent of each other. Exchange rate changes have regimes in both the mean and the volatility. To simplify notation, domestic and foreign regimes are denoted by \( \{q_t = j, q^*_t = k; q^*_t = c; q^*_{t+1} = d\} \), and the change in exchange rates is rewritten as

\[
\Delta s_{t+1} = r^d_t - r^c_t + \left( \Gamma_{t,t+1}^{jk} - \Gamma_{t,t+1}^{cd} \right) + \left( \frac{1}{2} \lambda^j_t \lambda^j_t - \frac{1}{2} \lambda^c_t \lambda^c_t \right) \lambda^j_t u^*_{t+1} - \lambda^c_t u^*_{t+1}. \] (2.25)

The exchange rate forward premium, denoted by \( fp \), is defined as the difference between the forward rate and the spot rate, i.e.,

\[ fp_t = \log F_t - \log S_t = f_t - s_t. \] (2.26)

Covered interest rate parity (CIP) postulates that the following two strategies are equivalent: either invest one US dollar in US treasury bills, or convert one US dollar into a foreign currency at the spot rate, invest in foreign treasury bills, and one period later, convert the payoff back into US dollars at the forward rate, i.e.,

\[ 1 \times (1 + r_t) = \frac{1}{S_t} \times (1 + r^*_t) \times F_t, \]
or equivalently, \[ fp_t = r_t - r_t^*. \] (2.27)

Following Fama (1984), the exchange rate risk premium, denoted by \( r_p \), is defined as the difference between the forward rate and the expected spot rate, i.e.,

\[
    r_p t \equiv \log F_t - E_t \left( \log S_{t+1} \right)
    = - \left[ \left( E_t (s_{t+1}) - s_t \right) - (f_t - s_t) \right]
    = - \left[ E_t (\Delta s_{t+1}) - fp_t \right].
\] (2.28)

That is, the risk premium is the expected excess return on taking a short forward position in the foreign currency, or equivalently, the expected excess return on borrowing the foreign currency and investing in domestic Treasury bills. The actual excess return, denoted by \( rx \), is calculated as

\[
    rx_{t+1} = - \left[ \Delta s_{t+1} - fp_t \right].
\] (2.29)

Given \( \{ q_t = j, q_{t+1} = k, q_t^* = c, q_{t+1}^* = d \} \), the excess return is equal to

\[
    rx_{t+1} = - \left[ \Delta s_{t+1} - fp_t \right] = - \left[ \Delta s_{t+1} - \left( r_j^* - r_t^* \right) \right] = \Gamma^{scd}_{t,t+1} - \Gamma^{jkc}_{t,t+1} + \frac{1}{2} \lambda^{t^*}_{sc} \lambda_t^{sc} - \frac{1}{2} \lambda_{t^*}^{j*} \lambda_t^{j*} + \lambda_{t^*}^{c*} u_{t+1}^* - \lambda_{t^*}^{j*} u_{t+1}. \] (2.30)

It follows that given \( \{ q_t = j, q_t^* = c \} \), the expected excess return (risk premium) and the variance of the excess return are given by

\[
    r_p t = E_t (rx_{t+1}) = \left( \sum_{d=L,H} \pi^{scd}_{cd,t} \Gamma^{scd}_{t,t+1} \right) - \left( \sum_{k=L,H} \pi^{jk}_{kj,t} \Gamma^{jk}_{t,t+1} \right) + \left( \frac{1}{2} \lambda^{jc}_{t^*} \lambda_t^{jc} \right) - \left( \frac{1}{2} \lambda_{t^*}^{j*} \lambda_t^{j*} \right)
\] (2.31)

\[
    Var_t (rx_{t+1}) = \pi^{sc}_{cd,t} \cdot \pi^{sc}_{cd,t} \cdot \left( \Gamma^{scL}_{t,t+1} - \Gamma^{scH}_{t,t+1} \right)^2 + \pi^{jL}_{jL,t} \cdot \pi^{jL}_{jL,t} \cdot \left( \Gamma^{jL}_{t,t+1} - \Gamma^{jL}_{t,t+1} \right)^2 + \left( \lambda^{j*}_{t^*} \lambda_t^{j*} \right) + \left( \lambda_{t^*}^{j*} \lambda_t^{j*} \right).
\] (2.32)

According to equation (2.31), the exchange rate risk premium is determined by the uncertainty over regime shifts \( \left( \sum \pi^{sc}_{cd,t} \Gamma_{t,t+1} \right) \) and the volatility of pricing kernels \( \left( \frac{1}{2} \lambda_{t^*} \lambda_t \right) \).

The uncertainty over regime shifts is non-negative, measuring the deviation of the transition probability of regimes under the physical measure from the risk-neutral measure, and is equal to zero when the probabilities under both measures are equal.
Generally speaking, an investor requires a positive risk premium on taking a short forward position in the foreign currency if in the foreign country there is a larger uncertainty over regime shifts and a higher volatility of the pricing kernel.

### 2.4 Model Estimation

#### 2.4.1 Minimum-chi-square estimation of RS-DTSM

The minimum-chi-square estimation proposed by Hamilton and Wu (2012) is applied to the estimation of the regime switching dynamic term structure model. To see the general procedure of this estimation method, let $\Theta_{\text{stru}}$ denote a vector consisting of the structural (theoretical) parameters in Section 2.3.1, i.e.,

$$
\Theta_{\text{stru}} = \{ \theta_{Pj}, \theta_{Qj}, \kappa_{Pj}, \kappa_{Q}, \Sigma_j, \delta_0, \delta_1, \eta_{jk}^0, \eta_{jk}^1, \pi_{jj} \},
$$

where $j, k = L, H$, $j \neq k$ denote regime labels.

It is assumed that there are $N$ zero coupon bonds that are priced without error, where $N$ is equal to the number of latent factors, and there are $M$ zeros coupon bonds that are priced with error. Specifically, let $Y_{1,t} = \{ y_{n_1,t}, y_{n_2,t}, \ldots, y_{n_N,t} \}'$ denote an $N \times 1$ vector of annualised bond yields that are priced without error, where the maturities of these bonds are months $\{ n_1, n_2, \ldots, n_N \}$, and let $Y_{2,t} = \{ y_{m_1,t}, y_{m_2,t}, \ldots, y_{m_M,t} \}'$ denote an $M \times 1$ vector of annualised bond yields that are priced with error, where the maturities of these bonds are months $\{ m_1, m_2, \ldots, m_M \}$. It is shown in Appendix 2.C that given $q_{t+1} = k$ and $q_t = j$, the process of $Y_{1,t}$ and $Y_{2,t}$ is

$$
Y_{1,t+1} = \alpha^k_1 + (Y_{1,t} - \alpha^j_1) - \kappa_{11}^j (Y_{1,t} - \alpha^j_1 - \theta^j_1) + \Sigma_j^1 u_{t+1}
$$

$$
Y_{2,t+1} = \alpha^j_2 + \phi_{21} (Y_{1,t+1} - \alpha^j_1) + \Sigma_j^2 v_{t+1},
$$

and the transition probability of regimes under the $P$-measure is a function of $Y_{1,t}$, i.e.,

$$
\pi_{jk,t}^P = p(q_{t+1} = k|q_t = j; Y_{1,t}) = \frac{1}{1 + e^{(\xi_0^k + \xi^j_{1Y_{1,t}})}}, \ j \neq k,
$$

where parameters $\{ \alpha_1, \theta_1, \kappa_{11}, \Sigma_1, \alpha_2, \phi_{21}, \xi_0, \xi_1 \}$ are functions of the structural parameters in (2.33), and the functional forms are given in Appendix 2.C. Equations (2.34)–(2.36) are the reduced-form models that characterise the process of the observable bond.
proximating the variance of \( \hat{\Theta} \) structural and reduced-form parameters. Hamilton and Wu (2012) also propose a procedure of solving \( \hat{\Theta} \) is just-identified, the minimum value attainable for this statistic is zero, and the procedure to motivate the estimate \( \hat{\Theta} \) is

\[
\Theta_{\text{redu}} = \left\{ \alpha^j_1, \theta^j_1, \kappa^j_1, \Sigma^j_1, \alpha^j_2, \phi_{21}, \zeta^j_0, \xi^j_1 \right\}, \tag{2.37}
\]

where \( j, k = L, H, j \neq k \) denote regime labels. The entire sample of bond yields is denoted by \( Y_T = \{ Y_{1,\tau}, Y_{2,\tau}; \tau \leq T \} \), where \( T \) is the sample size. Let \( \hat{\Theta}_{\text{redu}} \) denote the maximum likelihood (ML) estimate of \( \Theta_{\text{redu}} \) that maximises the log likelihood of \( Y_T \), denoted by \( \mathcal{L}(\Theta_{\text{redu}}; Y_T) \). Under the null hypothesis that reduced-form parameters are functions of structural parameters, i.e., \( \Theta_{\text{redu}} = g(\Theta_{\text{stru}}) \), the following statistic has an asymptotic chi-square distribution, i.e.,

\[
T \left[ \hat{\Theta}_{\text{redu}} - g(\Theta_{\text{stru}}) \right]' \hat{R} \left[ \hat{\Theta}_{\text{redu}} - g(\Theta_{\text{stru}}) \right] \sim \chi^2(q), \tag{2.38}
\]

where \( \hat{R} \) is a consistent estimate of the information matrix \( R = -T^{-1} E \left[ \frac{\partial^2 \mathcal{L}(\Theta_{\text{redu}}; Y_T)}{\partial \Theta_{\text{redu}} \partial \Theta_{\text{redu}}'} \right] \), and \( q \) is the dimension of \( \Theta_{\text{redu}} \). The minimum-chi-square estimation uses (2.38) to motivate the estimate \( \hat{\Theta}_{\text{stru}} \) that minimises this chi-square statistic. When the model is just-identified, the minimum value attainable for this statistic is zero, and the procedure of solving \( \hat{\Theta}_{\text{stru}} \) from \( \left[ \hat{\Theta}_{\text{redu}} - g(\hat{\Theta}_{\text{stru}}) = 0 \right] \) is termed the mapping between structural and reduced-form parameters. Hamilton and Wu (2012) also propose approximating the variance of \( \hat{\Theta}_{\text{stru}} \) with \( T^{-1} \left( \hat{\Gamma}' \hat{\Gamma} \right)^{-1} \) where \( \hat{\Gamma} = \frac{\partial g(\Theta_{\text{stru}})}{\partial \Theta_{\text{stru}}'} \bigg|_{\Theta_{\text{stru}} = \hat{\Theta}_{\text{stru}}} \).

The log likelihood function of the reduced-form models is

\[
\mathcal{L}(\Theta_{\text{redu}}; Y_T) = \sum_{t=2}^{T} \log f(Y_{1,t}, Y_{2,t}|Y_{t-1}; \Theta_{\text{redu}}), \tag{2.39}
\]

where \( Y_t = \{ Y_{1,\tau}, Y_{2,\tau}; \tau \leq t \} \) is the information set up to time \( t \). The conditional density of the observed yields is

\[
f(Y_{1,t}, Y_{2,t}|Y_{t-1}) = \sum_{q_t, q_{t-1}} f(Y_{1,t}, Y_{2,t}, q_t, q_{t-1}|Y_{t-1})
\]

\[
= \sum_{q_t, q_{t-1}} f(Y_{2,t}|Y_{1,t}, q_t) \cdot f(Y_{1,t}|Y_{t-1}, q_t, q_{t-1}) \cdot p(q_t|q_{t-1}, Y_{t-1}) \cdot p(q_{t-1}|Y_{t-1}),
\]

where \( Y_{1,t} \) and \( Y_{2,t} \) are conditionally normal according to equations (2.34) and (2.35), \( p(q_t|q_{t-1}, Y_{t-1}) \) is the transition probability of regimes, and \( p(q_{t-1}|Y_{t-1}) \) is the filtered probability. The filtered probability, \( p(q_t|Y_{t}) \), is the probability of the regime at time
CHAPTER 2. THE YEN RISK PREMIUMS

38

given the information set up to time \( t \), and is given by

\[
p(q_t|Y_t) = \frac{p(q_t, Y_{1,t}, Y_{2,t}|Y_{t-1})}{f(Y_{1,t}, Y_{2,t}|Y_{t-1})} = \frac{\sum_{q_{t-1}} p(q_t, q_{t-1}, Y_{1,t}, Y_{2,t}|Y_{t-1})}{f(Y_{1,t}, Y_{2,t}|Y_{t-1})} = \frac{\sum_{q_{t-1}} f(Y_{2,t}|Y_{1,t}, q_t) \cdot f(Y_{1,t}|Y_{t-1}, q_t, q_{t-1}) \cdot p(q_t|Y_{t-1}) \cdot p(q_{t-1}|Y_{t-1})}{f(Y_{1,t}, Y_{2,t}|Y_{t-1})}.
\]

The log likelihood function is obtained by iterating between \( f(Y_{1,t}, Y_{2,t}|Y_{t-1}) \) and \( p(q_t|Y_t) \), given a starting value, \( \pi_0^j = p(q_0 = j|Y_0), j = L, H \), where \( \pi_0^j \) is the probability of being in regime \( j \) at \( t = 0 \). We set \( \pi_0^H = 0.5 \) in model estimation.

Following the common practice in the regime switching literature, we use smoothed probabilities in classifying regimes. The smoothed probability, denoted by \( p(q_t|Y_T) \), is the probability of the regime at time \( t \) given the full sample information \( Y_T \). The smoothed probability is solved backwards from \( p(q_T|Y_T) \), which is equal to the filtered probability at time \( T \). Given the smoothed probability at time \( t \), \( p(q_t|Y_T) \), the probability at time \( t - 1 \) is given by

\[
p(q_{t-1}|Y_T) = \sum_{q_0} p(q_t, q_{t-1}|Y_T),
\]

where

\[
p(q_t, q_{t-1}|Y_T) = p(q_{t-1}|q_t, Y_T) \cdot p(q_t|Y_T) = p(q_{t-1}|q_t, Y_T) \cdot p(q_t|Y_T) = \frac{p(q_{t-1}|q_t, Y_T)}{p(q_t|Y_T)} \cdot p(q_t|Y_T).
\]

We classify the regime at time \( t \) according to the following rule:

\[
q_t = \begin{cases} 
H, & \text{if } p(q_t = H|Y_T) > 0.5 \\
L, & \text{if } p(q_t = H|Y_T) \leq 0.5 
\end{cases}
\]

That is, the regime at time \( t \) is classified as a high-volatility regime if time \( t \)’s smoothed probability of the high-volatility regime is greater than 0.5, and is classified as a low-volatility regime otherwise.

2.4.2 Identification of RS-DTSM

We adapt the normalisations and restrictions imposed by Dai, Singleton and Yang (2007) to the minimum-chi-square estimation procedure. They estimate a three-factor
regime switching model using maximum likelihood estimation, and impose on structural parameters the following normalizations:

\[ \theta^{PL} = 0_{3 \times 1} \]  
\[ \sqrt{12} \cdot \Sigma^L = I_3 \]  
\[ \kappa^{PL} \text{ is a lower triangular matrix} \]  
\[ \Sigma^H \text{ is a lower triangular matrix}, \]

and restrictions:

\[ \kappa^Q \text{ has a single regime} \]  
\[ \delta_1 \text{ has a single regime}. \]

Furthermore, given the difficulty in estimating the unconditional means of bond yields and short rates, they fix the long-run mean of short rates by fixing \( \{ \delta_0^L + \delta_1 \times \theta^{PL} \} \) and \( \{ \delta_0^H + \delta_1 \times \theta^{PH} \} \) at the sample mean of short rates in the low-volatility regime and in the high-volatility regime respectively, where \( \theta^{Pj} \), \( j = L, H \) is the mean of latent factors in each regime, and the mean in the low-volatility regime is normalised to be zero in equation (2.40). Finally, after a preliminary estimation of the model, they set \( \rho^{PL}(2, 1), \Sigma^H(2, 1), \Sigma^H(3, 1), \Sigma^H(3, 2), \lambda_0^H(1), \lambda_0^H(2), \lambda_0^H(3), \lambda_1^L(1, 1), \lambda_1^L(2, 1), \lambda_1^L(2, 2), \lambda_1^L(3, 2), \lambda_1^H(1, 3), \lambda_1^H(2, 3), \lambda_1^H(3, 2), \lambda_1^H(3, 3) \) equal to zero because these parameters are small relative to their estimated standard errors. As shown in equation (2.18), a zero \( \lambda_0^H \) actually implies restrictions on the relationship between \( \theta^{PH} \) and \( \theta^{QH} \) (the means of latent factors that are difficult to estimate), given \( \kappa^{PH} \) and \( \kappa^Q \) (the autoregressive coefficients that are relatively easier to estimate).

In order to adapt to the minimum-chi-square estimation procedure, it is necessary to amend the above normalisations and restrictions. Specifically, we keep normalisations (2.40) and (2.41) and restrictions (2.44) and (2.45). Instead of normalisations (2.42) and (2.43), we estimate \( \kappa^{PL} \) as a full matrix and normalise \( \Sigma^H \) to be a diagonal matrix, which is no more restrictive especially given that Dai, Singleton and Yang (2007) later set the off-diagonal elements of \( \Sigma^H \) to be zero in their model estimation. Another difference from their restrictions is in the ways of restricting the means of latent factors and bond yields. Dai, Singleton and Yang (2007) normalise \( \theta^{PL} \), fix the
long-run mean of short rates (equivalent to restricting parameters \( \{\delta^L_0, \delta^H_0, \theta^{PH}\} \)), restrict the means of latent factors, i.e., parameters \( \{\theta^{PH}, \theta^{QH}\} \), and estimate the other parameters \( \{\theta^{QL}, \pi^{Q}_{LL}, \pi^{Q}_{HH}\} \) as free parameters. In the minimum-chi-square estimation procedure, these structural parameters are mapped out from reduced-form parameters, so that in the first place we need to impose restrictions on reduced-form parameters. In equations (2.34) and (2.35), the reduced-form parameters relating to the means of bond yields and latent factors are parameters \( \{\alpha_j^1, \theta_j^1, \alpha_j^2\} \), where \( j = L, H \) denotes regime labels. The identification problem is in parameters \( \{\alpha_j^1, \theta_j^1\} \), where \( \theta_j^1 \) relates to the regime-dependent means of latent factors, and \( \alpha_j^1 \) and \( \theta_j^1 \) together determine the regime-dependent means of bond yields. This identification difficulty still exists after normalising \( \theta^{PL} \) to be zero in equation (2.40), which implies \( \theta_1^L = 0 \). We deal with this by further imposing the restriction

\[
\alpha_1^L = \alpha_1^H
\]  

(2.46)

in model (2.34) and estimate \( \theta_1^H \), therefore \( \theta^{PH} \), as free parameters. Finally, we fix the structural parameters \( \{\pi^{Q}_{LL}, \pi^{Q}_{HH}\} \) so that the mapping between reduced-form and structural parameters is just-identified, and parameters \( \{\delta^L_0, \delta^H_0, \theta^{QL}, \theta^{QH}\} \) are solved as free parameters. Specifically, we set \( \{\pi^{Q}_{LL}, \pi^{Q}_{HH}\} \) equal to the constant transition probabilities implied by the smoothed probabilities.

### 2.4.3 Exchange rate risk premium regressions

This section specifies the risk premium regressions where the realised exchange rate excess return is regressed on the expected excess return (risk premium) and its components as implied from term structure models. Firstly, for each country \( i = US, JP \), the model-implied regime risk premium, denoted by \( rp^i_r \), and the factor risk premium,
denoted by \( rp_j \), are calculated as

\[
\begin{align*}
\text{rp}^{US}_{\text{r,}t} &= \sum_{j=L,H} G_j^t \cdot \left( \sum_{k=L,H} \pi^P_{jk,t} \Gamma^j_{t,t+1} \right) \\
\text{rp}^{JP}_{\text{r,}t} &= \sum_{c=L,H} G^*_{ct} \cdot \left( \sum_{d=L,H} \pi^*_{cd,t} \Gamma^*_{t,t+1} \right) \\
\text{rp}^{US}_{\text{f,}t} &= \sum_{j=L,H} G_j^t \cdot \left( \frac{1}{2} \lambda_j^t \lambda_j^t \right) \\
\text{rp}^{JP}_{\text{f,}t} &= \sum_{c=L,H} G^*_{ct} \cdot \left( \frac{1}{2} \lambda^*_{ct} \lambda^*_{ct} \right)
\end{align*}
\]

where \( j, k, c, d = L, H \) are regime labels, i.e., \( \{ q_t = j, q_{t+1} = k, q^*_t = c, q^*_{t+1} = d \} \), and we use the smoothed probability at time \( t \) as the weight \( G^*_t \). It follows that the net regime risk premium, denoted by \( \text{rp}_r \), the net factor risk premium, denoted by \( \text{rp}_f \), and the total risk premium, denoted by \( \text{rp} \), are calculated as

\[
\begin{align*}
\text{rp}_r &= \text{rp}^{JP}_{r,t} - \text{rp}^{US}_{r,t} \\
\text{rp}_f &= \text{rp}^{JP}_{f,t} - \text{rp}^{US}_{f,t} \\
\text{rp} &= \text{rp}_r + \text{rp}_f.
\end{align*}
\]

The first group of regressions regress the excess return on the model-implied risk premiums, i.e.,

\[
\begin{align*}
rx_{t+1} &= l_{01} + l_1 \cdot \text{rp}_t + \varepsilon_{1,t+1} \\
rx_{t+1} &= l_{02} + l_2 \cdot \text{rp}_r,t + l_3 \cdot \text{rp}_f,t + \varepsilon_{2,t+1} \\
rx_{t+1} &= l_{03} + l_4 \cdot \text{rp}^{JP}_{r,t} + l_5 \cdot \left( - \text{rp}^{US}_{r,t} \right) + l_6 \cdot \text{rp}^{JP}_{f,t} + l_7 \cdot \left( - \text{rp}^{US}_{f,t} \right) + \varepsilon_{3,t+1},
\end{align*}
\]

where coefficients \( l_j, j = 1, \ldots, 7 \) are expected to be positive and equal to one. Two specifications of the error term \( \varepsilon_{i,t} \) are considered. The first specification assumes that the error term has a constant volatility, i.e., \( \varepsilon_{i,t} \sim i.i.d. N(0, \sigma^2_i) \), \( i = 1, 2, 3 \), and the model is estimated by ordinary least squares. The second specification assumes that the error term follows GARCH(1,1) process, which is motivated by the time-varying conditional variance of the excess return in equation (2.32). Specifically, for each error term in regressions (2.54)–(2.56), the variance equations are

\[
\begin{align*}
\varepsilon_t &= \sigma_t \cdot \varepsilon_t \\
\sigma^2_{t+1} &= w_0 + w_1 \cdot \varepsilon_t^2 + w_2 \cdot \sigma^2_t \\
\varepsilon_t &\sim i.i.d. N(0, 1),
\end{align*}
\]
and it is expected that $w_1$ and $w_2$ are jointly not equal to zero. Sarno, Schneider and Wagner (2012) estimate regression (2.54) using ordinary least squares while Brennan and Xia (2006) estimate GARCH-type models.

The second group of regressions include the forward premium as an additional regressor in regressions (2.54)–(2.56). This is motivated by the forward premium puzzle that the forward premium predicts the excess return. Fama (1984) regresses the excess return and the change in exchange rates on the forward premium, i.e.,

$$r_{x_{t+1}} = \gamma_0 + \gamma_1 \cdot f_{p_t} + \epsilon_{1,t+1}$$

$$\Delta s_{t+1} = \gamma_0 + \gamma_2 \cdot f_{p_t} + \epsilon_{2,t+1}.$$ 

These two regressions are equivalent where $\gamma_1 = 1 - \gamma_2$, and uncovered interest rate parity (UIP) predicts that $\gamma_1 = 0$ and $\gamma_2 = 1$. However, empirically, it is usually found that $\gamma_1 > 0$ and $\gamma_2 < 0$ for major currencies, which is known as the forward premium puzzle. We include the forward premium as an additional regressor in order to examine the extent to which model-implied risk premiums are able to account for the forward premium puzzle and whether they reflect any information beyond the forward premium about the exchange rate excess return.

### 2.5 Empirical Analysis

#### 2.5.1 Bond yield and foreign exchange data

The bond yield data is obtained from the dataset of Wright (2011). The sample period for the US data is from November 1971 to May 2009, and the sample period for the Japanese data is shorter, from January 1985 to May 2009. For the US bond yields, following Dai, Singleton and Yang (2007) who use bond yield data from 1972 to 2003, we estimate a three-factor model, assuming that the bonds with 6-, 24- and 120-month maturities are priced without error, and the bond with 60-month maturity is priced with error. For the Japanese bond yields, we estimate a parsimonious two-factor model, assuming that the bonds with 6- and 120-month maturities are priced without
error, and the bond with 60-month maturity is priced with error. The data source for foreign exchange spot rates and forward rates is Datastream, and the dates back to November 1983. We adopt a sample period from January 1985 to June 2009 consistent with the sample period of bond yields. Spot rates and forward rates are expressed as US dollars per Japanese yen. An increase in spot rates represents an appreciation of the Japanese yen and a depreciation of the US dollar.

Table 2.1 reports the summary statistics of the bond yields. In each country, the sample mean of the bond yields increases as the bond maturity increases, implying an upward-sloping yield curve. The sample standard deviation decreases as the maturity increases, i.e., long-term yields are less volatile than short-term yields. The autocorrelation coefficient of the bond yields is quite high, being greater than 0.98 for all maturities. The correlation across maturities is also very high—even for the least correlated 6-month and 10-year yields, the correlation is around 0.90. Comparing Panel B with Panel C shows that US bond yields have higher means, higher volatilities and slightly lower autocorrelations than Japanese yields. Panel C reports the correlation of bond yields across countries, which is generally lower than the correlation within each country. Japanese yields are more correlated with long-term US yields. Figure 2.1 plots the bond yields. In the US, bond yields are high in the period from the late 1970s to the early 1980s, and the minimum of yields is in the recent financial crisis. In Japan, bond yields are high from the beginning of the sample to the early 1990s, and have been low since the late 1990s. During the whole sample period, US yields are higher than Japanese yields except for the period from August 1990 to April 1993 where the US short-term 6-month yields are slightly lower.

Table 2.2 reports the summary statistics of the foreign exchange data, and Figure 2.2 plots the data. The forward premium is calculated as in equation (2.26), which is also equal to the difference between the domestic and the foreign short rates assuming CIP holds, i.e., $f_{t} = r_{tUS} - r_{tJP}$. As shown in Figure 2.2, the forward premium is more persistent than exchange rate changes. Figure 2.3 plots the forward premium.

\footnote{A principal component analysis on the Japanese bond yields of 6-, 24-, 60- and 120-month maturities shows that the marginal contribution of the first, second and third component to the total variance are 98.28\%, 1.59\%, 0.12\% respectively, in favour of a two-factor to a three-factor model. In model estimation, we choose 6-, 60- and 120-month to have spanned maturities and reduce the correlations between bond yields.}
with interest rate differentials. As shown in Panel B, the interest rate differentials are smoother and the forward premium is more volatile with several spikes in the sample. The positive spike in February 1991 and the negative spike in October 1994 do not correspond to any dramatic changes in interest rates, while the positive spike in September 2008 is partially due to a temporary increase in the US short rate. The exchange rate excess return is calculated as in equation (2.29), i.e., \( r_{x_{t+1}} = -[\Delta s_{t+1} - f_{p_t}] \). The sample mean of the forward premium is 0.25%—in other words, the US short rate is on average 0.25% higher than its Japanese counterpart. During the sample period, the US dollar on average depreciates 0.33% against the Japanese yen. The observation that the high interest rate currency (US dollar) depreciates nearly the same amount as the interest rate differential, or more precisely, the mean of the excess return is not significantly different from zero, is consistent with the unconditional version of UIP. However, UIP might not hold conditionally observing that the correlation between \( \Delta s_{t+1} \) and \( f_{p_t} \) is negative at -0.1308 and the correlation between \( r_{x_{t+1}} \) and \( f_{p_t} \) is positive at 0.1869, while UIP predicts a positive relationship between \( \Delta s_{t+1} \) and \( f_{p_t} \) and no relationship between \( r_{x_{t+1}} \) and \( f_{p_t} \). Another observation is that the volatility of the forward premium is comparable with the volatility of bond yields, and is much lower than the volatility of exchange rates, which reveals the empirical difficulty in matching the volatility of exchange rates using term structure models.

### 2.5.2 Latent factors

Tables 2.3 and 2.4 report the estimates of structural parameters in the US and the Japanese regime switching dynamic term structure model respectively. The estimates of reduced-form parameters are reported in Tables 2.C.1 and 2.C.2 in Appendix 2.C. As discussed earlier, we estimate a three-factor model for the US bond yields and a two-factor model for the Japanese bond yields. In both countries, regimes are driven by the conditional volatility of latent factors, which motivates our labelling of regimes. Specifically, the elements of parameters \( \Sigma \) are significantly different from zero and by a rough calculation using standard errors, the elements of \( \Sigma \) in the high volatility regime are more than two standard errors higher than their counterparts in the low-volatility regime. The process of the latent factors is highly persistent in both
regimes. In the US model, the process is more persistent in the low-volatility regime than in the high-volatility regime, with eigenvalues of \([0.9989, 0.9593, 0.9233]\) in the low-volatility regime and eigenvalues of \([0.9868, 0.8657, 0.7998]\) in the high-volatility regime. In the Japanese model, the process is slightly more persistent in the high-volatility regime, with eigenvalues of \([0.9906, 0.9550]\) in the low-volatility regime and eigenvalues of \([0.9953, 0.9282]\) in the high-volatility regime.

Figures 2.4 and 2.5 plot factor loadings of bond yields on the latent factors over different maturities, where bond yields depend on latent factors as in equation (2.12). In the left panel of each figure, the curve for the regime-dependent scalar \(a_n\) is upward sloping. Since we impose restriction (2.46) in the model estimation, the curves in the two regimes are close to each other. Then the cross-regime difference in the mean of bond yields is driven by the regime-dependent mean of latent factors, i.e., parameter \(\theta^P\) in Tables 2.3 and 2.4. In the right panel of each figure, factor loadings \(b_n\) are independent of regimes. Similar to the finding in Dai, Singleton and Yang (2007), the interpretation of \(b_n\) differs from the standard interpretation of factor loadings on level, slope and curvature factors. This is because the model allows flexible correlations between latent factors by assuming non-zero off-diagonal elements of autoregressive parameters, while the standard factors are obtained from principal component analysis. For ease of interpretation, we express the latent factors as a function of the bond yields that are priced without error. In the US model, the latent factors are

\[
\begin{bmatrix}
F_1 \\
F_2 \\
F_3
\end{bmatrix} = \text{const} +
\begin{bmatrix}
-62 & +52 & +99 \\
+110 & -292 & +228 \\
+277 & -220 & +85
\end{bmatrix} \times
\begin{bmatrix}
y_6 \\
y_{24} \\
y_{120}
\end{bmatrix}.
\] (2.60)

The first factor is approximated by \([y_{24} - y_6] + 2 \times y_{120}\), i.e., having a loading of \([-1, +1, +2]\) on bond yields, where \((y_{24} - y_6)\) is termed the short-term slope. The second factor has a curvature flavour with an approximate loading of \([+2, -5, +4]\) on bond yields. The third factor is similar to the first factor but places a negative weight on the short-term slope and is approximated by \([-3 \times (y_{24} - y_6) + y_{120}\], i.e., having a loading of \([+3, -3, +1]\) on bond yields. Table 2.5 reports the summary statistics of the model-implied latent factors, and Figure 2.6 plots the latent factors. As shown in Panel A of Table 2.5, the first and the third factor has a correlation coefficient
of 0.76. The positive correlation implies that these two factors are largely driven by the long-term yield. The second factor is less correlated with the other two, which is consistent with our interpretation of the second factor as a curvature factor. In the Japanese model, the latent factors are expressed as

\[
\begin{bmatrix}
F_1 \\
F_2
\end{bmatrix} = \text{const} + \begin{bmatrix}
+90 & -218 \\
+1883 & -196
\end{bmatrix} \times \begin{bmatrix}
y_6 \\
y_{120}
\end{bmatrix}. 
\] (2.61)

Both factors have a slope flavour while differing in the weight on the 6-month yield. The first factor can be approximated by \(-2 \times (y_{120} - y_6) - y_6\), i.e., having a loading of [+1, -2] on bond yields, where \((y_{120} - y_6)\) is termed the standard slope factor, which parallels the slope factor in the principal component analysis. The second factor is similar to the first factor but places more weight on the short-term yield and is approximated by \(- (y_{120} - y_6) + 9 \times y_6\), i.e., having a loading of [+10, -1] on bond yields. As shown in Panel B of Table 2.5, the two factors have a correlation coefficient of -0.78. The negative correlation implies that these two factors are largely driven by the short-term yield.

### 2.5.3 Regime probabilities

Figures 2.7 and 2.8 plot the filtered, smoothed and transition probabilities of regimes for the US and the Japanese model respectively. As shown in Figure 2.7, the Federal Reserve experiment from 1979 to 1982 is classified as in the high volatility regime, which is consistent with the literature where the regime is often linked to monetary policies. Similar to the finding in Dai, Singleton and Yang (2007) that the high-volatility regime extends beyond the end of 1982 to about 1985, we find that the episode around the mid 1980s is also classified as in the high-volatility regime. In addition, the volatility regimes are associated with business cycle fluctuations. There are six episodes in the sample that are classified by NBER as recessions, and in five of them, the filtered and the smoothed probability of being in the high-volatility regime are high. The exception is the early 1990s recession—only the aftermath of this recession is classified as in the high-volatility regime. Based on our classification of regimes using the smoothed probability, the high-volatility regime occurs more often than
NBER recessions: the economy spends 31% of its time in the high-volatility regime while 17% of its time in NBER recessions.

As shown in Figure 2.8, the volatility regimes in the Japanese bond market are less correlated with business cycle conditions indicated by the Economic Cycle Research Institute (ECRI). The episode from 1985 (the beginning of the sample) to the early 1990s is characterised by the high-volatility regime, which includes the monetary easing in 1986 and 1987 and the monetary tightening in 1989 and 1990. The expansionary monetary policy started from 1990 and persisted over the next decade. The first five years in this episode are characterised by the high-volatility regime while from 1995 to 2000, the classification of regimes switches frequently, characterising a period of uncertainty in the bond market. Bond yields have been low since 2000, and the economy is classified as being in the low-volatility regime until 2006, which coincides with the unconventional monetary policy of quantitative easing in Japan from 2001 to 2006 as documented in Kimura and Nakajima (2016). Our filtered and smoothed probability indicates the end of the high-volatility regime at the end of the sample period in 2009, which is also consistent with Kimura and Nakajima’s (2016) estimation of the unconventional monetary policy regime from 2010 to 2012. As a consequence, we interpret the high volatility regime as a conventional monetary policy regime, and the low volatility regime as an unconventional policy regime. Based on our regime classification, the high-volatility conventional-policy regime is more persistent than the low-volatility unconventional-policy regime—the economy spends 61% of its time in the former.

Table 2.5 reports the summary statistics of the transition probabilities, and Panel C of Figures 2.7 and 2.8 plot the transition probabilities, and shaded periods indicate the high-volatility regimes classified by the smoothed probability. Economic intuition suggests during a transition from the low-volatility regime to the high-volatility regime, the probability of switching to the high-volatility regime tends to increase while the probability of switching to the low-volatility regime tends to decrease, and vice versa. An implication is that the probability of switching from the low-volatility to the high-volatility regime (L2H) and the probability of switching from the high-volatility to the low-volatility regime (H2L) are expected to be negatively correlated. As shown in
Figure 2.7, the US transition probabilities do not vary dramatically between 0 and 1—in Table 2.5, the sample mean (standard deviation) of L2H and H2L is 0.0930 (0.0615) and 0.2186 (0.1536) respectively. Despite a lack of variation, the behaviour of the US transition probabilities as the economy moves in and out of regimes is generally consistent with economic intuition. In Figure 2.8, the Japanese transition probabilities vary more, and produce much sharper predictions in terms of the beginning and the end of regimes. The sample mean (standard deviation) of L2H and H2L is 0.5816 (0.4108) and 0.2236 (0.1998) respectively, and the correlation between L2H and H2L is −0.86.

During the high-volatility episode from 1985 (the beginning of the sample) to the early 1990s, the probability of switching to the low-volatility regime stays at zero. In the late 1990s as the expansionary monetary policy persists, the transition probability varies greatly, and the trend is a decrease in the probability of switching to the high-volatility regime and an increase in the probability of switching to the low-volatility regime. During the episode of the unconventional monetary policy from 2001 to 2006, the probability of switching to the other regime stays at the historically low level. In the recent high-volatility regime, the probability of switching to the high-volatility regime increases from 0.1 to 0.8 and then decreases back to 0.1, and the probability of switching to the low-volatility regime decreases from 0.4 to 0.2 and then increases back to 0.4, suggesting that the economy is moving out of the high-volatility regime at the end of the sample. Guided by the estimates of transition probability parameters in Tables 2.3, 2.4, 2.C.1 and 2.C.2, and the interpretation of latent factors in equations (2.60) and (2.61), we express the transition probability of regimes as a function of the bond yields that are priced without error. Specifically, in the US model, the probability of switching from the low-volatility regime at time \( t \) to the high-volatility regime at time \( t+1 \), \( \pi_{LH,t}^P \), and the probability of switching from the high-volatility to the low-volatility regime, \( \pi_{HL,t}^P \), are given by

\[
\pi_{LH,t}^P = \left[ 1 + \exp \left\{ 5 - 78 \times y_{6,t} - 116 \times (y_{120,t} - 1.41 \times y_{24,t}) \right\} \right]^{-1} \tag{2.62}
\]

\[
\pi_{HL,t}^P = \left[ 1 + \exp \left\{ 1 + 16 \times y_{6,t} - 157 \times (y_{24,t} - y_{6,t}) + 9 \times y_{120,t} \right\} \right]^{-1} \tag{2.63}
\]
and in the Japanese model,

\[
\pi_{LH,t}^P = \left[ 1 + \exp \left\{ 4 - 648 \times y_{6,t} - 66 \times (y_{120,t} - y_{6,t}) \right\} \right]^{-1} \tag{2.64}
\]

\[
\pi_{HL,t}^P = \left[ 1 + \exp \left\{ 1 + 162 \times y_{6,t} - 72 \times (y_{120,t} - y_{6,t}) \right\} \right]^{-1} \tag{2.65}
\]

Generally speaking, the transition probabilities are affected by the short-term yield \( y_6 \) and slope factors: the long-term slope \( (y_{120} - y_{24}) \) in (2.62), the short-term slope \( (y_{24} - y_6) \) in (2.63), and the standard slope \( (y_{120} - y_6) \) in (2.64) and (2.65). In terms of the magnitude of the effects, the transition probabilities are largely driven by the short-term yield \( y_6 \). Specifically, the probability of switching to the high-volatility regime, \( \pi_{LH,t}^P \), increases as the short-term yield increases, and the probability of switching to the low-volatility regime, \( \pi_{HL,t}^P \), increases as the short-term yield decreases. In the US model where the high-volatility regime is associated with economic recessions, this observation is consistent with business cycle intuition: central banks raise short rates due to concerns over inflation towards the end of economic expansions, and bring down short rates to stimulate the economy towards the end of recessions. In the Japanese model, the economy entered into the low-volatility unconventional-policy regime as the short-term yield was brought down to its historically low level around the early 2000s, and the economy entered into the high-volatility regime as the short-term yield increased in 2006.

### 2.5.4 Exchange rate risk premium regressions

Since we estimate the regime switching dynamic term structure model using bond yield data only, the scale of the model-implied risk premiums does not match exchange rate changes or currency excess returns. Mosburger and Schneider (2005) and Sarno, Schneider and Wagner (2012) point out that using exchange rate data in the estimation of international term structure models helps identify (the scale of) the price of risk. In this section, we show that in spite of the scale issue, our model-implied risk premiums do contain important information about currency excess returns.

Table 2.6 reports the summary statistics of the model-implied risk premiums. As predicted by the exchange rate model, an important observation is that the model-implied risk premiums, specifically, the Japanese risk premiums and the negative value
of the US risk premiums, are positively correlated with the realised currency excess return. The correlations are around 0.15, which are comparable with the 0.19 correlation between the excess return and the forward premium in the forward premium puzzle. Within each country, the correlation between the regime risk premium and the factor risk premium is 0.29 in the US, and is 0.61 in Japan. Across countries, the correlation between the US and the Japanese regime risk premium is low at 0.11 and the correlation between the US and the Japanese factor risk premium is slightly higher at 0.22. The correlations between model-implied risk premiums and the forward premium are around 0.30. The exception is the US factor risk premium, which has a correlation of 0.59 with the forward premium. The first-order autocorrelation coefficient of risk premiums is low at 0.2327–0.7803. When AIC is used to select the lag length in ADF test, a unit root is rejected at the 10% level for all the risk premiums except for the US regime risk premium; when BIC is used, a unit root is rejected at the 1% level except for the US regime risk premium; and a unit root is rejected for all the risk premiums in the Phillips-Perron (PP) test. For the forward premium, the autocorrelation is moderate at 0.8102, and a unit root cannot be rejected in ADF tests and is rejected in PP test. Given that the first-order autocorrelation of the risk premiums and the forward premium is moderate and the forecast horizon is short at one month, we include these variables on the right hand side of the exchange rate risk premium regressions without correcting for the estimates that are robust to the persistence of regressors. As shown in Figures 2.9 and 2.10, the regime risk premium generally increases during transitions of regimes, either moving from the low-volatility to the high-volatility regime or the other way around. The factor risk premium is more volatile in the high-volatility regime than in the low-volatility regime.

Table 2.7 reports the results of the risk premium regressions (2.54)–(2.56), where the exchange rate excess return is regressed on the model-implied risk premiums. In

\footnote{A stronger rejection of unit root in BIC is due to the fact that BIC penalises model complexity more heavily than AIC. As a result, fewer lags are chosen according to BIC, and since the first-order autocorrelation is low, the null is rejected. The strong rejection in PP test is consistent with the stronger rejection in BIC than in AIC, as PP test estimates a non-augmented DF test equation without adding lagged difference terms and controls for serial correlation using a nonparametric method. However, these test results reveal the uncertainty regarding the order of integration, which will be dealt with in future work.}
order to highlight the effects of regime risk premium on the excess return, we orthogonalise the regime risk premium $r_{p_r}$ and the factor risk premium $r_{p_f}$ by regressing $r_{p_f}$ on a constant and $r_{p_r}$, and use the residuals from this regression in place of $r_{p_f}$ in the final risk premium regressions. The coefficients on the risk premiums are predicted to be positive and equal to one. However, due to the scale issue discussed earlier, it is unlikely that the magnitude of the coefficients will be strictly equal or close to one, while the sign should remain positive. As shown in regressions 1–3, the currency excess return is predicted by the Japanese regime risk premium and the US factor risk premium, with the right signs. The adjusted $R^2$ increases from 1.16% to 2.00% and 5.86% as the total risk premium is decomposed into the regime and the factor risk premiums, and is further decomposed into the country-specific risk premiums. Regression 4–7 includes the forward premium as an additional regressor. As shown in regression 4, the estimated coefficient on the forward premium is positive and is significantly different from zero. This is referred to as the forward premium puzzle because UIP predicts that the coefficient should be zero. In regressions 5 and 6, the puzzle still exists, and the predictive ability of the model-implied risk premiums becomes weaker compared to those in regressions 1 and 2. In regression 7, however, when the country-specific risk premiums are included in the regression, the coefficient on the forward premium becomes smaller in magnitude and insignificantly different from zero, implying that the model-implied risk premiums are able to account for the forward premium puzzle. Table 2.8 reports the results of risk premium regressions with GARCH(1,1) effects as in equation (2.57). The estimated coefficients on the model-implied risk premiums and the forward premium are comparable to those in Table 2.7. In the variance equation, the coefficient on the lag of conditional variance is significantly different from zero, which is consistent with the time-varying variance of excess returns as shown in equation (2.32).

To assess the economic impact of changes in risk premiums, we examine a one-standard-deviation change in one predictor while keeping the other predictors constant. Using the coefficients reported for regressions 3 and 4 in Table 2.7, a one-standard-deviation (0.1478 in Table 2.6) increase in the Japanese regime risk premium predicts a monthly excess return of 0.46% ($= 0.0310 \times 0.1478$), and a one-standard-deviation (0.0694) decrease in the US factor risk premium predicts an excess return of 0.75%
These effects are comparable with the effect of the forward premium on excess returns: a one-standard-deviation (0.0019) increase in the forward premium predicts an excess return of 0.62% (= 3.2632 × 0.0019).

We then investigate whether the model-implied risk premiums contain any information beyond the forward premium. As shown in Tables 2.7 and 2.8, the adjusted $R^2$ of regression 7 doubles that of regression 4, in favour of additional information in the model-implied risk premiums. As a robustness check, we orthogonalise the forward premium and the US factor risk premium\(^5\), and Table 2.9 reports the results. As shown in Panel A, the estimated coefficient on the US factor risk premium remains significant after excluding the information in the forward premium from this variable, indicating that the US factor risk premium contains information beyond the forward premium. In contrast, Panel B shows that the estimated coefficient on the forward premium becomes smaller in magnitude and insignificantly different from zero after excluding the information in the US factor risk premium from this variable. We also highlight a period in 1995 when the Japanese regime risk premium was more successful than the forward premium in predicting the largest depreciation of yen. In August 1995, the Japanese yen depreciated against the US dollar by 10.24%—the largest depreciation of yen during the sample period from 1985 to 2009. The excess return on taking a short forward position in the Japanese yen is 10.72%. In the Japanese bond market, from June to July, the probability of switching from the high-volatility to the low-volatility regime increased from 0.2750 to 0.4079, and the probability of switching from the low-volatility to the high-volatility regime decreased from 0.9615 to 0.8594. As a result, the regime risk premium, which compensates for the risk of regime shifts per se rather than the direction of regime shifts, increased more than ten times from 0.0760 to 0.9758. A regime risk premium of 0.9758 predicts an excess return of 3.02% in the foreign exchange market, which is about one third of the realised excess return in the next month. In contrast, the forward premium was silent about the 10.72% excess return in the foreign exchange market: the forward premium in June and July was 4.3% and 4.7% respectively—no significant changes in the forward premium—and the 4.7% forward premium in July predicts an excess return of 1.54% in August. We

\(^5\)The US factor risk premium is chosen due to its moderately high correlation with the forward premium as shown in Table 2.6
then ask to what extent the regression results rely on such extreme events, and we win-
sorize the model-implied risk premiums at the upper 95th percentile—any data above the 95th percentile are set to the 95th percentile. The regression results are reported in Panel C of Table 2.9. Generally speaking, the estimated coefficients on the US factor risk premium remain highly significant, while the coefficients on the Japanese regime risk premium only become significant at 10% significance level, which suggests that the regime risk premium plays a role in predicting currency excess returns during important transitions of regimes.

2.6 Concluding Remarks

Uncovered interest rate parity (UIP) postulates that the excess return on taking a forward position in currencies is expected to be zero. The empirical violation of UIP is referred to as the forward premium puzzle, where taking a long forward position in the high interest rate currency earns a positive excess return, or equivalently, the excess return is predicted by the forward premium. Within the framework of Backus, Foresi and Telmer (2001), this excess return in the foreign exchange market is linked to the risk premiums implied from the international bond markets, and the literature often suggests a monetary channel based on Taylor rules. We contribute to the literature by documenting a new monetary mechanism, namely the shift of monetary policies, where the shift of policies (regimes) generates currency risk.

To capture the risk of regime shifts, we employ the term structure model of Dai, Singleton and Yang (2007). In this model, the risk of regime shifts is priced and compensated for by the regime risk premium. We study the USD-JPY currency pair and estimate the regime switching dynamic term structure model for the US and the Japanese bond yields. Empirical results reveal that the currency excess return is explained (predicted) by the Japanese regime risk premium and the US factor risk premium. The predictive ability of the forward premium in the forward premium puzzle weakens in the presence of model-implied risk premiums. Moreover, the Japanese regime risk premium plays a role in predicting currency excess returns during important transitions of regimes, and the US factor risk premium contains information
We highlight the predictive ability of the regime risk premium estimated from the Japanese bond market. Consistent with the extant literature, we find two regimes in the conditional volatility of the underlying latent factors—a low-volatility regime and a high-volatility regime. The high volatility regime that prevails the sample period is interpreted as a conventional monetary policy regime, and the low volatility regime from 2001 to 2006 is interpreted as an unconventional policy regime of quantitative easing. Our model reveals substantial regime uncertainty during the transition from the conventional to the unconventional policy regime from 1995 to 2000, implying substantial risk in Japanese yen, and it is in this period that the yen carry trade emerges. In August 1995, the Japanese yen depreciated against the US dollar by 10.24%—the largest depreciation of yen during the sample period from 1985 to 2009. In the Japanese bond market, from June to July, the risk premium that compensates for the risk of regime shifts increased more than ten times—exhibiting a clear signal for the depreciation in the next month. By contrast, the forward premium was silent about this depreciation—there were no significant changes in the forward premium in the period from June to August.

Having established the existence of a risk premium in the USD-JPY currency pair that compensates for the risk of regime shifts in monetary policies, we propose the following future work. Firstly, we would like to generalise our results to other currencies that show the forward premium puzzle. While we highlight the transition between conventional and unconventional monetary policies for the Japanese yen, our model can be readily applied to other transitions of regimes, such as a transition between monetary tightening and easing in conventional monetary policies or a transition between economic expansion and recession in business cycle models. Another example would be a transition of regimes in the response of monetary authorities based on Taylor-rule, as Backus, Gavazzoni, Telmer and Zin (2013) point out that the high-interest-rate Australian dollar tends to appreciate because the Australian monetary policy responses more to output gap and less to inflation relative to its US counterpart. Secondly, since the dataset of Wright (2011) ends in May 2009, the model does not tell much empirically about the recent financial crisis and the unconventional monetary
policies in the US and Japan. It is on the future agenda to include more contemporary
data, and a focal question would be the identification and the classification of regimes
in the recent data. Our empirical results currently show that the regimes in the US
and Japan are characterised by the difference in the conditional volatility of latent
factors and bond yields, and the 2001–2006 quantitative easing in Japan is classified
as in the low-volatility regime. However, it is not certain whether the quantitative
easing in the US or the recent quantitative easing in Japan are also characterised by
a low-volatility regime, or more critically, whether they are characterised by volatility
regimes. In this sense, we may need a third regime, and additional data other
than bond yields may help with the identification of regimes. These are mainly em-
pirical issues, and we leave them to future work. Thirdly, our model has important
implications for the effect of the recent unconventional monetary policies on exchange
rates. Since the end of 2008, central banks around the world has implemented several
rounds of quantitative easing purchases, and asset purchases were often accompanied
by strong currency reactions, while the direction of currency movements is mixed and
sometimes puzzling. Even though there are studies on the term structure of currency
risk premiums (see, e.g., Lustig, Stathopoulos and Verdelhan, 2016; Zviadadze, 2016),
no clear prediction is given by the academic side about how the purchase of long-term
assets will affect exchange rates. While our model does not explicitly model the effects
of quantitative easing on exchange rates either, it predicts an increase of uncertainty
as central banks enter into the unconventional monetary policy regime and therefore
a downward pressure on currencies. During the transition to the unconventional mon-
etary policy regime, the risk premium that compensates for regime shifts is expected
to increase, while the magnitude depends on the size and the implementation of asset
purchases, the signalling effect of quantitative easing, the expectation of market partic-
ipants, the credibility of monetary authorities, etc. Similarly, our model also predicts
an increase of uncertainty as central banks exit from the unconventional monetary
policy regime, thus calls for caution in the phasing out of quantitative easing. Finally,
extending beyond our assumption that the regimes in the two countries are independ-
ent, it would be interesting to consider a global economy where there are common
versus local regimes, so that regimes are expected to be more accurately identified,
and the underlying link between currency excess returns and regimes would be clearer.
CHAPTER 2. THE YEN RISK PREMIUMS

Bibliography


Table 2.1: Summary statistics of bond yield data

This table reports the summary statistics of bond yield data. The sample mean, standard deviation, maximum, minimum, autocorrelation coefficient and correlation matrix are reported. The US bond yield data is from 1971M11 to 2009M05 and the maturities are 6-month, 2-year, 5-year and 10-year. The Japanese bond yield data is from 1985M01 to 2009M05 and the maturities are 6-month, 5-year and 10-year. All data are not annualised and not in percentage.

<table>
<thead>
<tr>
<th></th>
<th>6m</th>
<th>2y</th>
<th>5y</th>
<th>10y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. US, 1971M11–2009M05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.0051</td>
<td>0.0055</td>
<td>0.0058</td>
<td>0.0062</td>
</tr>
<tr>
<td>std</td>
<td>0.0026</td>
<td>0.0025</td>
<td>0.0022</td>
<td>0.0020</td>
</tr>
<tr>
<td>max</td>
<td>0.0135</td>
<td>0.0132</td>
<td>0.0126</td>
<td>0.0124</td>
</tr>
<tr>
<td>min</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0013</td>
<td>0.0024</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.9838</td>
<td>0.9868</td>
<td>0.9896</td>
<td>0.9901</td>
</tr>
<tr>
<td>correlation</td>
<td>6m</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.9800</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5y</td>
<td>0.9362</td>
<td>0.9841</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>10y</td>
<td>0.8912</td>
<td>0.9525</td>
<td>0.9898</td>
</tr>
<tr>
<td><strong>B. US, 1985M01–2009M05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.0040</td>
<td>0.0044</td>
<td>0.0049</td>
<td>0.0054</td>
</tr>
<tr>
<td>std</td>
<td>0.0018</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0015</td>
</tr>
<tr>
<td>max</td>
<td>0.0079</td>
<td>0.0087</td>
<td>0.0095</td>
<td>0.0098</td>
</tr>
<tr>
<td>min</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0013</td>
<td>0.0024</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.9924</td>
<td>0.9889</td>
<td>0.9866</td>
<td>0.9854</td>
</tr>
<tr>
<td>correlation</td>
<td>6m</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.9760</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5y</td>
<td>0.9046</td>
<td>0.9715</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>10y</td>
<td>0.8126</td>
<td>0.9064</td>
<td>0.9792</td>
</tr>
<tr>
<td><strong>C. JP, 1985M01–2009M05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.0018</td>
<td>0.0023</td>
<td>0.0028</td>
<td></td>
</tr>
<tr>
<td>std</td>
<td>0.0020</td>
<td>0.0018</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>0.0069</td>
<td>0.0067</td>
<td>0.0064</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>–0.0000</td>
<td>0.0002</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td>autocorr</td>
<td>0.9965</td>
<td>0.9937</td>
<td>0.9935</td>
<td></td>
</tr>
<tr>
<td>correlation</td>
<td>6m</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5y</td>
<td>0.9750</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10y</td>
<td>0.9453</td>
<td>0.9904</td>
<td>1.0000</td>
</tr>
<tr>
<td><strong>D. cross-country correlation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1985M01–2009M05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>US</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>JP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6m</td>
<td>0.6218</td>
<td>0.6903</td>
<td>0.7812</td>
</tr>
<tr>
<td></td>
<td>5y</td>
<td>0.6728</td>
<td>0.7527</td>
<td>0.8414</td>
</tr>
<tr>
<td></td>
<td>10y</td>
<td>0.6605</td>
<td>0.7505</td>
<td>0.8442</td>
</tr>
</tbody>
</table>
Table 2.2: Summary statistics of foreign exchange data

This table reports the summary statistics of foreign exchange data. The sample mean, standard deviation, maximum, minimum, autocorrelation coefficient and correlation matrix are reported. Foreign exchange data includes exchange rate changes $\Delta s$, forward premia $fp$, and exchange rate excess returns $rx$. The sample period is from 1985M01 to 2009M06. All data are not annualised and not in percentage.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta s_{t+1}$</th>
<th>$fp_t$</th>
<th>$rx_{t+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0033</td>
<td>0.0025</td>
<td>−0.0008</td>
</tr>
<tr>
<td>std</td>
<td>0.0336</td>
<td>0.0019</td>
<td>0.0340</td>
</tr>
<tr>
<td>max</td>
<td>0.1554</td>
<td>0.0072</td>
<td>0.1072</td>
</tr>
<tr>
<td>min</td>
<td>−0.1024</td>
<td>−0.0043</td>
<td>−0.1510</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.0375</td>
<td>0.8102</td>
<td>0.0563</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>correlation</th>
<th>$\Delta s_{t+1}$</th>
<th>$fp_t$</th>
<th>$rx_{t+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta s_{t+1}$</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fp_t$</td>
<td>−0.1308</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>$rx_{t+1}$</td>
<td>−0.9984</td>
<td>0.1860</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
This table reports the estimates of structural parameters in the US three-factor regime switching dynamic term structure model. Standard errors are in brackets.

<table>
<thead>
<tr>
<th>regime L (low volatility)</th>
<th>regime H (high volatility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent factor under the $P$-measure, given $q_t = j$</td>
<td>$F_{t+1} = F_t - \kappa^P_j \cdot \left[ F_t - \theta^P_j \right] + \Sigma^P \cdot \delta^P_{t+1}$</td>
</tr>
<tr>
<td>$\theta^P$</td>
<td>0</td>
</tr>
<tr>
<td>$\kappa^P$</td>
<td>0.0288</td>
</tr>
<tr>
<td></td>
<td>[0.0235]</td>
</tr>
<tr>
<td></td>
<td>-0.0023</td>
</tr>
<tr>
<td></td>
<td>[0.0209]</td>
</tr>
<tr>
<td></td>
<td>-0.1331</td>
</tr>
<tr>
<td></td>
<td>[0.0392]</td>
</tr>
<tr>
<td>eigenvalue</td>
<td>0.0767</td>
</tr>
<tr>
<td>$\sqrt{\Sigma} \times \Sigma$</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>[0.1564]</td>
</tr>
<tr>
<td>Latent factor under the $Q$-measure, given $q_t = j$</td>
<td>$F_{t+1} = F_t - \kappa^Q_j \cdot \left[ F_t - \theta^Q_j \right] + \Sigma^Q \cdot \delta^Q_{t+1}$</td>
</tr>
<tr>
<td>$\theta^Q$</td>
<td>1.66</td>
</tr>
<tr>
<td>$\kappa^Q$</td>
<td>[19.54]</td>
</tr>
<tr>
<td></td>
<td>0.0296</td>
</tr>
<tr>
<td></td>
<td>[0.0077]</td>
</tr>
<tr>
<td></td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td>[0.0113]</td>
</tr>
<tr>
<td></td>
<td>-0.1615</td>
</tr>
<tr>
<td></td>
<td>[0.0234]</td>
</tr>
<tr>
<td>eigenvalue</td>
<td>0.0878</td>
</tr>
</tbody>
</table>
| Price of risk, given $q_t = j$ | $\lambda_j(q_j) = (\Sigma^j)^{-1} \left[ \lambda^j_0 + \lambda^j_1 F_1 \right]$,  
| where $\lambda^j_0 = \kappa^P_j \theta^P_j - \kappa^Q_j \theta^Q_j$, $\lambda^j_1 = \kappa^Q_j - \kappa^P_j$ |  
| $\lambda_0$ | -0.0149 | 0.0312 | -0.0548 | 0.0008 | 0.0563 | -0.0400 |
| | [0.0684] | [0.3389] | [0.7931] | [0.5855] | [1.9355] | [2.9748] |
| $\lambda_1$ | 0.0008 | -0.0061 | 0.0088 | -0.0670 | 0.0295 | 0.0275 |
| | [0.0224] | [0.0217] | [0.0127] | [0.0294] | [0.0487] | [0.0133] |
| | 0.0095 | 0.0023 | 0.0096 | 0.0121 | -0.1597 | -0.0077 |
| | [0.0191] | [0.0192] | [0.0120] | [0.0511] | [0.0868] | [0.0223] |
| | -0.0285 | -0.0242 | 0.0055 | -0.0358 | -0.1798 | 0.0062 |
| | [0.0217] | [0.0245] | [0.0129] | [0.0049] | [0.1559] | [0.0515] |
| One-period bond yield, given $q_t = j$ | $r_t = \delta_0^j + \delta_1^j F_1$  
| $12 \times \delta_0[1e-2]$ | 5.2237 | 4.9091 |
| | [8.56] | [8.55] |
| $12 \times \delta_1[1e-2]$ | 0.1212 | -0.2449 | 0.7142 |
| | [0.0781] | [0.0863] | [0.0353] |
| Transition probability of regimes, given $q_t = j$ and $q_{t+1} = k$ | $\pi^Q_{jk} = \text{const.}$, $\pi^P_{jk,t} = \frac{1}{1 + e^{(\pi^Q_{jk} + \pi^P_{jk,t})}}$, $j \neq k$. |
| $\pi^Q_{j1}$ | 0.9219 | 0.8301 |
| $\pi^P_{j1,t}$ | L2H | H2L |
| $\eta_0$ | 2.6258 | 1.3512 |
| $\eta_1$ | 0.365 | -0.4967 | -0.0765 | -0.5115 | 0.0867 | 0.4760 |
| | [0.2670] | [0.3106] | [0.1584] | [0.2584] | [0.3294] | [0.1668] |
Table 2.4: Parameter estimates in the Japanese term structure model

This table reports the estimates of structural parameters in the Japanese two-factor regime switching dynamic term structure model. Standard errors are in brackets.

<table>
<thead>
<tr>
<th></th>
<th>regime L (low volatility)</th>
<th>regime H (high volatility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent factor under the $P$-measure, given $q_t = j$ ( F_{t+1} = F_t - \kappa_P \cdot \left[ F_t - \theta_P \right] + \Sigma' \cdot u_{t+1}^P )</td>
<td>( \theta^P )</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[5.25]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0290]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0310]</td>
</tr>
<tr>
<td></td>
<td>eigenvalue</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>( \sqrt{\Sigma \times \Sigma} )</td>
<td>[0.1364]</td>
</tr>
<tr>
<td>Latent factor under the $Q$-measure, given $q_t = j$ ( F_{t+1} = F_t - \kappa_Q \cdot \left[ F_t - \theta_Q \right] + \Sigma' \cdot u_{t+1}^Q )</td>
<td>( \theta^Q )</td>
<td>-11.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[15.76]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0298]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0304]</td>
</tr>
<tr>
<td></td>
<td>eigenvalue</td>
<td>0.0251</td>
</tr>
<tr>
<td>Price of risk, given $q_t = j$ ( \lambda_1(q_t) = (\Sigma')^{-1} \left( \lambda_0^P + \lambda_1^P \right) F_t ), where $\lambda_0^P = \kappa_P \theta_P - \kappa_P \theta_Q$, $\lambda_1^P = \kappa_Q - \kappa_P$</td>
<td>( \lambda_0 )</td>
<td>0.0175</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0293]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0291]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0421]</td>
</tr>
<tr>
<td>One-period bond yield, given $q_t = j$ ( r_t = \delta_0 + \delta_1 F_t )</td>
<td>12 \times \delta_0 [1e-2]</td>
<td>0.1012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.30]</td>
</tr>
<tr>
<td>Transition probability of regimes, given $q_t = j$ and $q_{t+1} = k$ ( \pi^Q_{jk} = \text{const.}; \pi^P_{jk} = \frac{1}{1 + e^{(\eta_0^P q_{j} + \eta_1^P r_p)}} ), ( j \neq k )</td>
<td>( \pi^Q_{jj} )</td>
<td>0.8155</td>
</tr>
<tr>
<td></td>
<td>L2H</td>
<td>H2L</td>
</tr>
<tr>
<td></td>
<td>( \eta_0 )</td>
<td>1.9509</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.7391]</td>
</tr>
<tr>
<td></td>
<td>( \eta_1 )</td>
<td>0.6072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.3097]</td>
</tr>
</tbody>
</table>
Table 2.5: Summary statistics of transition probabilities and latent factors

This table reports the summary statistics of the model-implied transition probabilities and latent factors. The sample mean, standard deviation, maximum, minimum, autocorrelation coefficient and correlation matrix are reported. The transition probability \( j \rightarrow k, j, k = L, H, j \neq k \) is time \( t \)'s probability of switching from regime \( j \) at time \( t \) to regime \( k \) at time \( t + 1 \). The US sample is from 1971M11 to 2009M05 and the Japanese sample is from 1985M01 to 2009M05.

<table>
<thead>
<tr>
<th></th>
<th>A. US, 1971M11–2009M05</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2H</td>
<td>H2L</td>
<td>factor 1</td>
<td>factor 2</td>
</tr>
<tr>
<td>mean</td>
<td>0.0930</td>
<td>0.2186</td>
<td>0.5268</td>
<td>0.1798</td>
</tr>
<tr>
<td>std</td>
<td>0.0615</td>
<td>0.1536</td>
<td>2.2095</td>
<td>1.3156</td>
</tr>
<tr>
<td>max</td>
<td>0.4356</td>
<td>0.7245</td>
<td>7.0893</td>
<td>4.2086</td>
</tr>
<tr>
<td>min</td>
<td>0.0219</td>
<td>0.0005</td>
<td>−3.5013</td>
<td>−2.3005</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.8750</td>
<td>0.9437</td>
<td>0.9854</td>
<td>0.9323</td>
</tr>
<tr>
<td>correlation</td>
<td>L2H</td>
<td>H2L</td>
<td>factor 1</td>
<td>factor 2</td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>factor 1</td>
<td>0.6042</td>
<td>−0.2252</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>factor 2</td>
<td>0.8395</td>
<td>0.3754</td>
<td>0.4087</td>
<td>1.0000</td>
</tr>
<tr>
<td>factor 3</td>
<td>0.4546</td>
<td>−0.7418</td>
<td>0.7366</td>
<td>0.0186</td>
</tr>
</tbody>
</table>

|                  | B. US, 1985M01–2009M05 |                  |                  |                  |
|                  | L2H         | H2L         | factor 1 | factor 2 | factor 3 |
| mean             | 0.0844      | 0.2554      | −0.2734  | 0.2246   | −0.5687  |
| std              | 0.0496      | 0.1508      | 1.7457   | 1.3898   | 2.6489   |
| max              | 0.3381      | 0.7245      | 4.9724   | 4.2086   | 5.2768   |
| min              | 0.0219      | 0.0380      | −3.5013  | −2.3005  | −5.7716  |
| autocorr         | 0.8743      | 0.9478      | 0.9795   | 0.9543   | 0.9890   |
| correlation      | L2H         | H2L         | factor 1 | factor 2 | factor 3 |
|                  | 1.0000      | 1.0000      |          |          |          |
| factor 1         | 0.4674      | 0.0016      | 1.0000   |          |          |
| factor 2         | 0.9307      | 0.5837      | 0.3613   | 1.0000   |          |
| factor 3         | −0.0152     | −0.7290     | 0.6616   | −0.2405  | 1.0000   |

|                  | C. JP, 1985M01–2009M05 |                  |                  |                  |
|                  | L2H         | H2L         | factor 1 | factor 2 |
| mean             | 0.5816      | 0.2236      | 1.9063   | 32.5344  |
| std              | 0.4108      | 0.1998      | 2.3264   | 41.9420  |
| max              | 1.0000      | 0.6801      | 6.0556   | 139.0944 |
| min              | 0.0194      | 0.0000      | −2.3693  | −3.3117  |
| autocorr         | 0.9831      | 0.9861      | 0.9829   | 0.9962   |
| correlation      | L2H         | H2L         | factor 1 | factor 2 |
|                  | 1.0000      | −0.8599     | 1.0000   |          |
| factor 1         | 0.8982      | −0.7071     | 1.0000   |          |
| factor 2         | 0.7931      | −0.8447     | 0.7834   | 1.0000   |

|                  | D. cross-country correlation, 1985M01–2009M05 |                  |                  |                  |
|                  | US          |                  |                  |                  |
|                  | L2H         | H2L         | factor 1 | factor 2 | factor 3 |
| JP               | 0.3631      | −0.1759     | 0.7719   | 0.2710   | 0.6377   |
| H2L              | −0.5026     | 0.0716      | −0.6915  | −0.4197  | −0.5118  |
| factor 1         | 0.3889      | −0.0651     | 0.8625   | 0.2925   | 0.6124   |
| factor 2         | 0.5100      | −0.1496     | 0.8061   | 0.3932   | 0.6463   |
This table reports the summary statistics of the model-implied risk premiums with the actual currency excess return $r_x$ and the actual forward premium $f_p$. The total risk premium $r_p$ is decomposed into the regime risk premium $r_{p_r}$ and the factor risk premium $r_{p_f}$, and is further decomposed into the country-specific risk premiums. The sample mean, standard deviation, maximum, minimum, autocorrelation coefficient, p-values of unit root tests and correlation matrix are reported. A null hypothesis of unit root is tested in the level of a series, including an intercept in the test equation. In the Augmented Dickey-Fuller (ADF) test, the lag length is selected via AIC and BIC rules with a maximum lag of 15. The Phillips-Perron test estimates a non-augmented DF test equation, and use Bartlett kernel and Newey-West Bandwidth. The sample period is from 1985M01 to 2009M06.

<table>
<thead>
<tr>
<th></th>
<th>$r_{x_{t+1}}$</th>
<th>$f_p$</th>
<th>$r_p$</th>
<th>$r_{p_r}$</th>
<th>$r_{p_f}$</th>
<th>$r_{p_r}^{IP}$</th>
<th>$r_{p_f}^{IP}$</th>
<th>$r_{p_f}^{US}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.0008</td>
<td>0.0025</td>
<td>1.4769</td>
<td>0.1180</td>
<td>1.3589</td>
<td>0.1409</td>
<td>-0.0229</td>
<td>1.4192</td>
</tr>
<tr>
<td>std</td>
<td>0.0340</td>
<td>0.0019</td>
<td>3.0799</td>
<td>0.1584</td>
<td>2.9945</td>
<td>0.1478</td>
<td>0.0430</td>
<td>2.9767</td>
</tr>
<tr>
<td>max</td>
<td>0.1072</td>
<td>0.0072</td>
<td>22.3068</td>
<td>0.9734</td>
<td>21.3629</td>
<td>0.9758</td>
<td>-0.0000</td>
<td>21.4375</td>
</tr>
<tr>
<td>min</td>
<td>-0.1510</td>
<td>-0.0043</td>
<td>-0.4114</td>
<td>-0.3637</td>
<td>-0.2949</td>
<td>0.0013</td>
<td>-0.4698</td>
<td>0.0001</td>
</tr>
<tr>
<td>autocorr</td>
<td>0.0563</td>
<td>0.8102</td>
<td>0.2625</td>
<td>0.5647</td>
<td>0.2419</td>
<td>0.5522</td>
<td>0.4171</td>
<td>0.2327</td>
</tr>
<tr>
<td>unit root tests</td>
<td>ADF (AIC)</td>
<td>0.0002</td>
<td>0.3284</td>
<td>0.0875</td>
<td>0.0357</td>
<td>0.0838</td>
<td>0.0218</td>
<td>0.1701</td>
</tr>
<tr>
<td></td>
<td>ADF (BIC)</td>
<td>0.0000</td>
<td>0.3284</td>
<td>0.0034</td>
<td>0.0082</td>
<td>0.0029</td>
<td>0.0033</td>
<td>0.1701</td>
</tr>
<tr>
<td></td>
<td>Phillips-Perron</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>correlation</th>
<th>$r_{x_{t+1}}$</th>
<th>$f_p$</th>
<th>$r_p$</th>
<th>$r_{p_r}$</th>
<th>$r_{p_f}$</th>
<th>$r_{p_r}^{IP}$</th>
<th>$r_{p_f}^{IP}$</th>
<th>$r_{p_f}^{US}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_p$</td>
<td>0.1869</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_p$</td>
<td>0.1214</td>
<td>0.3503</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{p_r}$</td>
<td>0.1677</td>
<td>0.2966</td>
<td>0.5572</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{p_f}$</td>
<td>0.1160</td>
<td>0.3446</td>
<td>0.9990</td>
<td>0.5202</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{p_f}^{IP}$</td>
<td>0.1760</td>
<td>0.2375</td>
<td>0.5675</td>
<td>0.9629</td>
<td>0.5327</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-r_{p_f}^{US}$</td>
<td>0.0126</td>
<td>0.2763</td>
<td>0.1019</td>
<td>0.3738</td>
<td>0.0851</td>
<td>0.1095</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>$r_{p_f}^{IP}$</td>
<td>0.1117</td>
<td>0.3330</td>
<td>0.9987</td>
<td>0.5183</td>
<td>0.9997</td>
<td>0.5325</td>
<td>0.0790</td>
<td>1.0000</td>
</tr>
<tr>
<td>$-r_{p_f}^{US}$</td>
<td>0.2159</td>
<td>0.5861</td>
<td>0.2703</td>
<td>0.2131</td>
<td>0.2667</td>
<td>0.1471</td>
<td>0.2793</td>
<td>0.2450</td>
</tr>
</tbody>
</table>
Table 2.7: Risk premium regressions: ordinary least squares

This table reports the estimated coefficients in the risk premium regressions where the error term has a constant variance. In regressions 1–3, the exchange rate excess return is regressed on the model-implied risk premiums, and in regressions 4–7, the forward premium is also included as a regressor. The regime risk premium $r_p$ and the factor risk premium $r_f$ are orthogonalised by regressing $r_f$ on a constant and $r_p$, and the residual from this regression, denoted by $f_{p_f(r_p)}$, is used in place of $r_f$ in the risk premium regression. HAC (Newey-West) standard errors are in brackets. *, ** and *** reject the null hypothesis of no predictive ability at 10%, 5% and 1% level respectively. Adjusted R-squared is reported in percentage.

|   | $\text{const.}$ | $f_p$ | $r_p$ | $r_p(r_p)$ | $r_p^{IP}$ | $-r_p^{US}$ | $r_p^{IP}$ | $-r_p^{US}$ | $\bar{R}^2$(%)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0026</td>
<td>0.0012**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>[0.0023]</td>
<td>[0.0006]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.0046*</td>
<td>0.0319**</td>
<td>0.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>[0.0026]</td>
<td>[0.0129]</td>
<td>[0.0008]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.0054*</td>
<td>0.0310**</td>
<td>-0.0069</td>
<td>-0.0003</td>
<td>0.1077***</td>
<td></td>
<td></td>
<td></td>
<td>5.86</td>
</tr>
<tr>
<td></td>
<td>[0.0030]</td>
<td>[0.0135]</td>
<td>[0.0038]</td>
<td>[0.0007]</td>
<td>[0.0258]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.0090***</td>
<td>3.2632***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>[0.0028]</td>
<td>[0.9658]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.0090***</td>
<td>2.8711***</td>
<td>0.0007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.26</td>
</tr>
<tr>
<td></td>
<td>[0.0027]</td>
<td>[1.0183]</td>
<td>[0.0006]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.0104***</td>
<td>2.6956***</td>
<td>0.0234*</td>
<td>-0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.79</td>
</tr>
<tr>
<td></td>
<td>[0.0030]</td>
<td>[1.0138]</td>
<td>[0.0134]</td>
<td>[0.0007]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.0090***</td>
<td>1.4145</td>
<td>0.0286**</td>
<td>-0.0233</td>
<td>-0.0004</td>
<td>0.0872***</td>
<td></td>
<td></td>
<td>5.93</td>
</tr>
<tr>
<td></td>
<td>[0.0034]</td>
<td>[1.1591]</td>
<td>[0.0139]</td>
<td>[0.0407]</td>
<td>[0.0006]</td>
<td>[0.0301]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.8: Risk premium regressions: GARCH(1,1)

This table reports the estimated coefficients in the risk premium regressions where the error term follows GARCH(1,1) process as in equation (2.57). In regressions 1–3, the exchange rate excess return is regressed on model-implied risk premiums, and in regressions 4–7, the forward premium is also included as a regressor. The regime risk premium \( r_p \) and the factor risk premium \( r_{f_p} \) are orthogonalised by regressing \( r_{f_p} \) on a constant and \( r_p \), and the residual from this regression, denoted by \( f_{r_{f_p}}(r_p) \), is used in place of \( r_{f_p} \) in the risk premium regression. Standard errors are in brackets. *, ** and *** reject the null hypothesis of no predictive ability at 10%, 5% and 1% level respectively. Adjusted R-squared is reported in percentage.

<table>
<thead>
<tr>
<th></th>
<th>mean equation</th>
<th>variance equation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>const.</td>
<td>( f_{r_p} )</td>
<td>( r_p )</td>
</tr>
<tr>
<td>1</td>
<td>-0.0029</td>
<td>0.0012**</td>
<td>0.0006</td>
</tr>
<tr>
<td>2</td>
<td>-0.0045*</td>
<td>0.0297***</td>
<td>0.0004</td>
</tr>
<tr>
<td>3</td>
<td>-0.0055*</td>
<td>0.0295**</td>
<td>-0.0087</td>
</tr>
<tr>
<td></td>
<td>[0.0030]</td>
<td></td>
<td>0.0009</td>
</tr>
<tr>
<td>4</td>
<td>-0.0085**</td>
<td>3.1774***</td>
<td>0.0007</td>
</tr>
<tr>
<td>5</td>
<td>-0.0085**</td>
<td>2.7472**</td>
<td>0.0007</td>
</tr>
<tr>
<td>6</td>
<td>-0.0096***</td>
<td>2.6207**</td>
<td>0.0209*</td>
</tr>
<tr>
<td>7</td>
<td>-0.0083*</td>
<td>1.2254</td>
<td>0.0271**</td>
</tr>
</tbody>
</table>
Table 2.9: Risk premium regressions: robustness checks

This table reports the estimated coefficients in the risk premium regressions. Panels A and B orthogonalise the forward premium and the US factor risk premium. Panel C winsorizes risk premiums at their upper 95th percentile. In brackets are standard errors (Newey-West standard errors for OLS regressions). *, ** and *** reject the null hypothesis of no predictive ability at 10%, 5% and 1% level respectively. Adjusted R-squared is reported in percentage.

<table>
<thead>
<tr>
<th></th>
<th>mean equation</th>
<th>variance equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. orthogonalise ( r_{US}^P ) from ( f_P )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>const.</td>
<td>( f_{Pt} )</td>
</tr>
<tr>
<td>1</td>
<td>-0.0146***</td>
<td>3.1660***</td>
</tr>
<tr>
<td></td>
<td>[0.0032]</td>
<td>[1.0085]</td>
</tr>
<tr>
<td>2</td>
<td>-0.0140***</td>
<td>3.0308**</td>
</tr>
<tr>
<td></td>
<td>[0.0046]</td>
<td>[1.1871]</td>
</tr>
<tr>
<td></td>
<td>B. orthogonalise ( f_P ) from ( r_{US}^P )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>const.</td>
<td>( f_{Pt} (\langle r_{US}^P \rangle) )</td>
</tr>
<tr>
<td>3</td>
<td>-0.0001</td>
<td>1.3667</td>
</tr>
<tr>
<td></td>
<td>[0.0037]</td>
<td>[1.1786]</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>1.1927</td>
</tr>
<tr>
<td></td>
<td>[0.0034]</td>
<td>[1.3626]</td>
</tr>
<tr>
<td></td>
<td>C. winsorize risk premiums at the upper 95th percentile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>const.</td>
<td>( f_{Pt} )</td>
</tr>
<tr>
<td>5</td>
<td>-0.0055</td>
<td>0.0397*</td>
</tr>
<tr>
<td></td>
<td>[0.0040]</td>
<td>[0.0208]</td>
</tr>
<tr>
<td>6</td>
<td>-0.0054</td>
<td>0.0355*</td>
</tr>
<tr>
<td></td>
<td>[0.0039]</td>
<td>[0.0187]</td>
</tr>
<tr>
<td>7</td>
<td>-0.0078*</td>
<td>0.8189</td>
</tr>
<tr>
<td></td>
<td>[0.0044]</td>
<td>[1.5023]</td>
</tr>
<tr>
<td>8</td>
<td>-0.0067</td>
<td>0.5009</td>
</tr>
<tr>
<td></td>
<td>[0.0059]</td>
<td>[1.4188]</td>
</tr>
</tbody>
</table>
Figure 2.1: Bond yield data. This figure plots in A the US bond yields from November 1971 to May 2009, and in B the Japanese bond yields from January 1985 to May 2009. The maturities are 6-month, 2-year, 5-year and 10-year for the US data, and are 6-month, 5-year and 10-year for the Japanese data. Bond yields are annualised and in percentage.
Figure 2.2: Foreign exchange data. This figure plots in A the exchange rate change, in B the forward premium, and in C the exchange rate excess return. The sample period is from 1985M01 to 2009M06. All data are not annualised and not in percentage.
Figure 2.3: Forward premium and interest rate differentials. This figure plots in A the annualised US and Japanese short-term yields, and in C the forward premium and interest rate differentials (not annualised). The sample period is from 1985M01 to 2009M06. The 1-month bond yield is 1-month LIBOR rate from ICE Benchmark Administration (IBA) in Datastream, where the start date is 1986M01 for US data and 1986M07 for Japanese data. All data are not in percentage.
Figure 2.4: US factor loadings of bond yields. This figure plots loadings of bond yields on the latent factors as a function of bond maturities in months. Bond yields depend on the latent factor as in equation (2.12), i.e., $y_{n,t} = a_n + b_n' F_t$, where $a_n$ is a regime-dependent scalar, and $b_n$ is a regime-independent vector. Bond yields are annualised and not in percentage.
Figure 2.5: Japanese factor loadings of bond yields. This figure plots loadings of bond yields on the latent factors as a function of bond maturities in months. Bond yields depend on the latent factor as in equation (2.12), i.e., \( y_{n,t} = a_n + b_n^t F_t \), where \( a_n \) is a regime-dependent scalar, and \( b_n \) is a regime-independent vector. Bond yields are annualised and not in percentage.
Figure 2.6: Model-implied latent factors. This figure plots in A the latent factors implied by the US regime switching dynamic term structure model, and in B and C the first and the second latent factor implied by the Japanese model.
Figure 2.7: US filtered, smoothed and transition probabilities. This figure plots in A the filtered and in B the smoothed probabilities (shaded panels are NBER recessions), and in C the transition probabilities (shaded periods are the high-volatility regimes classified by the smoothed probability). The transition probability $j \rightarrow k, j, k = L, H, j \neq k$, is time $t$’s probability of switching from regime $j$ at time $t$ to regime $k$ at time $t+1$. 
Figure 2.8: Japanese filtered, smoothed and transition probabilities. This figure plots in A the filtered and in B the smoothed probabilities (shaded panels are ECRI recessions), and in C the transition probabilities (shaded periods are the high-volatility regimes classified by the smoothed probability). The transition probability $j_2k$, $j, k = L, H$, $j \neq k$, is time $t$’s probability of switching from regime $j$ at time $t$ to regime $k$ at time $t + 1$. 
Figure 2.9: US risk premiums. This figure plots in A the regime risk premium and in B the factor risk premium implied by the US regime switching dynamic term structure model. Shaded periods are the high-volatility regimes classified by the smoothed probability.
Figure 2.10: Japanese risk premiums. This figure plots in A the regime risk premium and in B the factor risk premium implied by the Japanese regime switching dynamic term structure model. Shaded periods are the high-volatility regimes classified by the smoothed probability.
Appendix

2.A Bond Prices

This section derives the dependence of bond prices on the latent factors. We consider a zero-coupon bond with a maturity of \((n + 1)\)-month at time \(t\). Given the current regime \(s_t = j\), the price of this bond, denoted by \(P_{n+1,t}\), is conjectured to depend on the latent factors \(F_t\) as

\[
P^{j}_{n+1,t} = e^{-A^j_{n+1} - B^j_{n+1} F_t},
\]  

(2.66)

where \(A\) is a regime-dependent scalar, and \(B\) is a regime-independent \(N \times 1\) vector.

The price of this bond is also equal to the expected price at time \(t + 1\) under the \(Q\)-measure discounted at the one-period risk free rate, i.e., given \(q_t = j\) and \(q_{t+1} = k\),

\[
P^{j}_{n+1,t} = E^Q_t\left[e^{-r^j_t P_{n,t+1} | q_t = j}\right]
= e^{-r^j_t} \sum_{k=L,H} \pi^Q_{jk} \cdot E^Q_t\left[P_{n,t+1} | q_t = j, q_{t+1} = k\right]
= e^{-r^j_t}\sum_{k=L,H} \pi^Q_{jk} \cdot E^Q_t\left[e^{-A^j_{n} - B^j_{n} F_{t+1}} | q_t = j, q_{t+1} = k\right]
= e^{-r^j_t}\left(\sum_{k=L,H} \pi^Q_{jk} \cdot e^{-A^j_{n}}\right) \cdot E^Q_t\left[e^{-B^j_{n} F_{t+1}} | q_t = j\right]
= e^{-r^j_t}\left(\sum_{k=L,H} \pi^Q_{jk} \cdot e^{-A^j_{n}}\right) \cdot e^{-B^j_{n} \mu_{q_t} + \frac{1}{2} B^j_{n} \Sigma^Q_{j} B^j_{n}}.
\]

(2.67)

Equating the two prices in equations (2.66) and (2.67) gives the relationships that \(A\) and \(B\) must satisfy under the assumption of no arbitrage.
2.B Price of Factor Risk

This section derives the relationship between the conditional means of the latent factors and the price of factor risk. We consider a security with payoff $e^{b'F_{t+1}}$. Given the current regime $q_t = j$, this security is priced under the $P$-measure and under the $Q$-measure, and the two prices should be equal, i.e.,

$$P_j = E_t^P \left[ e^{b'F_{t+1}M_{t,t+1}} \right] = e^{-r_j} E_t^Q \left[ e^{b'F_{t+1}} \right].$$

It follows that

$$E_t^P \left[ e^{b'F_{t+1}} e^{-r_j} \lambda'_j - \lambda'_u(t+1) \right] = e^{-r_j} E_t^Q \left[ e^{b'F_{t+1}} \right]$$

$$E_t^P \left[ e^{b'F_{t+1} - \frac{1}{2} \lambda'j' \lambda'_j} \right] = E_t^Q \left[ e^{b'F_{t+1}} \right]$$

$$E_t^P \left[ e^{b'F_{t+1} - \frac{1}{2} \lambda'j' \lambda'_j} \right] = E_t^Q \left[ e^{b'F_{t+1}} \right]$$

$$b' \mu^{P_j} - \frac{1}{2} \lambda'_j \lambda'_j + \frac{1}{2} \left( \lambda'_j - b' \Sigma \right)' \left( \lambda'_j - b' \Sigma \right) = b' \mu^{Q_j} + \frac{1}{2} b' \Sigma \left( b' \Sigma \right)'$$

$$b' \left( \mu^{P_j} - \Sigma \lambda'_j \right) = b' \left( \mu^{Q_j} \right)$$

which gives the relationship between the conditional mean of the latent factors under the physical and the risk neutral measure and the price of factor risk.

2.C Reduced-Form Models

This section derives the reduced-form models of the observable bond yields. According to equation (2.12), bond yields depend on the latent factors as

$$y_{n,t}(q_t) = \frac{A_n(q_t)}{n/12} + \frac{B'_n}{n/12} F_t = a_n(q_t) + b'_n F_t, \quad (2.12)$$

where $y_{n,t}$ is the annualised bond yield on a zero-coupon bond with a maturity of $n$ months; $F_t$ is an $N \times 1$ vector of latent factors; $a_n(q_t)$ is a scalar with two regimes; the factor loading $b_n$ is an $N \times 1$ vector with a single regime. It is shown by equations (2.10) and (2.11) that $a$ and $b$ are functions of structural parameters, i.e.,

$$a = f(\delta_0, \theta^Q, \Sigma, \pi^Q, \delta_1, \kappa^Q) \quad (2.68)$$

$$b = f(\delta_1, \kappa^Q), \quad (2.69)$$
where $a$ depends on $\{\delta_1, \kappa^Q\}$ through its dependence on $b$.

Let $Y_{1,t} = \{y_{n_1,t}, y_{n_2,t}, \ldots, y_{n_N,t}\}'$ denote an $N \times 1$ vector of annualised bond yields that are priced without error, where the maturities of these bonds are months $\{n_1, n_2, \ldots, n_N\}$. According to equation (2.12), the bond yield vector $Y_{1,t}$ is a function of the latent factors $F_t$, i.e., given the current regime $q_t = j$,

$$Y_{1,t} = \alpha^j_1 + \beta_1 F_t,$$

where

$$\alpha^j_1 = \left[ a^j_{n_1}, a^j_{n_2}, \ldots, a^j_{n_N} \right]'$$

$$\beta_1 = \left[ b_{n_1}, b_{n_2}, \ldots, b_{n_N} \right]' .$$

Specifically, $\alpha^j_1$ is an $N \times 1$ regime-dependent vector, and $\beta_1$ is an $N \times N$ matrix of factor loadings that has a single regime. For $\forall i = 1, \ldots, N$, $a^i_{n_i}$ is the regime-dependent scalar and $b_{n_i}$ is the $N \times 1$ regime-independent vector, such that the $n_i$-month bond yield depends on the latent factors as given in equation (2.12). Using equation (2.70) and the process of latent factors in equation (2.1), it can be shown that $Y_{1,t}$ follows an autoregressive process. Specifically, given the regimes in the next and the current periods, i.e., $q_{t+1} = k$ and $q_t = j$,

$$Y_{1,t+1} = \alpha^k_1 + \beta_1 \cdot F_{t+1}$$

$$= \alpha^k_1 + \beta_1 \cdot \left[ F_t - \kappa^j P_j \cdot \left( F_t - \theta^j P_j \right) + \Sigma^j u_{t+1} \right]$$

$$= \alpha^k_1 + \beta_1 \cdot \left[ \beta_1^{-1} (Y_{1,t} - \alpha^j_1) - \kappa^j P_j \cdot \left[ \beta_1^{-1} (Y_{1,t} - \alpha^j_1) - \theta^j P_j \right] + \Sigma^j u_{t+1} \right]$$

$$= \alpha^k_1 + (Y_{1,t} - \alpha^j_1) - \beta_1 \cdot \kappa^j P_j \cdot \beta_1^{-1} (Y_{1,t} - \alpha^j_1 - \beta_1 \cdot \theta^j P_j) + \beta_1 \cdot \Sigma^j u_{t+1} .$$

Define

$$\theta^j_1 \equiv \beta_1 \cdot \theta^j P_j$$

$$\kappa^j_{11} \equiv \beta_1 \cdot \kappa^j P_j \cdot \beta_1^{-1}$$

$$\Sigma^j_1 \equiv \beta_1 \cdot \Sigma^j ,$$

then given $q_{t+1} = k$ and $q_t = j$, the process of $Y_{1,t}$ is written as

$$Y_{1,t+1} = \alpha^k_1 + (Y_{1,t} - \alpha^j_1) - \kappa^j_{11} (Y_{1,t} - \alpha^j_1 - \theta^j_1) + \Sigma^j_1 u_{t+1} ,$$
CHAPTER 2. THE YEN RISK PREMIUMS

which is equation (2.34).

Let \( Y_{2,t} = \{y_{m_1,t}, y_{m_2,t}, \ldots, y_{m_M,t}\}' \) denote an \( M \times 1 \) vector of annualised bond yields that are priced with error, where the maturities of these bonds are months \( \{m_1, m_2, \ldots, m_M\} \). The bond yield vector \( Y_{2,t} \) depends on the latent factors according to equation (2.12) and since these bonds are priced with error, they have their own error terms, i.e., given the current regime \( q_t = j \),

\[
Y_{2,t} = \alpha_j^2 + \beta_2 F_t + \Sigma_j^2 v_t,
\]

where

\[
\alpha_j^2 = [a_{m_1}^j, a_{m_2}^j, \ldots, a_{m_M}^j]'
\]

(2.78)

\[
\beta_2 = [b_{m_1}, b_{m_2}, \ldots, b_{m_M}]'.
\]

(2.79)

Specifically, \( \alpha_j^2 \) is an \( M \times 1 \) regime-dependent vector, and \( \beta_2 \) is an \( M \times N \) matrix of factor loadings that has a single regime, where \{\( \alpha_1^2, \beta_1 \)\} in equations (2.71) and (2.72); \( \Sigma_j^2 \) is an \( M \times M \) constant and regime-dependent matrix; \( v_t \sim i.i.d. N(0, I_M) \). Rewrite equation (2.35) so that \( Y_{2,t} \) depends on \( Y_{1,t} \), i.e.,

\[
Y_{2,t} = \alpha_j^2 + \beta_2 F_t + \Sigma_j^2 v_{t+1}
\]

\[
= \alpha_j^2 + \beta_2 \cdot \beta_1^{-1}(Y_{1,t} - \alpha_1^2) + \Sigma_j^2 v_t.
\]

(2.80)

Define

\[
\phi_{21} = \beta_2 \cdot \beta_1^{-1},
\]

(2.81)

then given \( q_t = j \), the process of \( Y_{2,t} \) is written as

\[
Y_{2,t} = \alpha_j^2 + \phi_{21}(Y_{1,t} - \alpha_1^2) + \Sigma_j^2 v_t,
\]

which is equation (2.35).

Using equation (2.70), the transition probability of regimes under the \( P \)-measure can be expressed as a function of the bond yields that are priced without error. Specifically, given \( q_t = j \) and \( q_{t+1} = k, j \neq k \), the transition probability \( \pi_{jk,t}^P \) is written as

\[
\pi_{jk,t}^P = \frac{1}{1 + e^{(\eta_{0}^{jk} + \eta_{k'}^{j} F_t)}} = \frac{1}{1 + e^{(\eta_{0}^{jk} + \eta_{k'}^{j} F_t \cdot \beta_1^{-1}(Y_{1,t} - \alpha_1^2))}} = \frac{1}{1 + e^{(\eta_{0}^{jk} + \eta_{k'}^{j} F_t \cdot \beta_1^{-1} - \alpha_1^2 \cdot \alpha_1^2 + \eta_{k'}^{j} F_t \cdot \beta_1^{-1} Y_{1,t})}}.
\]

(2.82)
Define
\[
\xi_{0k}^j = \eta_{0k}^j - \eta_{1k}^j \cdot \beta_1^{-1} \cdot \alpha_1^j \\
\xi_{1k}^j = (\beta_1')^{-1} \cdot \eta_{1k}^j,
\]
then the transition probability can be written as
\[
\pi_{jk,t} = \frac{1}{1 + e^{(\xi_{0k}^j + \xi_{1k}^j Y_{1,t})}}
\]
which is equation (2.36).

Equations (2.34)–(2.36) are the reduced-form models that characterise the process of the observable bond yields \{Y_{1,t}, Y_{2,t}\}, and the vector of reduced-form parameters is given by
\[
\Theta_{\text{redu}} = \left\{ \alpha_1^j, \theta_1^j, \kappa_{11}^j, \Sigma_1^j, \alpha_2^j, \phi_{21}, \xi_{0k}^j, \xi_{1k}^j \right\},
\]
where \( j, k = L, H, j \neq k \), denote regime labels. The reduced-form parameters are functions of the structural parameters, and the functional form is given by equations (2.74), (2.75), (2.76), (2.81), (2.83) and (2.84).
Table 2.C.1: Estimates of the US reduced-form term structure model

This table reports the estimates of reduced-form parameters in the US three-factor regime switching dynamic term structure model. Standard errors are in brackets. Bond yields are annualised and not in percentage.

\[
\begin{align*}
\text{log likelihood} & = 19.9524 \\
\hline
\text{Regime L (low volatility)} & \text{Regime H (high volatility)} \\
\hline
\text{Bond yields priced without error, given } q_{t+1} = k \text{ and } q_t = j & \\
Y_{1,t+1} = \alpha_1^k + \left( Y_{1,t} - \alpha_1^j \right) - \kappa_{11}^j \left( Y_{1,t} - \alpha_1^j - \theta_1^j \right) + \Sigma_t^k u_{t+1} \\
\alpha_1 & 0.0533 \quad 0.0580 \quad 0.0676 \quad 0.0533 \quad 0.0580 \quad 0.0676 \\
& [0.1006] \quad [0.1306] \quad [0.1868] \quad [0.1006] \quad [0.1394] \quad [0.1867] \\
\theta_1 & 0 \quad 0 \quad 0 \quad 0.0269 \quad 0.0245 \quad 0.0210 \\
& [0.1457] \quad [0.1715] \quad [0.2030] \\
\kappa_{11} & 0.2226 \quad -0.3050 \quad 0.1086 \quad 0.1844 \quad -0.3097 \quad 0.1668 \\
& [0.0469] \quad [0.0734] \quad [0.0364] \quad [0.1781] \quad [0.3662] \quad [0.2396] \\
\Sigma_1^{1e-2} & 0.2294 \quad 0.0407 \quad 0.0011 \quad 0.0000 \quad 0.0470 \quad 0.9253 \\
& [0.0096] \quad [0.0503] \\
\text{Bond yields priced with error, given } q_t = j & \\
Y_{2,t} = a_2^j + \phi_2^j \left( Y_{1,t} - a_1^j \right) + \Sigma_1^j \epsilon_t \\
\alpha_2 & 0.0631 \quad 0.0626 \\
& [0.1756] \quad [0.1755] \\
\phi_2^1 & -0.1799 \quad 0.6699 \quad 0.5366 \\
& [0.0079] \quad [0.0120] \quad [0.0058] \\
\Sigma_2^{1e-2} & 0.0470 \quad 0.1364 \\
& [0.0021] \quad [0.0096] \\
\text{Transition probability of regimes, given } q_t = j \text{ and } q_{t+1} = k & \\
\pi_{jk,t}^P = \frac{1}{1 + e^{(\xi_0^{jk} + \xi_1^{jk} Y_{1,t})}} & j \neq k. \\
L2H & H2L \\
\xi_0 & 5.1386 \quad 5.6976 \\
& [1.4199] \quad [1.8662] \\
\xi_1^{1e2} & -0.7791 \quad 1.6376 \quad -1.1625 \quad 1.7319 \quad -1.5652 \quad 0.0939 \\
& [0.8203] \quad [1.2417] \quad [0.6289] \quad [0.7824] \quad [1.1895] \quad [0.7317] \\
\hline
\end{align*}
\]
Table 2.C.2: Estimates of the Japanese reduced-form term structure model

This table reports the estimates of reduced-form parameters in the Japanese two-factor regime switching dynamic term structure model. Standard errors are in brackets. Bond yields are annualised and not in percentage.

<table>
<thead>
<tr>
<th>log likelihood = 15.5806</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regime L (low volatility)</td>
</tr>
<tr>
<td>Bond yields priced without error, given $q_{t+1} = k$ and $q_t = j$</td>
</tr>
<tr>
<td>$Y_{1,t+1} = \alpha_k^j + \left( Y_{1,t} - \alpha_k^j \right) - \kappa_{11}^j \left( Y_{1,t} - \alpha_k^j - \theta_{k1}^j \right) + \Sigma_1^j \epsilon_{t+1}$</td>
</tr>
<tr>
<td>$\alpha_1$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\theta_1$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\kappa_{11}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>eigenvalue</td>
</tr>
<tr>
<td>$\Sigma_1[1e-2]$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Bond yields priced with error, given $q_t = j$</td>
</tr>
<tr>
<td>$Y_{2,t} = a_j^2 + \phi_{21}^j \left( Y_{1,t} - \alpha_1^j \right) + \Sigma_2^j \epsilon_t$</td>
</tr>
<tr>
<td>$\alpha_2$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\phi_{21}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\Sigma_2[1e-2]$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Transition probability of regimes, given $q_t = j$ and $q_{t+1} = k$</td>
</tr>
<tr>
<td>$\pi_{jk,t} = \frac{1}{1 + e^{(\xi_0^j + \xi_1^j Y_{1,t})}}$, $j \neq k.$</td>
</tr>
<tr>
<td>L2H</td>
</tr>
<tr>
<td>$\xi_0$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\xi_1[1e2]$</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Chapter 3

Correcting Estimation Bias in Regime Switching Dynamic Term Structure Models

Abstract

This work examines the small sample bias in the estimation of a regime switching dynamic term structure model. Using US data from 1971 to 2009, we document two regimes driven by the conditional volatility of bond yields and risk factors. In both regimes, the process of bond yields is highly persistent, which is the source of estimation bias when the sample size is small. After bias correction, the inference about expectations of future policy rates and long-maturity term premia changes dramatically in two high-volatility episodes: the 1979–1982 monetary experiment and the recent financial crisis. Empirical findings are supported by Monte Carlo simulation, which shows that correcting small sample bias leads to more accurate inference about expectations of future policy rates and term premia compared to before bias correction.
3.1 Introduction

Regime switching dynamic term structure models (RS-DTSMs) constitute an attractive class of models to capture the stochastic behaviour of interest rates and bond yields. These models offer insights into the dynamic system of bond yields and risk factors and how that system is affected by changes in business cycle conditions and monetary policies. One question that is central to the understanding of the dynamics is to what extent the variation in long-term bond yields is driven by the variation in expectations of future policy rates or the variation in term premia. The answer relies on the estimation of RS-DTSM, which is specified as a regime switching vector autoregression (RS-VAR). However, problems arise as the maximum likelihood (ML) estimates of RS-VAR models potentially suffer from two sources of bias: the bias in the estimation of vector autoregression (VAR) models and the bias in the estimation of regimes. While these two sources of bias have been well recognised, they are widely ignored in the RS-DTSM literature. In this study, we quantify the estimation bias in a regime switching dynamic term structure model with latent factors—a variant of Dai, Singleton and Yang (2007)—and assess the economic implications of bias correction. We document significant bias in the ML estimates of the RS-DTSM and show that correcting the bias dramatically changes the inference about expectations of future policy rates and term premia.

The estimation bias in VAR models parallels the well-documented estimation bias in autoregression (AR) models, and the bias in the estimation of regimes is also an AR-type bias where the persistence of regimes parallels the persistence of an AR(1) process (see Tauchen, 1986; Rouwenhorst, 1995). The ML estimates of AR models are biased towards a process that is less persistent than the true process, and this bias is particularly severe when the true process is highly persistent and the sample size is small. Unfortunately, in the empirical estimation of RS-DTSMs, the sample size is not sufficiently big: the typical data sample dating back to the 1970s has only a few business cycles. Moreover, bond yields are highly persistent, and the estimated regimes are usually very persistent. As a result, the persistence of bond yields and the persistence of regimes are likely to be under-estimated. The downward estimation bias in the persistence of regimes distorts the inference about regimes. The downward
estimation bias in the persistence of bond yields results in the bias in the inference about expectations of future policy rates: risk neutral rates will revert to their long-run mean more quickly than the true process suggests. Bias in risk neutral rates will further distort the inference about term premia.

Despite the intuitive implications of estimation bias, no study on RS-DTSMs directly accounts for the small sample bias, mostly because the computational burden of simulation-based bias correction methods is prohibitive. However, the recent work by Joslin, Singleton and Zhu (2011) and Hamilton and Wu (2012) proposes a two-step estimation procedure, which greatly reduces the computational burden and facilitates simulation-based bias correction methods. In this study, we apply the procedure of Hamilton and Wu (2012) to a regime switching framework. The estimation bias arises in the first step of this procedure where reduced-form models of bond yields (primarily a RS-VAR model) is estimated. We correct the bias in the first step using parametric bootstrap and then proceed with the second step where the cross section of bond yields is fitted.

Applying this bias correction procedure, we are able to quantify the estimation bias in the RS-DTSM and assess the economic implications of bias correction. Using US bond yield data from 1971 to 2009, we document two regimes driven by the conditional volatility of risk factors and bond yields: a low-volatility regime and a less frequent high-volatility regime. We find that the persistence of the high-volatility regime is under-estimated and the persistence of bond yields in the high-volatility regime is severely under-estimated. As a result, the estimates of risk neutral rates and term premia are severely distorted in two high-volatility episodes: the 1979–1982 monetary policy experiment and the recent financial crisis. After bias correction, the exceptionally high long-term forward rates in the 1979–1982 episode and the exceptionally low rates in the recent financial crisis are to a great extent driven by the expectations of future policy rates. While the term premia are high in these two periods, they belong to the normal business-cycle variation of term premia in the full sample period, without being exceptionally high in the 1979–1982 episode or being counter-intuitively low in the recent financial crisis as in the case before bias correction.
The remainder of the paper proceeds as follows. Section 3.2 reviews related literature with an emphasis on RS-DTSMs and estimation bias. Section 3.3 introduces the regime switching dynamic term structure model that is used in this study. Section 3.4 explains the procedure of model estimation and bias correction. Section 3.5 discusses the empirical results, including the magnitude of estimation bias and the economic implications of bias correction. Section 3.6 shows, via Monte Carlo simulation, that correcting estimation bias on average leads to more accurate estimates of risk neutral rates and term premia compared to before bias correction. Section 3.7 offers some concluding remarks.

### 3.2 Related Literature

Many empirical studies document regime switching behaviour in interest rates (see, among many others, Hamilton, 1988; Sola and Driffill, 1994; Garcia and Perron, 1996; Gray, 1996; Ang and Bekaert, 2002). Landén (2000), Dai and Singleton (2003), Evans (2003) propose RS-DTSMs that have closed-form solutions for zero-coupon bond prices. In terms of model specifications, the most closely related papers to this work are Bansal and Zhou (2002), Dai, Singleton and Yang (2007) and Ang, Bekaert and Wei (2008). Using bond yield data from 1964 to 1995, Bansal and Zhou (2002) document two volatility regimes in a two-factor model. They find that the regime switching model is supported by the data in the specification test compared to standard affine models, and only the regime switching model is able to duplicate the violations of the expectations hypothesis as documented in the data. Using bond yield data from 1972 to 2003, Dai, Singleton and Yang (2007) document two regimes driven by the conditional volatility of three latent factors, where the transition probability of regimes is time-varying. They show that compared to the regime switching model, a single-regime model would under-estimate the volatility of risk premia during transitions between regimes and over-estimate the volatility of risk premia in the low-volatility regime. Using a longer sample of bond yield and inflation data from 1952 to 2004, Ang, Bekaert and Wei (2008) document four regimes driven by the level and the volatility of real interest rates and inflation. In a three-factor model, they decompose the nominal
interest rate into a real interest rate, expected inflation and an inflation risk premium, and characterise the term structure of these components conditional and unconditional on regimes. A common observation in the RS-DTSM literature is that in all regimes, the estimated process of risk factors and bond yields is highly persistent. Therefore, the ML estimates of RS-DTSMs are likely to be biased, especially in less frequent regimes, which makes bias correction a highly relevant issue.

Our work is closely related to the literature on the estimation bias in autoregressive models. Several studies show, via simulation, the magnitude of estimation bias in single-regime DTSMs. For example, Ball and Torous (1996) document a downward estimation bias in the persistence of short rates, which is characterised by an AR process that has a near unit root. They show that the bias can be alleviated by imposing cross-section restrictions, which, however, is only achievable by assuming very simple forms for the market price of risk. Duffee and Stanton (2012) simulate one-factor and two-factor, Gaussian and square-root models with various assumptions about risk premia. They show that ML estimates are strongly biased when the model has a flexible specification of risk factors. However, they do not attempt to correct the bias or quantify the economic implications of the bias. The empirical work of Bekaert, Hodrick and Marshall (1997) shows that correcting the small sample bias leads to more consistent rejections of the expectations hypothesis. While they focus on the estimation bias in the slope coefficients in Campbell and Shiller’s (1991) regressions, our work studies the implications of estimation bias for the inference about risk neutral rates and term premia in a RS-DTSM. The most closely related papers are Phillips and Yu (2009) and Bauer, Rudebusch and Wu (2012), both of which correct the estimation bias in the vector autoregression of bond yields and examine the economic implications of bias correction. The two papers differ in the economic implications they examine: Phillips and Yu (2009) examine the price of contingent claims while Bauer, Rudebusch and Wu (2012) examine the decomposition of actual forward rates into risk neutral rates and term premia. Our study builds on the work of Bauer, Rudebusch and Wu (2012) and quantifies the estimation bias in a RS-DTSM. In the regime switching literature, several studies provide small sample evidence using simulation experiments based on stipulated data generating processes (DGPs). Psaradakis and Sola (1998) examine the mean, skewness and kurtosis of the ML estimates of autoregressive models.
with regimes in the mean and the conditional volatility of the process. They show that the persistence of autoregressive processes and the persistence of regimes are under-estimated. Droumaguet (2012) examines the mean squared error (MSE) of the ML estimates of univariate and multivariate autoregressive models with regimes in the intercept, the conditional volatility, and the autoregressive parameter of the process, and shows that more persistent processes lead to more accurate estimates of the autoregressive parameters. Since MSE measures the second moment of estimation errors, it is related to, but is fundamentally different from, the bias in the mean of estimates. While these simulation studies examine small sample properties of the ML estimates in RS-VAR models, it remains unclear the implication for RS-DTSMs estimated on real data. In contrast, we estimate a RS-DTSM using bond yield data and quantify the estimation bias in the model. The empirical work of Bekaert, Hodrick and Marshall (2001) examines whether the rejection of the expectations hypothesis in the US data is due to a peso problem, which is characterised by a high-interest-rate regime that occurs less frequently in the small sample of US data compared to its occurrence in the international data. While they focus on the estimation bias due to the frequency of regimes, we study the estimation bias driven by the high persistence of regimes, and bond yields as well. Bikbov and Chernov (2013) show that using the term structure of bond yields helps reduce MSE in the identification of monetary policy regimes. Different from their perspective, we examine the estimation bias in the persistence of regimes in the context of term structure models.

3.3 The Model

In the regime switching dynamic term structure model, the specification of latent factors and bond yields follows the two-regime term structure model of Dai, Singleton and Yang (2007). We deviate from their model in the specification of regimes: in their model, the transition probability of regimes under the physical measure is time-varying, and we assume that this probability is constant over time and is equal to the transition probability under the risk neutral measure. With this simplified specification of regimes, we are able to focus on the estimation bias in the process of latent
factors and bond yields while also dealing with the estimation bias in the persistence of regimes.

### 3.3.1 Latent factors

The process of latent state variables under the physical measure ($P$) and under the risk neutral measure ($Q$) is specified as

$$F_{t+1} = c^P(q_t) + \rho^P(q_t)F_t + \Sigma(q_t)u^P_{t+1}$$  \hspace{1cm} (3.1)  

$$F_{t+1} = c^Q(q_t) + \rho^Q F_t + \Sigma(q_t)u^Q_{t+1},$$  \hspace{1cm} (3.2)  

where $F_t$ is an $N \times 1$ vector of latent state variables; $q_t$ follows an observable\(^1\) two regime Markov process; the autoregressive parameter under the $Q$-measure, $\rho^Q$, is assumed to have a single regime; $u^P,Q_t \sim i.i.d.N(0,I_N)$; the conditional volatility, $\Sigma(q_t)$, is regime-dependent but does not depend on time. Let $\mu^P_t$ and $\mu^Q_t$ denote time $t$’s conditional mean of $F_{t+1}$ under the $P$-measure and under the $Q$-measure respectively. The conditional means are time-varying and regime-dependent, and are given by

$$\mu^P_t(q_t) = c^P(q_t) + \rho^P(q_t)F_t$$  \hspace{1cm} (3.3)  

$$\mu^Q_t(q_t) = c^Q(q_t) + \rho^Q F_t.$$  \hspace{1cm} (3.4)  

It is assumed that the transition probability of regimes under the $P$-measure, denoted by $\Pi^P$, is constant over time and is equal to the transition probability under the $Q$-measure, denoted by $\Pi^Q$. As will be shown in Section 3.5, the two regimes are driven by the conditional volatility of latent factors, characterising a low-volatility regime and a high-volatility regime, denoted by $q_t = j, j = L,H$. Using this label, the transition probability is written as,

$$\Pi^P = \Pi^Q = \begin{bmatrix}
\pi^{P,Q}_{LL} & \pi^{P,Q}_{LH} \\
\pi^{P,Q}_{HL} & \pi^{P,Q}_{HH}
\end{bmatrix},$$  \hspace{1cm} (3.5)  

where $\pi_{jk} = p(s_{t+1} = k | s_t = j)$ is the probability of being in regime $k$ next period given that the current regime is regime $j$.

\(^1\)The assumption that the regime is observable is also made by Dai, Singleton and Yang (2007). Specifically, the regime is observable to investors but is unobservable to econometricians. We make use of this assumption in bootstrap and simulation where the model is estimated assuming that the regime is observable.
3.3.2 Bond yields and forward rates

Let $P_{n,t}$ denote the price of a zero-coupon bond with a maturity of $n$ months at time $t$ and $y_{n,t} = \frac{-\log P_{n,t}}{n} \times 12$ is the annualised yield on this zero-coupon bond. It is assumed that the yield on the one-month zero-coupon bond (not annualised), denoted by $r_t$, depends on the latent factors $F_t$ and the current regime $q_t$ according to the function

$$r_t(q_t) = \delta_0(q_t) + \delta_1'F_t,$$  \hspace{1cm}  (3.6)

where $\delta_0$ is a regime-dependent scalar and the factor loading $\delta_1$ is an $N \times 1$ vector that does not depend on regimes. This functional form of $r_t$ gives closed-form solutions for the price of zero-coupon bonds. It can be shown that the price of an $n$-month maturity bond, $P_{n,t}$, satisfies

$$P_{n,t}(q_t) = e^{-A_n(q_t)-B_n'F_t},$$  \hspace{1cm}  (3.7)

where $A_n(q_t)$ is a scalar with two regimes and the factor loading $B_n$ is a $N \times 1$ vector with a single regime. Given the regimes in the current and the next periods, i.e., $q_t = j$ and $q_{t+1} = k$, the scalar $A_n(q_t)$ and the factor loading $B_n$ are given by

$$A^j_{n+1} = \delta^j_0 + (c_Q^j)'B_n - \frac{1}{2}B_n'\Sigma^j\Sigma^j'B_n - \log \left( \sum_{k=L,H} \pi^Q_{jk} \cdot e^{-A_k} \right),$$  \hspace{1cm}  (3.8)

$$B_{n+1} = \delta_1 + (\rho^Q)'B_n,$$  \hspace{1cm}  (3.9)

with initial conditions $A^j_0 = 0$ and $B_0 = 0_{N \times 1}$. The derivation of equations (3.7)–(3.9) is given in Appendix 2.A. According to equations (3.8) and (3.9), $A$ and $B$ are functions of

$$A = f_A(\delta_0, c_Q, \Sigma, \pi^Q, \delta_1, \rho^Q)$$  \hspace{1cm}  (3.10)

$$B = f_B(\delta_1, \rho^Q).$$  \hspace{1cm}  (3.11)

That is, the factor loading $B$ depends on parameters $\{\delta_1, \rho^Q\}$; the regime-dependent scalar $A$ depends on parameters $\{\delta_0, c_Q, \Sigma, \pi^Q\}$ and depends on $\{\delta_1, \rho^Q\}$ through its dependence on $B$. The actual bond price depends on the parameters under the risk neutral measure, i.e., $\{c_Q, \rho^Q, \pi^Q\}$.

Following equation (3.7), the annualised bond yield on a zero-coupon bond with a
CHAPTER 3. CORRECTING SMALL SAMPLE BIAS

maturity of \( n \) months is given by

\[
y_{n,t}(q_t) = \frac{A_n(q_t)}{n/12} + \frac{B_n'(q_t)}{n/12}F_t = a_n(q_t) + b_n'F_t. \tag{3.12}
\]

Following equations (3.8) and (3.9), \( a \) and \( b \) are functions of

\[
a = f_a(\delta_0, c^Q, \Sigma, \pi^Q, \delta_1, \rho^Q) \tag{3.13}
\]

\[
b = f_b(\delta_1, \rho^Q), \tag{3.14}
\]

where \( a \) depends on \( \{\delta_1, \rho^Q\} \) through its dependence on \( b \).

The forward rate, denoted by \( f_{n,t} \), is defined as the rate at which an investor enters into a contract at time \( t \) for a loan starting at time \( t + n - 1 \) and maturing at time \( t + n \). Since a forward contract can be synthesised from zero-coupon bonds, the forward rate is given by

\[
f_{n,t} = \log P_{n-1,t} - \log P_{n,t}. \tag{3.15}
\]

The forward rate can be decomposed into a risk neutral rate and a term premium. In a risk neutral world where bonds are priced by risk neutral investors, given the current regime \( q_t \), the risk neutral price of a zero-coupon bond with a maturity of \( n \) months, denoted by \( \ddot{P}_{n,t} \), satisfies

\[
\ddot{P}_{n,t}(q_t) = e^{-\ddot{A}_n(q_t)-\ddot{B}_n'(q_t)F_t}, \tag{3.16}
\]

where \( \ddot{A}_n(q_t) \) is a regime-dependent scalar and the factor loading \( \ddot{B}_n(q_t) \) is an \( N \times 1 \) vector that is also regime-dependent. Given the regimes in the current and the next periods, i.e., \( q_t = j \) and \( q_{t+1} = k \), the scalar \( \ddot{A}_n(q_t) \) and the factor loading \( \ddot{B}_n(q_t) \) are approximately given by

\[
\ddot{A}_{n+1}^j = \delta_0^j + \sum_{k=L,H} \pi_{jk}^P \cdot \left( \ddot{A}_n^k + (c^P)\ddot{B}_n^k - \frac{1}{2} \ddot{B}_n^{k'} \Sigma^j \Sigma^j' \ddot{B}_n^k \right) \tag{3.17}
\]

\[
\ddot{B}_{n+1}^j = \delta_1 + \sum_{k=L,H} \pi_{jk}^P \cdot (\rho^P)\ddot{B}_n^k, \tag{3.18}
\]

with initial conditions \( \ddot{A}_0^j = 0 \) and \( \ddot{B}_0^j = 0_{N \times 1} \). The derivation of equations (3.16)–(3.18) is given in Appendix 3.A. According to equations (3.17) and (3.18), \( \ddot{A} \) and \( \ddot{B} \) are functions of

\[
\ddot{A} = f_{\ddot{A}}(\delta_0, c^P, \Sigma, \pi^P, \delta_1, \rho^P) \tag{3.19}
\]

\[
\ddot{B} = f_{\ddot{B}}(\delta_1, \rho^P, \pi^P). \tag{3.20}
\]
That is, the regime-dependent factor loading $\tilde{B}$ depends on parameters $\{\delta_1, \rho_P, \pi^P\}$; the regime-dependent scalar $\tilde{A}$ depends on parameters $\{\delta_0, c^P, \Sigma, \pi^P\}$ and depends on $\{\delta_1, \rho^P\}$ through its dependence on $\tilde{B}$. The risk neutral bond price depends on the parameters under the physical measure, i.e., $\{c^P, \rho^P, \pi^P\}$. Analogous to the actual forward rate, the risk neutral forward rate is given by the risk neutral bond prices, i.e.,

\[
\tilde{f}_{n,t} = \log \tilde{P}_{n-1,t} - \log \tilde{P}_{n,t}.
\] (3.21)

The risk neutral forward rate $\tilde{f}_{n,t}$ reflects the expected one-month short rate at time $t + n - 1$, i.e., one month before the maturity date. The forward term premium is defined as the difference between the actual forward rate and the risk neutral forward rate, i.e.,

\[
ftp_{n,t} = f_{n,t} - \tilde{f}_{n,t}.
\] (3.22)

It will be shown in Section 3.4 that the small sample bias in parameter estimates distorts the decomposition of actual rates into risk neutral rates and term premia.

### 3.3.3 The pricing kernel

Given the current regime $q_t$, the pricing kernel is specified as

\[
\log \mathcal{M}_{t,t+1}(q_t) = -r_t(q_t) - \frac{1}{2} \lambda_t(q_t) \lambda_t(q_t) - \lambda'_t(q_t) u_{t+1},
\] (3.23)

where $\lambda_t(q_t)$ is the price of factor risk and is specified as

\[
\lambda_t(q_t) = \left( \Sigma(q_t) \right)^{-1} \left( \lambda_0(q_t) + \lambda_1(q_t) F_t \right).
\] (3.24)

It is shown in Appendix 2.B that $\lambda_t(q_t)$ satisfies

\[
\mu_t^Q(q_t) = \mu_t^P(q_t) - \Sigma(q_t) \lambda_t(q_t).
\] (3.25)

It follows that

\[
\lambda_0(q_t) = c^P(q_t) - c^Q(q_t)
\] (3.26)

\[
\lambda_1(q_t) = \rho^P(q_t) - \rho^Q.
\] (3.27)
To see the interpretation of $\lambda_t$, consider a security with payoff $e^{b^P_i t + 1}$. Given the current regime $q_t = j$, the expected excess return of this security is

$$\log \frac{E_t^P [e^{b^P_i t + 1}]}{P^j_t} - r_j^i = \log \frac{E_t^P [e^{b^P_i t + 1}]}{e^{-r_i^j E_t^Q [e^{b^Q_i t + 1}]}} - r_i^j = b^P_j \left( \mu^P_j - \mu^Q_j \right) = (b^j \Sigma^j) \lambda^j_t, \quad (3.28)$$

where $(b^j \Sigma^j)$ is interpreted as the exposure of this security to factor risk and $\lambda^j_t$ is interpreted as the price of factor risk.

### 3.4 Model Estimation and Estimation Bias

#### 3.4.1 Model estimation

The minimum-chi-square estimation proposed by Hamilton and Wu (2012) is applied to the estimation of the regime switching dynamic term structure model. To see the general procedure of this estimation method, let $\Theta_{stru}$ denote a vector consisting of the structural (theoretical) parameters, i.e.,

$$\Theta_{stru} = \{c^{Pj}, c^{Qj}, \rho^{Pj}, \rho^Q, \Sigma^j, \delta_0^j, \delta_1, \pi^{Pj}, \pi^{Qj} \}, \quad (3.29)$$

where $j = L, H$ denotes regime labels. The structural parameters characterise the process of latent factors under the physical and the risk neutral measure, and specifies how the short rate depends on the latent factors. There is also a vector of reduced-form parameters, denoted by $\Theta_{redu}$, that characterises the process of observable bond yields. It will be shown later that reduced-form parameters are functions of structural parameters, i.e., $\Theta_{redu} = g(\Theta_{stru})$. Under the null hypothesis that $\Theta_{redu} = g(\Theta_{stru})$, the following statistic has an asymptotic chi-square distribution, i.e.,

$$T \left[ \hat{\Theta}_{redu} - g(\Theta_{stru}) \right]' \hat{R} \left[ \hat{\Theta}_{redu} - g(\Theta_{stru}) \right] \sim \chi^2(q), \quad (3.30)$$

where $\hat{\Theta}_{redu}$ is an estimate of $\Theta_{redu}$, $q$ is the dimension of $\Theta_{redu}$, and $\hat{R}$ is a consistent estimate of the information matrix

$$R = -T^{-1} E \left[ \frac{\partial^2 L(\Theta_{redu}; Y_T)}{\partial \Theta_{redu} \partial \Theta_{redu}'} \right].$$
The minimum-chi-square estimation uses (3.30) to motivate the estimate \( \hat{\Theta}_{\text{stru}} \) that minimises this chi-square statistic. Hamilton and Wu (2012) also propose approximating the variance of \( \hat{\Theta}_{\text{stru}} \) with \( \left( \hat{\Gamma}' \hat{R} \hat{\Gamma} \right)^{-1} \) where

\[
\hat{\Gamma} = \left. \frac{\partial g(\Theta_{\text{stru}})}{\partial \Theta_{\text{stru}}} \right|_{\Theta_{\text{stru}} = \hat{\Theta}_{\text{stru}}}. 
\]

When the model is just-identified, the minimum value attainable for the statistic in (3.30) is zero, and the procedure of solving \( \hat{\Theta}_{\text{stru}} \) from \( \hat{\Theta}_{\text{redu}} - g(\hat{\Theta}_{\text{stru}}) = 0 \) is termed the mapping between structural and reduced-form parameters.

The reduced-form parameters (models) are derived as follows. It is assumed that there are \( N \) zero coupon bonds that are priced without error, where \( N \) is equal to the number of latent factors, and there are \( M \) zeros coupon bonds that are priced with error. Specifically, let \( Y_{1,t} = \{y_{n_1,t}, y_{n_2,t}, \ldots, y_{n_N,t}\}' \) denote an \( N \times 1 \) vector of annualised bond yields that are priced without error, where the maturities of these bonds are months \( \{n_1, n_2, \ldots, n_N\} \). According to equation (3.12), the bond yield vector \( Y_{1,t} \) is a function of the latent factors \( F_t \), i.e., given the current regime \( q_t = j \),

\[
Y_{1,t} = \alpha_1^j + \beta_1 F_t, \tag{3.31}
\]

where

\[
\alpha_1^j \quad \text{(N \times 1)} = \begin{bmatrix}
\alpha_{n_{i}}^j, \alpha_{n_{2}}^j, \ldots, \alpha_{n_{N}}^j
\end{bmatrix}', \tag{3.32}
\]

\[
\beta_1 \quad \text{(N \times N)} = \begin{bmatrix}
b_{n_{1}}, b_{n_{2}}, \ldots, b_{n_{N}}
\end{bmatrix}'. \tag{3.33}
\]

Specifically, \( \alpha_1^j \) is an \( N \times 1 \) regime-dependent vector and \( \beta_1 \) is an \( N \times N \) matrix of factor loadings that has a single regime. For \( \forall i = 1, \ldots, N \), \( \alpha_{n_i}^j \) is the regime-dependent scalar and \( b_{n_i} \) is the \( N \times 1 \) regime-independent vector, such that the \( n_i \)-month bond yield depends on the latent factors as given in equation (3.12). Using equation (3.31) and the process of latent factors in equation (3.1), it can be shown that \( Y_{1,t} \) follows an autoregressive process. Specifically, given the regimes in the next and the current
periods, i.e., \( q_{t+1} = k \) and \( q_t = j \),

\[
Y_{1,t+1} = \alpha^k_1 + \beta_1 \cdot F_{t+1} \\
= \alpha^k_1 + \beta_1 \cdot [c^{Pj} + \rho^{Pj} \cdot F_t + \Sigma^j u_{t+1}] \\
= \alpha^k_1 + \beta_1 \cdot [c^{Pj} + \rho^{Pj} \cdot \beta^{-1}_1 (Y_{1,t} - \alpha^j_1) + \Sigma^j u_{t+1}] \\
= \alpha^k_1 + \beta_1 \cdot c^{Pj} + \beta_1 \cdot \rho^{Pj} \cdot \beta^{-1}_1 (Y_{1,t} - \alpha^j_1) + \beta_1 \cdot \Sigma^j u_{t+1}. \tag{3.34}
\]

Define

\[
c^j_1 \equiv \beta_1 \cdot c^{Pj} \tag{3.35}
\]

\[
\phi^j_{11} \equiv \beta_1 \cdot \rho^{Pj} \cdot \beta^{-1}_1 \tag{3.36}
\]

\[
\Sigma^j_1 \equiv \beta_1 \cdot \Sigma^j, \tag{3.37}
\]

then given \( q_{t+1} = k \) and \( q_t = j \), the process of \( Y_{1,t} \) is written as

\[
Y_{1,t+1} = \alpha^k_1 + c^j_1 + \phi^j_{11} (Y_{1,t} - \alpha^j_1) + \Sigma^j_1 u_{t+1}. \tag{3.38}
\]

Let \( Y_{2,t} = \{y_{m_1,t}, y_{m_2,t}, \ldots, y_{m_M,t}\}' \) denote an \( M \times 1 \) vector of annualised bond yields that are priced with error, where the maturities of these bonds are months \( \{m_1, m_2, \ldots, m_M\} \). The bond yield vector \( Y_{2,t} \) depends on the latent factors according to equation (3.12), and since these bonds are priced with error, they have their own error terms, i.e., given the current regime \( q_t = j \),

\[
Y_{2,t} = \alpha^j_2 + \beta_2 F_t + \Sigma^j_2 \epsilon_t, \tag{3.39}
\]

where

\[
\alpha^j_2 \begin{pmatrix} M \times 1 \end{pmatrix} = \begin{bmatrix} a^{j}_{m_1}, a^{j}_{m_2}, \ldots, a^{j}_{m_M} \end{bmatrix}' \tag{3.40}
\]

\[
\beta_2 \begin{pmatrix} M \times N \end{pmatrix} = \begin{bmatrix} b_{m_1}, b_{m_2}, \ldots, b_{m_M} \end{bmatrix}'. \tag{3.41}
\]

Specifically, \( \alpha^j_2 \) is an \( M \times 1 \) regime-dependent vector and \( \beta_2 \) is an \( M \times N \) matrix of factor loadings that has a single regime, where \( \{\alpha^j_2, \beta_2\} \) has the same interpretation as \( \{\alpha^j_1, \beta_1\} \) in equation (3.31); \( \Sigma^j_2 \) is an \( M \times M \) constant and regime-dependent matrix; \( \epsilon_t \sim i.i.d. N(0, I_M) \). Rewrite equation (3.39) so that \( Y_{2,t} \) depends on \( Y_{1,t} \), i.e.,

\[
Y_{2,t} = \alpha^j_2 + \beta_2 F_t + \Sigma^j_2 \epsilon_{t+1} \\
= \alpha^j_2 + \beta_2 \cdot \beta^{-1}_1 (Y_{1,t} - \alpha^j_1) + \Sigma^j_2 \epsilon_t. \tag{3.42}
\]
Define
\[ \phi_{21} \equiv \beta_2 \cdot \beta_1^{-1}, \]  
then given \( q_t = j \), the process of \( Y_{2,t} \) is written as
\[ Y_{2,t} = \alpha_j^2 + \phi_{21} (Y_{1,t} - \alpha_1^j) + \Sigma_j^2 \epsilon_t. \]  

Equations (3.38) and (3.44) are the reduced-form models that characterise the process of the observable bond yields \( \{Y_{1,t}, Y_{2,t}\} \), and the vector of reduced-form parameters is given by
\[ \Theta_{\text{redu}} = \{\alpha_1^j, \epsilon_1^j, \phi_{11}^j, \Sigma_j^1, \phi_2, \phi_{21}, \pi_{jj}^j\}, \]  
where \( j = L, H \) denotes regime labels. The ML estimates of the reduced-form parameters are the estimates that maximise the likelihood function. The log likelihood function, denoted by \( \mathcal{L}(\Theta_{\text{redu}}, \mathbb{Y}_T) \), is given by
\[ \mathcal{L}(\Theta_{\text{redu}}, \mathbb{Y}_T) = \sum_{t=2}^{T} \log f(Y_{1,t}, Y_{2,t}|\mathbb{Y}_{t-1}; \Theta_{\text{redu}}), \]  
where \( \mathbb{Y}_t = \{Y_{1,t}, Y_{2,t}, \tau \leq t\} \) is the information set up to time \( t \), and \( T \) is the sample size. The conditional density of bond yields is
\[
f(Y_{1,t}, Y_{2,t}|\mathbb{Y}_{t-1}) = \sum_{q_t, q_{t-1}} f(Y_{1,t}, Y_{2,t}, q_t, q_{t-1}|\mathbb{Y}_{t-1}) \]
\[
= \sum_{q_t, q_{t-1}} f(Y_{2,t}|Y_{1,t}, q_t) \cdot f(Y_{1,t}|Y_{1,t-1}, q_t, q_{t-1}) \cdot p(q_t|q_{t-1}, \mathbb{Y}_{t-1}) \cdot p(q_{t-1}|\mathbb{Y}_{t-1}),
\]
where \( Y_{1,t} \) and \( Y_{2,t} \) are conditionally normal according to equations (3.38) and (3.44), \( p(q_t|q_{t-1}, \mathbb{Y}_{t-1}) \) is the transition probability of regimes, and \( p(q_{t-1}|\mathbb{Y}_{t-1}) \) is the filtered probability. The filtered probability, \( p(q_t|\mathbb{Y}_t) \), is the probability of the regime at time \( t \) given the information set up to time \( t \), and is given by
\[
p(q_t|\mathbb{Y}_t) = \frac{p(q_t, Y_{1,t}, Y_{2,t}|\mathbb{Y}_{t-1})}{\sum_{q_{t-1}} p(q_{t-1}, Y_{1,t}, Y_{2,t}|\mathbb{Y}_{t-1})} \]
\[
= \frac{\sum_{q_{t-1}} p(q_{t-1}, Y_{1,t}, Y_{2,t}|\mathbb{Y}_{t-1})}{\sum_{q_{t-1}} f(Y_{2,t}|Y_{1,t}, q_t) \cdot f(Y_{1,t}|Y_{1,t-1}, q_t, q_{t-1}) \cdot p(q_t|q_{t-1}, \mathbb{Y}_{t-1}) \cdot p(q_{t-1}|\mathbb{Y}_{t-1})}.
\]
The log likelihood function is obtained by iterating between \( f(Y_{1,t}, Y_{2,t}|\mathbb{Y}_{t-1}) \) and \( p(q_t|\mathbb{Y}_t) \), given a starting value \( p(q_0 = j|\mathbb{Y}_0) \), \( j = L, H \), which is set equal to the
unconditional probability of being in regime \( j \) implied by the transition probability of regimes.

Following the common practice in the regime switching literature, we use smoothed probabilities in classifying regimes. The smoothed probability, denoted by \( p(q_t|Y_T) \), is the probability of the regime at time \( t \) given the full sample information \( Y_T \). The smoothed probability is solved backwards from \( p(q_T|Y_T) \), which is equal to the filtered probability at time \( T \). Given the smoothed probability at time \( t \), \( p(q_t|Y_T) \), the probability at time \( t-1 \) is

\[
p(q_{t-1}|Y_T) = \sum_q p(q_t, q_{t-1}|Y_T),
\]

(3.47)

where

\[
p(q_t, q_{t-1}|Y_T) = p(q_{t-1}|q_t, Y_T) \cdot p(q_t|Y_T)
\]

(3.48)

We classify the regime at time \( t \) according to the following rule:

\[
q_t = \begin{cases} 
H, & \text{if } p(q_t = H|Y_T) > 0.5 \\
L, & \text{if } p(q_t = H|Y_T) \leq 0.5
\end{cases}. 
\]

(3.49)

That is, the regime at time \( t \) is classified as a high-volatility regime if time \( t \)'s smoothed probability of being in the high-volatility regime is greater than 0.5, and is otherwise classified as a low-volatility regime.

### 3.4.2 Model identification

In summary, the vectors of structural and reduced-form parameters are given by

\[
\Theta_{stru} = \{c^P_j, c^Q_j, \rho^P_j, \rho^Q_j, \Sigma^j, \delta_0^j, \delta_1^j, \pi^P_{jj}, \pi^Q_{jj}\}
\]

\[
\Theta_{redu} = \{\alpha_1^j, c_1^j, \phi_{11}^j, \Sigma_1^j, \alpha_2^j, \phi_{21}^j, \pi^P_{jj}\},
\]

where \( j = L, H \) denotes regime labels. The estimates of structural parameters are mapped out from the estimates of reduced-form parameters through a number of linear
and non-linear equations. Firstly, we solve structural parameters \( \{ \delta_1, \rho^Q, \Sigma^j \} \) numerically from reduced-form parameters \( \{ \Sigma^j_1, \phi_{21} \} \) through their non-linear relationship as given in equations (3.14), (3.33), (3.41), (3.37) and (3.43), i.e.,

\[
\begin{align*}
    b_n &= f_b(\delta_1, \rho^Q, \psi) \quad \text{(3.14)} \\
    \beta_1 &= \left[ b_{n1}, b_{n2}, \ldots, b_{nN} \right]' \quad \text{(3.33)} \\
    \Sigma_j^1 &= \beta_1 \cdot \Sigma^j \quad \text{(3.37)} \\
    \phi_{21} &= \beta_2 \cdot \beta_1^{-1}. \quad \text{(3.43)}
\end{align*}
\]

Next, since the transition probability of regimes under the \( Q \)-measure is equal to that under the \( P \)-measure as assumed in equation (3.5), \( \pi_{jj}^Q \) is set equal to \( \pi_{jj}^P \). Then given the solved structural parameters \( \{ \delta_1, \rho^Q, \Sigma^j, \pi_{jj}^Q \} \), we solve structural parameters \( \{ \delta_0^j, c^{Qj} \} \) numerically from reduced-form parameters \( \{ \alpha_1^j, c_1^j, \alpha_2^j \} \) through their non-linear relationship as given in equations (3.13), (3.32) and (3.40), i.e.,

\[
\begin{align*}
    a_n &= f_a(\delta_0, c^Q, \Sigma, \pi^Q, \delta_1, \rho^Q). \quad \text{(3.13)} \\
    \alpha_1^j &= \left[ a_{n1}^j, a_{n2}^j, \ldots, a_{nN}^j \right]'. \quad \text{(3.32)} \\
    \alpha_2^j &= \left[ a_{m1}^j, a_{m2}^j, \ldots, a_{mM}^j \right]'. \quad \text{(3.40)}
\end{align*}
\]

Finally, we solve structural parameters \( \{ c^{Pj} \} \) and \( \{ \rho^{Pj} \} \) analytically from reduced-form parameter \( \{ c_1^j \} \) and \( \{ \phi_{11}^j \} \) respectively, i.e.,

\[
\begin{align*}
    c_1^j &= \beta_1 \cdot c^{Pj} \quad \text{(3.35)} \\
    \phi_{11}^j &= \beta_1 \cdot \rho^{Pj} \cdot \beta_1^{-1}. \quad \text{(3.36)}
\end{align*}
\]

where \( \beta_1 \) is given by the solved structural parameters \( \{ \delta_1, \rho^Q \} \) as in equations (3.14) and (3.33).

In the estimation of the regime switching dynamic term structure model, we adapt the normalisations and restrictions imposed by Dai, Singleton and Yang (2007) to the minimum-chi-square estimation. They estimate a three-factor regime switching model
using maximum likelihood estimation and impose the following normalizations:

\[ c^{PL} = 0_{3 \times 1} \]  
\[ \sqrt{12} \cdot \Sigma^L = I_3 \]  
\[ \rho^{PL} \text{ is a lower triangular matrix} \]  
\[ \Sigma^H \text{ is a lower triangular matrix}, \]  

and restrictions:

\[ \rho^Q \text{ has a single regime} \]  
\[ \delta_1 \text{ has a single regime}. \]

Furthermore, given the difficulty in estimating the means of bond yields and short rates, they fix the mean of short rates by fixing \( \{\delta^L_0 + \delta^L_1 \cdot \bar{\mu}^L\} \) and \( \{\delta^H_0 + \delta^H_1 \cdot \bar{\mu}^H\} \) at the sample mean of short rates in the low-volatility regime and in the high-volatility regime respectively, where \( \bar{\mu}^j = (I_3 - \rho^{Pj})^{-1} \cdot c^{Pj}, \ j = L, H \) is the mean of latent factors in each regime. Normalisation (3.50) implies that \( \bar{\mu}^L = 0 \). Finally, after a preliminary estimation of the model, they set \( \rho^{PL}(2,1), \Sigma^H(2,1), \Sigma^H(3,1), \Sigma^H(3,2), \lambda^H_0(1), \lambda^H_0(2), \lambda^H_0(3), \lambda^H_L(1,1), \lambda^H_L(2,1), \lambda^H_L(2,2), \lambda^H_L(3,2), \lambda^H_L(1,3), \lambda^H_L(2,3), \lambda^H_L(3,2), \lambda^H_L(3,3) \) equal to zero because these parameters are small relative to their estimated standard errors.

In order to adapt to the minimum-chi-square estimation, it is necessary to amend the above normalisations and restrictions. Specifically, we keep normalisations (3.50) and (3.51) and restrictions (3.54) and (3.55). Instead of normalisations (3.52) and (3.53), we estimate \( \rho^{PL} \) as a full matrix and normalise \( \Sigma^H \) to be a diagonal matrix, which is no more restrictive than (3.52) and (3.53), especially given that Dai, Singleton and Yang (2007) later set the off-diagonal elements of \( \Sigma^H \) to be zero in their model estimation. Dai, Singleton and Yang (2007) deal with the difficulty in estimating the means by fixing the means of short rates, and we deal with this difficulty by fixing the means of the latent factors. Specifically, we impose

\[ c^{PH} = 0_{3 \times 1}. \]  

Together with normalisation (3.50), by imposing (3.56), we actually assume that the mean of latent factors is zero in both regimes. The implications for the reduced-form
parameters is shown in equation (3.35), i.e.,
\[
\begin{align*}
  c^L_1 &= 0_{3 \times 1}^L \\
  c^H_1 &= 0_{3 \times 1}^H.
\end{align*}
\] (3.57) (3.58)

After imposing \( c^j_1 = 0, \ j = L, H \), the reduced-form parameter \( \alpha^j_1, \ j = L, H \) in equation (3.38) has clear economic interpretations as the mean of bond yields in the low-volatility regime and in the high-volatility regime respectively. In the estimation of the RS-DTSM, we also find that a few elements of the price-of-risk parameters (\( \lambda s \)) are estimated to be insignificantly different from zero. However, we still estimate them as free parameters instead of setting them to zeros as in Dai, Singleton and Yang (2007). A summary of the mapping procedure, including the imposed normalisations and restrictions, is given in Appendix 3.B.

3.4.3 Estimation bias and bias correction

There are two sources of bias in the ML estimates of the reduced-form models: the bias in the estimates of regimes and the bias in the estimates of VAR models. Both are AR-type bias where the ML estimates are biased towards a process that is less persistent than the true process, and this bias is particularly severe when the true process is highly persistent and the sample size is small. Empirically, bond yields are highly persistent and the estimated regimes are usually very persistent. As a result, the persistence of bond yields and the persistence of regimes are likely to be underestimated. The downward bias in the estimated persistence of regimes distorts the inference about regimes, and the downward bias in the estimated persistence of bond yields distorts the inference about risk neutral rates and term premia.

The inference about risk neutral rates and term premia is closely related to the estimated persistence of bond yields. Specifically, as shown in Section 3.3.2, the persistence of risk neutral rates is determined by the persistence of latent factors under the physical measure, i.e., the autoregressive parameter \( \rho^{Pj}, \ j = L, H \). This parameter, according to equation (3.36), is mapped out from the autoregressive parameter \( \phi^j_{11} \) of bond yields and has the same persistence as \( \phi^j_{11} \). A downward estimation bias in the persistence of bond yields implies a downward bias in the persistence of latent factors.
CHAPTER 3. CORRECTING SMALL SAMPLE BIAS

of the same magnitude. As a result, the persistence of risk neutral rates is under-estimated: the estimated risk neutral rates revert to the mean more quickly than the true process suggests. Therefore, the volatility of risk neutral rates is under-estimated, and it is predicted that after bias correction, the volatility of risk neutral rates will increase. Note that the actual forward rate is estimated without bias as it is determined by the structural parameters under the risk neutral measure, which can be accurately pinned down by the cross section of bond yields. As a consequence, the term premium, defined as the difference between the actual rate and the risk neutral rate, is distorted by the estimation bias in the risk neutral rate. Since the estimation bias indirectly distorts the term premium, there is no clear prediction of how the volatility of term premia will change after bias correction, as this depends on the correlation structure among actual rates, risk neutral rates and term premia.

A parametric bootstrap approach is implemented to correct the estimation bias in the reduced-form parameters. Let $\hat{\Theta}_{\text{redu}}^{mle}$ denote the ML estimates of $\Theta_{\text{redu}}$ before bias correction, and let $\hat{\Theta}_{\text{redu}}^{bc}$ denote the estimates after bias correction. We generate $B = 1000$ samples of regimes and bond yields using $\hat{\Theta}_{\text{redu}}^{mle}$ as data generating parameters. Regimes are generated according to equation (3.5) and bond yields are generated according to equations (3.38) and (3.44). The sample size is equal to that of the historical dataset, i.e., $T = 451$ months, and in each sample, the first observation of bond yields that are priced without error is randomly sampled from the historical dataset. We impose several restrictions on the samples of regimes so that bootstrapped samples represent the historical dataset. Specifically, we require that there are at least 2 transitions from the low-volatility to the high-volatility regime and at least 2 transitions from the high-volatility to the low-volatility regime in the sample, so that the estimates of transition probability parameters are more reliable. During the sample period from November 1971 to May 2009, there are 6 transitions from the expansion to the recession regime and 5 transitions from the recession to the expansion regime according to NBER business cycle classifications. Based on our classification of regimes, the number of transitions from the low-volatility to the high-volatility regime is 4 before bias correction and is 3 after bias correction, and the number of transitions from the high-volatility to the low-volatility regime is 3 both before and after bias correction. We also require that there are at least 30 observations in the high-volatility
regime, so that the parameters of bond yields for the high-volatility regime are more reliably estimated. In the 451-month historical sample, there are 77 months of NBER recessions. Based on our classification of regimes, the number of months in the high-volatility regime is 97 before bias correction and is 138 after bias correction. Finally, we require that there are more observations in the low-volatility regime than in the high-volatility regime, which is consistent with the estimated transition probabilities, where the low-volatility regime is more persistent than the high-volatility regime, both before and after bias correction.

Given \( B = 1000 \) bootstrapped samples, for each sample \( b \), we estimate reduced-form models, and the ML estimates are denoted by \( \hat{\Theta}^{mle}_{redu} \). In the next step, all the parameter estimates in \( \hat{\Theta}^{mle}_{redu} \) are corrected for estimation bias, if there is any. However, for ease of demonstration, we focus on parameter \( \phi_{11} \), \( j = L, H \), which is the source of estimation bias in the process of bond yields. Let \( \hat{\phi}^{mle}_{11} \) denote the estimates of \( \phi_{11} \) before bias correction, let \( \hat{\phi}^{bc}_{11} \) denote the estimates after bias correction, and let \( \hat{\phi}^{mle}_{11} \) denote the ML estimates of \( \phi_{11} \) in each bootstrapped sample. In the estimation of \( \hat{\phi}^{mle}_{11} \), since economic theories strongly suggest that the process of bond yields is stationary, we impose stationarity by using the stationarity adjustment method suggested by Kilian (1998), which is also used by Bauer, Rudebusch and Wu (2012) in their bias correction procedure. Specifically, if the maximum eigenvalue of \( \hat{\phi}^{mle}_{11} \) is greater than 0.9999, we shrink \( \hat{\phi}^{mle}_{11} \) until the maximum eigenvalue is smaller than 0.9999. Then given stationary \( \hat{\phi}^{mle}_{11} \), \( b = 1, 2, \ldots, B \), we calculate the mean of parameter estimates over the bootstrapped samples, i.e.,

\[
\hat{\phi}^{boot\_mle}_{11} = \frac{1}{B} \sum_{b=1}^{B} \hat{\phi}^{mle}_{11}_b. \tag{3.59}
\]

If there is no bias in ML estimates, \( \hat{\phi}^{boot\_mle}_{11} \) should be very close to the data-generating parameter \( \hat{\phi}_{11}^{mle} \). If \( \hat{\phi}^{boot\_mle}_{11} \) deviates from \( \hat{\phi}_{11}^{mle} \), there is bias in ML estimates, implying bias in \( \hat{\phi}_{11}^{mle} \). The difference between \( \hat{\phi}^{boot\_mle}_{11} \) and \( \hat{\phi}_{11}^{mle} \) characterises the magnitude of the bias in \( \hat{\phi}_{11}^{mle} \), and the bias-corrected estimates are obtained by adding the difference back, i.e.,

\[
\hat{\phi}^{diff}_{11} = \hat{\phi}_{11}^{mle} - \hat{\phi}^{boot\_mle}_{11} \tag{3.60}
\]

\[
\hat{\phi}^{bc}_{11} = \hat{\phi}_{11}^{mle} + \hat{\phi}^{diff}_{11}. \tag{3.61}
\]
If after bias correction, the maximum eigenvalue of $\hat{\phi}_{11}^{bc}$ is greater than 0.9999, we shrink $\hat{\phi}_{11}^{bc}$ towards $\hat{\phi}_{11}^{mle}$ until the maximum eigenvalue is smaller than 0.9999. In the first equation, a positive $\hat{\phi}_{11}^{diff}$ means that there is a downward bias in ML estimates, then in the second equation, the ML estimates $\hat{\phi}_{11}^{mle}$ are adjusted upwards. The same mechanism applies to a negative $\hat{\phi}_{11}^{diff}$. The underlying assumption of this bias correction approach is that the estimation bias when the true parameter is $\hat{\phi}_{11}^{bc}$ is equal to the estimation bias when the true parameter is $\hat{\phi}_{11}^{mle}$. As a result, if we bootstrap samples using $\hat{\phi}_{11}^{bc}$ as data generating parameters, the mean of parameter estimates over the bootstrapped samples, denoted by $\hat{\phi}_{11}^{bc}$, should be equal to or close to $\hat{\phi}_{11}^{mle}$, which are the estimates from the historical data. This bias correction approach is less accurate if the bias depends on parameter $\phi_{11}$—in other words, the estimation bias when the true parameter is $\hat{\phi}_{11}^{bc}$ is not equal to the estimation bias when the true parameter is $\hat{\phi}_{11}^{mle}$—which motivates the indirect inference approach proposed by Bauer, Rudebusch and Wu (2012). In this study, we use the simple bootstrap approach, and find this approach does a reasonably good job. Finally, for the full vector of parameter estimates $\hat{\Theta}_{redu}$, the bias correction procedure parallels equations (3.59)–(3.61) and is given by

\[
\hat{\Theta}_{redu}^{boot_{mle}} = \frac{1}{B} \sum_{b=1}^{B} \left( \hat{\Theta}_{redu}^{mle} \right)_b
\]

(3.62)

\[
\hat{\Theta}_{redu}^{diff} = \hat{\Theta}_{redu}^{mle} - \hat{\Theta}_{redu}^{boot_{mle}}
\]

(3.63)

\[
\hat{\Theta}_{redu}^{bc} = \hat{\Theta}_{redu}^{mle} + \hat{\Theta}_{redu}^{diff}
\]

(3.64)

### 3.5 Empirical Analysis

Following Dai, Singleton and Yang (2007), it is assumed that the bonds with 6-, 24-, 120-month maturities are priced without error, and the bond with 60-month maturity is priced with error. The dataset of Wright (2011) is used and the sample period is from November 1971 to May 2009. Table 3.1 reports the summary statistics and Figure 3.1 plots the data. As shown in Table 3.1, the sample mean of bond yields increases as the maturity increases, implying an upward sloping yield curve. The sample standard deviation decreases as the maturity increases, i.e., the long-term yields are less volatile.
than the short-term yields. The volatility curve, defined by the standard deviation of changes in yields, is downward sloping. As documented in the literature, while the volatility curve has different shapes over different periods of time, the volatility of short-term yields is always high, which is consistent with the summary statistics in our sample. As shown in Figure 3.1, the maximum of yields is in the period from the late 1970s to the early 1980s, and the minimum of yields is in the recent financial crisis. As shown in the last row of Table 3.1, the autocorrelation coefficients of bond yields are quite high, being greater than 0.98 for all maturities, and the coefficient increases as the maturity increases. It is this high persistence of bond yields that induces estimation bias in autoregressive models, especially when the sample size is small. Even though the autocorrelation coefficient is close to one and a unit root cannot be rejected as is often found in the literature, economic arguments strongly suggest that the process of bond yields is stationary—bond yields have long swings in the sample, however, are not explosive and stay within a range (see, e.g., a comment of Cochrane, 2005, pp. 199). Therefore, empirical studies of term structure models almost invariably impose stationarity. In this paper, we impose stationarity in both model estimation and the procedure of bias correction.

### 3.5.1 Results before bias correction

Tables 3.2 and 3.3 report, before bias correction, the estimates of reduced-form and structural parameters respectively. As shown in these two tables, regimes are driven by the conditional volatility of latent factors and bond yields: the reduced-form parameters \( \{ \Sigma_1, \Sigma_2 \} \) in Table 3.2 and the structural parameters \( \{ \Sigma \} \) in Table 3.3 are significantly different from zero, and by a rough calculation using standard errors, the elements of \( \Sigma \) in the high volatility regime are more than two standard errors away from their counterparts in the low-volatility regime, and are about 1.5–4.5 times the volatility in the low-volatility regime. As indicated by the estimates of transition probability parameters, both regimes are very persistent. The average durations of the low-volatility regime and the high-volatility regime are 82 months and 30 months respectively. The high persistence of regimes induces a downward bias in the estimated persistence of regimes.
Tables 3.2 and 3.3 also report the eigenvalues of the autoregressive parameters. Since the persistence of bond yields is equal to the persistence of risk factors by construction, we only discuss the persistence of risk factors. The process of risk factors is highly persistent in both regimes, being more persistent in the low-volatility regime than in the high-volatility regime, with eigenvalues of \([0.9920, 0.9540, 0.8818]\) in the low-volatility regime and eigenvalues of \([0.9791, 0.8420, 0.5078]\) in the high-volatility regime. The finding that risk factors are highly persistent in both regimes is consistent with the literature. For example, using bond yield data from 1964 to 1995, Bansal and Zhou (2002) document two volatility regimes in a two-factor model. For one factor, the maximum eigenvalue is around 0.97 in both regimes and this factor is more persistent in the low-volatility regime. For the other factor, the maximum eigenvalue is around 0.99 in both regimes and this factor is more persistent in the high-volatility regime. Using bond yield data from 1972 to 2003, Dai, Singleton and Yang (2007) document two regimes driven by the conditional volatility of three latent factors. The maximum eigenvalue of their estimated autoregressive parameters is around 0.98 in both regimes, and the process is more persistent in the high-volatility regime than in the low-volatility regime. Using a longer quarterly sample of bond yield and inflation data from 1952 to 2004, Ang, Bekaert and Wei (2008) document four regimes in the process of one inflation factor and two latent factors. They assume that the autoregressive parameter has a single regime, and the maximum eigenvalue of the estimates is about 0.97, which would translate to a higher number if the dataset were at monthly frequency. The first-order autocorrelation coefficient of their estimated one-quarter expected inflation is 0.89, which implies a monthly autocorrelation coefficient of 0.96 under the null hypothesis of an AR(1) process. In summary, while the empirical findings are mixed in terms of in which regime the process is more persistent, the general consensus is that risk factors are highly persistent in all regimes, which is the source of estimation bias in the autoregressive parameters.

Figure 3.2 plots the filtered and the smoothed probability of regimes. There are mainly three episodes in the sample that are classified as in the high-volatility regime: the oil price shock in the mid 1970s, the monetary experiment in 1979–1982 and the recent financial crisis. Consistent with the finding in Dai, Singleton and Yang (2007) that the high-volatility regime extends beyond the end of 1982 to about 1985, we find
The episode around the mid 1980s is classified as in the high-volatility regime. The 1984–1985 high-volatility episode is then treated as an extension of the monetary experiment in 1979–1982 and the two episodes are jointly referred to as the monetary experiment in 1979–1982. The economy has been classified as in the low-volatility regime since the mid 1980s until the recent financial crisis. The recent financial crisis differs from the 1979–1982 episode in terms of the mean of bond yields: the 1979–1982 episode is characterised by a high-mean-high-volatility regime while the recent financial crisis is a low-mean-high-volatility regime. In this sense, the four-regime model of Ang, Bekaert and Wei (2008) may fit the data better. However, since the focus of this work is to examine, in the presence of regime shifts, the estimation bias in the process of bond yields, we stick to simple specifications of regimes as long as the model fits the data reasonably well.

Figure 3.3 plots factor loadings of bond yields on the latent factors as a function of bond maturities, where bond yields depend on the latent factors as in equation (3.12). In the left panel, the curve for the regime-depend scalar $a_n$ is upward sloping, implying an upward-sloping yield curve in both regimes. The curve for the high-volatility regime is about 1% higher, and the higher $a_n$ is largely due to the 1979–1982 high-volatility episode where the mean of bond yields is also high. In the right panel, factor loadings $b_n$ are independent of regimes. Similar to the finding in Dai, Singleton and Yang (2007), the interpretation of $b_n$ differs from the standard interpretation of factor loadings on level, slope and curvature factors. This is because the model allows flexible correlations between latent factors by assuming non-zero off-diagonal elements of autoregressive parameters, while the factors in the standard interpretation are obtained from principal component analysis. For ease of interpretation, we express the latent factors as a function of the bond yields that are priced without error, i.e.,

\[
\begin{bmatrix}
F_1 \\
F_2 \\
F_3
\end{bmatrix} = \text{const} + \begin{bmatrix} +67 & -69 & -76 \\ +87 & -227 & +188 \\ +211 & -139 & +48 \end{bmatrix} \times \begin{bmatrix} y_6 \\ y_{24} \\ y_{120} \end{bmatrix}.
\] (3.65)

The first factor is approximated by $-[(y_{24} - y_6) + y_{120}]$, i.e., having a loading of $[+1, -1, -1]$ on the bond yields, where $(y_{24} - y_6)$ is termed the short-term slope. The second factor has a curvature flavour with an approximate loading of $[+2, -5, +4]$ on
the bond yields. The third factor is similar to the first factor but places more weight on the short-term yield \( y_6 \) and is approximated by \([y_6 - (y_24 - y_6)]\), i.e., having a loading of \([+2, -1, 0]\) on the bond yields.

### 3.5.2 Results after bias correction

Tables 3.4 and 3.5 report, after bias correction, the estimates of reduced-form and structural parameters respectively. Compared with the parameter estimates before bias correction in Tables 3.2 and 3.3, the estimated process of bond yields and latent factors becomes more persistent, as indicated by the increase in the maximum eigenvalues of the estimated autoregressive parameters. The process of regimes also becomes more persistent, as indicated by the increase in the estimated transition probability parameters.

To highlight the bias correction procedure, Table 3.6 summarises the estimates of key parameters in the different stages of bias correction, where notations follow those in Section 3.4.3. In the first row of Panel A, \( \hat{\phi}_{11}^{mle} \) is the estimate of \( \phi_{11} \) before bias correction, which has a maximum eigenvalue of 0.9920 in the low-volatility regime and a maximum eigenvalue of 0.9791 in the high-volatility regime. In the second row, as given in equation (3.59), \( \hat{\phi}_{11}^{boot,mle} \) is the mean of ML estimates over the bootstrapped samples that are generated using \( \hat{\phi}_{11}^{mle} \) as data generating parameters. If there is no bias in ML estimates, \( \hat{\phi}_{11}^{boot,mle} \) should be close to \( \hat{\phi}_{11}^{mle} \). However, this is not the case as indicated by the maximum eigenvalues: the persistence of bond yields is underestimated in both regimes, and the bias is severe in the high-volatility regime. In the third row, \( \hat{\phi}_{11}^{bc} \) is the estimate of \( \phi_{11} \) after bias correction, where the maximum eigenvalue is adjusted from 0.9920 to 0.9936 in the low-volatility regime and is adjusted from 0.9791 to 0.9955 in the high-volatility regime. In the fourth row, \( \hat{\phi}_{11}^{boot,bc} \) is the mean of ML estimates over the bootstrapped samples that are generated using \( \hat{\phi}_{11}^{bc} \) as data generating parameters. Ideally, \( \hat{\phi}_{11}^{boot,bc} \) should be close to \( \hat{\phi}_{11}^{mle} \) in the first row, so that on average, the bias-corrected estimate \( \hat{\phi}_{11}^{bc} \) produces the sample estimate \( \hat{\phi}_{11}^{mle} \). However, as shown in Panel A, the eigenvalues of \( \hat{\phi}_{11}^{boot,bc} \) and \( \hat{\phi}_{11}^{mle} \) are not identical. A more persistent \( \hat{\phi}_{11}^{mle} \) implies that the true \( \phi_{11} \) is more persistent than the bias-corrected estimate \( \hat{\phi}_{11}^{bc} \)—in other words, the bias correction procedure...
corrects the bias but does not do so fully. As discussed in Section 3.4.3, this is likely because the estimation bias depends on the value or the persistence of parameter $\phi_{11}$, which cannot be accounted for using the traditional bootstrap bias correction method. However, we show later that even though the bootstrap bias correction is not perfect, it offers an insight into the economic implications of bias correction. Any more refined bias correction methods are expected to strengthen our findings.

Panel B of Table 3.6 summarises the estimates of transition probability parameters in each stage of the bias correction procedure. The ML estimates are slightly downward biased so that after bias correction, both regimes become more persistent. Specifically, the average duration of the low-volatility regime increases from 82 months before bias correction to 92 months after bias correction, and the average duration of the high-volatility regime increases from 30 months to 38 months. Moreover, $\hat{\pi}_{jj}^{\text{boot}_c}$ is close to $\hat{\pi}_{jj}^{\text{mle}}$—the bias-corrected estimates on average produce the sample estimates—which indicates that the bias correction procedure fully corrects the estimation bias in transition probability parameters. As shown in Figure 3.2, after bias correction, the 1983–1984 period is also classified as being in the high-volatility regime by the smoothed probability. In Dai, Singleton and Yang (2007), the whole 1982–1985 period is classified as being in the high-volatility regime by the filtered probability, and this high persistence is the result of assuming time-varying transition probability of regimes under the physical measure.

To assess the economic implications of bias correction, we examine the implications of bias correction for the inference about the risk neutral forward rates and the forward term premium. As shown in Section 3.3.2, the $n$-month actual forward rate, $f_{n,t}$, is decomposed into a risk neutral forward rate, $\bar{f}_{n,t}$, and a forward term premium, $ftp_{n,t}$. The risk neutral forward rate reflects the expected one-month policy rate at month $t + n - 1$, i.e., one month before the maturity date, and the forward term premium is the associated risk premium. As discussed in Section 3.4.3, the estimation bias in the autoregressive parameters distorts the decomposition of actual forward rates into risk neutral forward rates and forward term premia: it under-estimates the volatility of risk neutral forward rates and thus produces misleading inference about forward term premia.
Table 3.7 reports the volatility of the actual forward rates, the risk neutral forward rates and the forward term premia, implied by the estimates of structural parameters before and after bias correction. Four maturities are considered: 6-month, 2-year, 5-year and 10-year, which are the maturities used in model estimation. Consistent with the analysis in Section 3.4.3, bias correction is irrelevant to the estimates of actual rates and only affects the decomposition of actual rates into risk neutral rates and term premia. As predicted, bias correction increases the volatility of risk neutral forward rates: it increases the volatility by about one half for the 5-year risk neutral rates, and doubles the volatility for the 10-year risk neutral rates. The effect of bias correction is negligible for the short-term 6-month forward rates, which is consistent with the intuition that the estimation bias in the persistence of yields is more relevant to the long-run expectations of policy rates. As shown in the column for term premia, bias correction decreases the volatility of term premia. Since the inference about term premia is indirectly corrected, there is no clear prediction of how the volatility will change after bias correction. For example, Bauer, Rudebusch and Wu (2012) find that the volatility of term premia actually increases after bias correction.

Figures 3.4–3.7 plot the actual forward rates, the risk neutral forward rates and the forward term premia, before and after bias correction, for the 6-month, 2-year, 5-year and 10-year forward rates respectively. For the short-term 6-month forward rates in Figure 3.4, the effects of bias correction on risk neutral rates and term premia are negligible in most sample periods, which is consistent with the finding in Table 3.7. However, even for the 6-month forward rates, there is a hint of difference between before and after bias correction in the 1979–1982 episode and in the recent financial crisis. Specifically, after bias correction, risk neutral rates are higher in the 1979–1982 episode and are lower in the recent financial crisis. As the maturity of forward rates increases in Figures 3.5–3.7, the difference between before and after bias correction becomes more significant. Throughout these figures, in the 1979–1982 episode, before bias correction, the higher-than-ever actual forward rate is decomposed into a historically high risk neutral rate and an exceptionally high term premium. In contrast, after bias correction, it turns out that the higher-than-ever actual forward rate is decomposed into a higher-than-ever risk neutral rate and a moderate term premium. In the recent financial crisis, before bias correction, the lower-than-ever actual forward
rate is decomposed into a historically low risk neutral rate and a modest term premium. In contrast, after bias correction, it turns out that the lower-than-ever actual forward rate is decomposed into a lower-than-ever risk neutral rate and a high term premium. In summary, in these two high-volatility episodes, after bias correction, the exceptionally high and the exceptionally low actual forward rates are interpreted as being driven by the higher-than-ever and the lower-than-ever expectations of future policy rates. While term premia are high in these two episodes, they belong to the normal business-cycle variation of term premia in the full sample period, i.e., neither being exceptionally high in the 1979–1982 episode nor being counter-intuitively low in the recent financial crisis. In this sense, bias correction decreases the volatility of term premia.

3.6 Monte Carlo Simulation

This section uses Monte Carlo simulation to address two related concerns. The first concern is that the bias correction procedure is designed to correct the estimation bias in model parameters and thus only indirectly corrects the estimation bias in risk neutral rates and term premia. The second concern is that these two rates depend on model parameters in a highly non-linear way. As a result, it remains unclear how effective the bias correction procedure is in terms of correcting the estimation bias in risk neutral rates and term premia. Using Monte Carlo simulation, this section shows that correcting the estimation bias in model parameters on average produces more accurate estimates of risk neutral rates and term premia, compared to the estimates of these rates before bias correction.

3.6.1 Simulation design

The simulation procedure parallels the bootstrap bias correction procedure in Section 3.4.3 in the general set-ups. Let \( \hat{\Theta}_{mle}^{redu} \) denote the sample estimates of reduced-form parameters before bias correction as reported in Table 3.2, let \( \hat{\Theta}_{bc}^{redu} \) denote the sample estimates after bias correction as reported in Table 3.4, and let
\( \hat{\Theta}_{stru}^{bc} \) \(_{\text{sample}} \) denote the sample estimates of structural parameters after bias correction as reported in Table 3.5. We use the bias-corrected estimates as data generating parameters, i.e., \((\Theta_{redu})_{dgp} = (\hat{\Theta}_{redu}^{bc})_{\text{sample}} \), or equivalently, \((\Theta_{stru})_{dgp} = (\hat{\Theta}_{stru}^{bc})_{\text{sample}} \).

We generate \( M = 1000 \) samples of regimes and bond yields according to equations (3.5), (3.38) and (3.44). The sample size is \( T = 451 \) months\(^2\), and in each sample, the first observation of bond yields that are priced without error is randomly sampled from the historical dataset. We impose the same restrictions on samples of regimes as in the bootstrap procedure. Then for each simulated sample \( m \), we estimate reduced-form models, and the ML estimates are denoted by \((\hat{\Theta}_{redu}^{mle})_{m} \). The bias in \((\hat{\Theta}_{redu}^{mle})_{m} \) is corrected by adding back the bias calculated from the bootstrap bias correction procedure, i.e.,

\[
(\hat{\Theta}_{redu}^{bc})_{m} = (\hat{\Theta}_{redu}^{mle})_{m} + \hat{\Theta}_{redu}^{diff},
\]

where \( \hat{\Theta}_{redu}^{diff} \), given in equation (3.63), is the estimation bias when the data generating parameter is \((\hat{\Theta}_{redu}^{mle})_{sample} \). The underlying assumption of step (3.66) is that the estimation bias when the true parameter is \((\hat{\Theta}_{redu}^{bc})_{sample} \) is equal to the estimation bias when the true parameter is \((\hat{\Theta}_{redu}^{mle})_{sample} \), which is exactly the underlying assumption of the bootstrap bias correction procedure.

Then for each simulated sample \( m \), given the estimates of reduced-form parameters before and after bias correction, we estimate the structural parameters and therefore the actual forward rates \((f)\), the risk neutral rates \((\hat{f})\) and the forward term premia \((ftp)\). The estimated time series of forward rates before and after bias correction are denoted by \( F^{mle} \) and \( F^{bc} \) respectively, where \( F = \{f, \hat{f}, ftp\} \). The true forward

\(^2\)In simulation, the sample size is set equal to that in the data, i.e., \( T = 451 \) month—a little more than 37 years. It is in the future agenda to show via simulation the magnitude of estimation bias if the sample size is increased to, e.g., 100 years. While the estimation bias is expected to decrease as the sample size increases, the literature does not often examine this relationship. For example, Bauer, Rudebusch and Wu (2012) only consider a sample size of \( T = 216 \) months—18 years as in their dataset; Duffee and Stanton (2012) only consider \( T = 1000 \) weeks—a little more than 19 years. Droumaguet (2012) shows via simulation that the mean squared error of estimates in RS-VARs generally decreases substantially when the sample size is increased to \( T = 800 \), while the economic meaning of \( T = 800 \) is vague given the chosen parameter values: matching his \( \phi = 0.9 \) with bond yield data, a sample size of \( T = 800 \) roughly corresponds to 200–800 years of data, while other parameters are less comparable with what is typically estimated in empirical RS-DTSMs.
CHAPTER 3. CORRECTING SMALL SAMPLE BIAS

rates, denoted by \( F_{\text{dgp}} \), is calculated using the data generating structural parameters, \((\Theta_{\text{stru}})_{\text{dgp}}\). We then calculate the volatility (sample standard deviation) of forward rates, denoted by \( \sigma(F_{\text{dgp}}), \sigma(F_{\text{mle}}), \sigma(F_{\text{bc}}) \), and the root-mean-square error (RMSE) of forward rates, denoted by \( \text{RMSE}(F_{\text{mle}}), \text{RMSE}(F_{\text{bc}}) \), where

\[
\text{RMSE}(F^i) = \sqrt{\frac{\sum_{t=1}^{T}(F^i_t - F_{\text{dgp}})^2}{T}}, \quad i = \text{mle}, \text{bc}.
\] (3.67)

Finally, given \( \sigma(F_{\text{dgp}}), \sigma(F_{\text{mle}}), \sigma(F_{\text{bc}}) \) and \( \text{RMSE}(F_{\text{mle}}), \text{RMSE}(F_{\text{bc}}) \) for \( m = 1, 2, \ldots, M \), we calculate the mean and the median of sample volatilities and sample RMSEs over the \( M = 1000 \) simulated samples.

3.6.2 Simulation results

Tables 3.8–3.11 report the simulation results for the four maturities respectively. Consistent with the empirical findings, after bias correction, the volatility of risk neutral rates increases, and the volatility of term premia decreases. The simulation results support the empirical results in showing that before bias correction, the volatility of risk neutral rates is indeed under-estimated compared with the true volatility, and after bias correction, the volatility of both risk neutral rates and term premia moves towards the true volatility. The empirical results are also supported in terms of RMSE. As shown in these tables, actual rates are estimated accurately with an RMSE less than 0.1 percentage point. Risk neutral rates and term premia are estimated with much less accuracy. An important observation is that for both risk neutral rates and term premia, RMSE is lower after bias correction, which means that the bias correction procedure on average produces better estimates of both rates. The gain in RMSE from bias correction is comparable to that reported in Bauer, Rudebusch and Wu (2012): they use 4-year forward rates and the gain in RMSE is about 0.1 percentage point, and our gain in RMSE for 2-year and 5-year forward rates is also about 0.1 percentage point in the case of full sample. Another observation from these tables is that the gain in RMSE differs across regimes. Specifically, the gain is moderate in the low-volatility regime, at about 0.05 percentage point for the long-term forward rates. This is consistent with the finding in Table 3.6 that the estimates of the autoregressive parameter in the low-volatility regime is less biased. In contrast, the gain
in RMSE is much higher in the high-volatility regime, at about 0.1–0.3 percentage point for long-maturity forward rates, which is consistent with the empirical finding that the inference about risk neutral rates and risk premia changes dramatically in two high-volatility episodes after bias correction.

It can also be seen that while the bias correction procedure corrects the bias it does not do so fully. Before bias correction, the volatility of risk neutral rates is on average under-estimated by 6%, 23%, 45% and 65% for the 6-month, 1-year, 5-year and 10-year maturity respectively. After bias correction, the volatility is still under-estimated by 1%, 3%, 8% and 11% respectively, and the gain in RMSE actually decreases as the maturity increases from 5-year to 10-year. As discussed earlier, this is likely because the estimation bias in parameter $\phi_{11}$ depends on the value or the persistence of $\phi_{11}$ itself, which cannot be accounted for using the traditional bootstrap bias correction method, and we leave more refined bias correction procedures, e.g., the indirect inference approach proposed by Bauer, Rudebusch and Wu (2012), to future work.

In summary, simulation results show that the gain from bias correction prevails across maturities and regimes and in both economic measures. The positive simulation results lend confidence to the empirical findings.

### 3.7 Concluding Remarks

Correcting the small sample bias in the estimation of the regime switching dynamic term structure model has important implications for the estimated persistence of regimes and bond yields and for the inference about risk neutral rates and term premia, especially in the high-volatility regimes that is less frequent in the sample. After bias correction, the exceptionally high long-term forward rates in the 1979–1982 episode and the exceptionally low rates in the recent financial crisis are to a great extent driven by expectations of future policy rates. While term premia are high in these two periods, they belong to the normal business-cycle variation of term premia in the full sample period, without being exceptionally high in the 1979–1982 episode or being counter-intuitively low in the recent financial crisis as in the case before bias.
correction. The empirical results are supported by Monte Carlo simulation where we show that correcting small sample bias leads to more accurate estimates of risk neutral rates and term premia compared to before bias correction.
CHAPTER 3. CORRECTING SMALL SAMPLE BIAS

Bibliography


CHAPTER 3. CORRECTING SMALL SAMPLE BIAS

Table 3.1: Summary statistics of bond yields

This table reports the summary statistics of bond yields. The sample mean, the standard deviation of yields, the standard deviation of yield changes, the maximum, the minimum, the autocorrelation coefficient and the statistics of unit root test are reported for four bond maturities: 6-month, 2-year, 5-year and 10-year. In the Augmented Dickey-Fuller test (ADF), a null hypothesis of unit root is tested in the level of a series, including an intercept in the test equation, and the lag length is selected via BIC rules with a maximum lag of 15. The sample period is from 1971M11 to 2009M05. Bond yields are annualised and in percentage.

<table>
<thead>
<tr>
<th></th>
<th>6m</th>
<th>2y</th>
<th>5y</th>
<th>10y</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean(y)</td>
<td>6.16</td>
<td>6.60</td>
<td>7.00</td>
<td>7.41</td>
</tr>
<tr>
<td>std(y)</td>
<td>3.11</td>
<td>2.94</td>
<td>2.69</td>
<td>2.44</td>
</tr>
<tr>
<td>std(Δy)</td>
<td>0.56</td>
<td>0.48</td>
<td>0.39</td>
<td>0.35</td>
</tr>
<tr>
<td>max(y)</td>
<td>16.22</td>
<td>15.78</td>
<td>15.18</td>
<td>14.89</td>
</tr>
<tr>
<td>min(y)</td>
<td>0.33</td>
<td>0.57</td>
<td>1.56</td>
<td>2.88</td>
</tr>
<tr>
<td>autocorr(y)</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>ADF test, t-stats</td>
<td>-1.58</td>
<td>-1.29</td>
<td>-1.36</td>
<td>-1.22</td>
</tr>
<tr>
<td>p-value</td>
<td>0.49</td>
<td>0.64</td>
<td>0.61</td>
<td>0.67</td>
</tr>
</tbody>
</table>
### Table 3.2: Estimates of reduced-form parameters, before bias correction

This table reports the estimates of reduced-form parameters before bias correction. Standard errors are in brackets. Bond yields are annualised and not in percentage.

\[
\text{log likelihood } = 19.6126
\]

<table>
<thead>
<tr>
<th>Regime L (low volatility)</th>
<th>Regime H (high volatility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond yields priced without error, given ( q_{t+1} = k ) and ( q_t = j )</td>
<td></td>
</tr>
<tr>
<td>( Y_{1,t+1} = \alpha_1^k + \phi_{11}^j \left( Y_{1,t} - \alpha_1^j \right) + \Sigma_1^j u_{t+1} )</td>
<td></td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.0598</td>
</tr>
<tr>
<td></td>
<td>[0.0173]</td>
</tr>
<tr>
<td>( \phi_{11} )</td>
<td>0.8802</td>
</tr>
<tr>
<td></td>
<td>[0.0332]</td>
</tr>
<tr>
<td></td>
<td>0.0282</td>
</tr>
<tr>
<td></td>
<td>[0.0521]</td>
</tr>
<tr>
<td></td>
<td>0.0623</td>
</tr>
<tr>
<td></td>
<td>[0.0502]</td>
</tr>
<tr>
<td>eigenvalue</td>
<td>0.9920</td>
</tr>
<tr>
<td>( \Sigma_1 [1 \text{e-2}] )</td>
<td>0.2544</td>
</tr>
<tr>
<td></td>
<td>[0.0078]</td>
</tr>
<tr>
<td></td>
<td>0.2638</td>
</tr>
<tr>
<td></td>
<td>[0.0136]</td>
</tr>
<tr>
<td></td>
<td>0.1430</td>
</tr>
<tr>
<td></td>
<td>[0.0130]</td>
</tr>
</tbody>
</table>

| Bond yields priced with error, given \( q_t = j \) | | |
| \( Y_{2,t} = \alpha_2^j + \phi_{21}^j \left( Y_{1,t} - \alpha_1^j \right) + \Sigma_2^j \epsilon_t \) | | |
| \( \alpha_2 \) | 0.0689 | 0.0756 | | | | |
| | [0.0145] | | | | | |
| \( \phi_{21} \) | -0.2002 | 0.7306 | 0.4735 | | | |
| | [0.0103] | [0.0157] | [0.0075] | | | |
| \( \Sigma_2 [1 \text{e-2}] \) | 0.0661 | 0.1226 | | | | |
| | [0.0025] | | | | | |

| Transition probability of regimes, given \( q_t = j \) and \( q_{t+1} = k \) | | |
| \( \pi_{jk}^P = \text{const.} \) | | |
| \( \pi_{jj}^P \) | 0.9878 | 0.9672 | | | | |
| | [0.0216] | [0.0214] | | | | |
CHAPTER 3. CORRECTING SMALL SAMPLE BIAS

Table 3.3: Estimates of structural parameters, before bias correction

This table reports the estimates of structural parameters before bias correction. Standard errors are in brackets. Even though price-of-risk parameters are implied by other parameters, their standard errors are also reported.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Regime L (low volatility)</th>
<th>Regime H (high volatility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c^P$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho^P$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c^Q$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho^Q$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price of risk, given $q_t = j$</td>
<td>$\lambda_t(q_t) = (\Sigma^j)^{-1} \left( \lambda_0^j + \lambda_1^j \right)$, where $\lambda_0^j = c^Pj + \rho^Pj F_t + \Sigma^j u_t^P + 1$</td>
<td>$\lambda_t(q_t) = (\Sigma^j)^{-1} \left( \lambda_0^j + \lambda_1^j \right)$, where $\lambda_0^j = c^Pj + \rho^Pj$</td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>$-0.0024$</td>
<td>$0.0115$</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$-0.0361$</td>
<td>$-0.0216$</td>
</tr>
<tr>
<td>One-period bond yield, given $q_t = j$</td>
<td>$r_t = \delta_0^j + \delta_1^j F_t$</td>
<td>$r_t = \delta_0^j + \delta_1^j F_t$</td>
</tr>
<tr>
<td>$12 \times n_0[1e-2]$</td>
<td>$5.7931$</td>
<td>$7.0014$</td>
</tr>
<tr>
<td>$[1.7517]$</td>
<td>$[1.7529]$</td>
<td></td>
</tr>
<tr>
<td>$12 \times n_1[1e-2]$</td>
<td>$-0.0616$</td>
<td>$-0.1754$</td>
</tr>
<tr>
<td>$[0.0780]$</td>
<td>$[0.0899]$</td>
<td>$[0.0244]$</td>
</tr>
<tr>
<td>Transition probability of regimes, given $q_t = j$ and $q_{t+1} = k$</td>
<td>$\pi_{j,k}^Q = \pi_{j,k}^P = \text{const.}$</td>
<td>$\pi_{j,k}^Q = \pi_{j,k}^P = \text{const.}$</td>
</tr>
<tr>
<td>$\pi_{j,j}^P$</td>
<td>$0.9878$</td>
<td>$0.9672$</td>
</tr>
<tr>
<td>$[0.0216]$</td>
<td>$[0.0214]$</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4: Estimates of reduced-form parameters, after bias correction

This table reports the estimates of reduced-form parameters after bias correction. Standard errors are in brackets. Bond yields are annualised and not in percentage.

\[
\text{log likelihood = 19.5472}
\]

<table>
<thead>
<tr>
<th></th>
<th>Regime L (low volatility)</th>
<th>Regime H (high volatility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond yields priced without error, given ( q_{t+1} = k ) and ( q_t = j )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_{1,t+1} = \alpha_1^k + \phi_{11}(\hat{Y}<em>{1,t} - \alpha_1^j) + \Sigma_1^j u</em>{t+1} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.0593 0.0638 0.0725</td>
<td>0.0706 0.0723 0.0800</td>
</tr>
<tr>
<td></td>
<td>[0.0215] [0.0206] [0.0162]</td>
<td>[0.0220] [0.0215] [0.0166]</td>
</tr>
<tr>
<td>( \phi_{11} )</td>
<td>0.8752 0.1746 -0.0627</td>
<td>0.3392 0.9286 -0.3569</td>
</tr>
<tr>
<td></td>
<td>[0.0273] [0.0426] [0.0234]</td>
<td>[0.1992] [0.4160] [0.2805]</td>
</tr>
<tr>
<td></td>
<td>0.0180 0.9604 0.0171</td>
<td>-0.2230 1.2689 -0.0672</td>
</tr>
<tr>
<td></td>
<td>[0.0485] [0.0710] [0.0352]</td>
<td>[0.1673] [0.3483] [0.2329]</td>
</tr>
<tr>
<td></td>
<td>0.0513 -0.0615 1.0046</td>
<td>-0.2001 0.3544 0.7954</td>
</tr>
<tr>
<td></td>
<td>[0.0498] [0.0682] [0.0298]</td>
<td>[0.0679] [0.1390] [0.0919]</td>
</tr>
<tr>
<td>( \Sigma_1[1\text{e-2}] )</td>
<td>0.2168</td>
<td>0.9852</td>
</tr>
<tr>
<td></td>
<td>[0.0055]</td>
<td>[0.0620]</td>
</tr>
<tr>
<td></td>
<td>0.2314 0.1637 0.6808</td>
<td>0.2999</td>
</tr>
<tr>
<td></td>
<td>[0.0117] [0.0056] [0.0443]</td>
<td>[0.0254]</td>
</tr>
<tr>
<td></td>
<td>0.1118 0.1666 0.1359</td>
<td>0.3000 0.2121 0.2753</td>
</tr>
<tr>
<td></td>
<td>[0.0121] [0.0103] [0.0050]</td>
<td>[0.0392] [0.0328] [0.0278]</td>
</tr>
<tr>
<td>Bond yields priced with error, given ( q_t = j )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_{2,t} = \alpha_2^j + \phi_{21}(\hat{Y}_{1,t} - \alpha_1^j) + \Sigma_2^j \epsilon_t )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.0684 0.0756</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0184] [0.0192]</td>
<td></td>
</tr>
<tr>
<td>( \phi_{21} )</td>
<td>-0.2005 0.7308 0.4734</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0107] [0.0161] [0.0076]</td>
<td></td>
</tr>
<tr>
<td>( \Sigma_2[1\text{e-2}] )</td>
<td>0.0666</td>
<td>0.1256</td>
</tr>
<tr>
<td></td>
<td>[0.0025]</td>
<td>[0.0077]</td>
</tr>
<tr>
<td>Transition probability of regime, given ( q_t = j ) and ( q_{t+1} = k )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \pi_{jk}^P )</td>
<td>0.9891 0.9738</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0168] [0.0134]</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.5: Estimates of structural parameters, after bias correction

This table reports the estimates of structural parameters after bias correction. Standard errors are in brackets. Even though price-of-risk parameters are implied by other parameters, their standard errors are also reported.

<table>
<thead>
<tr>
<th></th>
<th>regime L (low volatility)</th>
<th>regime H (high volatility)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Latent factor under the P-measure, given ( q_t = j )</td>
<td>( F_{t+1} = c^{P,j} + \rho^{P,j} F_t + \Sigma^j u^F_{t+1} )</td>
</tr>
<tr>
<td>( c^P )</td>
<td>0.0019 0.0108 0.0868</td>
<td>0.0037 0.0227 0.0800</td>
</tr>
<tr>
<td>( \rho^P )</td>
<td>[0.0152] [0.0150] [0.0590]</td>
<td>[0.0208] [0.0170] [0.0844]</td>
</tr>
<tr>
<td></td>
<td>0.9633 0.336 0.9859</td>
<td>0.0090 0.0033 0.9282</td>
</tr>
<tr>
<td>( \rho^Q )</td>
<td>[0.0097] [0.0063] [0.0028]</td>
<td>[0.0106] [0.0081] [0.0032]</td>
</tr>
<tr>
<td></td>
<td>0.9994 0.9718 0.9062</td>
<td>0.9994 0.9718 0.9062</td>
</tr>
<tr>
<td></td>
<td>Price of risk, given ( q_t = j )</td>
<td>( \lambda_t(q_t) = (\Sigma^j)^{-1} \left( \lambda^j_0 + \lambda^j_1 F_t \right) ), where ( \lambda^j_0 = c^{P,j} - c^{Q,j} ), ( \lambda^j_1 = \rho^{P,j} - \rho^{Q} )</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>-0.0019 0.0108 -0.0868</td>
<td>0.0337 -0.0227 -0.0800</td>
</tr>
<tr>
<td></td>
<td>[0.0152] [0.0150] [0.0590]</td>
<td>[0.0208] [0.0170] [0.0844]</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>-0.0294 -0.0175 -0.0109</td>
<td>-0.0232 0.0014 -0.0162</td>
</tr>
<tr>
<td></td>
<td>[0.0261] [0.0239] [0.0116]</td>
<td>[0.0332] [0.0761] [0.0118]</td>
</tr>
<tr>
<td></td>
<td>-0.0137 -0.0377 -0.0024</td>
<td>-0.1095 -0.3096 -0.0415</td>
</tr>
<tr>
<td></td>
<td>[0.0239] [0.0257] [0.0129]</td>
<td>[0.0973] [0.1013] [0.0210]</td>
</tr>
<tr>
<td></td>
<td>0.0758 -0.0179 0.0300</td>
<td>-0.2951 -0.4920 -0.1412</td>
</tr>
<tr>
<td></td>
<td>[0.0153] [0.0225] [0.0068]</td>
<td>[0.1528] [0.2441] [0.0657]</td>
</tr>
<tr>
<td>One-period bond yield, given ( q_t = j )</td>
<td>( r_t = \delta^0_q + \delta^1_q F_t )</td>
<td></td>
</tr>
<tr>
<td>( 12 \times \delta^0_q )</td>
<td>5.7440 [2.1544]</td>
<td>7.0067 [2.1693]</td>
</tr>
<tr>
<td>( 12 \times \delta^1_q )</td>
<td>0.0399 [-0.1599] 0.7398</td>
<td>[0.0624] [0.0597] [0.0177]</td>
</tr>
<tr>
<td>Transition probability of regime, given ( q_t = j ) and ( q_{t+1} = k )</td>
<td>( \pi^Q_{jk} = \pi^P_{jk} = \text{const.} )</td>
<td></td>
</tr>
<tr>
<td>( \pi^P_{jk} )</td>
<td>0.9891 0.9738</td>
<td>[0.0168] [0.0134]</td>
</tr>
</tbody>
</table>
Table 3.6: Parameter estimates in the procedure of bootstrap bias correction

This table summarises the eigenvalues of the estimated autoregressive parameters (Panel A) and the estimates of transition probability parameters (Panel B) in the different stages of the bootstrap bias correction procedure. Please refer to Section 3.4.3 for a detailed explanation of notations.

### A. eigenvalues of bond-yield autoregressive parameters

<table>
<thead>
<tr>
<th></th>
<th>low-volatility regime</th>
<th>high-volatility regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\phi}_{11} ) (MLE)</td>
<td>0.9920 0.9540 0.8818</td>
<td>0.9791 0.8420 0.5078</td>
</tr>
<tr>
<td>( \hat{\phi}_{11} ) (Boot MLE)</td>
<td>0.9807 0.9352 0.8706</td>
<td>0.9320 0.7845 0.4636</td>
</tr>
<tr>
<td>( \hat{\phi}_{11} ) (BC)</td>
<td>0.9936 0.9627 0.8840</td>
<td>0.9955 0.8722 0.5358</td>
</tr>
<tr>
<td>( \hat{\phi}_{11} ) (Boot BC)</td>
<td>0.9828 0.9436 0.8729</td>
<td>0.9470 0.8162 0.4908</td>
</tr>
</tbody>
</table>

### B. transition probabilities of regimes

<table>
<thead>
<tr>
<th></th>
<th>low-volatility regime</th>
<th>high-volatility regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\pi}_{jj} ) (MLE)</td>
<td>0.9878 0.9672</td>
<td>0.9865 0.9605</td>
</tr>
<tr>
<td>( \hat{\pi}_{jj} ) (Boot MLE)</td>
<td>0.9891 0.9738</td>
<td>0.9877 0.9672</td>
</tr>
</tbody>
</table>
Table 3.7: Volatility of forward rates before and after bias correction

This table reports the volatility (sample standard deviation) of the actual forward rates ($f$), the risk neutral forward rates ($\hat{f}$), and the forward term premia ($ftp$), implied by the estimates of structural parameters before and after bias correction. The maturities are 6-month, 2-year, 5-year and 10-year. Forward rates are annualised and in percentage.

<table>
<thead>
<tr>
<th></th>
<th>A. 6m</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f$</td>
<td>$\hat{f}$</td>
<td>$ftp$</td>
<td></td>
</tr>
<tr>
<td>Before BC</td>
<td>3.0720</td>
<td>2.8047</td>
<td>0.5052</td>
<td></td>
</tr>
<tr>
<td>After BC</td>
<td>3.0716</td>
<td>2.9122</td>
<td>0.3993</td>
<td></td>
</tr>
</tbody>
</table>

|       |       | B. 2y |       |       |       |
|-------|-------|-------|-------|-------|
|       | $f$   | $\hat{f}$ | $ftp$ |       |
| Before BC | 2.7834 | 2.1589 | 0.8524 |       |
| After BC  | 2.7836 | 2.5517 | 0.5720 |       |

|       |       | C. 5y |       |       |       |
|-------|-------|-------|-------|-------|
|       | $f$   | $\hat{f}$ | $ftp$ |       |
| Before BC | 2.3849 | 1.3388 | 1.2475 |       |
| After BC  | 2.3835 | 1.9555 | 0.8294 |       |

|       |       | D. 10y |       |       |       |
|-------|-------|-------|-------|-------|
|       | $f$   | $\hat{f}$ | $ftp$ |       |
| Before BC | 2.1957 | 0.6671 | 1.6502 |       |
| After BC  | 2.1982 | 1.3196 | 1.1871 |       |
Table 3.8: Simulation results: 6-month forward rates

This table reports the simulated economic implications of bias correction for the inference about actual forward rates \((f)\), risk neutral forward rates \((\tilde{f})\) and forward term premia \((ftp)\). Panel A compares the volatility (sample standard deviation) of the forward rates implied by the data generating parameters and the parameter estimates before and after bias correction. Panel B compares the root-mean-square error (RMSE) of the forward rates implied by the parameter estimates before and after bias correction in the full sample and in regimes. For each economic measure, the mean and the median are calculated over simulated samples. Forward rates are annualised and in percentage.

<table>
<thead>
<tr>
<th></th>
<th>A. volatility, (\sigma(\cdot))</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ((f)) ((\tilde{f})) ((ftp))</td>
<td>median ((f)) ((\tilde{f})) ((ftp))</td>
<td></td>
</tr>
<tr>
<td>DGP</td>
<td>2.6498 2.5760 0.3357</td>
<td>2.5104 2.4521 0.3319</td>
<td></td>
</tr>
<tr>
<td>Before BC</td>
<td>2.6501 2.4312 0.5115</td>
<td>2.5145 2.3161 0.4943</td>
<td></td>
</tr>
<tr>
<td>After BC</td>
<td>2.6501 2.5593 0.4471</td>
<td>2.5144 2.4468 0.4314</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B. root-mean-square error, RMSE((\cdot))</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ((f)) ((\tilde{f})) ((ftp))</td>
<td>median ((f)) ((\tilde{f})) ((ftp))</td>
<td></td>
</tr>
<tr>
<td>I. full sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before BC</td>
<td>0.0090 0.3501 0.3511</td>
<td>0.0077 0.3358 0.3366</td>
<td></td>
</tr>
<tr>
<td>After BC</td>
<td>0.0088 0.3234 0.3243</td>
<td>0.0075 0.3057 0.3062</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>II. low-volatility regime</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ((f)) ((\tilde{f})) ((ftp))</td>
<td>median ((f)) ((\tilde{f})) ((ftp))</td>
<td></td>
</tr>
<tr>
<td>Before BC</td>
<td>0.0074 0.1753 0.1757</td>
<td>0.0062 0.1529 0.1531</td>
<td></td>
</tr>
<tr>
<td>After BC</td>
<td>0.0074 0.1647 0.1653</td>
<td>0.0062 0.1457 0.1471</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>III. high-volatility regime</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ((f)) ((\tilde{f})) ((ftp))</td>
<td>median ((f)) ((\tilde{f})) ((ftp))</td>
<td></td>
</tr>
<tr>
<td>Before BC</td>
<td>0.0126 0.6230 0.6251</td>
<td>0.0099 0.5916 0.5941</td>
<td></td>
</tr>
<tr>
<td>After BC</td>
<td>0.0119 0.5730 0.5748</td>
<td>0.0096 0.5230 0.5228</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.9: Simulation results: 2-year forward rates

This table reports the simulated economic implications of bias correction for the inference about actual forward rates ($f$), risk neutral forward rates ($\hat{f}$) and forward term premia ($ftp$). Panel A compares the volatility (sample standard deviation) of the forward rates implied by the data generating parameters and the parameter estimates before and after bias correction. Panel B compares the root-mean-square error (RMSE) of the forward rates implied by the parameter estimates before and after bias correction in the full sample and in regimes. For each economic measure, the mean and the median are calculated over simulated samples. Forward rates are annualised and in percentage.

<table>
<thead>
<tr>
<th>A. volatility, $\sigma(\cdot)$</th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>($f$) ($\hat{f}$) ($ftp$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGP</td>
<td>2.3246</td>
<td>2.1969</td>
</tr>
<tr>
<td>Before BC</td>
<td>2.3245</td>
<td>1.6989</td>
</tr>
<tr>
<td>After BC</td>
<td>2.3243</td>
<td>2.1229</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. root-mean-square error, RMSE(\cdot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. full sample</td>
</tr>
<tr>
<td>mean</td>
</tr>
<tr>
<td>($f$) ($\hat{f}$) ($ftp$)</td>
</tr>
<tr>
<td>Before BC</td>
</tr>
<tr>
<td>After BC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. low-volatility regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
</tr>
<tr>
<td>($f$) ($\hat{f}$) ($ftp$)</td>
</tr>
<tr>
<td>Before BC</td>
</tr>
<tr>
<td>After BC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. high-volatility regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
</tr>
<tr>
<td>($f$) ($\hat{f}$) ($ftp$)</td>
</tr>
<tr>
<td>Before BC</td>
</tr>
<tr>
<td>After BC</td>
</tr>
</tbody>
</table>
Table 3.10: Simulation results: 5-year forward rates

This table reports the simulated economic implications of bias correction for the inference about actual forward rates \((f)\), risk neutral forward rates \((\hat{f})\) and forward term premia \((ftp)\). Panel A compares the volatility (sample standard deviation) of the forward rates implied by the data generating parameters and the parameter estimates before and after bias correction. Panel B compares the root-mean-square error (RMSE) of the forward rates implied by the parameter estimates before and after bias correction in the full sample and in regimes. For each economic measure, the mean and the median are calculated over simulated samples. Forward rates are annualised and in percentage.

### A. volatility, \(\sigma(\cdot)\)

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th></th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((f))</td>
<td>((\hat{f}))</td>
<td>((ftp))</td>
<td>((f))</td>
<td>((\hat{f}))</td>
</tr>
<tr>
<td>DGP</td>
<td>1.8937</td>
<td>1.7190</td>
<td>0.7250</td>
<td>1.7788</td>
<td>1.6263</td>
</tr>
<tr>
<td>Before BC</td>
<td>1.8954</td>
<td>0.9382</td>
<td>1.2205</td>
<td>1.7782</td>
<td>0.7569</td>
</tr>
<tr>
<td>After BC</td>
<td>1.8939</td>
<td>1.5763</td>
<td>1.0835</td>
<td>1.7771</td>
<td>1.4939</td>
</tr>
</tbody>
</table>

### B. root-mean-square error, \(\text{RMSE}(\cdot)\)

#### I. full sample

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th></th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((f))</td>
<td>((\hat{f}))</td>
<td>((ftp))</td>
<td>((f))</td>
<td>((\hat{f}))</td>
</tr>
<tr>
<td>Before BC</td>
<td>0.0193</td>
<td>1.2033</td>
<td>1.2047</td>
<td>0.0172</td>
<td>1.1295</td>
</tr>
<tr>
<td>After BC</td>
<td>0.0193</td>
<td>1.0598</td>
<td>1.0608</td>
<td>0.0173</td>
<td>0.9857</td>
</tr>
</tbody>
</table>

#### II. low-volatility regime

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th></th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((f))</td>
<td>((\hat{f}))</td>
<td>((ftp))</td>
<td>((f))</td>
<td>((\hat{f}))</td>
</tr>
<tr>
<td>Before BC</td>
<td>0.0169</td>
<td>1.0795</td>
<td>1.0800</td>
<td>0.0147</td>
<td>0.9710</td>
</tr>
<tr>
<td>After BC</td>
<td>0.0169</td>
<td>1.0044</td>
<td>1.0045</td>
<td>0.0147</td>
<td>0.8901</td>
</tr>
</tbody>
</table>

#### III. high-volatility regime

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th></th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((f))</td>
<td>((\hat{f}))</td>
<td>((ftp))</td>
<td>((f))</td>
<td>((\hat{f}))</td>
</tr>
<tr>
<td>Before BC</td>
<td>0.0236</td>
<td>1.4187</td>
<td>1.4216</td>
<td>0.0204</td>
<td>1.3218</td>
</tr>
<tr>
<td>After BC</td>
<td>0.0235</td>
<td>1.1297</td>
<td>1.1323</td>
<td>0.0208</td>
<td>1.0156</td>
</tr>
</tbody>
</table>
Table 3.11: Simulation results: 10-year forward rates

This table reports the simulated economic implications of bias correction for the inference about actual forward rates \( (f) \), risk neutral forward rates \( \hat{f} \) and forward term premia \( ftp \). Panel A compares the volatility (sample standard deviation) of the forward rates implied by the data generating parameters and the parameter estimates before and after bias correction. Panel B compares the root-mean-square error (RMSE) of the forward rates implied by the parameter estimates before and after bias correction in the full sample and in regimes. For each economic measure, the mean and the median are calculated over simulated samples. Forward rates are annualised and in percentage.

<table>
<thead>
<tr>
<th>A. volatility, ( \sigma(\cdot) )</th>
<th>mean</th>
<th>median</th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( (f) )</td>
<td>( \hat{f} )</td>
<td>( ftp )</td>
<td>( (f) )</td>
</tr>
<tr>
<td>DGP</td>
<td>1.7215</td>
<td>1.2108</td>
<td>0.9823</td>
<td>1.6363</td>
</tr>
<tr>
<td>Before BC</td>
<td>1.7183</td>
<td>0.4289</td>
<td>1.4186</td>
<td>1.6293</td>
</tr>
<tr>
<td>After BC</td>
<td>1.7219</td>
<td>1.0730</td>
<td>1.2716</td>
<td>1.6318</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. root-mean-square error, RMSE(( \cdot ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. full sample</td>
</tr>
<tr>
<td>mean</td>
</tr>
<tr>
<td>Before BC</td>
</tr>
<tr>
<td>After BC</td>
</tr>
</tbody>
</table>

| II. low-volatility regime                   |
| mean | \( (f) \) | \( \hat{f} \) | \( ftp \) | median | \( (f) \) | \( \hat{f} \) | \( ftp \) |
| Before BC | 0.0459 | 1.3881 | 1.3894 | 0.0391 | 1.1939 | 1.1944 |
| After BC | 0.0451 | 1.3450 | 1.3466 | 0.0387 | 1.1701 | 1.1712 |

| III. high-volatility regime                 |
| mean | \( (f) \) | \( \hat{f} \) | \( ftp \) | median | \( (f) \) | \( \hat{f} \) | \( ftp \) |
| Before BC | 0.0615 | 1.5379 | 1.5364 | 0.0534 | 1.3793 | 1.3813 |
| After BC | 0.0606 | 1.4044 | 1.4013 | 0.0531 | 1.1743 | 1.1616 |
Figure 3.1: Bond yield data. This figure plots the US bond yields from 1971M11 to 2009M05. Bond maturities are 6-month, 2-year, 5-year and 10-year. Bond yields are annualised and in percentage.
Figure 3.2: Filtered and smoothed probabilities of being in the high-volatility regime. This figure plots the filtered and smoothed probability of being in the high-volatility regime, implied by the estimates of reduced-form parameters before and after bias correction. Shaded areas are NBER recessions.
Figure 3.3: Factor loadings of bond yields. This figure plots loadings of bond yields on the latent factors as a function of bond maturities in months. Bond yields depend on the latent factor as in equation (3.12), i.e., \( y_{n,t} = a_n + b_n' F_t \), where \( a_n \) is a regime-dependent scalar and \( b_n \) is a regime-independent vector. Bond yields are annualised and not in percentage.
Figure 3.4: 6-month forward rates before and after bias correction. This figure plots actual forward rates (Panel A), risk neutral forward rates (Panel B) and forward term premia (Panel C), implied by the estimates of structural parameters before and after bias correction. Forward rates are annualised and in percentage. Shaded areas are the high-volatility regimes classified by the smoothed probability after bias correction.
Figure 3.5: 2-year forward rates before and after bias correction. This figure plots actual forward rates (Panel A), risk neutral forward rates (Panel B) and forward term premia (Panel C), implied by the estimates of structural parameters before and after bias correction. Forward rates are annualised and in percentage. Shaded areas are the high-volatility regimes classified by the smoothed probability after bias correction.
Figure 3.6: 5-year forward rates before and after bias correction. This figure plots actual forward rates (Panel A), risk neutral forward rates (Panel B) and forward term premia (Panel C), implied by the estimates of structural parameters before and after bias correction. Forward rates are annualised and in percentage. Shaded areas are the high-volatility regimes classified by the smoothed probability after bias correction.
Figure 3.7: 10-year forward rates before and after bias correction. This figure plots actual forward rates (Panel A), risk neutral forward rates (Panel B) and forward term premia (Panel C), implied by the estimates of structural parameters before and after bias correction. Forward rates are annualised and in percentage. Shaded areas are the high-volatility regimes classified by the smoothed probability after bias correction.
Appendix

3.A Risk Neutral Bond Prices

This section derives the dependence of the risk neutral bond prices on the latent factors. We consider a zero-coupon bond with a maturity of \((n+1)\)-month at time \(t\). Given the current regime \(s_t = j\), the risk neutral price of this bond, denoted by \(\tilde{P}_{n+1,t}\), is conjectured to depend on the latent factors \(F_t\) as

\[
\tilde{P}_{n+1,t}^j = e^{-\tilde{A}_{n+1}^j - \tilde{B}_{n+1}^j F_t},
\]

where \(\tilde{A}\) is a regime-dependent scalar; \(\tilde{B}\) is an \(N \times 1\) vector that is also regime dependent. The risk neutral price of this bond is also equal to the expected price at time \(t+1\) under the \(P\)-measure discounted at the one-period risk free rate, i.e., given \(q_t = j\) and \(q_{t+1} = k\),

\[
\tilde{P}_{n+1,t}^j = E_t^P \left[ e^{-r_t^j \tilde{P}_{n+1,t+1}^j | q_t = j} \right] \\
= \sum_{k=L,H} \left( \pi_{jk}^P \cdot e^{-r_t^j} \cdot E_t^P \left[ \tilde{P}_{n+1,t+1}^k | q_t = j, q_{t+1} = k \right] \right) \\
= \sum_{k=L,H} \left( \pi_{jk}^P \cdot e^{-r_t^j} \cdot E_t^P \left[ e^{-A_k^j - B_{k+1}^j F_{t+1}^j} \cdot e^{-\frac{1}{2} B_{k+1}^j \Sigma_{k+1} B_{k+1}^j} \right] \right) \\
= \sum_{k=L,H} \left( \pi_{jk}^P \cdot e^{-r_t^j} \cdot e^{-A_k^j} \cdot e^{-\frac{1}{2} \sum_k B_{k+1}^j \Sigma_{k+1} B_{k+1}^j} \right), \tag{3.69}
\]

Equating the two prices in equations (3.68) and (3.69) gives

\[
e^{-\tilde{A}_{n+1}^j - \tilde{B}_{n+1}^j F_t} = \sum_{k=L,H} \left( \pi_{jk}^P \cdot e^{-r_t^j} \cdot e^{-A_k^j} \cdot e^{-\frac{1}{2} \sum_k B_{k+1}^j \Sigma_{k+1} B_{k+1}^j} \right) \]

\[
1 = \sum_{k=L,H} \left( \pi_{jk}^P \cdot e^{-r_t^j} \cdot e^{-A_k^j} \cdot e^{-\frac{1}{2} \sum_k B_{k+1}^j \Sigma_{k+1} B_{k+1}^j} \cdot e^{\tilde{A}_{n+1}^j + \tilde{B}_{n+1}^j F_t} \right).
\]
CHAPTER 3. CORRECTING SMALL SAMPLE BIAS

Using the approximation $e^y \approx 1 + y$ gives

$$1 \approx \sum_{k=L,H} \pi^P_{jk} \cdot \left(1 - r^j_t - \ddot{A}_n^k - \ddot{B}_n^k \mu^P_t + \frac{1}{2} \ddot{B}_n^{k'} \sum j' \ddot{B}_n^k + \ddot{A}_{n+1}^j + \ddot{B}_{n+1}^j F_t \right)$$

$$-\ddot{A}_{n+1}^j - \ddot{B}_{n+1}^j F_t = \sum_{k=L,H} \pi^P_{jk} \cdot \left(-r^j_t - \ddot{A}_n^k - \ddot{B}_n^k \mu^P_t + \frac{1}{2} \ddot{B}_n^{k'} \sum j' \ddot{B}_n^k \right),$$

where

$$\mu^P_t = c^P_j + \rho^P_j F_t$$

$$r^j_t = \delta^j_0 + \delta^j_1 F_t. \quad (3.3)$$

It follows that

$$\ddot{A}_{n+1}^j = \delta^j_0 + \sum_{k=L,H} \pi^P_{jk} \cdot \left(\ddot{A}_n^k + (c^P_j)' \ddot{B}_n^k - \frac{1}{2} \ddot{B}_n^{k'} \sum j' \ddot{B}_n^k \right)$$

$$\ddot{B}_{n+1}^j = \delta_1 + \sum_{k=L,H} \pi^P_{jk} \cdot (\rho^P_j)' \ddot{B}_n^k,$$

which are equations (3.17) and (3.18). The initial conditions are $\ddot{A}_0^j = 0$ and $\ddot{B}_0^j = 0_{N \times 1}.$
### 3.B The Mapping between Structural and Reduced-Form Parameters

This tabular shows the detailed steps in the mapping. In parentheses are numbers of elements of parameters. In the last column, N stands for numerical solution and A stands for analytical solution.

<table>
<thead>
<tr>
<th>step</th>
<th>equation</th>
<th>reduced-form parameters/solved structural parameters</th>
<th>structural parameters</th>
<th>normalisation</th>
<th>restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.14, 3.33, 3.41, 3.43, 3.37</td>
<td>{\Sigma^L_i(6), \Sigma^H_i(6), \phi_{21}(3)} \cdot (15)</td>
<td>{\delta_1(3), \rho^Q(9), \Sigma^H(3)} \cdot (15) + \Sigma^L</td>
<td>\Sigma^L = I_3/\sqrt{12}</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>3.14, 3.33</td>
<td>{\delta_1, \rho^Q} from step 1</td>
<td>{\beta_1, \beta_2}</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.36</td>
<td>{\phi_{11}(9), \phi_{12}(9)} \cdot (18)</td>
<td>{\rho^{PL}(9), \rho^{PH}(9)} \cdot (18) + {\beta_1}</td>
<td>\rho^{PL} full matrix</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>3.27</td>
<td>{\rho^Q, \rho^{PL}, \rho^{PH}}</td>
<td>{\lambda_1^L(9), \lambda_1^H(9)}</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.5</td>
<td>{\pi_{LL}^Q(1), \pi_{HH}^Q(1)} \cdot (2)</td>
<td>{\pi_{LL}^Q(1), \pi_{HH}^Q(1)} \cdot (2)</td>
<td>\pi_{jj}^P = \pi_{jj}^Q, j = L, H</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>3.13, 3.32, 3.40</td>
<td>{\alpha_1^L(3), \alpha_1^H(3), \alpha_2^L(3), \alpha_2^H(1)} \cdot (8)</td>
<td>{\delta_0^L(1), \delta_0^H(1), \epsilon^{QL}(3), \epsilon^{QH}(3)} \cdot (8) + {\pi_{LL}^Q, \pi_{HH}^Q} from step 5 + {\delta_1, \rho^Q, \Sigma^H} from step 1 + \Sigma^L</td>
<td>c^{PL} = 0_{3\times1}</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>3.26</td>
<td>{c^{PL}, c^{PH}, c^{QL}, c^{QH}}</td>
<td>{\lambda_0^L(3), \lambda_0^H(3)}</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4

Learning about the Persistence of Recessions under Ambiguity Aversion

Abstract

We incorporate ambiguity aversion into the process of parameter learning and assess the asset pricing implications of the model. Ambiguity is characterised by the unknown parameter that governs the persistence of recessions, and the representative investor learns about this parameter while being ambiguity averse towards parameter uncertainty. We examine model-implied conditional moments and simulated moments of asset prices and returns, and document an uncertainty effect that characterises the difference between learning under ambiguity aversion and learning under standard recursive utility. This uncertainty effect is asymmetric across economic expansions and recessions, and this asymmetry generates in simulation a sharp increase in the equity premium at the onset of recessions, as in the recent financial crisis.
4.1 Introduction

Learning about the persistence of recessions is difficult due to their less frequent nature. However, the persistence of recessions has important implications for asset prices and returns. For instance, there were fears in the early stages of the recent financial crisis that it might develop into a depression, where a key concern was how long the crisis would persist. On the academic side, Collin-Dufresne, Johannes and Lochstoer (2016) show that the uncertainty over the persistence of recessions leads to the biggest welfare loss and therefore has the largest asset pricing impact, compared to the uncertainty over the magnitude and the volatility of consumption shocks in recessions. In addition, it is well known that the equity premium is countercyclical (Campbell and Shiller, 1988a, 1988b; Fama and French, 1989). Recently, Martin (2016) estimates that in November 2008—the height of the recent financial crisis, the (monthly) annualised expected return on the US stock market peaked at 55.0%, and this is in sharp contrast to the Great Moderation period from 2004 to 2006 when the average expected return was 1.86%. In this paper, we incorporate ambiguity aversion into the process of learning about the persistence of recessions and assess the asset pricing implications of the model. We reproduce in simulation the sharp increase in the equity premium at the onset of recessions, and show that ambiguity aversion models offer desirable flexibility in matching the equity premium compared to standard recursive preferences.

There are two ingredients in the model: the source of ambiguity and the attitude towards ambiguity. The source of ambiguity describes a situation in which there is uncertainty over the distribution of random outcomes, which is different from “risk” where the distribution is known. In an economy with two regimes characterising business cycle fluctuations and labelled as “recession” and “expansion” respectively, the source of ambiguity refers to the uncertainty over the unknown parameter that governs the persistence of the recession regime, or more precisely, the uncertainty over the distribution of (random) regimes in the next period given that the current period is in recession. As an illustrative example, suppose that the representative investor enters into the recession regime and is confronted with two possible economic outcomes: the average duration of this recession is either 2 years or 5 years. She then learns from historical samples and forms Bayesian probabilities of the two outcomes. The term
“risk” describes a situation where the probability and therefore the distribution of the two outcomes is assumed to be known, and the source of risk is the randomness in the realisation of average durations. In contrast, “ambiguity” describes a situation where the representative investor also concerns about the uncertainty over the probability of the on-average short and long recessions. This source of ambiguity is motivated by the observation that learning about the persistence of recessions is difficult due to their less frequent occurrence. For example, Collin-Dufresne, Johannes and Lochstoer (2016) show that after 200 years of learning about an average-4-year-duration and once-per-century recession, the 5th and 95th percentile of the duration distribution are about 1 quarter and 12 years respectively.

The attitude towards ambiguity determines whether the representative investor is averse to ambiguity and if she is, the magnitude of ambiguity aversion. The basic intuition for ambiguity aversion is that compared to an ambiguity-neutral investor, an ambiguity-averse investor places more weight on the scenario that has lower continuation value, as if the probability of this bad state is distorted endogenously to be higher than the probability updated via Bayes’ rule. Using the previous illustrative example, suppose that the Bayesian probability of an average-5-year-duration recession is 0.5. If the representative investor is averse to ambiguity, she has concerns over this probability and places more weight on the scenario of long recessions as if the probability of long recessions is higher than 0.5. The extant literature studies the uncertainty over the persistence of recessions under constant relative risk aversion (CRRA) utility (Cecchetti, Lam and Mark, 2000; Cogley and Sargent, 2008) and recursive preferences (Collin-Dufresne, Johannes and Lochstoer, 2016), and we are the first to examine this uncertainty under ambiguity aversion, employing the generalized recursive smooth ambiguity preferences of Ju and Miao (2012).

The asset pricing implications of the model are assessed from two perspectives. Firstly, we solve the model using the projection method and examine the conditional moments of asset prices and returns. Secondly, we simulate the model and examine the long-sample moments before and at the onset of recessions. In both analyses, we compare parameter learning under ambiguity aversion with parameter learning under recursive preferences (Epstein and Zin, 1989—henceforth, EZ), and document
an uncertainty effect that characterises the difference between the two. Specifically, an ambiguity-averse investor requires higher compensation for parameter uncertainty compared to an EZ investor. More importantly, this uncertainty effect, driven by the distorted probability in the ambiguity model, is asymmetric across regimes—the size of the uncertainty effect is moderate in the expansion regime and is substantial in the recession regime. This asymmetry plays a key role in matching the sharp increase in the equity premium at the onset of recessions. Empirically, Martin (2016) estimates that the equity premium at the height of the recent financial crisis is 30 times as much as the equity premium during the 2004–2006 Great Moderation at the one-month horizon and is 10 times as much at the one-year horizon. Our simulation is designed to capture these two periods in history, where after 100 years of learning\footnote{A sample length of 100 years is comparable with the 1890–1994 sample of consumption data used in Cecchetti, Lam and Mark (2000).}, the onset of the next recession represents the height of the recent financial crisis and the preceding period represents the Great Moderation. At the annual frequency, our simulation results show that at the onset of the next recession, ambiguity models are able to generate 10–30 times the equity premium in the preceding period, and increasing the magnitude of ambiguity aversion will further increase the ratio. By contrast, recursive utility models generate 3–12 times the equity premium in the preceding period, and increasing the magnitude of risk aversion does not help increase the ratio. In this sense, ambiguity models offer desirable flexibility in matching the sharp increase in the equity premium at the onset of recessions.

The remainder of the paper proceeds as follows. Section 4.2 reviews related literature with an emphasis on subjective beliefs about the persistence of recessions and learning under ambiguity aversion. Section 4.3 introduces the model, including learning about the persistence of recessions and the generalized recursive ambiguity preferences of Ju and Miao (2012). Sections 4.4 and 4.5 examine the conditional moments and long-sample moments of asset prices and returns and compare parameter learning under ambiguity aversion and parameter learning under EZ preferences. Section 4.6 offers some concluding remarks and suggestions for future research.
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERTION

4.2 Related Literature

This study is closely related to the literature on subjective beliefs about the persistence of recessions or rare disaster events. Cecchetti, Lam and Mark (2000) consider an endowment economy where the representative investor has CRRA preferences and consumption growth follows a Markov switching process with two regimes in the mean, characterising economic expansions and contractions. Regimes are observable, but the true transition probabilities of regimes are unknown to the investor. The subjective belief about the persistence of regimes is systematically distorted, and the distorted belief is time-varying and persistent. Their model is able to match the first and second moments of the risk free rate and the equity premium. Since Cecchetti, Lam and Mark (2000) assume permanent distortion of beliefs, there is no role for learning about the persistence of regimes. Cogley and Sargent (2008) consider the same economy as in Cecchetti, Lam and Mark (2000), except that the investor updates her belief about the persistence of contractions via Bayes’ rule with a pessimistic prior. The pessimistic prior is induced by using a short training sample ending with the 1929–1933 Great Depression and using a robust calculation that twists the prior pessimistically. Simulation results show that sufficient parameter pessimism generates a high market price of risk. Collin-Dufresne, Johannes and Lochstoer (2016) examine the asset pricing implications of parameter learning under EZ preferences, including learning about the persistence of rare disasters. They show that parameter learning is a martingale process and contributes a subjective long run risk when the investor prefers an early resolution of uncertainty. In their learning-about-rare-disaster model, consumption growth follows a two-regime Markov switching process with the disaster regime happening once per century and mimicking the 1929–1933 Great Depression in terms of magnitude and persistence, and the EZ investor learns about the parameter that governs the persistence of disasters. Their simulation results show that learning about the persistence of rare disasters contributes to the first and second moments of asset returns. This effect decays over time but still remains even after 300 years of learning. Gillman, Kejak and Pakos (2015) examine learning about a highly persistent recession, namely a lost decade. Specifically, they consider a three-regime Markov
chain where two regimes characterise two recessions of the same magnitude but different persistence. The average persistence of the shorter recession is slightly above one year, and the average persistence of the longer recession is ten years, characterising a lost decade. Their model matches a wide range of consumption and asset pricing phenomena as observed in the postwar US sample. While the focus of these studies is to match the long-sample moments of asset prices and returns, the focus of this paper is to match the sharp increase in the equity premium at the onset of recessions.\footnote{Cogley and Sargent (2008), Ju and Miao (2012) and this paper adopt the same parameters of regime and consumption growth as in Cecchetti, Lam and Mark (2000), where the contraction regime has an annual consumption decline of \(-6.785\%\), and on average persists 2 years and occurs twice per century. While a contraction is not an ordinary recession, we use the term “recession” for this regime as in Ju and Miao (2012). The rare disaster regime in Collin-Dufresne, Johannes and Lochstoer (2016) mimics the 1929–1933 Great Depression in terms of magnitude and persistence, i.e., having a consumption decline of \(-4.6\%\) and an average persistence of 4 years. This rare disaster regime is assumed to occur once per century on average.} Despite the difference in focus, our model does not require a rare occurrence of recessions as in Collin-Dufresne, Johannes and Lochstoer (2016), or a possibility of highly persistent recessions as in Cogley and Sargent (2008) and Gillman, Kejak and Pakos (2015). Collin-Dufresne, Johannes and Lochstoer (2016) also examine asset prices and returns at the onset of recessions under EZ preferences, however, we show that ambiguity models offer desirable flexibility in matching the equity premium at the onset of recessions compared to recursive utility models.

This paper is also closely related to the literature on learning under ambiguity aversion. Using an example where the representative investor with robustness concern\footnote{The robustness theory developed by Hansen (2007), Hansen and Sargent (2001, 2007, 2008) shares the intuition of ambiguity aversion where the investor fears model misspecification and deals with this uncertainty by distorting probabilities pessimistically towards the model that has low continuation value.} learns between the long run risk model of Bansal and Yaron (2004) and a model with much less persistent consumption growth, Hansen and Sargent (2010) show that the belief about models is distorted towards the long run risk model and this distortion contributes a countercyclical price of risk. In a related work, Collard, Mukerji, Sheppard and Tallon (2016) consider a parameter learning model where the representative investor learns about the persistence parameter in the mean of consumption growth in the long run risk model of Bansal and Yaron (2004). The ambiguity aversion of the investor is characterised by the smooth ambiguity model of Klibanoff, Marinacci and Mukerji (2005, 2009). Using a sample from 1930 to 2007 (1930–1977 as training
sample), their model successfully generates a counter-cyclical conditional equity premium with a pro-cyclical conditional volatility. Chen, Ju and Miao (2014) study the implications of learning about competing models for asset allocation. The ambiguity-averse investor learns between two models of excess market return—an IID model and a predictive model with dividend yield as the predictor. They show that the difference in asset allocation between ambiguity and recursive utility models is substantial when the IID model is unlikely and the dividend yield is high. As in Hansen and Sargent (2010), Collard, Mukerji, Sheppard and Tallon (2016), we find that the equity premium is counter-cyclical, and we go further by investigating the sharp increase in the equity premium at the onset of recessions. Similar to Chen, Ju and Miao (2014), we find that the difference in equity premium between ambiguity and recursive utility models is substantial when the current regime is recession and the probability of long recessions is low.

4.3 The Model

4.3.1 Learning about the persistence of recessions

We consider a pure-exchange economy. At each time period $t$, a representative investor consumes one consumption good $C_t$, and trades two financial assets: a risk free bond with zero net supply and equity that pays dividend $D_t$. As in Cecchetti, Lam and Mark (2000), Cogley and Sargent (2008), Ju and Miao (2012), we consider a parsimonious two-regime Markov switching model for consumption growth, i.e.,

$$
\Delta c_t = \mu_c(s_t) + \sigma_c \cdot \varepsilon_t,
$$

(4.1)

where $\Delta c_t = \ln \left( \frac{C_t}{C_{t-1}} \right)$ is the log of consumption growth, and $\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$. There are two regimes in the mean of consumption growth: $s_t = 1, 2$ follows an observable two-regime Markov process, where regime $1 = \{\mu^l\}$ characterises an economic recession with low consumption growth, and regime $2 = \{\mu^h\}$ characterises an economic expansion with high consumption growth. The transition probability of regimes is
assumed to be constant over time, and is given by

\[ \Pi = \begin{bmatrix} \pi_{11} & 1 - \pi_{11} \\ 1 - \pi_{22} & \pi_{22} \end{bmatrix}, \quad (4.2) \]

where \( \pi_{jk} = p(s_{t+1} = k|s_t = j) \) is the probability of being in regime \( k \) next period given that the current regime is regime \( j \). Following Ju and Miao (2012), the process of dividend growth is specified as

\[ \Delta d_t = \mu_d + \lambda \cdot \Delta c_t, \quad (4.3) \]

where \( \Delta d_t = \ln \left( \frac{D_t}{D_{t-1}} \right) \) is the log of dividend growth, and \( \lambda \) can be interpreted as the leverage ratio of dividend growth on consumption growth as in Abel (1999). The leverage parameter helps capture the empirical observation that dividend growth has a much higher volatility than consumption growth. Since \( \mu_d \) is assumed to have a single regime, dividend growth is regime-dependent only through its dependence on consumption growth.

Due to the less frequent nature of the recession regime, it is assumed that the investor is uncertain about the persistence of this regime, i.e., parameter \( \pi_{11} \), and only learns about this parameter gradually as she observe more recessions. Learning about the unknown parameter \( \pi_{11} \) is modelled as learning between two possible values, i.e.,

\[ \pi_{11} = \begin{cases} \pi_{11}^l, & \text{if the persistence of recessions is low (short recession)} \\ \pi_{11}^h, & \text{if the persistence of recessions is high (long recession)} \end{cases} \quad (4.4) \]

where \( \pi_{11}^l \) characterises a shorter recession with \( \pi_{11}^l < \pi_{11}^h \). This two-point learning method is also used by Collard, Mukerji, Sheppard and Tallon (2016) where the investor learns between two possible values for the persistence parameter in the mean of consumption growth. Since two-point learning is specified as learning between two models that are only different in the parameter(s) of interest, it is a special case of learning between competing models, a common set-up in the robustness theory developed by Hansen (2007), Hansen and Sargent (2001, 2007, 2008), where the investor has concerns about model misspecification.

The probabilities of the two parameter values are updated via Bayes’ rule. Since regimes are assumed to be observable, the probability is updated using the observation
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERSION 150

of regimes only. Let \( s^t = \{ s_\tau; \tau \leq t \} \) denote the history of regimes up to time \( t \), and let \( p_t \) denote the probability of the long-recession parameter value conditional on the information set \( s^t \), i.e.,

\[
p_t \equiv p(\pi_{11} = \pi_{11}^h | s^t).
\]

(4.5)

Given an initial guess \( p_0 = p(\pi_{11} = \pi_{11}^h | s^0) \), the probability of the long-recession parameter value is updated via Bayes’ rule, i.e.,

\[
p_{t+1} = p(\pi_{11} = \pi_{11}^h | s^{t+1}) = p(\pi_{11} = \pi_{11}^h | s_{t+1}, s^t) = \frac{p(s_{t+1}, \pi_{11} = \pi_{11}^h | s^t)}{p(s_{t+1} | s^t)}
\]

\[
= \frac{p(s_{t+1}, \pi_{11} = \pi_{11}^h | s^t)}{\sum_{i=l,h} p(s_{t+1} | s^t, \pi_{11} = \pi_{11}^i) \cdot p_t}
\]

\[
= p(s_{t+1} | s^t, \pi_{11} = \pi_{11}^l) \cdot (1 - p_t) + p(s_{t+1} | s^t, \pi_{11} = \pi_{11}^h) \cdot p_t,
\]

(4.6)

where \( p(s_{t+1} | s^t, \pi_{11} = \pi_{11}^i), i = l, h \) is the transition probability of regimes conditional on the parameter value \( \pi_{11}^i \). Given the regimes in the current and the next periods, i.e., \( \{ s_t, s_{t+1} \} \), the updating process can be rewritten as

\[
p_{t+1} = \begin{cases} 
\pi_{11}^h \cdot p_t & \text{if } s_t = 1, \ s_{t+1} = 1 \\
\pi_{11}^l \cdot (1 - p_t) + \pi_{11}^h \cdot p_t & \text{if } s_t = 1, \ s_{t+1} = 2 \\
(1 - \pi_{11}) \cdot (1 - p_t) + (1 - \pi_{11}) \cdot p_t & \text{if } s_t = 2.
\end{cases}
\]

(4.7)

That is, if the current regime is the recession regime (\( s_t = 1 \)), the probability of the long-recession parameter value increases as the economy spends more time in the recession regime, and decreases as the economy switches from the recession to the expansion regime. If the current regime is the expansion regime (\( s_t = 2 \)), there is no update on the probability of the long-recession parameter value, i.e., there is no learning about the persistence of recessions in the expansion regime.

4.3.2 Learning under ambiguity aversion

We incorporate the generalized recursive smooth ambiguity model of Ju and Miao (2012) into the above learning procedure. The continuation value of the representative
investor, denoted by $V_t$, is given by

$$V_t = \left\{ (1 - \beta)C_t^{1-\rho} + \beta \left[ R_t(V_{t+1}) \right]^{1-\rho} \right\}^{\frac{1}{1-\rho}}$$

(4.8)

$$R_t(V_{t+1}) = \left\{ E_{p_t} \left[ \left( E_{\pi_{11},t} \left( V_{t+1}^{1-\gamma} \right) \right)^{\frac{1}{1-\gamma}} \right] \right\}^{\frac{1}{1-\gamma}}$$

(4.9)

where $C_t$ is consumption at time $t$; $R_t(V_{t+1})$ is time $t$’s certainty equivalent of the continuation value at time $t+1$; $\pi_{11}$ is the unknown parameter, and $p_t$ is the Bayesian probability of parameter values; $E_{\pi_{11},t}$ calculates the expectation conditional on each possible value for $\pi_{11}$ and conditional on the history up to time $t$, and $E_{p_t}$ calculates the expectation given the probability of parameter values; $\beta \in (0, 1)$ is time discount factor; $1/\rho$ is the elasticity of intertemporal substitution (EIS); $\gamma$ is the risk aversion coefficient; $\eta$ is the ambiguity aversion coefficient. The pricing kernel is given by

$$M_{\pi_{11},t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left( \frac{V_{t+1}}{R_t(V_{t+1})} \right)^{\rho-\gamma} \left( \frac{E_{\pi_{11},t} \left( V_{t+1}^{1-\gamma} \right)}{R_t(V_{t+1})} \right)^{\frac{1}{1-\gamma}} \right)^{-\eta}$$

(4.10)

and the gross return of any traded assets at time $t$, denoted by $R_{i,t+1}$, satisfies the Euler equation

$$E_t \left[ M_{\pi_{11},t+1} \cdot R_{i,t+1} \right] = 1,$$

(4.11)

where the expectation $E_t$ is taken over the predictive distribution conditional on the history up to time $t$. The gross risk free rate at time $t$, denoted by $R_{f,t}$, is the reciprocal of the expectation of the pricing kernel, i.e.,

$$R_{f,t} = \frac{1}{E_t \left[ M_{\pi_{11},t+1} \right]}.$$

(4.12)

An investor is ambiguity averse if and only if $\eta > \gamma$. She is ambiguity neutral if $\eta = \gamma$, then the third term in equation (4.10) is equal to one, and the pricing kernel reduces to the pricing kernel under EZ recursive utility. The basic intuition for ambiguity aversion is that compared to an ambiguity-neutral investor, an ambiguity-averse investor places more weight on the scenario that has lower continuation value, as if the probability of this bad state is distorted endogenously to be higher than the Bayesian probability. This mechanism can be shown by decomposing the pricing
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERSION

kernel into an EZ component and an ambiguity-aversion component,\(^4\) i.e.,

\[ M_{\pi_{11},t+1} = M_{t+1}^{EZ} \cdot M_{\pi_{11}}^{AA}, \] \tag{4.13}

where

\[ M_{t+1}^{EZ} \equiv \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left( \frac{V_{t+1}}{R_{t}(V_{t+1})} \right)^{\rho-\gamma} \] \tag{4.14}

\[ M_{\pi_{11}}^{AA} \equiv \left( \frac{E_{\pi_{11}}\left[ V_{t+1}^{1-\gamma} \right]}{R_{t}(V_{t+1})} \right)^{-(\eta-\gamma)}. \] \tag{4.15}

Using this decomposition, the Euler equation (4.11) can be written as

\[ \left[ p_t \cdot M_{\pi_{11}}^{AA} \right] \cdot E_{\pi_{11}}^{h,t} \left[ M_{t+1}^{EZ} \cdot (R_{i,t+1} - R_{f,t}) \right] \]
\[ + \left[ (1 - p_t) \cdot M_{\pi_{11}}^{AA} \right] \cdot E_{\pi_{11}}^{l,t} \left[ M_{t+1}^{EZ} \cdot (R_{i,t+1} - R_{f,t}) \right] = 0, \] \tag{4.16}

where \( p_t \) is the Bayesian probability of the long-recession parameter value. There exists a distorted belief, denoted by \( \tilde{p}_t \), that satisfies

\[ \tilde{p}_t \cdot E_{\pi_{11}}^{h,t} \left[ M_{t+1}^{EZ} \cdot (R_{i,t+1} - R_{f,t}) \right] + (1 - \tilde{p}_t) \cdot E_{\pi_{11}}^{l,t} \left[ M_{t+1}^{EZ} \cdot (R_{i,t+1} - R_{f,t}) \right] = 0, \] \tag{4.17}

where

\[ \tilde{p}_t = \begin{cases} 
\frac{p_t \cdot E_{\pi_{11}}^{h,t} \left[ V_{t+1}^{1-\gamma} \right]}{p_t \cdot E_{\pi_{11}}^{h,t} \left[ V_{t+1}^{1-\gamma} \right] + (1 - p_t) \cdot E_{\pi_{11}}^{l,t} \left[ V_{t+1}^{1-\gamma} \right]} & \text{if } s_t = 1 \\
p_t & \text{if } s_t = 2. \end{cases} \] \tag{4.18}

In the recession regime \((s_t = 1)\), the distorted belief is greater than the Bayesian probability if the investor is ambiguity averse towards parameter uncertainty, and the two probabilities are equal if and only if the investor is ambiguity neutral \((\eta = \gamma)\).

By distorting the Bayesian probability of the long-recession parameter value, \( p_t \), to a higher value, \( \tilde{p}_t \), an ambiguity-averse investor pessimistically places more weight on the scenario of the long-recession parameter value, which characterises a lower continuation value compared to the short-recession parameter value. In the expansion regime \((s_t = 2)\), there is no distortion from the Bayesian probability.

\(^4\)This decomposition is only an approximation as \( M_{t+1}^{EZ} \) is not identical to the EZ pricing kernel—\( M_{t+1}^{EZ} \) can be obtained for various values of \( \eta \) as long as \( \eta \geq \gamma \) while the EZ pricing kernel is obtained only when \( \eta = \gamma \), and the continuation value \( V_{t+1} \) varies slightly with \( \eta \). Despite being an approximation, this decomposition, and the distorted probability shown later, account for the majority effect of ambiguity aversion.
4.3.3 Choice of parameters

We adopt the parameters of regime and consumption growth from Cecchetti, Lam and Mark (2000), who estimate model (4.1) using annual consumption data from 1890 to 1994. Their parameter estimates are reported in Panel A of Table 4.1. As shown, there are two regimes in the mean of consumption growth. The mean in regime 1 is $-6.785\%$, characterising an economic recession, and the mean in regime 2 is $2.251\%$, characterising an expansion. The standard deviation of consumption growth has a single regime and is equal to $3.127\%$. The expansion regime is very persistent with $\pi_{22} = 0.978$ while the recession regime is much less persistent with $\pi_{11} = 0.516$. The economy spends most of time in the expansion regime: the unconditional probability of being in expansion is $0.96$ ($= \frac{1-0.516}{2-0.516-0.978}$). The transition probability of the recession regime corresponds to a 2-year ($= \frac{1}{1-0.516}$) recession every 50 years ($= \frac{2}{1-0.96}$). For the parameters of dividend growth, we follow Abel (1999) and set the leverage parameter equal to $2.74$. The parameter $\mu_d$ is set equal to $-3.23\%$ such that dividend growth has the same unconditional mean as consumption growth as in Bansal and Yaron (2004).

In our model, the choice of the two possible values for the transition probability parameter is guided by the estimate of $\pi_{11}$ in Cecchetti, Lam and Mark (2000). Their estimate $\hat{\pi}_{11} = 0.516$ is not accurate with a t-value of 1.95—from a rough calculation using the t-value and the normal asymptotic approximation, the 90% confidence interval is $(0.0820, 0.9500)$. We set the short-recession parameter value equal to their estimate, i.e., $\pi^l_{11} = 0.516$, and set the long-recession parameter value at $\pi^h_{11} = 0.8$, which is well within the 90% interval and corresponds to an 5-year recession every 50.45 years. The choice of $\pi^h_{11} = 0.8$ is also comparable with the pessimistic prior for learning about parameter $\pi_{11}$ in Cogley and Sargent (2008). Using the same parameters as in Cecchetti, Lam and Mark (2000) and a short training sample ending with the 1929–1933 Great Depression, Cogley and Sargent (2008) estimate a prior belief about $\pi_{11}$ to be 0.805, and further distort the prior to 0.915 and 0.952 using robust methods. Since our choice of 0.8 is the upper bound of the parameter value, our prior about the persistence of recessions is much less pessimistic than those in Cogley and Sargent (2008).

We choose the preference parameters following the convention in the literature and
set $\beta = 0.9750$ and $\rho = 1/2$. We calibrate the risk aversion coefficient at $\gamma = 2$ such that in simulation, after 100 years of learning with a prior of $p_0 = 0.5$, the mean of the equity premium in the preceding period of the next recession is close to 2.07–2.81%, as estimated for the Great Moderation period in Martin (2016); we calibrate the ambiguity aversion coefficient at $\eta = 15$ such that the mean of the equity premium at the onset of the next recession is close to the estimated equity premium of 21.5% at the height of the recent financial crisis.\footnote{See Section 4.5 for details of the simulation procedure, which is designed to capture these two periods in history.} The literature uses a wide range of the ambiguity aversion coefficient. For instance, in a state (regime) learning model, Ju and Miao (2012) use an ambiguity aversion coefficient of $\eta = 8.864$ as the benchmark value and examine a range of $[\gamma, 15]$ in their comparative statics analysis. In a parameter learning model, Chen, Ju and Miao (2014) set $\eta = 80$ as the benchmark value. We take into account the wide range of the coefficient by considering three other values ($\eta = 5, 30, 60$) in comparative statics. To have an economic sense about the selected ambiguity aversion coefficients, we use the thought experiment in Ju and Miao (2012), which is related to the Ellsberg (1961) paradox. Suppose subjects confront two urns: urn 1 contains 50 black and 50 white balls, while urn 2 contains either 100 black balls or 100 white balls with equal probability. If a subject picks a black ball from an urn, she wins a prize, and if she picks a white ball, she does not win or lose anything. Camerer (1999) and Halevy (2007) show that most subjects prefer betting on urn 1 rather than urn 2. This paradox cannot be explained by standard expected utility models, and is able to be explained by the static version of Ju and Miao (2012) if subjects are ambiguity aversion. The difference between the certainty equivalents of urn 1 and urn 2 for an ambiguity-averse subject is a measure of ambiguity premium. Using power functions and given a risk aversion coefficient of $\gamma = 2$ and a prize-wealth ratio of 1%, the ambiguity premium is calculated to be 3.23% of the expected prize value for an ambiguity aversion coefficient of $\eta = 15$. Increasing the prize-wealth ratio and the ambiguity aversion coefficient will further increase the ambiguity premium, which is $\{0.75\%, 6.94\%, 14.22\%\}$ for $\eta = \{5, 30, 60\}$ respectively. Camerer (1999) shows that the ambiguity premium is 10–20% of the expected prize. Compared to his estimate, our benchmark value for the ambiguity aversion coefficient is small
but is still being reasonable and consistent with experimental findings. An obvious limitation of thought experiments is that the plausible range of ambiguity aversion coefficients is rather wide, and we leave more refined calibration and estimation of the ambiguity aversion coefficient, e.g., the method of detection-error probability in Jahan-Parvar and Liu (2014) and the structural estimation in Gallant, Jahan-Parvar and Liu (2015), to future work.

4.4 Results: Conditional Moments

In this section, we examine the moments of asset prices and returns conditional on the state variables. In the models with parameter learning, the state variable is \( \{p_t, s_t\} \), where \( p_t \) is time \( t \)'s probability of the long-recession parameter value and \( s_t \) is the observable regime (expansion or recession) at time \( t \). We then solve the models numerically using the projection method where the value function and the P/D ratio are conjectured to be functions of the state variables. The full details are given in the Appendices 4.A and 4.B. We document an uncertainty effect that characterises the difference between learning under ambiguity aversion and learning under EZ preferences. This uncertainty effect, driven by the distorted probability in the ambiguity model, is asymmetric across regimes—the size of the uncertainty effect is moderate in the expansion regime and is substantial in the recession regime. As will be shown in Section 4.5, this asymmetry plays a key role in matching the sharp increase in the equity premium at the onset of recessions.

4.4.1 Distorted probability

Before a formal examination of the conditional moments of asset prices and returns, we plot in Figure 4.1 the distorted probability of the long-recession parameter value in the ambiguity model where the risk aversion coefficient and the ambiguity aversion coefficient are set at their benchmark values: \( \gamma = 2 \) and \( \eta = 15 \). The x-axis is the Bayesian probability of the long-recession parameter value, which is also the belief of an ambiguity-neutral investor. As shown in equation (4.17), if the investors is ambiguity
averse towards parameter uncertainty, in the recession regime, the ambiguity-aversion component in the pricing kernel distorts the Bayesian probability of the long-recession parameter value to be higher pessimistically. As shown in Figure 4.1, in the recession regime, the Bayesian probability is distorted upwards, leading to a more pessimistic belief about the persistence of recessions. For instance, if the Bayesian probability of the long-recession parameter value is 0.20 and 0.05,\(^6\) the distorted belief is 0.52 and 0.12 respectively, suggesting that even if the probability of the long-recession parameter value is low from a Bayesian perspective, an ambiguity-averse investor attaches a much higher probability to the long-recession parameter value. Since the distorted belief is the source of the uncertainty effect of learning under ambiguity aversion, it is expected that the uncertainty effect is much higher in the recession regime than in the expansion regime, as there is no distortion from the Bayesian probability in the expansion regime. We quantify the uncertainty effect next and highlight the asymmetry in the size of the effect across regimes.

### 4.4.2 Conditional moments

The conditional moments we consider are the risk free rate, the expected excess market return (the equity premium), the market price of risk, the P/D ratio and the volatility of the pricing kernel and the market return. Starting from Figure 4.2, we plot the model-implied moments of asset prices and returns as functions of the (Bayesian) probability of the long-recession parameter value and conditional on the recession regime, the expansion regime, and unconditional on regime. By “unconditional on regime” we mean that the moments of asset prices and returns do not depend on the regime at time \(t\), however, the moments are still conditional on the probability of the long-recession parameter value. The moments unconditional on regime are calculated by taking a weighted average of the moments conditional on regime, where the weights are the stationary probability of regimes calculated with \(\pi_{11} = 0.516\) and \(\pi_{22} = 0.978\). Since the economy spends most of time in the expansion regime, the moments conditional on the expansion regime have a weight of 0.96 \(\left(= \frac{1-0.516}{2-0.516-0.978}\right)\),

\(^6\)As will be shown in Table 4.2, \(p = 0.20\) and \(p = 0.05\) are roughly the mean of the probability after 200 years of learning from a prior of \(p_0 = 0.5\) and \(p_0 = 0.1\) respectively
which is 24 times the weight of 0.04 for the recession regime. To help interpret these figures, we consider a real-world example where the short-recession parameter value represents the true persistence of recessions while the representative investor has a high prior belief about the long-recession parameter value. As the investor gradually learns the true persistence, the probability of the long-recession parameter value will decrease from \( p = 1 \) to \( p = 0 \)—we move along the x-axis from the right to the left when we interpret these figures.

Figures 4.2–4.4 show the effects of parameter learning on asset prices and returns. We consider four model specifications: (i) the parameter of transition probability is known at \( \pi_{11} = 0.516 \) with \( \gamma = 2 \); (ii) the parameter is unknown and the investor learns about the parameter under EZ preferences with \( \gamma = 2 \); (iii) the parameter is unknown and the investor learns about the parameter under ambiguity aversion with \( \gamma = 2 \) and \( \eta = 15 \); (iv) the parameter is known at \( \pi_{11} = 0.8 \) with \( \gamma = 2 \). For the models with parameter certainty (i and iv), in each regime, the conditional moments of asset prices and returns are constant. For the models with parameter learning (ii and iii), the conditional moments are functions of the probability of the long-recession parameter value.

There are two common features in the models with parameter learning. Firstly, if the probability is very high at \( p = 1 \) or very low at \( p = 0 \), the conditional moments implied by the models with parameter learning are equal to those implied by the model with parameter certainty at \( \pi_{11} = 0.8 \) and \( \pi_{11} = 0.516 \) respectively. In other words, the curves for learning under ambiguity aversion and learning under EZ preferences share the same start point at \( p = 1 \) and the same end point at \( p = 0 \). Secondly, as the probability decreases from \( p = 1 \) to \( p = 0 \), the curves behave differently across regimes. Consider the market price of risk in Figure 4.2 as an example. In the expansion regime, as the investor gradually learns the true parameter value, the price of risk implied by the models with parameter learning decreases monotonically and is within the bounds implied by the models with parameter certainty. In contrast, in the recession regime, as the probability decrease from \( p = 1 \) to \( p = 0 \), the price of risk firstly increases and then decreases, and in most areas of the domain, is even higher than what is implied.
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERTION 158

by the model with parameter certainty at $\pi_{11} = 0.8$. The underlying intuition is that in the expansion regime, the investor cares more about the parameter value per se and requires less compensation for risk as the long-recession parameter value is less likely, while in the recession regime where learning occurs, she cares more about the uncertainty over the parameter value and requires a high compensation when she feels uncertain about the true persistence of recessions. Specifically, suppose that the economy is currently in the recession regime, one more year of recessions will increase the probability of the long-recession parameter value. To an investor who is either ambiguity neutral or ambiguity averse, one more year of recessions is bad news and she therefore requires a higher price of risk. However, the change in the price of risk also depends on, or is dominated by, her prior probability of the long-recession parameter value. If her prior is high, e.g., $p = 0.9$, an increase in the probability is bad news however reduces the uncertainty over parameter value, and the net effect is a decrease in the price of risk. If her prior probability is low, e.g., $p = 0.1$, an increase in the probability is bad news while also increases the uncertainty over parameter value, and the total effect is an increase in the price of risk. For an EZ investor with $\gamma = 2$, a probability of $p = 0.5$ corresponds to the maximum uncertainty and the largest price of risk, while for an ambiguity-averse investor with $\gamma = 2$ and $\eta = 15$, the probability is lower at about $p = 0.2$. This is because an ambiguity-averse investor places more weight on the scenario of long recessions and behaves as if the probability of the long-recession parameter value is higher than the Bayesian probability. As shown in Figure 4.1, when the Bayesian probability of the long-recession parameter value is $p = 0.20$, the distorted belief is approximately at $\bar{p} = 0.50$, characterising the maximum uncertainty for an ambiguity-averse investor with $\gamma = 2$ and $\eta = 15$.

We highlight this uncertainty effect of learning under ambiguity aversion, which is measured by the difference between the curves for learning under ambiguity aversion and for learning under EZ preferences. An important observation is that the uncertainty effect of learning under ambiguity aversion is asymmetric across regimes—the size of the effect is moderate in the expansion regime and is substantial in the recession regime. This is not surprising as it is in the recession regime that the investor learns

---

7An exception is the P/D ratio, which increases monotonically and is within the bounds implied by the models with parameter certainty in both regimes.
about the parameter value. While the size of the uncertainty effect is asymmetric across regimes, its behaviour is similar in both regimes: as the probability decreases from $p = 1$ to $p = 0$, the uncertainty effect of learning under ambiguity aversion firstly increases and then decreases, and is maximised in the area of $p = (0.2, 0.4)$. The maximised difference characterises a hard time of learning for an ambiguity-averse investor compared to an EZ investor.

The right panel of Figures 4.2–4.4 plot the moments of asset prices and returns unconditional on regime. Since the moments conditional the recession regime have only a weight of 0.04 due to the less frequent occurrence of the recession regime, the size of the moments unconditional on regime is mostly driven by the moments conditional on the expansion regime. Despite the substantial uncertainty effect of the ambiguity model in the recession regime, the moments conditional on the recession regime only affect the curvature of the curve in the right panel. As a result, the uncertainty effect of parameter learning under ambiguity aversion is expected to be moderate unconditional on regime.

4.4.3 Comparative statics

Figures 4.5–4.7 show the effects of the ambiguity aversion coefficient on asset prices and returns. Since the literature uses a wide range of the ambiguity aversion coefficient, we consider four values that cover the range of $\eta$ in the literature, i.e., $\eta = \{5, 15, 30, 60\}$, where $\eta = 15$ is the benchmark value. The risk aversion coefficient is set at the benchmark value of $\gamma = 2$. As shown in these figures, given a fixed value of the risk aversion coefficient, changes in the ambiguity aversion coefficient have asymmetric effects across regimes, which is consistent with our previous finding that the uncertainty effect of leaning under ambiguity aversion is much stronger in the recession regime than in the expansion regime. Specifically, in the expansion regime, changes in the ambiguity aversion coefficient change the shape more than the level of the curve. As the ambiguity aversion coefficient increases from 5 to 60, the curve becomes “more concave” for the price of risk, the equity premium and the volatility of the pricing kernel and the market return, and the curve becomes “more convex” for the risk free rate and the P/D ratio. By contrast, in the recession regime, as the ambiguity aversion coefficient increases,
the change in the curve is substantial. The curve peaks at much higher levels and at smaller probabilities of the long-recession parameter value—an ambiguity-averse investor with a higher ambiguity aversion coefficient is more sensitive to a small possibility of long recessions, and requires much more compensation for the uncertainty over the small probability. In summary, due to this asymmetric effect of increasing the ambiguity aversion coefficient across regimes, the cross-regime difference in asset prices and returns also increases as the ambiguity aversion coefficient increases. This is a desirable feature of the ambiguity model in producing, e.g., a sharp increase in the equity premium at the onset of recessions as in the recent financial crisis.

Figures 4.8–4.10 show the effects of the risk aversion coefficient on asset prices and returns. We consider three values of the risk aversion coefficient, i.e., $\gamma = \{2, 3, 5\}$, where $\gamma = 2$ is the benchmark value. As shown in these figures, an important observation is that changes in the risk aversion coefficient change the level of the curve in both regimes—an EZ investor requires compensation for the uncertainty over the recession parameter even in the expansion regime. Specifically, as the risk aversion coefficient increases from 2 to 5, the changes in the level of the curve in the expansion regime and in the recession regime are comparable. As a result, the asymmetric effect observed in ambiguity models is small or even negligible in recursive utility models. On the other side, similar to the uncertainty effect of learning under ambiguity aversion, as the risk aversion coefficient increases, the curve in the expansion regime changes its curvature, and the curve in the recession regime peaks at higher levels and at smaller probabilities of the long-recession parameter value. In this sense, it remains a question whether the conditional moments of asset prices and returns implied by ambiguity models are able to be re-produced by recursive utility models with higher risk aversion coefficients, and we address this question next.

Figures 4.11–4.13 compare the effects of the ambiguity aversion coefficient and the risk aversion coefficient on asset prices and returns. We consider an ambiguity model where the ambiguity aversion coefficient and the risk aversion coefficient are set at their benchmark values ($\gamma = 2, \eta = 15$) and recursive utility models where the risk aversion coefficients are equal to or higher than 2 ($\gamma = \{2, 3, 5\}$). As shown in these figures, the conditional moments implied by the ambiguity model cannot be re-produced by
increasing the risk aversion coefficient in the recursive utility model. A notable ob-
servation is that the ambiguity model is able to raise the compensation for parameter
uncertainty substantially in the recession regime while keeping the compensation in
the expansion regime at the low level as implied by the recursive utility model with
$\gamma = 2$.\footnote{This effect is less significant for the volatility of the market return.} This feature is desirable as it offers great flexibility in producing a sharp
increase in the equity premium at the onset of recessions—the goal of this paper.

4.5 Simulation Results

In Section 4.4, we examine the conditional moments of asset prices and returns and
document an uncertainty effect of learning under ambiguity aversion, which is mod-
erate in the expansion regime but is substantial in the recession regime. Relating to
these findings, there are mainly two concerns. Firstly, if the investor learns the true
persistence very quickly, the effects of parameter learning would also decay quickly.
Secondly, in spite of some evidence of the difference between learning under ambiguity
aversion and learning under EZ preferences, the asset pricing implications of learning
under ambiguity aversion are still not obvious. In this section, we show via simulation
that learning about the persistence of recessions is slow and difficult, and learning
under ambiguity aversion has important asset pricing implications at the onset of
recessions.

4.5.1 Simulated probability

In simulation, we use the short-recession parameter value as the true persistence of
the recession regime and generate 5000 samples of regimes. In each simulated sample,
the representative investor learns about the persistence of recessions and updates the
probability of the long-recession parameter value via Bayes’ rule as she observes more
recessions. In other words, the recession regime on average persists 2 years and occurs
twice per century, while the investor learns about the possibility of a 5-year recession.
We consider three prior probabilities of the long-recession parameter value, i.e., $p_0 = \ldots$
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVersion

0.9, \( p_0 = 0.5 \) and \( p_0 = 0.1 \). If the investor indeed learns the true persistence very quickly, the probability would decrease to zero quickly. We examine the distribution of the updated probabilities over the 5000 simulated samples at the end of 100, 200 and 500 years of learning, and the results are reported in Table 4.2. As shown, since the investor will eventually learn the true persistence as she observes more recessions, the mean and the median of the distribution decrease from the prior over time. However, the less frequent nature of recessions makes learning difficult. The standard deviation of the distribution does not decrease significantly over time. The distribution has a long tail on the right and fat tails, and the skewness and the excess kurtosis increase over time. Moreover, as shown in section 4.4.1, even a small probability such as 0.05 (the mean after 200 years of learning from a prior of \( p_0 = 0.1 \)) is not negligible—an ambiguity-averse investor with \( \gamma = 2 \) and \( \eta = 15 \) views this probability as doubled in the recession regime.

4.5.2 Before and at the onset of recessions

We aim to reproduce in simulation the sharp increase in the equity premium and the conditional volatility of returns at the onset of recessions. The empirical study of Martin (2016) estimates that in November 2008—the height of the recent financial crisis, the (monthly) annualised expected return on the US stock market peaked at 55.0\%, and the annualised SVIX index for market volatility exceeded 70\%. These are in sharp contrast to the Great Moderation period from 2004 to 2006 when the average expected return was 1.86\% and the SVIX index was about 14\%. At the one-year horizon, the annualised expected market return was 21.5\% and the annualised SVIX index exceeded 45\% at the height of the crisis, while the average expected return was 2.07–2.81\% and the annualised SVIX index was about 15\% in the Great Moderation period.\(^9\) At the one-month horizon, the equity premium and the SVIX index at the

\(^9\) Martin (2016) does not report the SVIX index for these two periods, and we imply the index from how he calculates equity premium from SVIX index, i.e., \( E_t(R) - R_{f,t} = R_{f,t} \cdot SVIX_t^2 \), where \( E_t(R) - R_{f,t} \) is the equity premium and \( R_{f,t} \) is the risk free rate. Our calculation of the index is consistent with his plot of VXO index—an index similar to the SVIX index. At the one-year horizon, he does not report the average expected return for the Great Moderation period, and we got a range of 2.07–2.81\% from his Table I, which are the 10\% and the 25\% quantiles of the equity premium at the one-year horizon, as the reported return of 1.86\% falls into this range of quantiles at the one-month horizon in the table.
height of the crisis are 30 times and 5 times as much as their counterparts during the Great Moderation. At the one-year horizon, they are about 10 times and 3 times as much as the averages in the Great Moderation period. Since our model parameters are for annual data, we match the risk premium and the volatility at the one-year horizon.

Our simulation is designed to capture these two periods in history. As in Section 4.5.1, we use the short-recession parameter value as the true persistence of the recession regime and the representative investor learns about the persistence of recessions via Bayes’ rule as she observes more recessions. We assume that the investor has no stronger prior belief about either parameter value, i.e., she updates the probability of the long-recession parameter value from a prior of $p_0 = 0.5$. After 100 years of learning—comparable length with the 1890–1994 sample used in Cecchetti, Lam and Mark (2000)—we calculate the conditional moments of asset prices and returns at the onset of the next recession and in the preceding period. The onset of the next recession represents the height of the recent financial crisis and the preceding period represents the Great Moderation.

Panel I of Table 4.3 reports the results. Model 1 is the full information model where the persistence of recessions is known at the true value. In model 2, the true persistence is unknown and the investor learns about the parameter under EZ preferences, where the risk aversion coefficient is set at the benchmark value $\gamma = 2$. Model 3 is our benchmark model where the investor learns about the parameter under ambiguity aversion with $\gamma = 2$ and $\eta = 15$. Models 4–6 consider other values of $\eta$, and models 7 and 8 consider other values of $\gamma$. In the benchmark model, after 100 years of learning, at the onset of the next recession, the expected excess market return increases from 1.37% to 24.98%, and the conditional volatility (standard deviation) of the market return increases from 12.11% to 57.65%. These results are comparable to Martin’s (2016) finding that the expected excess market return is estimated to increase from 2.07–2.81% to 21.5% from the Great Moderation period to the height of the recent financial crisis, and the SVIX index for volatility increases from 15% to 45%. Comparing between models 1–3 shows the effects of parameter learning on the increase in the equity premium and the volatility. In the full information model, at the onset of the next recession, the equity premium increases from 0.86% to 4.18%, and the conditional
volatility increases from 10.54% to 24.87%. Introducing parameter learning under EZ preferences in model 2 dramatically increases the equity premium (to 13.41%) and the volatility (to 46.67%) at the onset of the next recession. An increase from 24.87% to 46.67% in the volatility is consistent with the parameter-learning literature where learning about parameters generates excess volatility (see, for example, Timmermann, 1993, 1996). Introducing ambiguity aversion in model 3 further increases the equity premium at the onset of the next recession to 24.98%, which nearly doubles the equity premium of 13.41% in the recursive utility model.

Comparing between models 4 and 5 shows that an increase in the ambiguity aversion coefficient from $\eta = 15$ to $\eta = 30$ increases the equity premium at the onset of the next recession from 24.98% to 33.78%, and more importantly, the ratio of the equity premium at the onset of the next recession to that in the preceding period increases from 18 to 22. By contrast, even though increasing the risk aversion coefficient from model 2 to models 7 and 8 also increase dramatically the risk premium at the onset of the next recession, the ratio almost remains the same or even decreases, and increasing the risk aversion coefficient beyond $\gamma = 5$ does not increase the ratio. The ability of ambiguity models to generate high equity premium ratios is desirable. For instance, at the one-month horizon, Martin (2016) estimates that the equity premium at the height of the crisis is actually 30 times as much as the average equity premium during the Great Moderation, and based on our analysis so far, ambiguity models have the potential to produce a high equity premium ratio such as 30.

Panels II and III of Table 4.3 report the results after 200 and 500 years of learning. As the investor gradually learns the true parameter value, i.e., moving from Panel I to II and III, the equity premium and the volatility generally decreases. The exceptions are the models with very high ambiguity aversion coefficient or very high risk aversion coefficient (models 6 and 8), whose equity premium in the recession regime actually increase slightly after 200 years of learning. This is consistent with the comparative statics in Section 4.4.3—as the investor is more ambiguity averse or more risk averse, the uncertainty effect of parameter learning in the recession regime peaks at smaller probabilities of the long recession value. Tables 4.4 and 4.5 report the results for a prior probability of $p_0 = 0.9$ and $p_0 = 0.1$ respectively. Relative to Table 4.3, after
100 years of learning, the equity premium in the benchmark model is smaller for both prior probabilities. This is due to the hard time effect as shown in Section 4.4.2—an ambiguity-averse investor with $\gamma = 2$ and $\eta = 15$ requires the highest compensation for parameter uncertainty in the area of $p = (0.2, 0.4)$. As shown in Table 4.2, after 100 years of learning, the mean of the simulated probability is 0.7447, 0.3274 and 0.0704 for a prior probability of 0.9, 0.5 and 0.1 respectively, where 0.3274 is within the range of $p = (0.2, 0.4)$. More importantly, throughout these tables, the ratio of the equity premium at the onset of the next recession to that in the preceding period ranges from 10 to 30 for ambiguity models, and ranges from 3 to 12 for recursive utility models. The ratio for the conditional volatility of the market return ranges from 3 to 5 in both ambiguity and recursive utility models, which is consistent with the observation that the SVIX index at the height of the recent financial crisis is about 5 times as much as the index during the Great Moderation.

Tables 4.6–4.8 report the simulated long-sample moments unconditional on regime, with the prior probability of the long-recession parameter value equal to $p_0 = 0.5$, $p_0 = 0.9$ and $p_0 = 0.1$ respectively. As shown in Table 4.6, at the end of 100 years of learning, the mean of the risk free rate and the excess market return in our benchmark model is 3.08% and 3.26% respectively, with a standard deviation of 1.13% and 20.01%. Using data from 1871 to 1993, Cecchetti, Lam and Mark (2000) report the sample mean of the risk free rate and the excess market return as 2.66% and 5.75% respectively, with a sample standard deviation of 5.13% and 19.02%. Our benchmark model roughly matches the level of the risk free rate and the volatility of the excess market return, while under-estimates the level of the excess market return and the volatility of the risk free rate. The lower level of the excess market return in the benchmark model is due to our calibration of the model where after 100 years of learning, the expected excess market return in the expansion regime is calibrated to be close to 2.07–2.81% as estimated from the Great Moderation period, which drives down the level of the excess market return. As for the low simulated volatility of the risk free rate, Campbell (1999) points out that the high historical volatility of the risk free rate is probably driven by the unanticipated inflation in the interwar period. Using postwar data, he reports a volatility of 1.8%, which our benchmark model matches better.
It can also be seen that the contribution of ambiguity models to the moments unconditional on regime is moderate, which is not surprising as it is shown in Section 4.4.2 that the size of the moments unconditional on regime is mostly driven by the moments conditional on the expansion regime due to the less frequent occurrence of the recession regime, and the uncertainty effect of learning under ambiguity aversion is only moderate in the expansion regime.

4.6 Concluding Remarks

Learning about the persistence of recessions is difficult due to their less frequent nature. However, the persistence of recessions has important implications for asset prices and returns. We are the first in the literature to incorporate ambiguity aversion into the process of learning about the persistence of recessions. There are two ingredients in the model: the source of ambiguity and the attitude towards ambiguity. In our benchmark model, ambiguity is characterised by the unknown parameter that governs the persistence of recessions, and the representative investor learns about this parameter while being ambiguity averse towards parameter uncertainty.

We assess the asset pricing implications of the model from two related perspectives: the conditional moments and the simulated long-sample moments of asset prices and returns. In both analyses, we show that the asset pricing implications of the ambiguity model cannot be re-produced by increasing the risk aversion coefficient in the recursive utility model. Specifically, we document an uncertainty effect that characterises the difference between the two models. More importantly, this uncertainty effect, driven by the distorted probability in the ambiguity model, is asymmetric across regimes—the size of the uncertainty effect is moderate in the expansion regime and is substantial in the recession regime. This asymmetry plays a key role in matching the sharp increase in the equity premium at the onset of recessions. Empirically, Martin (2016) estimates that the equity premium at the height of the recent financial crisis is 30 times as much as the equity premium during the 2004–2006 Great Moderation at the one-month horizon and is 10 times as much at the one-year horizon. At the annual frequency, our simulation results show that at the onset of the next recession, ambiguity
models are able to generate 10–30 times the equity premium in the preceding period, and increasing the magnitude of ambiguity aversion will further increase the ratio. By contrast, recursive utility models generate 3–12 times the equity premium in the preceding period, and increasing the magnitude of risk aversion does not help increase the ratio. In this sense, ambiguity models offer desirable flexibility in matching the sharp increase in the equity premium at the onset of recessions.

Having established the existence of an important asymmetric uncertainty effect due to learning under conditions of ambiguity aversion via simulations, future work should consider actual sample data and attempt both the historical sample moments and the time variation of asset prices and returns present in the data. Another consideration is that since the moments of asset prices and returns due to ambiguity aversion depends on the belief about the persistence of recessions in a flexible way, it would be interesting to consider an economy where ambiguity-averse investors have heterogeneous beliefs about parameters.
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERSION

Bibliography


CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERSION 170


This table reports the parameters of consumption growth in equation (4.1), the parameters of regime in equation (4.2), the parameters of dividend growth in equation (4.3), and the preference parameters. We adopt the parameters of regime and consumption growth from Cecchetti, Lam and Mark (2000), and their estimates are reported in Panel A. The two values for the transition probability of the recession regime, $\pi_{11}$, is reported in Panel B. Panels C and D report the dividend growth and the preference parameters respectively. All parameters are for annual data, and the mean and the standard deviation of consumption and dividend growths are in percentage.

### A. consumption growth parameters

(estimates in Cecchetti, Lam and Mark, 2000)

<table>
<thead>
<tr>
<th></th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\sigma$</th>
<th>$\pi_{11}$</th>
<th>$\pi_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimate</td>
<td>-6.785</td>
<td>2.251</td>
<td>3.127</td>
<td>0.516</td>
<td>0.978</td>
</tr>
<tr>
<td>(t-value)</td>
<td>(-3.60)</td>
<td>(6.87)</td>
<td>(13.00)</td>
<td>(1.95)</td>
<td>(50.94)</td>
</tr>
</tbody>
</table>

### B. choice of two values for the transition probability parameter

<table>
<thead>
<tr>
<th>$\pi_{11}^l$</th>
<th>$\pi_{11}^h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.516</td>
<td>0.8</td>
</tr>
</tbody>
</table>

### C. dividend growth parameters

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\mu_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.74</td>
<td>-3.23</td>
</tr>
</tbody>
</table>

### D. preference parameters

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\rho$ (1/EIS)</th>
<th>$\gamma$</th>
<th>$\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9750</td>
<td>1/2</td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>
Table 4.2: Moments of simulated probability

This table reports the moments of the simulated probability of the long-recession parameter value. In simulation, we use the short-recession parameter value as the true persistence of the recession regime, and generate 5000 samples. In each simulated sample, investors update the probability of the long-recession parameter value via Bayes’ rule from a prior of $A. p_0 = 0.9$, $B. p_0 = 0.5$ and $C. p_0 = 0.1$. At the end of 100, 200 and 500 years of learning, we calculate the mean, median, minimum, maximum, standard deviation, skewness and excess kurtosis of the 5000 updated probabilities.

<table>
<thead>
<tr>
<th></th>
<th>A. $p_0 = 0.9$</th>
<th></th>
<th>B. $p_0 = 0.5$</th>
<th></th>
<th>C. $p_0 = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100y</td>
<td>200y</td>
<td>500y</td>
<td>100y</td>
<td>200y</td>
</tr>
<tr>
<td>mean</td>
<td>0.7447</td>
<td>0.5775</td>
<td>0.2385</td>
<td>0.3274</td>
<td>0.2209</td>
</tr>
<tr>
<td>median</td>
<td>0.7779</td>
<td>0.5898</td>
<td>0.0910</td>
<td>0.2802</td>
<td>0.1377</td>
</tr>
<tr>
<td>min</td>
<td>0.0258</td>
<td>0.0029</td>
<td>0.0000</td>
<td>0.0029</td>
<td>0.0003</td>
</tr>
<tr>
<td>max</td>
<td>0.9994</td>
<td>0.9999</td>
<td>0.9996</td>
<td>0.9945</td>
<td>0.9990</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.1828</td>
<td>0.2675</td>
<td>0.2851</td>
<td>0.1976</td>
<td>0.2123</td>
</tr>
<tr>
<td>skewness</td>
<td>-1.1866</td>
<td>-0.3339</td>
<td>1.2316</td>
<td>0.7549</td>
<td>1.3989</td>
</tr>
<tr>
<td>kurtosis</td>
<td>1.1182</td>
<td>-1.0038</td>
<td>0.2839</td>
<td>0.3564</td>
<td>1.5070</td>
</tr>
</tbody>
</table>
This table reports the simulated conditional moments of asset prices and returns before and at the onset of recessions. In simulation, the investor updates the probability of the long-recession parameter value from a prior of \( p_0 = 0.5 \). After 100, 200 and 500 years of learning, we calculate the conditional mean and standard deviation of the excess market return and the conditional Sharpe ratio and price of risk, in the preceeding period of the next recession (A), at the onset of the next recession (B), and the ratio of the equity premium and the volatility of excess market return at the onset of the next recession to those in the preceding period (C).

In model 1, the persistence of recessions is known at the true value. In model 2, the investor learns about the persistence under EZ preferences, where \( \gamma = 2 \) is the benchmark value for the ambiguity aversion coefficient. Models 4–6 consider other values of \( \gamma \) and models 7 and 8 consider other values of \( \eta \). Mean and standard deviation of returns are in percentage.

<table>
<thead>
<tr>
<th>( E_t[R_{m}^c] )</th>
<th>( \sigma_t[R_{m}^c] )</th>
<th>( SR_t )</th>
<th>( PR_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma ) &amp; ( \eta ) &amp; Mean before &amp; Mean onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma = 2 ) &amp; ( \eta = \gamma ) &amp; 0.86 &amp; 10.54 &amp; 0.08 &amp; 0.08 &amp; 4.18 &amp; 24.87 &amp; 0.17 &amp; 0.17 &amp; 5 &amp; 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 &amp; ( \gamma = 2 ) &amp; ( \eta = 15 ) &amp; 1.23 &amp; 11.64 &amp; 0.11 &amp; 0.11 &amp; 33.78 &amp; 63.05 &amp; 0.53 &amp; 1.47 &amp; 22 &amp; 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 &amp; ( \gamma = 2 ) &amp; ( \eta = 5 ) &amp; 2.66 &amp; 12.43 &amp; 0.21 &amp; 0.23 &amp; 32.44 &amp; 65.48 &amp; 0.49 &amp; 0.50 &amp; 12 &amp; 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 &amp; ( \gamma = 3 ) &amp; ( \eta = \gamma ) &amp; 8.86 &amp; 13.29 &amp; 0.67 &amp; 0.83 &amp; 50.29 &amp; 69.69 &amp; 0.72 &amp; 0.74 &amp; 6 &amp; 5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I. after 100y learning

II. after 200y learning

III. after 500y learning

<table>
<thead>
<tr>
<th>( E_t[R_{m}^c] )</th>
<th>( \sigma_t[R_{m}^c] )</th>
<th>( SR_t )</th>
<th>( PR_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma ) &amp; ( \eta ) &amp; Mean before &amp; Mean onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma = 2 ) &amp; ( \eta = \gamma ) &amp; 0.94 &amp; 10.78 &amp; 0.09 &amp; 0.09 &amp; 6.46 &amp; 30.48 &amp; 0.20 &amp; 0.20 &amp; 7 &amp; 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 &amp; ( \gamma = 2 ) &amp; ( \eta = 15 ) &amp; 0.99 &amp; 10.96 &amp; 0.09 &amp; 0.09 &amp; 9.72 &amp; 34.57 &amp; 0.25 &amp; 0.38 &amp; 10 &amp; 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 &amp; ( \gamma = 2 ) &amp; ( \eta = 5 ) &amp; 1.06 &amp; 11.24 &amp; 0.09 &amp; 0.10 &amp; 15.44 &amp; 40.76 &amp; 0.32 &amp; 0.94 &amp; 15 &amp; 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 &amp; ( \gamma = 3 ) &amp; ( \eta = \gamma ) &amp; 1.75 &amp; 11.25 &amp; 0.15 &amp; 0.16 &amp; 15.29 &amp; 39.50 &amp; 0.36 &amp; 0.36 &amp; 9 &amp; 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 &amp; ( \gamma = 5 ) &amp; ( \eta = \gamma ) &amp; 5.88 &amp; 12.68 &amp; 0.46 &amp; 0.54 &amp; 45.81 &amp; 62.78 &amp; 0.72 &amp; 0.74 &amp; 8 &amp; 5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERSION 174

Table 4.4: Before and at the onset of recessions: \( p_0 = 0.9 \)

This table reports the simulated conditional moments of asset prices and returns before and at the onset of recessions. In simulation, the investor updates the probability of the long-recession parameter value from a prior of \( p_0 = 0.9 \). After 100, 200 and 500 years of learning, we calculate the conditional mean and standard deviation of the excess market return and the conditional Sharpe ratio and price of risk, in the preceding period of the next recession (A), at the onset of the next recession (B), and the ratio of the equity premium and the volatility of excess market return at the onset of the next recession to those in the preceding period (C). In model 1, the persistence of recessions is known at the true value. In model 2, the true persistence is unknown and the investor learns about the parameter under EZ preferences, where \( \gamma = 2 \) is the benchmark value for the risk aversion coefficient. In model 3, the investor learns about the parameter under ambiguity aversion, where \( \eta = 15 \) is the benchmark value for the ambiguity aversion coefficient. Models 4–6 consider other values of \( \eta \) and models 7 and 8 consider other values of \( \gamma \). Mean and standard deviation of returns are in percentage.

<table>
<thead>
<tr>
<th></th>
<th>A. before recessions</th>
<th>B. at the onset of recessions</th>
<th>C. onset before</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( E_t[R_m^r] )</td>
<td>( \sigma_t[R_m^r] )</td>
<td>( SR_t )</td>
</tr>
<tr>
<td>I. after 100y learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, ( \gamma = 2 )</td>
<td>0.86</td>
<td>10.54</td>
<td>0.08</td>
</tr>
<tr>
<td>2 ( \gamma = 2 ) ( \eta = \gamma )</td>
<td>1.60</td>
<td>12.45</td>
<td>0.13</td>
</tr>
<tr>
<td>3 ( \gamma = 2 ) ( \eta = 15 )</td>
<td>1.71</td>
<td>12.76</td>
<td>0.13</td>
</tr>
<tr>
<td>4 ( \gamma = 2 ) ( \eta = 5 )</td>
<td>1.64</td>
<td>12.55</td>
<td>0.13</td>
</tr>
<tr>
<td>5 ( \gamma = 2 ) ( \eta = 30 )</td>
<td>1.77</td>
<td>12.87</td>
<td>0.14</td>
</tr>
<tr>
<td>6 ( \gamma = 2 ) ( \eta = 60 )</td>
<td>1.81</td>
<td>12.91</td>
<td>0.14</td>
</tr>
<tr>
<td>7 ( \gamma = 3 ) ( \eta = \gamma )</td>
<td>3.59</td>
<td>13.19</td>
<td>0.27</td>
</tr>
<tr>
<td>8 ( \gamma = 5 ) ( \eta = \gamma )</td>
<td>10.35</td>
<td>13.43</td>
<td>0.77</td>
</tr>
<tr>
<td>II. after 200y learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, ( \gamma = 2 )</td>
<td>0.86</td>
<td>10.54</td>
<td>0.08</td>
</tr>
<tr>
<td>2 ( \gamma = 2 ) ( \eta = \gamma )</td>
<td>1.44</td>
<td>12.08</td>
<td>0.12</td>
</tr>
<tr>
<td>3 ( \gamma = 2 ) ( \eta = 15 )</td>
<td>1.58</td>
<td>12.51</td>
<td>0.13</td>
</tr>
<tr>
<td>4 ( \gamma = 2 ) ( \eta = 5 )</td>
<td>1.48</td>
<td>12.19</td>
<td>0.12</td>
</tr>
<tr>
<td>5 ( \gamma = 2 ) ( \eta = 30 )</td>
<td>1.68</td>
<td>12.76</td>
<td>0.13</td>
</tr>
<tr>
<td>6 ( \gamma = 2 ) ( \eta = 60 )</td>
<td>1.76</td>
<td>12.88</td>
<td>0.14</td>
</tr>
<tr>
<td>7 ( \gamma = 3 ) ( \eta = \gamma )</td>
<td>3.22</td>
<td>12.89</td>
<td>0.25</td>
</tr>
<tr>
<td>8 ( \gamma = 5 ) ( \eta = \gamma )</td>
<td>9.76</td>
<td>13.37</td>
<td>0.73</td>
</tr>
<tr>
<td>III. after 500y learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, ( \gamma = 2 )</td>
<td>0.86</td>
<td>10.54</td>
<td>0.08</td>
</tr>
<tr>
<td>2 ( \gamma = 2 ) ( \eta = \gamma )</td>
<td>1.10</td>
<td>11.21</td>
<td>0.10</td>
</tr>
<tr>
<td>3 ( \gamma = 2 ) ( \eta = 15 )</td>
<td>1.20</td>
<td>11.55</td>
<td>0.10</td>
</tr>
<tr>
<td>4 ( \gamma = 2 ) ( \eta = 5 )</td>
<td>1.12</td>
<td>11.28</td>
<td>0.10</td>
</tr>
<tr>
<td>5 ( \gamma = 2 ) ( \eta = 30 )</td>
<td>1.31</td>
<td>11.94</td>
<td>0.11</td>
</tr>
<tr>
<td>6 ( \gamma = 2 ) ( \eta = 60 )</td>
<td>1.49</td>
<td>12.49</td>
<td>0.12</td>
</tr>
<tr>
<td>7 ( \gamma = 3 ) ( \eta = \gamma )</td>
<td>2.27</td>
<td>11.89</td>
<td>0.19</td>
</tr>
<tr>
<td>8 ( \gamma = 5 ) ( \eta = \gamma )</td>
<td>7.58</td>
<td>13.06</td>
<td>0.58</td>
</tr>
</tbody>
</table>
This table reports the simulated conditional moments of asset prices and returns before and at the onset of recessions. In simulation, the investor updates the probability of the long-recession parameter value from a prior of $p_0 = 0.1$. After 100, 200 and 500 years of learning, we calculate the conditional mean and standard deviation of the excess market return and the conditional Sharpe ratio and price of risk, in the preceding period of the next recession (A), at the onset of the next recession (B), and the ratio of the equity premium and the volatility of excess market return at the onset of the next recession to those in the preceding period (C).

In model 1, the persistence of recessions is known at the true value. In model 2, the investor learns about the parameter under ambiguity aversion, where the true persistence is unknown and the investor learns about the parameter under EZ preferences, where $\gamma = 2$ is the benchmark value for the ambiguity aversion coefficient. Models 4–6 consider other values of $\gamma$. In model 3, the investor learns about the parameter under EZ preferences, where $\eta = 15$ is the benchmark value for the risk aversion coefficient.

### Table 4.5: Before and at the onset of recessions: $p_0 = 0.1$

<table>
<thead>
<tr>
<th>A. before recessions</th>
<th>B. at the onset of recessions</th>
<th>C. onset before</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_t[R_m]$</td>
<td>$\sigma_t[R_m]$</td>
</tr>
<tr>
<td>I. after 100y learning</td>
<td>1 full info, $\gamma = 2$</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>2 $\gamma = 2$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td></td>
<td>3 $\gamma = 2$</td>
<td>$\eta = 15$</td>
</tr>
<tr>
<td></td>
<td>4 $\gamma = 2$</td>
<td>$\eta = 5$</td>
</tr>
<tr>
<td></td>
<td>5 $\gamma = 2$</td>
<td>$\eta = 30$</td>
</tr>
<tr>
<td></td>
<td>6 $\gamma = 2$</td>
<td>$\eta = 60$</td>
</tr>
<tr>
<td></td>
<td>7 $\gamma = 3$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td></td>
<td>8 $\gamma = 5$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td>II. after 200y learning</td>
<td>1 full info, $\gamma = 2$</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>2 $\gamma = 2$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td></td>
<td>3 $\gamma = 2$</td>
<td>$\eta = 15$</td>
</tr>
<tr>
<td></td>
<td>4 $\gamma = 2$</td>
<td>$\eta = 5$</td>
</tr>
<tr>
<td></td>
<td>5 $\gamma = 2$</td>
<td>$\eta = 30$</td>
</tr>
<tr>
<td></td>
<td>6 $\gamma = 2$</td>
<td>$\eta = 60$</td>
</tr>
<tr>
<td></td>
<td>7 $\gamma = 3$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td></td>
<td>8 $\gamma = 5$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td>III. after 500y learning</td>
<td>1 full info, $\gamma = 2$</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>2 $\gamma = 2$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td></td>
<td>3 $\gamma = 2$</td>
<td>$\eta = 15$</td>
</tr>
<tr>
<td></td>
<td>4 $\gamma = 2$</td>
<td>$\eta = 5$</td>
</tr>
<tr>
<td></td>
<td>5 $\gamma = 2$</td>
<td>$\eta = 30$</td>
</tr>
<tr>
<td></td>
<td>6 $\gamma = 2$</td>
<td>$\eta = 60$</td>
</tr>
<tr>
<td></td>
<td>7 $\gamma = 3$</td>
<td>$\eta = \gamma$</td>
</tr>
<tr>
<td></td>
<td>8 $\gamma = 5$</td>
<td>$\eta = \gamma$</td>
</tr>
</tbody>
</table>
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERSION

Table 4.6: Long-sample moments: \( p_0 = 0.5 \)

This table reports the simulated long-sample moments of asset prices and returns. In simulation, the investor updates the probability of the long-recession parameter value from a prior of \( p_0 = 0.5 \). At the end of 100, 200, and 500 years of learning, we calculate the mean and the standard deviation of risk free rate, market return and excess market return, the Sharpe ratio, the price of risk and the P/D ratio. In model 1, the persistence of recessions is known at the true value. In model 2, the true persistence is unknown and the investor learns about the parameter under EZ preferences, where \( \gamma = 2 \) is the benchmark value for the risk aversion coefficient. In model 3, the investor learns about the parameter under ambiguity aversion, where \( \eta = 15 \) is the benchmark value for the ambiguity aversion coefficient. Models 4–6 consider other values of \( \eta \) and models 7 and 8 consider other values of \( \gamma \). Mean and standard deviation of returns are in percentage.

\[
\begin{array}{cccccccc}
E[R_f] & \sigma[R_f] & E[R_m] & \sigma[R_m] & E[R_m'] & \sigma[R_m'] & SR & P/D \\
\hline
\text{I. at the end of 100y learning} \\
1 & \text{full info, } \gamma = 2 & 3.30 & 0.58 & 4.35 & 11.57 & 1.05 & 11.60 & 0.09 & 58.19 \\
2 & \gamma = 2 \ \eta = \gamma & 3.15 & 0.94 & 5.70 & 17.32 & 2.55 & 17.66 & 0.11 & 45.69 \\
3 & \gamma = 2 \ \eta = 15 & 3.08 & 1.13 & 6.32 & 19.54 & 3.26 & 20.01 & 0.13 & 40.50 \\
4 & \gamma = 2 \ \eta = 5 & 3.14 & 1.00 & 5.85 & 17.90 & 2.72 & 18.27 & 0.11 & 44.44 \\
5 & \gamma = 2 \ \eta = 30 & 3.03 & 1.20 & 6.73 & 20.64 & 3.71 & 21.18 & 0.14 & 36.66 \\
6 & \gamma = 2 \ \eta = 60 & 3.01 & 1.13 & 6.88 & 20.71 & 3.87 & 21.22 & 0.15 & 34.38 \\
7 & \gamma = 3 \ \eta = \gamma & 2.74 & 1.24 & 7.69 & 21.99 & 4.97 & 22.56 & 0.22 & 29.08 \\
8 & \gamma = 5 \ \eta = \gamma & 1.01 & 1.23 & 12.42 & 22.71 & 11.41 & 23.29 & 0.67 & 12.51 \\
\hline
\text{II. at the end of 200y learning} \\
1 & \text{full info, } \gamma = 2 & 3.30 & 0.59 & 4.28 & 11.59 & 0.97 & 11.61 & 0.09 & 58.17 \\
2 & \gamma = 2 \ \eta = \gamma & 3.19 & 0.87 & 5.07 & 15.29 & 1.88 & 15.50 & 0.10 & 49.38 \\
3 & \gamma = 2 \ \eta = 15 & 3.13 & 1.07 & 5.57 & 17.30 & 2.44 & 17.65 & 0.12 & 44.92 \\
4 & \gamma = 2 \ \eta = 5 & 3.18 & 0.91 & 5.18 & 15.76 & 2.00 & 16.00 & 0.11 & 48.41 \\
5 & \gamma = 2 \ \eta = 30 & 3.06 & 1.24 & 6.11 & 19.21 & 3.05 & 19.68 & 0.13 & 40.25 \\
6 & \gamma = 2 \ \eta = 60 & 3.01 & 1.24 & 6.51 & 20.17 & 3.49 & 20.69 & 0.14 & 35.77 \\
7 & \gamma = 3 \ \eta = \gamma & 2.83 & 1.20 & 6.70 & 19.25 & 3.87 & 19.69 & 0.20 & 32.89 \\
8 & \gamma = 5 \ \eta = \gamma & 1.26 & 1.40 & 11.53 & 22.03 & 10.26 & 22.64 & 0.61 & 13.64 \\
\hline
\text{III. at the end of 500y learning} \\
1 & \text{full info, } \gamma = 2 & 3.28 & 0.63 & 4.02 & 12.08 & 0.71 & 12.13 & 0.09 & 58.10 \\
2 & \gamma = 2 \ \eta = \gamma & 3.24 & 0.76 & 4.37 & 13.77 & 1.10 & 13.92 & 0.09 & 54.83 \\
3 & \gamma = 2 \ \eta = 15 & 3.21 & 0.87 & 4.62 & 15.10 & 1.39 & 15.34 & 0.10 & 52.95 \\
4 & \gamma = 2 \ \eta = 5 & 3.23 & 0.78 & 4.42 & 14.03 & 1.16 & 14.19 & 0.09 & 54.45 \\
5 & \gamma = 2 \ \eta = 30 & 3.16 & 1.02 & 4.94 & 16.78 & 1.75 & 17.14 & 0.11 & 50.10 \\
6 & \gamma = 2 \ \eta = 60 & 3.08 & 1.27 & 5.42 & 19.01 & 2.29 & 19.48 & 0.12 & 44.06 \\
7 & \gamma = 3 \ \eta = \gamma & 2.99 & 1.01 & 5.40 & 16.49 & 2.38 & 16.82 & 0.16 & 40.40 \\
8 & \gamma = 5 \ \eta = \gamma & 1.86 & 1.52 & 9.37 & 21.54 & 7.46 & 22.18 & 0.47 & 17.78 \\
\end{array}
\]
CHAPTER 4. PARAMETER LEARNING UNDER AMBIGUITY AVERSION

Table 4.7: Long-sample moments: $p_0 = 0.9$

This table reports the simulated long-sample moments of asset prices and returns. In simulation, the investor updates the probability of the long-recession parameter value from a prior of $p_0 = 0.9$. At the end of 100, 200, and 500 years of learning, we calculate the mean and the standard deviation of risk free rate, market return and excess market return, the Sharpe ratio, the price of risk and the P/D ratio. In model 1, the persistence of recessions is known at the true value. In model 2, the true persistence is unknown and the investor learns about the parameter under EZ preferences, where $\gamma = 2$ is the benchmark value for the risk aversion coefficient. In model 3, the investor learns about the parameter under ambiguity aversion, where $\eta = 15$ is the benchmark value for the ambiguity aversion coefficient. Models 4-6 consider other values of $\eta$ and models 7 and 8 consider other values of $\gamma$. Mean and standard deviation of returns are in percentage.

<table>
<thead>
<tr>
<th></th>
<th>$E[R_f]$</th>
<th>$\sigma[R_f]$</th>
<th>$E[R_m]$</th>
<th>$\sigma[R_m]$</th>
<th>$E[R^e_m]$</th>
<th>$\sigma[R^e_m]$</th>
<th>$SR$</th>
<th>$P/D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. at the end of 100y learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, $\gamma = 2$</td>
<td>3.30</td>
<td>0.58</td>
<td>4.35</td>
<td>11.57</td>
<td>1.05</td>
<td>11.60</td>
<td>0.09</td>
<td>58.19</td>
</tr>
<tr>
<td>2 $\gamma = 2$, $\eta = \gamma$</td>
<td>3.05</td>
<td>0.97</td>
<td>6.65</td>
<td>20.07</td>
<td>3.61</td>
<td>20.50</td>
<td>0.13</td>
<td>36.66</td>
</tr>
<tr>
<td>3 $\gamma = 2$, $\eta = 15$</td>
<td>3.02</td>
<td>1.02</td>
<td>6.86</td>
<td>20.57</td>
<td>3.85</td>
<td>21.04</td>
<td>0.14</td>
<td>34.55</td>
</tr>
<tr>
<td>4 $\gamma = 2$, $\eta = 5$</td>
<td>3.04</td>
<td>0.99</td>
<td>6.72</td>
<td>20.26</td>
<td>3.69</td>
<td>20.71</td>
<td>0.14</td>
<td>36.00</td>
</tr>
<tr>
<td>5 $\gamma = 2$, $\eta = 30$</td>
<td>3.00</td>
<td>1.01</td>
<td>6.92</td>
<td>20.63</td>
<td>3.92</td>
<td>21.09</td>
<td>0.15</td>
<td>33.69</td>
</tr>
<tr>
<td>6 $\gamma = 2$, $\eta = 60$</td>
<td>3.00</td>
<td>0.97</td>
<td>6.94</td>
<td>20.58</td>
<td>3.95</td>
<td>21.02</td>
<td>0.15</td>
<td>33.30</td>
</tr>
<tr>
<td>7 $\gamma = 3$, $\eta = \gamma$</td>
<td>2.51</td>
<td>1.09</td>
<td>8.80</td>
<td>23.69</td>
<td>6.30</td>
<td>24.23</td>
<td>0.28</td>
<td>23.11</td>
</tr>
<tr>
<td>8 $\gamma = 5$, $\eta = \gamma$</td>
<td>0.55</td>
<td>0.85</td>
<td>13.30</td>
<td>22.12</td>
<td>12.76</td>
<td>22.51</td>
<td>0.76</td>
<td>11.11</td>
</tr>
<tr>
<td>II. at the end of 200y learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, $\gamma = 2$</td>
<td>3.30</td>
<td>0.59</td>
<td>4.28</td>
<td>11.59</td>
<td>0.97</td>
<td>11.61</td>
<td>0.09</td>
<td>58.17</td>
</tr>
<tr>
<td>2 $\gamma = 2$, $\eta = \gamma$</td>
<td>3.08</td>
<td>0.99</td>
<td>6.01</td>
<td>18.46</td>
<td>2.92</td>
<td>18.82</td>
<td>0.13</td>
<td>40.17</td>
</tr>
<tr>
<td>3 $\gamma = 2$, $\eta = 15$</td>
<td>3.03</td>
<td>1.10</td>
<td>6.41</td>
<td>19.77</td>
<td>3.37</td>
<td>20.21</td>
<td>0.14</td>
<td>36.80</td>
</tr>
<tr>
<td>4 $\gamma = 2$, $\eta = 5$</td>
<td>3.07</td>
<td>1.02</td>
<td>6.12</td>
<td>18.85</td>
<td>3.05</td>
<td>19.23</td>
<td>0.13</td>
<td>39.25</td>
</tr>
<tr>
<td>5 $\gamma = 2$, $\eta = 30$</td>
<td>3.01</td>
<td>1.11</td>
<td>6.59</td>
<td>20.19</td>
<td>3.58</td>
<td>20.65</td>
<td>0.14</td>
<td>34.82</td>
</tr>
<tr>
<td>6 $\gamma = 2$, $\eta = 60$</td>
<td>3.00</td>
<td>1.05</td>
<td>6.68</td>
<td>20.19</td>
<td>3.67</td>
<td>20.63</td>
<td>0.15</td>
<td>33.70</td>
</tr>
<tr>
<td>7 $\gamma = 3$, $\eta = \gamma$</td>
<td>2.59</td>
<td>1.19</td>
<td>8.07</td>
<td>22.42</td>
<td>5.47</td>
<td>22.95</td>
<td>0.26</td>
<td>25.40</td>
</tr>
<tr>
<td>8 $\gamma = 5$, $\eta = \gamma$</td>
<td>0.72</td>
<td>1.06</td>
<td>12.71</td>
<td>21.87</td>
<td>11.98</td>
<td>22.33</td>
<td>0.72</td>
<td>11.65</td>
</tr>
<tr>
<td>III. at the end of 500y learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, $\gamma = 2$</td>
<td>3.28</td>
<td>0.63</td>
<td>4.02</td>
<td>12.08</td>
<td>0.71</td>
<td>12.13</td>
<td>0.09</td>
<td>58.10</td>
</tr>
<tr>
<td>2 $\gamma = 2$, $\eta = \gamma$</td>
<td>3.17</td>
<td>0.89</td>
<td>4.90</td>
<td>16.30</td>
<td>1.69</td>
<td>16.57</td>
<td>0.10</td>
<td>49.65</td>
</tr>
<tr>
<td>3 $\gamma = 2$, $\eta = 15$</td>
<td>3.12</td>
<td>1.05</td>
<td>5.20</td>
<td>17.81</td>
<td>2.04</td>
<td>18.17</td>
<td>0.11</td>
<td>46.44</td>
</tr>
<tr>
<td>4 $\gamma = 2$, $\eta = 5$</td>
<td>3.16</td>
<td>0.93</td>
<td>4.97</td>
<td>16.68</td>
<td>1.78</td>
<td>16.97</td>
<td>0.11</td>
<td>48.93</td>
</tr>
<tr>
<td>5 $\gamma = 2$, $\eta = 30$</td>
<td>3.07</td>
<td>1.19</td>
<td>5.50</td>
<td>19.18</td>
<td>2.39</td>
<td>19.61</td>
<td>0.12</td>
<td>42.86</td>
</tr>
<tr>
<td>6 $\gamma = 2$, $\eta = 60$</td>
<td>3.00</td>
<td>1.29</td>
<td>5.94</td>
<td>20.62</td>
<td>2.88</td>
<td>21.11</td>
<td>0.14</td>
<td>37.92</td>
</tr>
<tr>
<td>7 $\gamma = 3$, $\eta = \gamma$</td>
<td>2.82</td>
<td>1.19</td>
<td>6.32</td>
<td>19.93</td>
<td>3.45</td>
<td>20.39</td>
<td>0.20</td>
<td>34.27</td>
</tr>
<tr>
<td>8 $\gamma = 5$, $\eta = \gamma$</td>
<td>1.35</td>
<td>1.50</td>
<td>10.77</td>
<td>22.69</td>
<td>9.36</td>
<td>23.31</td>
<td>0.58</td>
<td>14.58</td>
</tr>
</tbody>
</table>
Table 4.8: Long-sample moments: $p_0 = 0.1$

This table reports the simulated long-sample moments of asset prices and returns. In simulation, the investor updates the probability of the long-recession parameter value from a prior of $p_0 = 0.1$. At the end of 100, 200, and 500 years of learning, we calculate the mean and the standard deviation of risk free rate, market return and excess market return, the Sharpe ratio, the price of risk and the P/D ratio. In model 1, the persistence of recessions is known at the true value. In model 2, the true persistence is unknown and the investor learns about the parameter under EZ preferences, where $\gamma = 2$ is the benchmark value for the risk aversion coefficient. In model 3, the investor learns about the parameter under ambiguity aversion, where $\eta = 15$ is the benchmark value for the ambiguity aversion coefficient. Models 4–6 consider other values of $\eta$ and models 7 and 8 consider other values of $\gamma$. Mean and standard deviation of returns are in percentage.

<table>
<thead>
<tr>
<th>$E[R_f]$</th>
<th>$\sigma[R_f]$</th>
<th>$E[R_m]$</th>
<th>$\sigma[R_m]$</th>
<th>$E[R^n_m]$</th>
<th>$\sigma[R^n_m]$</th>
<th>$SR$</th>
<th>$P/D$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. at the end of 100y learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, $\gamma = 2$</td>
<td>3.30</td>
<td>0.58</td>
<td>4.35</td>
<td>11.57</td>
<td>1.05</td>
<td>11.60</td>
<td>0.09</td>
</tr>
<tr>
<td>2 $\gamma = 2$, $\eta = \gamma$</td>
<td>3.26</td>
<td>0.73</td>
<td>4.79</td>
<td>13.67</td>
<td>1.54</td>
<td>13.84</td>
<td>0.09</td>
</tr>
<tr>
<td>3 $\gamma = 2$, $\eta = 15$</td>
<td>3.21</td>
<td>0.91</td>
<td>5.16</td>
<td>15.51</td>
<td>1.96</td>
<td>15.79</td>
<td>0.10</td>
</tr>
<tr>
<td>4 $\gamma = 2$, $\eta = 5$</td>
<td>3.25</td>
<td>0.77</td>
<td>4.86</td>
<td>14.02</td>
<td>1.62</td>
<td>14.21</td>
<td>0.09</td>
</tr>
<tr>
<td>5 $\gamma = 2$, $\eta = 30$</td>
<td>3.13</td>
<td>1.18</td>
<td>5.80</td>
<td>18.17</td>
<td>2.67</td>
<td>18.61</td>
<td>0.11</td>
</tr>
<tr>
<td>6 $\gamma = 2$, $\eta = 60$</td>
<td>3.04</td>
<td>1.33</td>
<td>6.62</td>
<td>20.51</td>
<td>3.59</td>
<td>21.12</td>
<td>0.14</td>
</tr>
<tr>
<td>7 $\gamma = 3$, $\eta = \gamma$</td>
<td>2.99</td>
<td>1.06</td>
<td>6.07</td>
<td>17.16</td>
<td>3.09</td>
<td>17.55</td>
<td>0.17</td>
</tr>
<tr>
<td>8 $\gamma = 5$, $\eta = \gamma$</td>
<td>1.67</td>
<td>1.46</td>
<td>10.88</td>
<td>22.36</td>
<td>9.22</td>
<td>23.06</td>
<td>0.53</td>
</tr>
<tr>
<td>II. at the end of 200y learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, $\gamma = 2$</td>
<td>3.30</td>
<td>0.59</td>
<td>4.28</td>
<td>11.59</td>
<td>0.97</td>
<td>11.61</td>
<td>0.09</td>
</tr>
<tr>
<td>2 $\gamma = 2$, $\eta = \gamma$</td>
<td>3.27</td>
<td>0.69</td>
<td>4.48</td>
<td>12.63</td>
<td>1.20</td>
<td>12.71</td>
<td>0.09</td>
</tr>
<tr>
<td>3 $\gamma = 2$, $\eta = 15$</td>
<td>3.24</td>
<td>0.81</td>
<td>4.69</td>
<td>13.82</td>
<td>1.45</td>
<td>13.97</td>
<td>0.10</td>
</tr>
<tr>
<td>4 $\gamma = 2$, $\eta = 5$</td>
<td>3.26</td>
<td>0.71</td>
<td>4.51</td>
<td>12.85</td>
<td>1.24</td>
<td>12.94</td>
<td>0.09</td>
</tr>
<tr>
<td>5 $\gamma = 2$, $\eta = 30$</td>
<td>3.19</td>
<td>1.01</td>
<td>5.10</td>
<td>15.79</td>
<td>1.91</td>
<td>16.08</td>
<td>0.10</td>
</tr>
<tr>
<td>6 $\gamma = 2$, $\eta = 60$</td>
<td>3.07</td>
<td>1.33</td>
<td>5.99</td>
<td>19.01</td>
<td>2.92</td>
<td>19.51</td>
<td>0.13</td>
</tr>
<tr>
<td>7 $\gamma = 3$, $\eta = \gamma$</td>
<td>3.03</td>
<td>0.95</td>
<td>5.45</td>
<td>15.01</td>
<td>2.42</td>
<td>15.24</td>
<td>0.16</td>
</tr>
<tr>
<td>8 $\gamma = 5$, $\eta = \gamma$</td>
<td>1.86</td>
<td>1.47</td>
<td>9.93</td>
<td>20.96</td>
<td>8.06</td>
<td>21.59</td>
<td>0.48</td>
</tr>
<tr>
<td>III. at the end of 500y learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 full info, $\gamma = 2$</td>
<td>3.28</td>
<td>0.63</td>
<td>4.02</td>
<td>12.08</td>
<td>0.71</td>
<td>12.13</td>
<td>0.09</td>
</tr>
<tr>
<td>2 $\gamma = 2$, $\eta = \gamma$</td>
<td>3.27</td>
<td>0.68</td>
<td>4.11</td>
<td>12.58</td>
<td>0.82</td>
<td>12.65</td>
<td>0.09</td>
</tr>
<tr>
<td>3 $\gamma = 2$, $\eta = 15$</td>
<td>3.26</td>
<td>0.73</td>
<td>4.19</td>
<td>13.03</td>
<td>0.91</td>
<td>13.13</td>
<td>0.09</td>
</tr>
<tr>
<td>4 $\gamma = 2$, $\eta = 5$</td>
<td>3.27</td>
<td>0.69</td>
<td>4.13</td>
<td>12.67</td>
<td>0.84</td>
<td>12.75</td>
<td>0.09</td>
</tr>
<tr>
<td>5 $\gamma = 2$, $\eta = 30$</td>
<td>3.24</td>
<td>0.82</td>
<td>4.36</td>
<td>13.89</td>
<td>1.10</td>
<td>14.06</td>
<td>0.09</td>
</tr>
<tr>
<td>6 $\gamma = 2$, $\eta = 60$</td>
<td>3.18</td>
<td>1.04</td>
<td>4.87</td>
<td>16.54</td>
<td>1.67</td>
<td>16.91</td>
<td>0.10</td>
</tr>
<tr>
<td>7 $\gamma = 3$, $\eta = \gamma$</td>
<td>3.08</td>
<td>0.85</td>
<td>4.80</td>
<td>13.92</td>
<td>1.70</td>
<td>14.09</td>
<td>0.15</td>
</tr>
<tr>
<td>8 $\gamma = 5$, $\eta = \gamma$</td>
<td>2.25</td>
<td>1.38</td>
<td>8.06</td>
<td>19.55</td>
<td>5.77</td>
<td>20.10</td>
<td>0.38</td>
</tr>
</tbody>
</table>
Figure 4.1: Distorted probability of the long-recession parameter value under ambiguity aversion. This figure plots the distorted probability of the long-recession parameter value implied by the ambiguity model, for the recession regime ($s_t = 1$) and the expansion regime ($s_t = 2$). The risk aversion coefficient and the ambiguity aversion coefficient are set at their benchmark values, i.e., $\gamma = 2$ and $\eta = 15$. The x-axis is the Bayesian probability of the long-recession parameter value.
Figure 4.2: Effects of parameter learning on risk free rate and market price of risk. This figure plots the risk free rate (top) and the market price of risk (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes (i) parameter is known at $\pi_{11} = 0.516$; (ii) parameter learning under EZ preferences; (iii) parameter learning under ambiguity aversion (AA); (iv) parameter is known at $\pi_{11} = 0.8$. The risk aversion coefficient and the ambiguity aversion coefficient are set at the benchmark values: $\gamma = 2$ and $\eta = 15$. 
Figure 4.3: Effects of parameter learning on equity premium and P/D ratio. This figure plots the equity premium (top) and the P/D ratio (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes (i) parameter is known at $\pi_{11} = 0.516$; (ii) parameter learning under EZ preferences; (iii) parameter learning under ambiguity aversion (AA); (iv) parameter is known at $\pi_{11} = 0.8$. The risk aversion coefficient and the ambiguity aversion coefficient are set at the benchmark values: $\gamma = 2$ and $\eta = 15$. 
Figure 4.4: Effects of parameter learning on volatilities. This figure plots the volatility of the pricing kernel (top) and the market return (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes (i) parameter is known at $\pi_{11} = 0.516$; (ii) parameter learning under EZ preferences; (iii) parameter learning under ambiguity aversion (AA); (iv) parameter is known at $\pi_{11} = 0.8$. The risk aversion coefficient and the ambiguity aversion coefficient are set at the benchmark values: $\gamma = 2$ and $\eta = 15$. 
Figure 4.5: Effects of ambiguity aversion coefficient on risk free rate and market price of risk. This figure plots the risk free rate (top) and the market price of risk (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under ambiguity aversion (AA) with four values for the ambiguity aversion coefficient: $\eta = \{5, 15, 30, 60\}$, where $\eta = 15$ is the benchmark value. The risk aversion coefficient is set at the benchmark value of $\gamma = 2$. 
Figure 4.6: Effects of ambiguity aversion coefficient on equity premium and P/D ratio. This figure plots the equity premium (top) and the P/D ratio (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under ambiguity aversion (AA) with four values for the ambiguity aversion coefficient: $\eta = \{5, 15, 30, 60\}$, where $\eta = 15$ is the benchmark value. The risk aversion coefficient is set at the benchmark value of $\gamma = 2$. 
Figure 4.7: Effects of ambiguity aversion coefficient on volatilities. This figure plots the volatility of the pricing kernel (top) and the market return (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under ambiguity aversion (AA) with four values for the ambiguity aversion coefficient: $\eta = \{5, 15, 30, 60\}$, where $\eta = 15$ is the benchmark value. The risk aversion coefficient is set at the benchmark value of $\gamma = 2$. 
Figure 4.8: Effects of risk aversion coefficient on risk free rate and market price of risk. This figure plots the risk free rate (top) and the market price of risk (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under EZ preferences with five values for the risk aversion coefficient: $\gamma = \{2, 3, 5\}$, where $\gamma = 2$ is the benchmark value.
Figure 4.9: Effects of risk aversion coefficient on equity premium and P/D ratio. This figure plots the equity premium (top) and the P/D ratio (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under EZ preferences with five values for the risk aversion coefficient: $\gamma = \{2, 3, 5\}$, where $\gamma = 2$ is the benchmark value.
Figure 4.10: Effects of risk aversion coefficient on volatilities. This figure plots the volatility of the pricing kernel (top) and the market return (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under EZ preferences with five values for the risk aversion coefficient: $\gamma = \{2, 3, 5\}$, where $\gamma = 2$ is the benchmark value.
Figure 4.11: Effects of ambiguity and risk aversion coefficients on risk free rate and market price of risk. This figure plots the risk free rate (top) and the market price of risk (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under ambiguity aversion (AA) with $\gamma = 2$ and $\eta = 15$, and parameter learning under EZ preferences with $\gamma = \{2, 3, 5\}$. 
Figure 4.12: Effects of ambiguity and risk aversion coefficients on equity premium and P/D ratio. This figure plots the equity premium (top) and the P/D ratio (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under ambiguity aversion (AA) with $\gamma = 2$ and $\eta = 15$, and parameter learning under EZ preferences with $\gamma = \{2, 3, 5\}$. 
Figure 4.13: Effects of ambiguity and risk aversion coefficients on volatilities. This figure plots the volatility of the pricing kernel (top) and the market return (bottom) as a function of the probability of the long-recession parameter value, conditional on the recession regime (left), the expansion regime (middle), and unconditional on regime (right). Each panel assumes parameter learning under ambiguity aversion (AA) with $\gamma = 2$ and $\eta = 15$, and parameter learning under EZ preferences with $\gamma = \{2, 3, 5\}$. 
Appendix

4.A Value Function

Conjecture the value function as

\[ V_t = \begin{cases} 
G_1(p_t) \cdot C_t, & \text{if } s_t = 1 \\
G_2(p_t) \cdot C_t, & \text{if } s_t = 2,
\end{cases} \]

where \( p_t \) is the probability of the long-recession parameter value and \( G_1, G_2 \) are functions to be solved. Equation (4.8) and (4.9) imply that function \( G \) satisfies

\[
V_t = \left( (1 - \beta)C_t^{1-\rho} + \beta \left[ R_t(V_{t+1}) \right]^{1-\rho} \right)^{\frac{1}{1-\rho}}
\]

\[
\frac{V_t}{C_t} = \left( (1 - \beta) + \beta \left[ R_t \left( \frac{V_{t+1}}{C_t} \right) \right]^{1-\rho} \right)^{\frac{1}{1-\rho}}
\]

\[
G(p_t) = \left( (1 - \beta) + \beta \left[ R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right) \right]^{1-\rho} \right)^{\frac{1}{1-\rho}}
\]

\[
G(p_t) = \left( (1 - \beta) + \beta \left\{ E \left[ \left( E \left[ G(p_{t+1})^{1-\gamma} \left( \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right] \right)^{\frac{1-\gamma}{1-\gamma}} \right] \right\}^{\frac{1-\gamma}{1-\gamma}} \right)^{\frac{1}{1-\gamma}}
\]

The numerical method has the following steps.

1. Chebyshev polynomial

   (a) Choose the degree of the approximating Chebyshev polynomial, i.e., \( (N-1) \) for \( p \). We choose \( N = 400 \) to ensure a nicely-behaved function at the boundary where \( p \) approaches zero. Approximate function \( G_i(p), i = 1, 2 \)
by

\[ G_t(p) = \exp \left\{ \sum_{n=0}^{N-1} c_n^i T_n(\psi) \right\}. \]

where \( \psi \) are Chebyshev interpolation nodes; \( T \) are Chebyshev polynomials; \( c \) are the coefficients of polynomials that are going to be solved. For \( \forall \psi \), \( T_n(\psi) \) is calculated recursively by

\[
\begin{align*}
T_0(\psi) &= 1 \\
T_1(\psi) &= \psi \\
T_{n+1}(\psi) &= 2\psi T_n(\psi) - T_{n-1}(\psi).
\end{align*}
\]

(b) The \( n \)th Chebyshev interpolation node for \( p \) is

\[ \psi = \cos \left( \frac{N - n + 0.5}{N} \pi \right), \]

and \( \psi \in [-1, 1] \) is adjusted to state variable \( p \in [0, 1] \) by

\[ p = 0.5 + 0.5 \cdot \psi. \]

[Adjust \( \psi \in [-1, 1] \) to \( z \in [lb, ub] \) by \( z = \frac{lb+ub}{2} + \frac{ub-lb}{2} \cdot \psi = \frac{(\psi+1)(ub-lb)}{2} + lb \).]

(c) Given state variable \( p \), \( \psi \in [-1, 1] \) is transformed from \( p \in [0, 1] \) by

\[ \psi = -1 + 2 \cdot p. \]

[Adjust \( z \in [lb, ub] \) to \( \xi \in [-1, 1] \) by \( \xi = -\frac{lb+ub}{ub-lb} + \frac{2}{ub-lb} \cdot z = \frac{2(z-lb)}{ub-lb} - 1 \).]

(d) There are \( (N \times 2) \) unknowns coefficients of \( c_n^i \) to be solved with \( (N \times 2) \) equations. Equation \( (n, i) \) is the Euler equation given the \( n \)th node of \( p \) and the current observable regime \( s_t = i, \ i = 1, 2 \). Without loss of generality, Step 2 and 3 calculate Euler equation \( (n, 1) \), \( \forall n = 1, 2, \ldots, N \) and Euler equation \( (n, 2) \), \( \forall n = 1, 2, \ldots, N \), respectively.

2. The current observable regime is \( s_t = 1 \) (recession regime).

(a) At time \( t \), given state variable \( p_t \), calculate function \( G_1(p_t) \) as

\[ G_1(p_t) = \exp \left\{ \sum_{n=0}^{N-1} c_n^1 T_n(\psi) \right\}. \]
(b) Update \( p_t \) to \( p_{t+1} \) by equation (4.7), i.e.,

\[
p_{t+1} = \begin{cases} 
\frac{\pi^h_{11} \cdot p_t}{\pi^l_{11} \cdot (1 - p_t) + \pi^h_{11} \cdot p_t} > p_t, & \text{if } s_t = 1, \ s_{t+1} = 1 \\
\frac{(1 - \pi^h_{11}) \cdot (1 - p_t) + (1 - \pi^h_{11}) \cdot p_t}{\pi^l_{11} \cdot (1 - p_t) + (1 - \pi^h_{11}) \cdot p_t} < p_t, & \text{if } s_t = 1, \ s_{t+1} = 2 . 
\end{cases}
\]

(c) Calculate function \( G_i(p_{t+1}) \), \( i = 1, 2 \) by

\[
G_i(p_{t+1}) = \exp \left\{ \sum_{n=0}^{N-1} c_n^i T_n(\psi) \right\} .
\]

where \( \psi \) is transformed from \( p_{t+1} \) when \( s_{t+1} = i \).

(d) The risk aggregator \( R_t \left( \frac{V_{t+1}}{C_{t+1}} \right) = R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right) \) is calculated by

\[
R_t \left( \frac{V_{t+1}}{C_t} \right) = \left[ (p_t) \cdot \left( E_{\pi_{11}=\pi_{11}^h} \right)^{1-\eta} + (1 - p_t) \cdot \left( E_{\pi_{11}=\pi_{11}^l} \right)^{1-\eta} \right]^{\frac{1}{1-\eta}} , \quad (4.19)
\]

where \( \left( E_{\pi_{11}=\pi_{11}^j} \right), \ j = l, h \) is

\[
\left( E_{\pi_{11}=\pi_{11}^j} \right) = \left( E \left[ G(p_{t+1})^{1-\gamma} \left( \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}}
\]

\[
= \left\{ \left( \pi_{11}^l \cdot \left( G_1(p_{t+1}) \right)^{1-\gamma} \cdot e^{(1-\gamma)\mu_1 + \frac{1}{2}(1-\gamma)^2\sigma_1^2} \right) + \left( 1 - \pi_{11}^l \right) \cdot \left( G_2(p_{t+1}) \right)^{1-\gamma} \cdot e^{(1-\gamma)\mu_2 + \frac{1}{2}(1-\gamma)^2\sigma_2^2} \right\}^{\frac{1}{1-\gamma}} . \quad (4.20)
\]

(e) The Euler equation is

\[
G_1(p_t) = \left\{ (1 - \beta) + \beta \cdot \left[ R_t \left( \frac{V_{t+1}}{C_t} \right) \right]^{1-\rho} \right\}^{\frac{1}{1-\rho}} .
\]

3. The current observable regime is \( s_t = 2 \) (expansion regime).

(a) At time \( t \), given state variable \( p_t \), calculate function \( G_1(p_t) \) as

\[
G_2(p_t) = \exp \left\{ \sum_{n=0}^{N-1} c_n^2 T_n(\psi) \right\} .
\]

(b) Update \( p_t \) to \( p_{t+1} \) by equation (4.7), i.e.,

\[
p_{t+1} = p_t, \text{ if } s_t = 2 .
\]
(c) Calculate function $G_i(p_t), i = 1, 2$ by

$$G_i(p_{t+1}) = \exp \left\{ \sum_{n=0}^{N-1} c_n^i T_n(\psi) \right\}.$$  

where $\psi$ is transformed from $p_{t+1}$ when $s_{t+1} = i$.

(d) The risk aggregator $R_t \left( \frac{V_{t+1}}{C_t} \right) = R_t \left( \frac{G(p_{t+1})C_{t+1}}{C_t} \right)$ is calculated by

$$R_t \left( \frac{V_{t+1}}{C_t} \right) = \left( E \left[ G(p_{t+1})^{1-\gamma} \left( \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}}$$

$$= \left\{ (1 - \pi_{22}) \cdot \left( G_1(p_{t+1})^{1-\gamma} \cdot e^{(1-\gamma)\mu_1 + \frac{1}{2}(1-\gamma)\sigma_1^2} \right) 
+ (\pi_{22}) \cdot \left( G_2(p_{t+1})^{1-\gamma} \cdot e^{(1-\gamma)\mu_2 + \frac{1}{2}(1-\gamma)\sigma_2^2} \right) \right\}^{\frac{1}{1-\gamma}}. \quad (4.21)$$

(e) The Euler equation is

$$G_2(p_t) = \left\{ (1 - \beta) + \beta \cdot \left[ R_t \left( \frac{V_{t+1}}{C_t} \right) \right]^{\frac{1}{1-\gamma}} \right\}^{-\frac{1}{1-\gamma}}.$$

### 4.B P/D Ratio

Let $\varphi_i(p_t), i = 1, 2$ denote the P/D ratio at time $t$ when the current observable regime is $s_t = i$, and this function is approximated by 1-dimensional, $(N - 1)$ degree Chebyshev polynomials. The P/D ratio satisfies

$$1 = E_t \left[ M_{t+1} \cdot R_{m,t+1} \right],$$

where $R_{m,t+1}$ is the gross stock market return and is given by

$$R_{m,t+1} = \frac{P_{t+1} + D_{t+1}}{D_t} = \frac{\varphi(\mu_{t+1}) + 1}{\varphi(\mu_t)} \cdot \frac{D_{t+1}}{D_t},$$

where

$$\frac{D_{t+1}}{D_t} = \left( \frac{C_{t+1}}{C_t} \right)^\lambda \cdot e^{\mu_d}.$$
The pricing kernel is

\[ M_{t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left( \frac{V_{t+1}}{R_t(V_{t+1})} \right)^{-\gamma} \left( \frac{\left( E \left[ V_{t+1}^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}}}{R_t(V_{t+1})} \right)^{-(\eta-\gamma)} \]

\[ = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left( \frac{G(p_{t+1}) \frac{C_{t+1}}{C_t}}{R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right)} \right)^{\rho-\gamma} \left( \frac{\left( E \left[ \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}}}{R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right)} \right)^{-(\eta-\gamma)} \]

where value function is solved by \( V_t = G_i(p_t) \cdot C_t, \ i = 1, 2 \) in Appendix 4.A.

When the current observable regime is \( s_t = 1 \) (recession regime), the risk aggregator \( R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right) \) is given by (4.19) and \( \left( E \left[ \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}} \) is given (4.20). When the current observable regime is \( s_t = 2 \) (expansion regime),

\[ R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right) = \left( E \left[ \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}}, \]

and the pricing kernel reduces to

\[ M_{t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left( \frac{G(p_{t+1}) \frac{C_{t+1}}{C_t}}{R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right)} \right)^{\rho-\gamma}, \]

where the risk aggregator \( R_t \left( G(p_{t+1}) \frac{C_{t+1}}{C_t} \right) \) is given by (4.21).
Chapter 5

Conclusions

Uncertainty is a fundamental theme in the field of asset pricing. The specification of the uncertainty for which investors require compensation is central to every asset pricing model. In empirical studies, researchers cannot control the random shocks of data due to the nature of social science, and utilise financial econometrics for estimation and inference. The inconvenient truth is that there are uncertainties over parameter estimates and even model specifications. This thesis examines three uncertainties that have an impact on the determination or the estimation of asset prices and returns, namely the uncertainty over monetary policy regime shifts, the statistical bias in parameter estimates, and the uncertainty over the persistence of recessions.

Chapter 2 examines how the uncertainty over monetary policy regime shifts implied from bond markets is able to explain the excess return on the USD-JPY currency pair. Applying the regime switching dynamic term structure model of Dai, Singleton and Yang (2007) to a two-country setting, I find that the shift of monetary policies generates currency risk: the yen excess return is predicted by the Japanese regime shift premium. During the transition from the conventional monetary policy regime to the unconventional policy regime of quantitative easing from 1995 to 2000, there is substantial uncertainty over regime shifts in the Japanese bond market, implying substantial risk in Japanese yen, and it is in this period that the yen carry trade emerges. One of the limitations of this empirical work is that it studies only one currency pair in the forward premium puzzle, and I would like to generalise the results
to other currencies. For example, Backus, Gavazzoni, Telmer and Zin (2013) point out that the high-interest-rate Australian dollar tends to appreciate because the Australian monetary policy responses more to output gap and less to inflation relative to its US counterpart, and it would be convenient to model a transition of regimes in the response of monetary authorities as in, e.g., Bikbov and Chernov (2013). Another limitation of this study is that the model does not tell much empirically about the recent financial crisis and the unconventional monetary policies in the US and Japan, as the dataset of Wright (2011) ends in May 2009. It is on the future agenda to include more contemporary data, and the implications of recent regime shifts for exchange rates are going to be tested. An extension of the model is motivated by its prediction about unconventional monetary policies of quantitative easing: it predicts an increase of uncertainty as central banks both enter into and exit from the unconventional monetary policy regime, and thus calls for caution in the phasing out of quantitative easing. A more structured model explicitly incorporating monetary policy regimes may be considered, which would involve both theoretical and empirical attempts.

Chapter 3 examines the small sample bias in the estimation of a regime switching dynamic term structure model and quantifies the effects of bias correction on the inference about expectations of future policy rates and term premia. Using US data from 1971 to 2009, I document two regimes driven by the conditional volatility of bond yields and risk factors. In both regimes, the process of bond yields is highly persistent, which is the source of estimation bias when the sample size is small. After bias correction, the inference about expectations of future policy rates and long-maturity term premia changes dramatically in two high-volatility periods: the 1979–1982 monetary experiment and the recent financial crisis. It turns out that the trends in long-term forward rates are to a great extent driven by the expectations of future policy rates, while the term premia exhibit normal business-cycle variations. Empirical findings are further supported by Monte Carlo simulation. I then propose the following future work. Firstly, since simulation results also indicate that the bootstrap bias correction procedure corrects the bias but does not do so fully, future work may consider more refined bias correction procedures, e.g., the indirect inference approach proposed by Bauer, Rudebusch and Wu (2012). Secondly, it would be of interest to examine, e.g., how large a sample is sufficient for a substantial decrease in estimation bias; what is
the role of regime switches in the estimation bias in the persistence of bond yields, because in simulation, the bias in the low-volatility regime is trivial compared to an affine model with similar persistence of bond yields and similar sample size.

Chapter 4 examines the implications of learning about the uncertainty over the persistence of recessions under ambiguity aversion for asset prices and returns, in particular, the equity premium at the onset of recessions. In the benchmark model, the parameter that governs the persistence of recessions is unknown, and the representative investor learns about this parameter while being ambiguity averse towards parameter uncertainty. I document an uncertainty effect that characterises the difference between learning under ambiguity aversion and learning under recursive preferences. Moreover, this uncertainty effect is asymmetric across regimes—the size of the effect is moderate in the expansion regime and is substantial in the recession regime. At the annual frequency, our simulation results show that at the onset of the next recession, ambiguity models are able to generate 10–30 times the equity premium in the preceding period, and increasing the magnitude of ambiguity aversion will further increase the ratio. Having established the existence of an important asymmetric uncertainty effect due to learning under conditions of ambiguity aversion via simulations, future work should consider actual sample data and attempt both the historical sample moments and the time variation of asset prices and returns present in the data, as in Collard, Mukerji, Sheppard and Tallon (2016). Another strand of empirical studies on ambiguity aversion proposes measures of Knightian uncertainty (ambiguity) using survey data, see, e.g., Rossi, Sekhposyan and Soupre (2016). It would be of interest to look into any survey data that contains forecasts about the duration of recessions. Another consideration is that since the moments of asset prices and returns due to ambiguity aversion depends on the belief about the persistence of recessions in a flexible way, it would be interesting to consider an economy where ambiguity-averse investors have heterogeneous beliefs about parameters.

In summary, a wide range of uncertainties affect asset prices and returns and the way we interpret economic and financial phenomena. Future work would consider more structured specification of uncertainties and deal with the uncertainty in model estimation, which would involve both theoretical and empirical attempts.
Bibliography


