The Term Structure of Interest Rates and the Economic Value of its Predictability in the UK

Thesis submitted for the degree of

Doctor of Philosophy

at the

University of Leicester

by

Kavita Sirichand B.Sc. M.Sc.

Department of Economics

University of Leicester

December 2009

To my beloved family,

Papaji, Choko Daddy, Mimi, Dad, Mum, Sandhya and Rohini

Itni shiddat se maine tumhe paane ki koshish ki hai, ke har zarre ne mujhe tumse milaane ki saazish ki hai...kehte hai agar kisi cheez ko dil se chaho tho puri kayanaat usse tumse milaane ki koshish mein lag jaati hai.

Om Shanti Om, 2007

The Term Structure of Interest Rates and the Economic Value of its Predictability in the UK

by

Kavita Sirichand

Abstract

The term structure of interest rates describes the relationship between short- and long-term rates and embeds the market's expectation of future interest rates. This has led to a large literature concerned with modelling the term structure and hence attempting to extract this information.

This thesis is concerned with both modelling and forecasting the UK term structure, with a focus on the application of density forecasting and decision-based forecast evaluation. We test the Expectations Hypothesis of the term structure and more generally, examine if the term structure is best described by a statistical or theory informed model.

Interest rate forecasts are essential for policymakers and practitioners alike. Since density forecasts provide the entire distribution about the forecast, we argue that they are appropriate for an investor concerned with the uncertainties about future asset returns.

We find economic theory to have explanatory power for the term structure and the UK money market to be consistent with the Expectations Hypothesis. Further, we demonstrate how density forecasting techniques can be applied to forecast asset returns and inform portfolio allocation decisions; and how these optimal allocations are sensitive to the forecast uncertainties about the expected future returns and the assumptions made regarding return predictability.

Furthermore, given the importance of forecast evaluation, our results highlight the need to judge forecasts in the decision making context for which they are ultimately intended. That is, our findings advocate the use of decision-based criteria that assess forecasts from the user's perspective, i.e. in terms of economic value, rather than conventional statistical measures. Under decision-based methods, we find that the investor may gain in terms of wealth by assuming returns are predictable and using a theory informed model to forecast.

In short, we find economic theory to be significant for both modelling and forecasting the term structure.

1

Acknowlegdements

First and foremost I thank my dear family, my grandparents, parents and sisters. Your unwavering love and support have been a pillar of strength to me, words are inadequate to express my gratitude. Without you this would not have been possible and to you this thesis I dedicate.

To my supervisors, Prof. Kevin Lee and Prof. Stephen Hall, I am truly thankful for all your guidance, inspiration, kindness and patience. You have generously shared your knowledge and time with me. It has been an honour and a privilege to have been your student.

I would like to thank Prof. Giorgio Valente for all his advice, encouragement and help. I am grateful to my thesis committee chair Prof. Gianni De Fraja, and thankful to Prof. Panicos Demetriades and Dr Ali al-Nowaihi for inspiring me to pursue a doctorate. I am indebted to the University of Leicester for giving me the opportunity to do my PhD and for their financial support. Further, I thank my thesis examiners Prof. Wojciech Charemza and Prof. Chris Brooks.

I give thanks to Dr. Kalvinder Shields and Dr. Emi Mise for introducing me to and assisting me with GAUSS, and to Dr. Ludovic Renou for his help with SWP. I appreciate the useful comments and guidance given by Prof. Gary Koop, Dr. Rodney Strachan and Dr. Sebastiano Manzan at the early stages of my thesis. Also, I thank Sebastian O'Halloran, Eve Kilbourne, Ladan Baker and Sam Hill.

My dearest friends Dr. Simeon Coleman, Andrea Oterová and Sahar Qaqeesh, I thank you for your continued encouragement and support, your friendship has been invaluable. I extend my thanks to the many fellow students I have had the pleasure of meeting, I will always remember the time we have shared with great fondness.

I give my heartfelt thanks to all of those who have helped me see this thesis to fruition and for this I am eternally grateful.

Contents

Appendices **268**

List of Figures

List of Tables

Chapter 1

Introduction

The term structure of interest rates has been the subject of much research, unsurprisingly, given that it embeds the market's expectations of future interest rates. Because of the information contained in the term structure, having an explanation of the term structure is important for market players and provides a means to extract this information. In turn, being able to accurately model the term structure is key for forecasting interest rates.

This thesis contributes to this important literature by modelling and forecasting the UK term structure at the short end. We consider the importance of economic theory for modelling the term structure and further, examine if there is economic value to using a theory based model to forecast the term structure. We focus on the application of density forecasting techniques for predicting interest rates and stock returns and the use of these forecasts for investment decision making. Typically, forecasts are evaluated using statistical measures which ignore how the forecasts will be used and the preferences of the user. Granger and Pesaran (1996, 2000) argue that given forecasts are used to inform decisions, the evaluation criteria should depend on the decision making environment. Here, we evaluate forecasts using both conventional statistical measures and the not so frequently used decision-based methods.

In this chapter, we begin by discussing the term structure, the importance of the information it contains, density forecasts and decision-based forecast evaluation. Section 1.1 describes the motivation of the thesis, Section 1.2 summarises the three empirical chapters and Section 1.3 outlines the contributions this thesis makes.

The term structure of interest rates describes the relationship between the interest rate and term to maturity of a bond, where the yield curve plots these interest rates against their terms[1]. The shape of the yield curve reflects the market's expectation of future interest rates, given the current market conditions. The most frequently observed upward sloping curve, reflects the expectation of future economic growth and a risk of future inflation. Given that the term structure of interest rates embodies the market's anticipation of future events, having an explanation of the term structure provides a way to extract this important information, Cox, Ingersoll and Ross (1985). As such, there is a vast literature concerned with examining the term structure.

The leading theory of the term structure is the Expectations Hypothesis (EH), which postulates that the long term rate is a weighted average of expected future short term rates. In their key study, Campbell and Shiller (CS, 1991) consider if the slope, given by the spread, of the term structure has predictive power for future interest rate changes, and if this predictive power complies with the EH. They argue that these questions are key for forecasting interest rates and explaining yield curve shifts. The empirical evidence in favour of the EH is found to be sensitive to the types of interest rate data used, the time period considered, the monetary policy in operation and the testing method employed. However, many report strong comovement between the theoretical spread, i.e. that predicted by the EH and the actual spread, thereby making it difficult to reject a hypothesis that is economically significant. Other theories include the Market Segmentation Hypothesis and the Liquidity Preference Hypothesis. We discuss the theories and review the empirical literature in Chapter 2.

---

[1]The term structure refers to a particular yield curve, that for zero-coupon bonds.

In this thesis, we begin by modelling the term structure at the short end. The first empirical chapter provides a time series analysis of the UK term structure of interest rates, and further examines if the EH provides a good characterisation of the UK money market. We then turn our attention from modelling to forecasting the term structure. We generate density forecasts of interest rates and judge these forecasts using statistical and decision-based forecast evaluation criteria.

Density forecasts provide every possible outcome of a chosen variable for a particular future date, thereby giving a complete description of the uncertainty surrounding the forecast. From this, the likelihood of a given event being realised can be computed. This is a more useful means of presenting forecasts, since it communicates the uncertainty about the forecast in a clear and coherent manner[2], unlike point forecasts that only give a single value. This method of forecasting is potentially of great use with those including the Bank of England, JP Morgan and Reuters using density forecasts to convey the uncertainty about their projections.

We explore this potential in the second and third empirical chapters, by examining how an investor optimally allocates his portfolio when faced with the uncertainties about the future asset returns. Given portfolio composition depends on both the expected return and the associated risk of each asset, density forecasting is an appropriate way of conveying the risk and return information of each asset. More specifically, by using density forecasts of future returns to inform his decision making, the investor utilises the entire distribution about the forecast, i.e. expected return and the uncertainty (risk) about this expectation as captured by the variance of the forecast. The expected return distribution of each asset can be compared to derive suggestions for an optimal portfolio.

Further, we evaluate the forecasts using both statistical and decision-based criteria. Recently leading researchers have argued in favour of decision-based forecast

---

[2] Chapter 2 provides a more detailed discussion of density forecasting and types of forecast uncertainties.

14

evaluation[3], where forecasts are evaluated by their economic value to the user. Rather than in terms of conventional statistical measures based on forecast errors, that do not take into account the objectives of the user, Leitch and Tanner (1991) and Pesaran and Skouras (2004). We argue that in this investment decision making context, it is important to consider (a) the risk and return of the asset and (b) the investor's feeling about risk and how valuable the forecasts are to him. We show that density forecasting is appropriate for (a) and decision-based forecast evaluation for (b).

1.1   Motivation of Thesis

The research conducted in this thesis is motivated by the importance of the term structure and the information it embeds. The term structure embeds information about the market's expectation of future events. Because of this information, having knowledge of and being able to explain the term structure is important for market players including (1) policy makers for the transmission of monetary policy, (2) practitioners concerned with pricing instruments and (3) investors interested in allocating their portfolio. Our objective is to both model and forecast the UK term structure. Chapter 3 is concerned with modelling the term structure, which is important for all market players. Chapters 4 and 5 are concerned with forecasting the term structure and we examine this from the perspective of an investor.

We observe a parallel between the term structure and forecasting literature. That is, under statistical measures the EH is often rejected. However, CS argue that the EH is economically significant if the theoretical spread can explain the majority of the variation in the actual spread. Similarly, in the forecasting literature under statistical forecast evaluation measures, theory informed models are largely rejected in favour of

---

[3]"Quite simply, good forecasts help produce good decisions. Recognition and awareness of the decision-making environment is the key to effective...evaluation of forecasting models", Diebold (2004, pp. 32).

a simple no change model[4]. However, recent research shows that when decision-based measures are used support for theory informed models is found, see Della Corte et al (2008) and Garratt and Lee (2009). This observation motivates the research presented in this thesis, where Chapter 3 tests the EH using standard statistical techniques employed by the literature and then Chapters 4 and 5 examine interest rate predictability using both statistical and decision-based methods. That is, in Chapters 4 and 5 we extend this notion of economic significance to forecasting. We do this by considering if the forecasts generated from a theory informed model are economically significant, i.e. if the investor is better off in terms of wealth by assuming returns are predictable and using a theory informed model to forecast. By comparing the forecasts of different models, we are examining if the forecasts improve as we augment the information set used to produce them. This is important, since the quality of the forecast is dependent on the quality and quantity of information used to produce it, Diebold (2004).

The questions addressed in this thesis are:

1. Does the EH hold for this recent sample of UK data?

2. Do we find more support using this dataset than previous UK studies, that test the EH using data over a period when interest rates were comparatively more volatile?

3. Is a statistical or theory informed model best placed to describe the UK money market? Then continuing with the comparison between statistical and theory informed models to the forecasting stage, we consider:

4. Is an atheoretic random walk model or a model that embeds the long-run relationships implied by the EH better at forecasting the term structure? Given that we examine the forecasts in an investment decision making context:

---

[4]For exchange rates see Abhyankar et al (2005) for a discussion. For interest rates, mixed support is reported for the EH as will be discussed in Chapter 2, see also Carriero et al (2006) and Guidolin and Thornton (2008).

5. Is the investor's allocation sensitive to the investment horizon, parameter uncertainty and the assumption made regarding predictability? Given that both statistical and decision-based forecast evaluation criteria are used:

6. Which model performs best?

7. Is this performance sensitive to the evaluation criteria employed?

8. Do these forecasts have an economic value?

These questions are explored using a Vector Autoregressive modelling framework, where density forecasts of the asset returns are generated through simulation methods, as detailed in Chapter 2. In the following section we summarise each of the empirical chapters.

1.2 Summary of Empirical Chapters

Our first empirical chapter, Chapter 3, examines the time series properties of the UK Term Structure over 1997 to 2004, using a range of models for weekly 1-, 3-, 6- and 12-month yields. The range of models embed varying degrees of economic theory. On the one hand we consider the statistical Autoregressive and Vector Autoregressive in Differences models, which allow for no or limited interaction between the yields. On the other hand we consider the theory informed VAR in Transformed Interest Rates and Vector Error Correction models, that embed the long-run relationships between the yields as implied by the EH. The aim of this chapter is, first to test the EH using cointegration analysis and the VAR methodology. Second, to determine if a statistical or theory informed model has greater explanatory power for the UK term structure at the short end. We find support for the EH under both approaches, suggesting that the EH provides a good representation of the UK money market. Further, we find the theory informed models to have greater explanatory power.

17

Chapter 4 moves to forecasting the UK term structure at the short end. Given in Chapter 3 we find support for modelling the term structure using a theory informed model, Chapter 4 forecasts yields using a Multivariate VAR in Transformed Interest Rates (MVART) model, which embeds the cointegration relations implied by the EH. In particular, we compute the optimal portfolio allocation for a buy-and-hold investor with power utility over terminal wealth for two assets, the 1-month and the n-month T-bill, for $n = 3, 6, 12$ over investment horizons of up to 2 years using weekly UK data over 1997 to 2007. We use two models that make opposing assumptions regarding return predictability to forecast returns. That is, the investor uses a random walk with drift model to forecast returns and inform his allocation decisions, if he believes returns are not predictable. However, if he believes returns are predictable he uses the MVART model. We generate density forecasts of the returns from the two alternate models. The aim of this chapter is first to consider if parameter uncertainty and predictability influence how the investor allocates. Second, to see if there is economic value to interest rate predictability, in that does the investor gain in terms of higher wealth from assuming returns are predictable and using a theory informed model to forecast. We find the allocation to be influenced by parameter uncertainty and the assumption made regarding predictability. Further, some evidence of economic value to interest rate predictability is found, suggesting that the investor may gain in terms of wealth from assuming returns are predictable.

In Chapter 5, we use the asset allocation framework of Chapter 4 and extend it by considering a risky asset. That is, we compute the optimal portfolio allocation for a buy-and-hold investor with power utility over terminal wealth, for two assets the 1-month T-bill and the FTSE All-Share Index using weekly UK data over 1997 to 2007. Here we use four models that assume varying degrees of bond and stock return predictability to forecast returns. Such that, if returns are assumed predictable, then key stock and term structure variables are believed to have explanatory power. We

generate density forecasts of the returns from all models. The aim of this chapter is first to examine the effects of predictability and parameter uncertainty on how the investor optimally allocates. Second, to see if there is any economic value to the investor of bond and stock returns being predictable. We find that an investor who assumes returns are predictable allocates differently to one who assumes returns are not predictable. Further, we find evidence of economic value to bond and stock return predictability. Whereby, the investor gains in terms of a higher terminal wealth by assuming returns are predictable and gains further still by modelling both returns together.

1.3 Contributions of Thesis

In this section we detail the empirical contributions this thesis makes to both the term structure and the financial economics forecasting literature.

Chapter 3 presents new support for the Expectations Hypothesis of the term structure of interest rates using recent UK data over 1997 to 2004. Our results are more supportive of the EH than previous UK studies that use data at the short end. The interest rates post-1992, in particular after 1997 have been noticeably less volatile than those observed in earlier decades. Such that, we argue our favourable findings are a result of interest rates being sufficiently volatile for the EH to hold, but not too volatile as to invalidate the EH with a constant term premium. Where the weaker support of earlier UK studies may be due to them using pre-1997 data, when interest rates were significantly more volatile. The reduction in the volatility of interest rates observed after 1992 could be in part due to the changes in the monetary policy regime, with inflation targeting being adopted in 1992 and the Bank of England being granted independence in 1997. Or as appears more likely now, this change in the volatility was due to the stable global economic climate observed until recently, Hall and Henry (2000)

and King (2008)[5].

Hence our findings for the UK imply that the volatility of the interest rates is important for the EH. This result adds to those reported for the US, Germany and Denmark, that suggest favourable evidence for the EH is more likely under some monetary regimes than others, because the regime in operation has an impact on the volatility of the interest rates, Mankiw and Miron (1986), Cuthbertson et al (2000b) and Christiansen et al (2003).

Further, when modelling the term structure more generally using a range of time series models, Chapter 3 finds the theory informed models have greater in-sample explanatory power than the statistical models. The results presented in this chapter demonstrate the importance of economic theory in explaining the term structure.

Chapters 4 and 5 explain the importance of economic theory for forecasting asset returns. The existing literature that examines asset return predictability, as will be reviewed in Chapter 2, primarily focus on stock return predictability, with some recent studies also considering exchange rate predictability. But the investigation of interest rate predictability in the context of asset allocation, with the use of decision-based forecast evaluation methods has largely been neglected by the empirical literature. This is where Chapters 4 and 5 make a contribution.

The innovation of Chapter 4 is that we consider (1) predictability and parameter uncertainty in asset allocation, (2) generate density forecasts to capture the risk as well as the return about the asset and (3) the economic value to the investor of these return forecasts, all in the context of interest rates. In Chapter 5, we extend earlier studies that examine stock return predictability by allowing for predictability in bond returns too, whereas these earlier studies assume the T-bill rate is constant, e.g. Barberis (2000). The innovations of this chapter are that we also model bond returns, further we model

---

[5]The Governor of the Bank of England Mervyn King, describes the steady growth and low inflation enjoyed by the UK since 1997 until the recent downturn as the "nice decade". Nice being an acronym for "non-inflationary consistent expansion".

the two returns separately and jointly, and furthermore evaluate predictability using decision-based methods. This joint modelling framework allows for the possibility of feedbacks between term structure and stock variables. Hence the contribution of this chapter is that we consider (1), (2) and (3) in the context of interest rates and stock returns.

Two key findings emerge from Chapters 4 and 5, first that the investor's allocation is sensitive to the investment horizon, whether he assumes returns are predictable and to parameter uncertainty. Second, we find support for the predictive power of theory informed models of interest rates and stock returns, this support is found to be sensitive to the evaluation criteria used. In that, when the conventional root mean squared errors (RMSE) measure is used our results correspond to those reported by the forecasting literature, that the random walk is difficult to beat. However, support for the theory informed models is found when they are judged in terms of the terminal wealth gained by the investor. That is, an investor gains in terms of wealth by assuming returns are predictable and using a theory informed model to forecast returns, as opposed to assuming they are not predictable and using a random walk model. These findings are new for interest rate predictability, and fall in line with those reported by studies examining stock and exchange rate predictability.

Our findings highlight the importance of using an appropriate forecast evaluation criterion, one that reflects the economic value of the forecasts to the user i.e. a decision-based criterion. Here the user is an investor, he is concerned with maximising his wealth and hence wants the forecast that will achieve this. When the forecasts are indeed judged under a decision-based criterion, support for using a model that embeds economic theory over a statistical model is found.

This thesis tests the EH, models and forecasts the term structure using a theory informed model that embeds the cointegration implied by the EH, uses these forecasts to derive optimal portfolio allocations and assess asset return predictability using statisti-

cal and decision-based methods. This work brings together and contributes to the areas of the term structure, density forecasting, asset allocation and forecast evaluation.

In summary, we find economic theory to have explanatory power for the term structure and the volatility of the interest rates to be important when testing the validity of the EH. Further, this thesis demonstrates the importance of considering the distribution about the predicted returns when making allocation decisions; and the sensitivity of these allocations to forecast uncertainties and assumptions made regarding return predictability. Furthermore, our results highlight the importance of evaluating forecasts in the decision making context for which they are ultimately intended.

The organisation of this thesis is as follows, Chapter 2 provides a review of the empirical literature to which this thesis contributes and summarises the modelling and forecasting techniques used by the empirical chapters. This is followed by the three empirical chapters; Chapter 3 models the UK Term Structure 1997 to 2004, Chapter 4 examines interest rate predictability using decision-based forecast evaluation, Chapter 5 considers the economic value of interest rate and stock predictability and finally Chapter 6 concludes.

Chapter 2

Modelling and Forecasting Interest Rates and Other Asset Returns; An Overview

2.1 Introduction

This thesis is concerned with modelling the UK term structure, testing the Expectations Hypothesis of the term structure, using time series models of interest rates to forecast and assessing the predictability of the term structure in an investment decision making context. The aim of this chapter is to review the theories, modelling techniques and existing literature that are relevant to the empirical chapters that follow. This chapter broadly falls into two parts, the first part examines the term structure of interest rates and the second discusses recent innovations in and the use of forecasting in financial economics.

More specifically, the first part of this chapter discusses the importance of the term structure, the UK term structure and how it has evolved over recent decades, Sections 2.2 to 2.2.1. This is followed by a summary of term structure theories, the methods

by which they are assessed, together with a review of the empirical literature, Sections 2.2.2 to 2.2.4. We then provide a more general discussion of how interest rates can be modelled using a range of time series models, that can all be summarised by a standard Vector Autoregressive modelling framework in Section 2.3.

In the second part of this chapter we turn our attention from modelling the term structure to forecasting. In particular, we describe how to generate the conventionally used point forecasts in Section 2.4.1 and the not so frequently used density forecasts, which provide an entire probability distribution of all possible future outcomes in Section 2.4.2. We then review the literature concerned with forecasting interest rates in Section 2.4.3 and lastly Section 2.4.4 considers how forecasts are evaluated, both statistically and using decision-based methods.

2.2 Term Structure of Interest Rates: Theories and Studies

The term structure of interest rates refers to the relationship that exists between the interest payable on bonds of varying terms to maturity. The yield curve is an illustration of the interest rates of these bonds plotted against their terms.

The shape of the yield curve reflects the markets expectations of future interest rates, it can be upward sloping, flat or downward sloping. An upward sloping yield curve reflects the expectation of future economic growth, together with an increase in the risk of inflation rising. With this risk, is the expectation of short rates increasing as the monetary authorities try to control inflation. Hence investors demand a premium for longer maturities because of the uncertainty about future inflation and the implications for future income, this resulting in an upward curve. Generally, a positively sloped curve is observed with longer rates being higher than short term rates, where flat or inverted curves are less frequently observed.

Term structure theories provide an explanation of how the yield changes with maturity, these include the Expectations Hypothesis (EH), the Market Segmentation Hypothesis (MSH) and the Liquidity Preference Hypothesis (LPH). The main theory is the EH, which states that the long term rate is a weighted average of expected future short term rates. In its pure form the EH assumes that the term premium is zero, which implies that the investor is indifferent between holding one long term bond and the equivalent in short term bonds. This pure form is applicable to the instruments of the money market, because they are close in maturity. The LPH is considered an offshoot of the pure EH, in that the long rate is composed of future short rates, but here a premium is paid for holding the long bond. Under the MSH/Preferred Habitat Hypothesis investors need a premium to entice them out of their preferred maturity habitat. In this overview we focus our discussion on the EH, which given the vast amount of empirical research that has been conducted on it, is the leading hypothesis of the term structure. In contrast, very little attention is paid to the MSH and the LPH by the literature, so here we provide only a brief description of each.

The term structure contains information about the market's anticipation of future events. Hence by having an explanation of the term structure provides a means to extract this information and to predict how changes in the underlying variables will affect the yield curve, Cox et al (1985). The term structure and the shape of the yield curve is important to market players including policy makers concerned with the transmission of monetary policy, practitioners who want to know how to price instruments and investors who want to know how to allocate their portfolio.

For policy makers the relationship between long and short rates may be of interest if they want to influence real economic activity. If real activity is related to the long rate and the central bank can manipulate the short rates, having knowledge of the relationship between long and short rates will inform the government on how it can influence real activity. For instance, if the EH accurately describes the term structure,

then the long rate and hence real activity can be influenced through manipulating the short rates in the market. Further, being able to satisfactorily describe and model the term structure is key for forecasting interest rates, where these forecasts can be used to inform for instance asset allocation decisions as examined here.

Much attention has been paid to the term structure by the academic literature because of the information it embeds and the importance of this information for market players. In the sections that follow we begin with a look at the UK term structure and then discuss term structure theories focusing on the EH. This is followed by a review of the empirical literature, paying particular attention to UK studies that test the EH.

2.2.1 UK Term Structure and History

Here we detail how the UK monetary policy regime changed in the 1990s, the role of the Monetary Policy Committee, how these reforms have served to increase both transparency and credibility of monetary policy in the UK and the potential implications of these changes for interest rates.

Inflation targeting was introduced in the UK in 1992, with the Chancellor responsible for setting interest rates. This change in the monetary policy regime followed the UK's exit from the Exchange Rate Mechanism. In 1997 the new Labour government granted the Bank of England (BoE, the Bank) operational independence. This giving the BoE the freedom to set the monetary instrument, in order to achieve the inflation target set for the Bank by the government, Tootell (2002). The current target is 2%. Both the adoption of an inflation target and the subsequent independence of the central bank have served to increase transparency[1] of monetary policy in the UK and brought macroeconomic stability, Bean (2003) and Mariscal and Howells (2007).

---

[1] This increased transparency has come in the form of the objective of monetary policy being clear, the MPC publishing the minutes of its meetings. And further, the Bank in its Inflation Report presenting its forecasts of inflation, together with an evaluation of the current economic climate, Chadha and Nolan (2001).

The Monetary Policy Committee (MPC) of the BoE is responsible for setting interest rates. As detailed in the Bank of England Act 1998 (cited in Kim et al (2008)) the objectives of monetary policy is to maintain price stability and subject to maintaining stability, support the governments economic policy. Economic policy includes the governments employment and growth objectives. The Bank provides a density forecast of inflation up to 2 years ahead in its quarterly Inflation Report, known as the 'fan chart' as illustrated in Figure (2-1)[2]. "The fan chart portrays a probability distribution that approximates to the MPC's subjective assessment of inflationary pressures evolving through time, based on a central view and the risks surrounding it," this assessment is based on judgment and statistics[3], Britton et al (1998, pp. 31)[4].
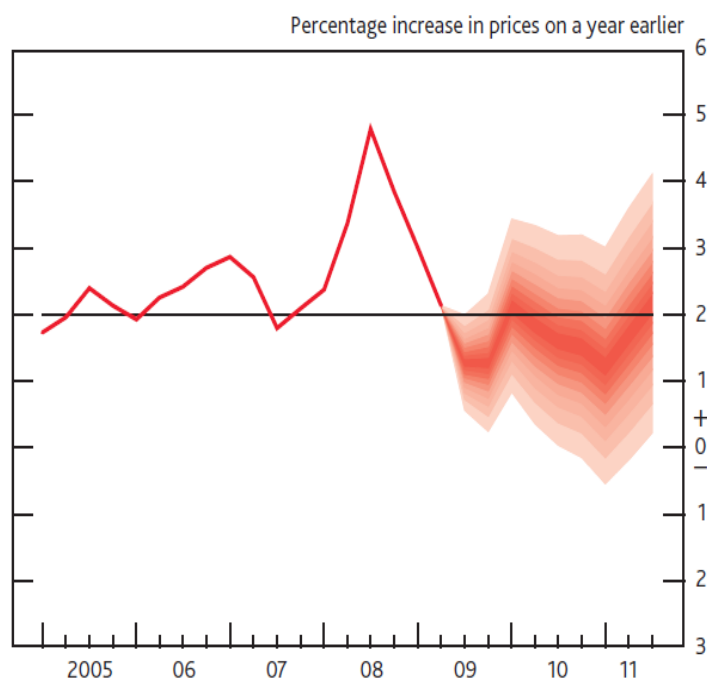
Figure 2-1: Bank of England's Inflation Fan Chart August 2009

---

[2] CPI Inflation Projection based on constant nominal interest rates at 0.5%. Chart obtained from the Bank of England's Inflation Report August 2009.

[3] These forecasts assume UK short-term rates are unchanged during the forecast period, i.e. does not consider uncertainty about interest rates.

[4] See for more detail on the fan chart and how it is constructed.

The darkest region of the chart depicts the highest probability path, and contains the point forecast. There is a 10% probability that inflation will fall into this central band at any given horizon. The lighter regions correspond to the paths that are less likely. The width of the fan at a particular forecast horizon, gives with a 90% probability, that inflation will fall into that range of possible outcomes. The fan shape of the chart reflects that, the further into the future the projection is made, the greater the uncertainty about that projection. The greater the uncertainty about the forecast, the wider the fan. The uncertainty about the central projection may not necessarily be symmetric[5], Vickers (1998). Britton et al (1998) argue that by presenting their forecasts of inflation in this way, helps communicate that monetary policy decisions are being made in an uncertain world and that the MPC does not know with certainty what inflation will be in the future.

If inflation fails to meet its target by more than one percent either side, the Governor is required to write to the Chancellor explaining why, what action will be taken, how long this action is expected to take effect and how this action meets the monetary policy objectives of the government, Kim et al (2008). King (1997, cited in Kim et al (2008)) notes that by writing this open letter the MPC must explain in public how it intends to react to large shocks.

The Bank's independence "generated an immediate credibility gain as long-term inflation expectations fell" Bean (2003, pp. 482). Bean (2003, 2004) argues that it has been a combination of structural reforms to labour and product markets, first introduced by Thatcher's government and later consolidated by both parties in the 1990s, together with the adoption of an inflation target that have kept inflation low and stable[6], and unemployment low.

From this time series plot of the official bank rate over 1975 to 2007, Figure (2-2).

---

[5]This suggesting that it is more likely that the forecast error will be in one direction than the other, Britton et al (1998).

[6]Prices have remained stable over the period 1992 to 2007, although recently prices have become more volatile, Hammond (2009), see for a summary of the UK experience of inflation targeting.
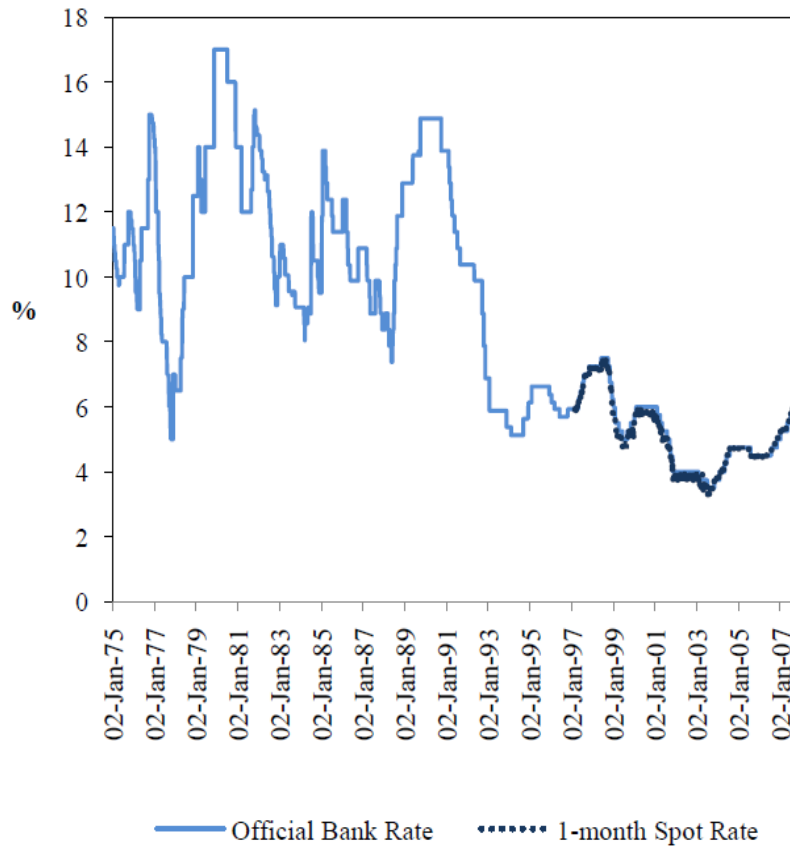
Figure 2-2: UK Official Bank Rate and Spot Rate 1975 to 2007

It is apparent that post-1992 the interest rates are both lower in level[7] and significantly less volatile than the rates observed during the ERM years and further back. These changes coincide with the governments inflation targeting regime and the reforms that followed. By adopting an anti-inflation stance, granting the central bank independence and the MPC making policy decisions in a transparent way, as described above, has led to gains in credibility. These gains may explain the relatively less volatile interest rates.

Although the changes to how monetary policy is conducted has been credited for

---

[7]The current rate of 0.5% is the lowest observed since 1694, Hammond (2009).

the low levels of inflation, unemployment and interest rates enjoyed over the last 15 years or so in the UK, Hall and Henry (2000) argue that it is difficult to disentangle how much of this macroeconomic success is down to policy reforms and how much is due to stable global economic conditions. Further, the Governor of the Bank Mervyn King describes the steady growth and low inflation observed in the UK since 1997 until the recent downturn as the "nice decade" [8]. This suggests that the reduction in the volatility of interest rates may partly be explained by the anti-inflation stance taken by the Bank of England, but is largely due to the benevolent economic climate of the nice decade.

The data at the short end of the UK yield curve does not stretch as far back as the official rate data, but from Figure (2-2) it can be seen for the data available, the market determined 1-month spot rate follows the official rate very closely. Such that the official rate can be used as an indication of how the spot rates have evolved over the last few decades. These changes in the interest rates are relevant to us because we use data over the period 1997 to 2007 in the empirical chapters that follow.

More precisely, the data employed in this thesis[9] is official BoE data on the Government liability nominal yield curve at the short end. These zero coupon yields (spot rates) are calculated using gilt prices and General Collateral (GC) repo rates. A gilt is a UK government coupon paying bond and is considered to be a safe investment, since the UK government is unlikely to default. General Collateral Sale and Repurchase agreements (GC repo) refers to the sale and repurchase of gilts, "gilt repo", this is where gilts are used as collateral for short-term borrowing. The GC repo rate is that for the repurchase of the gilt, which should be close to the true risk-free rate. As detailed in the Bank's data notes[10], these repo contracts are actively traded for maturities

---

[8]Nice being an acronym for "non-inflationary consistent expansion".

[9]More detail of the actual data used is provided in the empirical chapter itself and in the Data Appendix.

[10]BoE, "Yield curve terminology and concepts, Notes on the Bank of England Yield Curves": http://www.bankofengland.co.uk/statistics/yieldcurve/index.htm

of up to one year and the repo rates are very similar to the yields on conventional gilts of comparable maturity.

The GC repo agreements market began in January 1996, after March 1997 the BoE began using gilt repos to conduct Open Market Operations, this led to GC repos becoming "a more satisfactory indicator of expectations of future interest rates". Prior to March 1997 there was limited information available at the short-end of the government curve, because the only short-term assets available were Treasury bills, "which do not have an active secondary market and the prices of which are affected by banks' liquidity requirements", BoE notes[11]. This limitation of short-end data has resulted in, as we will see, early studies examining the longer end of the UK term structure.

As can be seen from Figure (2-2) interest rates post-1992, in particular after 1997 are noticeably less volatile than the rates observed prior to the adoption of the inflation target and central bank independence. And it is data after these reforms, over 1997 to 2004, that we use to test the Expectations Hypothesis. Whereas, previous UK studies that examine the EH at the short end use pre-1997 data.

2.2.2   Expectations Hypothesis

The leading theory of the term structure is the Expectations Hypothesis (EH). There are many competing yet related hypotheses referred to as the EH, Cox et al (1985). However, Campbell (1986) demonstrates that they can all be approximated by a linear form of the EH. The form we refer to and describe here is that of Campbell and Shiller (1991). We will discuss how the fundamental equations of the EH are derived in terms of spot yields, as presented in Patterson (2000) and Cuthbertson and Nitzsche (2004).

We largely adopt the notation used by the literature, where $r_t^{(n)}$ is the yield to maturity of a n-period bond at time t, which matures in $t+n$. This return is continuously

---

[11]BoE FAQs: http://www.bankofengland.co.uk/statistics/yieldcurve/faq.htm#q12

compounded and expressed as an annual rate. $r_{t+i}^{(n)}$ is the yield to maturity of a n-period bond at time $t + i$, that matures in n periods time at $t + i + n$. $s_t^{(n,m)} = r_t^{(n)} - r_t^{(m)}$ describes the yield spread at time t between a long n-period and a short m-period bond, and $k = n/m$ gives the number of "short" m periods within a "long" n period.

Consider two strategies available to an investor who wants to invest £I for n periods at t: Strategy A invest in a long n-period bond which pays a known return of $R_t^{(n)}$, or Strategy B a 'rolling' investment in a sequence of k m-period bonds. Such that under Strategy B, at t a m-period bond is purchased with actual return $R_t^{(m)}$, together with a 'basket' of contracts in the forward market[12]. Contracts are made in t for the following $t + (k-1)m$ periods, giving a total of k contracts with rates $R_t^{(m)}, F_{t+m}^{(m)}, F_{t+2m}^{(m)}, \ldots, F_{t+(k-1)m}^{(m)}$. Under this second strategy the amount invested plus the interest earned is reinvested into the next m-period bond. Where $F_{t+j}^{(m)}$ is the m-period bond forward rate and $R_{t+j}^{(m)}$ is the actual rate, it is not necessarily true that the contracted rate will equal the prevailing actual rate. The value of the investment V ignoring transactions costs, at the end of the investment period under each strategy is

$$V_A = £I \left[1 + R_t^{(n)}\right]^n \tag{2.1}$$

$$V_B = £I \left[1 + R_t^{(m)}\right]^m \cdot \left[1 + F_{t+m}^{(m)}\right]^m \ldots \left[1 + F_{t+(k-1)m}^{(m)}\right]^m \tag{2.2}$$

where the Efficient Market Hypothesis implies that the expected return under both strategies should be equal

$$\left[1 + R_t^{(n)}\right]^n = \left[1 + R_t^{(m)}\right]^m \cdot \left[1 + F_{t+m}^{(m)}\right]^m \ldots \left[1 + F_{t+(k-1)m}^{(m)}\right]^m \tag{2.3}$$

---

[12] In this contract the investor agrees to buy in $t + m$ a m-period bond which matures at the end of m periods with fixed rate $F_{t+m}^{(m)}$. This is the m-period forward rate agreed at t for $t + m$. Further, he agrees to purchase another m-period bond in $t + 2m$ that matures at the end of $t + 2m$, with the forward rate $F_{t+2m}^{(m)}$ fixed in the contract for this bond. In total $t + (k-1)m$ such purchases are made.

taking logs, lower case referring to the log of the variable

$$n \ln \left[1 + R_t^{(n)}\right] = m \left\{ \ln \left[1 + R_t^{(m)}\right] + \ln \left[1 + F_{t+m}^{(m)}\right] + \dots + \ln \left[1 + F_{t+(k-1)m}^{(m)}\right] \right\} \quad (2.4)$$

if rates are continuously compounded then

$$n r_t^{(n)} = m \left\{ r_t^{(m)} + f_{t+m}^{(m)} + f_{t+2m}^{(m)} + \dots + f_{t+(k-1)m}^{(m)} \right\}$$

$$r_t^{(n)} = \frac{1}{k} \left[\sum_{i=1}^{k} f_{t+(i-1)m}^{(m)} \right] \quad (2.5)$$

From equation (2.5) the actual n-period bond yield to maturity is given by the average of the actual m-period yield to maturity and the forward m-period rates. With an efficient market there is no expectation of a profit, as any profit will be arbitraged away. The forward rate which describes the rate at which a future loan is made, can be written in terms of the expected spot rate

$$f_{t+j}^{(m)} = E_t\{r_{t+j}^{(m)}\} + \tau_{t+j}^{(m)} \quad (2.6)$$

where the m-period forward rate $f_{t+j}^{(m)}$ for $t+j$ is given by the actual anticipated rate $E_t(r_{t+j}^{(m)})$ for $t+j$ and the term premium $\tau_{t+j}^{(m)}$. A term premium may arise because future interest rates are uncertain, hence there is a risk associated with the fixing of a forward rate for $t+j$ in t. Under the Pure Expectations Hypothesis (PEH)[13] this term

---

[13]The literature sometimes refers to this as the EH under risk neutrality, because the investor is indifferent between investing in long term bonds or short term ones so does not require a term premium.

premium is zero and constant under the EH. From (2.5) and (2.6)

$$r_t^{(n)} = \frac{1}{k}\left[\sum_{i=1}^{k} E_t\left\{r_{t+(i-1)m}^{(m)}\right\}\right] + c^{(n,m)} \tag{2.7}$$

with the term premium $c^{(n,m)} = \frac{1}{k}\left[\sum_{i=1}^{k}\tau_{t+j}^{(m)}\right]$. Equation (2.7) is known as the fundamental equation of the EH, this describes the long rate as a weighted average of current and future expected short rates and a term premium, the weights are equal and sum to one. The value of the investment in the long bond is known because it is held to maturity, whereas the value of the investment in a sequence of short bonds is subject to uncertainty because the future m-period rates are unknown. If short rates are anticipated to increase then this leads to an increase in the long rate. Hurn et al (1995) state that the term premium summarises the effects of factors other than the expectations of future short rates, arguing that a term premium may arise due to factors like liquidity preference and hedging behaviour[14].

From equation (2.7) the second fundamental equation of the EH can be deduced

$$s_t^{(n,m)} = \sum_{i=1}^{k-1}\left(\frac{k-i}{k}\right) E_t\left\{\Delta r_{t+im}^{(m)}\right\} + c^{(n,m)} \tag{2.8}$$

where the spread is a function of expected changes in the short rate over the life of the long bond and a term premium. The m-period change is $\Delta r_{t+im}^{(m)} = r_{t+im}^{(m)} - r_{t+(i-1)m}^{(m)}$. A spread between the long and short rate arises if the expected changes in the short rate over the next n periods is non-zero, or if there is a term premium. If no change in the short rate is expected then the long rate should only differ from the short by the

---

[14]That is, in the long-term securities market investors are exposed to price volatility and thus a capital risk, in which case a liquidity premium maybe paid. Further, "...income risk might discourage investors with long-term horizons from holding shorter dated assets, suggesting hedging behaviour could give rise to term premium..." Hurn et al (1995, pp. 420).

term premium. If the short rate is expected to increase, then the long rate and hence the spread will also increase, as implied by both fundamental equations of the EH, equations (2.7) and (2.8). The assumption of expectations being formed rationally is not necessary for (2.8) to hold, as it only requires that the assumed expectations process generates errors that are stationary.

A "perfect foresight spread" (pfs) can be identified from (2.8)

$$pfs_t^{(n,m)} = \sum_{i=1}^{k-1} \left(\frac{k-i}{k} \right) \Delta r_{t+im}^{(m)} \tag{2.9}$$

Campbell and Shiller (1991) describe this as the spread obtained if we had "perfect foresight" about future interest rates, i.e. if we are able to predict future rates correctly and do not need to form expectations. Substituting (2.9) into (2.8)

$$s_t^{(n,m)} = E_t\{pfs_t^{(n,m)}\} + c^{(n,m)} \tag{2.10}$$

If expectations are formed rationally then

$$r_{t+im}^{(m)} = E_t\{r_{t+im}^{(m)}\} + \varepsilon_{t+im}^{(m)} \qquad for \ i > 0 \tag{2.11}$$

according to the Rational Expectations Hypothesis (REH) $E_t\{r_{t+im}^{(m)}\}$ is an optimal predictor of $r_{t+im}^{(m)}$. Where the expectational error $\varepsilon_{t+im}^{(m)}$ has an expected value zero, can not be forecasted with information available at t and $r_t^{(m)} = E_t\{r_t^{(m)}\}$ for $i = 0$.

Using equation $(2.11)^{15}$

$$E_t\{\Delta r_{t+im}^m\} = \Delta r_{t+im}^m - \Delta \varepsilon_{t+im}^m \qquad (2.12)$$

substituting (2.12) into (2.8)

$$s_t^{(n,m)} = \sum_{i=1}^{k-1} \left(\frac{k-i}{k}\right) \Delta r_{t+im}^{(m)} + \varphi_t^{(m)} + c^{(n,m)} \qquad (2.13)$$

From (2.13), the spread is given by observable changes in the m-period rates from $t + m$ to $t + (k - 1)m$, a weighted average of expectations errors $\varphi_t^{(m)}$ which are assumed to be stationary and the term premium $c^{(n,m)}$. The term premium is assumed to be stationary, its structure and form varies with the model of the term structure used. The Pure Expectations Hypothesis assumes it takes a value of zero, whereas the Expectations Hypothesis assumes it is constant but varies across maturities. The Liquidity Preference Hypothesis implies that the term premium is constant for a given maturity, but increases with the term to maturity and the Time Varying Hypothesis proposes that is varies with time as well as maturity.

Assessing the Expectations Hypothesis

There are several ways to assess the validity of the EH, these include the cointegration method, the two single equation tests that test the predictive power of the spread for future changes in interest rates, and the VAR methodology. Each of these will now be briefly discussed in turn.

---

[15]Rearranging equation (2.11) gives (a) $E_t\{r_{t+im}^{(m)}\} = r_{t+im}^{(m)} - \varepsilon_{t+im}^{(m)}$, lag by 1 gives (b) $E_t\{r_{t+(i-1)m}^{(m)}\} = r_{t+(i-1)m}^{(m)} - \varepsilon_{t+(i-1)m}^{(m)}$. Equation (a) minus (b) gives $E_t\{r_{t+im}^{(m)}\} - E_t\{r_{t+(i-1)m}^{(m)}\} = r_{t+im}^{(m)} - r_{t+(i-1)m}^{(m)} - \left[\varepsilon_{t+im}^{(m)} - \varepsilon_{t+(i-1)m}^m\right]$ which is equivalent to equation (2.12).

Stationary spreads and cointegration amongst the yields   Lopes and Monteiro (2007) point out that nominal interest rates are bounded below by zero, and from a theoretical stance it is difficult to justify interest rates as being non-stationary.   But given that they are highly persistent and slowly mean reverting, this has led to many researchers treating rates as having a unit root and using cointegration analysis.

If the yields contain a stochastic trend and the term premium is stationary, then the RHS of equation (2.13) is a linear combination of stationary variables.   Hence from equation (2.13), a theoretical implication of the EH is that the spread must be a stationary process.   Further implications of the EH are:

1. Assuming the yields share a common stochastic trend, then there should exist $(q-1)$ cointegrating vectors in a set of q non-stationary yields, as implied by stationary bivariate spreads.

2. Each of the n-month yields are cointegrated with the m-month yield, such that the cointegrating vector is of the form $\left(1, -1, c^{(n,m)}\right)'$, with the term premium free from restriction.

3. The PEH can be tested through the imposition of the restrictions that the premia are zero, such that the cointegrating vector is now $(1, -1, 0)'$.

These three implications can be tested using the Johansen maximum likelihood estimation procedure, Johansen (1988, 1991) and Johansen and Juselius (1990).   The next two methods assess the EH within a single equation framework. Given that EH implies that the spread is an optimal forecast of changes in the future interest rates, they seek to ascertain if the spread has predictive power for future short rates and changes in the long rate.

Spread predicting future changes in the short rate   This method is concerned with seeing if the actual spread, $s_t^{(n,m)}$ is able to forecast the $pfs_t^{(n,m)}$.   Some rearrange-

ment of equation (2.13) is first required. Substituting (2.9) into (2.13) gives

$$s_t^{(n,m)} = pfs_t^{(n,m)} + \varphi_t^{(m)} + c^{(n,m)} \qquad (2.14)$$

assuming the term premia is constant i.e. $c^{(n,m)} = \alpha_0$, then

$$s_t^{(n,m)} = \alpha_0 + \alpha_1 pfs_t^{(n,m)} + \varphi_t^{(m)} \qquad (2.15)$$

or

$$pfs_t^{(n,m)} = \beta_0 + \beta_1 s_t^{(n,m)} + \varphi_t^{(m)*} \qquad (2.16)$$

where the EH + REH suggests that $\alpha_1 = \beta_1 = 1$ [16]. It maybe that the term premium, $c^{(n,m)}$ is not constant, but as Patterson (2000) notes as long as it is stationary and cannot be modelled e.g. as a function of other variables, then the bivariate regression interpretation of equations (2.15) and (2.16) holds. The literature indicates a preference for estimating (2.16) and testing the null hypothesis that $\beta_1 = 1$ against the alternative that $\beta_1 \neq 1$, as opposed to the equivalent test of $\alpha_1 = 1$ in equation (2.15). This is because in equation (2.15) $\varphi_t^{(m)}$ and $pfs_t^{(n,m)}$ are positively correlated[17] which results in the OLS estimator of α_0 being inconsistent. Whereas in (2.16) the spread term dated at time t has no correlation with the future expectational errors, therefore OLS estimation yields consistent estimates.

---

[16]For example, the 12- and 6-mth spread at t should predict the sum of the future changes in the 6-month rate over the coming 6 months, Rossi (1996).

[17]From equation (2.11) a shock to the expectational errors $\varepsilon_{t+im}^{(m)}$ is translated into a shock to $r_{t+im}^{(m)}$. Since $pfs_t^{(n,m)}$ is a function of $\Delta r_{t+im}^{(m)}$, and $\varphi_t^{(m)}$ is a function of $\varepsilon_{t+im}^{(m)}$, there is a positive correlation between them.

Spread predicting changes in the long rate   Campbell and Shiller (1991, CS) suggest that another implication of equation (2.7) is that the spread is able to predict changes in the long rate. Equation (2.7) states that the yield on a n-period bond at time t is given by a weighted sum of the expected m-period bond rates over n periods, with the addition of a predictable excess returns term.   From Patterson $(2000)^{18}$ equation (2.7), omitting constants, can be manipulated to give

$$\left(\frac{m}{n-m}\right) s_t^{(n,m)} = E_t\left\{r_{t+m}^{(n-m)}\right\} - r_t^{(n)} \tag{2.17}$$

the LHS of (2.17) is a multiple of the spread, which is given by an expected m period change in the long n-period rate. If expectations are formed rationally then from (2.11)

$$E_t\left\{r_{t+m}^{(n-m)}\right\} = r_{t+m}^{(n-m)} - \varepsilon_{t+m}^{(n-m)}$$

substituting this into (2.17) and rearranging

$$r_{t+m}^{(n-m)} - r_t^{(n)} = \left(\frac{m}{n-m}\right) s_t^{(n,m)} + \varepsilon_{t+m}^{(m)} \tag{2.18}$$

this suggests that the m period change in the long n-period rate should equal a multiple of the spread with the addition of a stationary error term. Where if the n-period yield is expected to rise over the coming m periods, then we would expect the yield on the

---

[18] In one periods time there are $n-1$ periods left on the n-period bond, thus the yield on a (n-1)-period bond is given by $r_{t+1}^{(n-1)}$.   In m periods time there are $n-m$ periods left on the n-period bond, hence the yield on a (n-m)-period bond is $r_{t+m}^{(n-m)}$. The m period change in the n period rate is $r_{t+m}^{(n-m)} - r_t^{(n)} = m$ period change in the long rate, where the expected m period change in the long rate is given by $E_t\left\{r_{t+m}^{(n-m)}\right\} - r_t^{(n)}$.

n-period bond to be higher than that on a m-period bond[19]. Equation (2.18) can be estimated to test the EH+REH model through the regression

$$r_{t+m}^{(n-m)} - r_t^{(n)} = \beta_0 + \beta_1 \left(\frac{m}{n-m} \right) s_t^{(n,m)} + \varepsilon_{t+m}^{(m)} \tag{2.19}$$

where the EH + REH suggests[20] that $\beta_1 = 1$.

VAR Methodology The Campbell-Shiller VAR methodology tests the EH by statistically and economically assessing the deviation of the actual observed yield spread from the theoretical spread as implied by the EH. They suggest some alternative metrics to assess the economic significance of the EH. A more detailed discussion of the VAR methodology is provided in Chapter 3 [21], but a brief summary is given below.

If $\mathbf{z}_t = \left(s_t^{(n,m)}, \Delta r_t^{(m)} \right)'$ is stationary, then there exists a bivariate Wold representation which can be approximated by a vector autoregression of order p, in companion form

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \boldsymbol{\epsilon}_t \tag{2.20}$$

if $s_t^{(n,m)} = \mathbf{e1}'\mathbf{z}_t$ and $\Delta r_t^{(m)} = \mathbf{e2}'\mathbf{z}_t$, see Campbell and Shiller (1987, 1991), then from equation (2.8) the following VAR non-linear cross-equation restrictions are obtained

$$\mathbf{e1}' - \mathbf{e2}'\mathbf{A}\left[I - (m/n)\left(I - \mathbf{A}^n \right)\left(I - \mathbf{A}^m \right)^{-1} \right]\left(I - \mathbf{A} \right)^{-1} = 0 \tag{2.21}$$

---

[19] For example, the 12- and 6-month spread at t, should predict the gap between the 6-month rate 6 months ahead and the 12-month rate at t, i.e. the 6 month yield change on the 12-month bond, Rossi (1996).

[20] The value of β_1 in both of these regressions (2.16) and (2.19) indicates by how much the future spot rates change for a given value of the spread at t. Where the null that $\beta_1 = 1$ suggests a "one-to-one relationship between the current spread and changes in future spot rates (unbiasedness)". The null that $\beta_1 = 0$ "indicates whether future spot rates are at least related to the current spread (information content)", Rossi (1996, pp. 12).

[21] See also Campbell and Shiller (1987, 1991) and Cuthbertson et al (1996 and 2000a).

where the theoretical spread $s_t^{(n,m)*}$ as implied by the EH is

$$s_t^{(n,m)*} = \mathbf{e2'A} \left[I - (m/n)\left(I - \mathbf{A}^n\right)\left(I - \mathbf{A}^m\right)^{-1} \right] \left(I - \mathbf{A}\right)^{-1} \mathbf{z}_t \qquad (2.22)$$

This theoretical spread is given by the weighted sum of the optimal forecasts of the changes in the short rates. If the EH holds, the theoretical spread should equal the actual, this hypothesis that $s_t^{(n,m)*} = s_t^{(n,m)}$ can be formally tested by imposing and testing the VAR parameter restrictions in equation (2.21). However, CS note that even small deviations from this null may lead to a rejection of the EH under these formal tests of the VAR restrictions. But Campbell and Shiller (1987) argue that this does not imply that the EH is economically insignificant. In that, the EH may explain the majority of the variation in the actual spread even though it is statistically rejected. They hence propose some alternate metrics[22] to measure the degree of comovement between the actual and theoretical spread, to ascertain the extent to which the predictions made by the EH are close to actual observations. These include time series plots, a measure of the correlation between the actual and theoretical spread $Corr\left(s_t^{(n,m)*}, s_t^{(n,m)} \right)$ and the ratio of their standard deviations[23] $\sigma\left(s_t^{(n,m)*} \right) / \sigma\left(s_t^{(n,m)} \right)$. Both the correlation coefficient and the ratio should equal unity under the EH, if the ratio is less than unity then the actual spread is more volatile than that predicted by the EH.

2.2.3 Other Term Structure Theories

We have thus far discussed how the fundamental equations of the EH are derived, followed by the methods by which the EH can be assessed. This section discusses

---

[22]As stated in Cuthbertson et al (1996) these alternate metrics test the EH+REH under weakly rational expectations.

[23]In the studies discussed below some authors compute the ratio of the variances, we state when this is the case. Further, some compute the ratio of $\sigma\left(s_t^{(n,m)} \right) / \sigma\left(s_t^{(n,m)*} \right)$, in this case we report the inverse of these ratios so that the results are comparable.

two alternative views of the term structure, the Liquidity Preference Hypothesis (LPH) and the Market Segmentation Hypothesis (MSH). Since these views have not been the subject of much empirical investigation their discussion is kept brief.

Liquidity Preference Hypothesis

As noted in Cox et al (1985) the LPH put forward by Hicks (1946) emphasises the "risk preferences of market participants". Under the LPH there is a preference for short term bonds, as these are less exposed to interest rate risks. Such that, a term or liquidity premium is offered for holding the longer term asset. This premium is constant for a particular maturity n, but increases with n. This reflects the need for the investor to be compensated more as the term to maturity increases.

Market Segmentation Hypothesis

The MSH proposed by Culbertson (1957, cited in Cox at al (1985)) provides an alternative explanation of the term premium. Here investors have strong preferences for particular maturities, choosing to trade only in their 'preferred habitat'. This gives rise to separate markets for bonds of different maturities, each with their own demand, supply and prevailing price. These prices are not influenced by the prices prevailing in other markets, if participants are reluctant to trade outside their preferred habitat[24]. In this case any excess returns existing in that market will not be arbitraged away, this giving rise to non-zero term premiums. Empirical examinations of the MSH are rare, particularly using UK data, see Taylor (1992).

Both the LPH and the MSH discuss a preference for a maturity. The MSH is more restrictive, assuming that the investors stick to their preferred habitat. Whereas under the LPH, investors generally prefer short-term bonds and require compensation for investing in longer term bonds.

---

[24]This theory assumes that bonds of close maturity are not close substitutes. Thus market conditions in one will not affect the other. Cox et al (1985) argue that this is a strong assumption to make.

2.2.4 Empirical Studies of the Term Structure

In this section we review the empirical findings of the term structure literature, with a focus on studies of the EH. Specifically, we consider early studies of the term structure, UK studies of the EH, other findings for the EH, stylised facts and why the EH may be rejected and lastly, other studies and recent developments.

Early Studies of the Term Structure

The origins of the EH dates back to the early works of Fisher (1930), Lutz (1940) and Hicks (1946). The empirical literature has been concerned with examining if future expected short rates are fundamental to determining the current long rate. Various approaches have been taken to test the EH, including Campbell and Shiller (1991) who test if the spread has predictive power for changes in future spot rates. And studies by Fama (1984, 1990), Fama and Bliss (1987), Mishkin (1988) and Rossi (1996) who examine if forward rates have predictive power for future spot rates.

As is evident from Froot's opening statement[25] "If the attractiveness of an economic hypothesis is measured by the number of papers which statistically reject it, the expectations theory of the term structure is a knockout.", the early empirical literature on the term structure, conducted primarily using US data, rejects the EH. These studies test if the spread can predict the direction of future short rates include Fama (1984) who uses monthly US T-bills over 1959 to 1982, Mankiw and Summers (1984) using quarterly data 1963 to 1983, Mankiw (1986) who considers data on Canada, UK and Germany too; Mankiw and Miron (1986) who argue that it is more likely support for the EH will be found under a policy of monetary targeting than one of interest rate smoothing, Fama and Bliss (1987) and Campbell and Shiller (1987).

However, Fama (1984), Mankiw and Miron (1986), Mishkin (1988) do find evidence of the spread having predictive power for future interest rates. Further, Fama and Bliss

---

[25]Froot (1989, pp. 283).

(1987) find that this power, although low, increases with the forecast horizon. They attribute this increasing predictive power to the mean-reverting behaviour of interest rates, this implies that rates are easier to forecast in the longer run than the short.

Campbell and Shiller (1987) were the first to apply cointegration techniques to the term structure. In this key paper they propose an informal way to measure the "fit" of the model, suggesting if the EH is true then the theoretical and actual spread should be equal. They recommend the use of time series plots, correlation coefficients and ratio of the standard deviations to gauge this fit. Although they statistically reject the EH, given the degree of comovement between actual and theoretical spreads, they conclude these deviations are transitory and not economically significant.

In another key paper Campbell and Shiller (1991) consider US data for maturities between one month and 10 years[26] over 1952 to 1987. For the regression of the spread predicting future long rates all coefficients are found to be of the wrong sign and significantly different from one. In the regression of the theoretical spread on the actual, the coefficients are significantly different from one at the short end, but not the long. So they find a paradox[27] in that the slope incorrectly forecasts the direction of the short term change in the long rate, but correctly forecasts the direction of the long term change in the short rate. Under the VAR methodology the correlation is nearly always positive and frequently quite high, and the standard deviation ratio takes a value around one half indicating the actual spread to be excessively volatile compared to what the EH predicts. CS suggest that this paradox might be explained by an overreaction model of the spread. Explaining that the "long rate differs from the short rates in the direction implied by the" EH, but this long and short rate spread is larger than "can be justified by rational expectations of future short rate changes". Or put another way, the long rate underreacts to short rates, where the long rate does not

---

[26]1- , 2-, 3-, 4-, 6- and 9-months and 1-, 2-, 3-, 4-, 5- and 10-years. They define the short end of the term structure as less than 1 year and long end as more than 1 year.

[27]This paradox is also referred to as the puzzle by the literature.

react to the current short by as much as the EH predicts.

Shea (1992) and Hall, Anderson and Granger (1992), like CS (1987) also use coin-tegration. At the time these papers were published the application of Johansen's cointegration techniques was novel. They find evidence of cointegration amongst the yields, which they suggest implies that there is single non-stationary common factor that underlies the time series behaviour of each yield and that risk premia are station-ary. Further, during the periods in their sample when the Fed targeted short rates, the cointegrating vector was defined by the yield spread. This finding, in support of Mankiw and Miron (1986), suggests that the success of the EH may be sensitive to the monetary policy regime in operation. Additionally, Lanne (1999) finds the persistence of spreads differs between when there was a regime of interest rate targeting in the US 1952 to 1979 and that of the remaining sample 1979 to 1991. This change in persistence during the considered sample being a potential reason for rejections of the EH in the US.

UK Studies of the EH

Now considering studies conducted using UK data, we begin with earlier studies that use data at the long end, followed by those who examine data at the short end. MacDonald and Speight (1988) use quarterly data on 3-month, 5-, 10- and 20-year government bonds during 1963 to 1987, to find cointegration amongst the yields, spreads Granger cause short run changes in the T-bill rate and vice versa, variance ratios that imply excess volatility of the actual spread relative to the theoretical and that the VAR restrictions can not be rejected. In contrast to US studies this early examination of UK data finds strong support for the EH. Mills (1991) examines subperiods of quarterly data[28] over 1870 to 1988, his findings are consistent with MacDonald and Speight

---

[28]For the subperiods prior to 1939 the 3-month prime bank bill rate and a consol is considered. During the post war periods the 3-month risk-free T-bill rate with a richer set of long rates, 5- and 20-year gilts and a perpetuity are used. Mills shows that the means and variances of the rates and

45

(1988), with stronger support found in the post war periods. MacDonald and Speight (1991) find mixed results in their multi-country[29] study from 1964 to 1986, with excess volatility of the actual spread observed for all countries, the correlation between the actual and theoretical spread ranging between 0.77 and 0.92, and the VAR restrictions rejected for all but the UK.

Hardouvelis (1994) considers quarterly 3-month and 10-year post war yield data up until 1992 for the G7[30]. He highlights the puzzle reported by the literature, where the coefficient in the regression of the long rate change on the spread is less than its theoretical value of one or negative, the puzzle being the negative sign. Consistent with these findings he reports a negative coefficient for the US, UK, Germany, Japan and Canada[31]. Further, he finds the correlation ranges between 0.75 to 0.99 (0.83 for the UK), variance ratios between 0.33 to 1.24 (0.84 for the UK), the coefficient in the regression of the theoretical spread on the actual taking values between 0.20 and 1.20 (0.83 for the UK) and the VAR restrictions are not rejected for any of the countries. Hardouvelis finds support for the EH for all countries except the US, where the VAR methodology results show any departures from the EH are not economically significant.

Less encouraging results for the EH are presented by Taylor (1992), who uses weekly data on 3-month, 10-, 15- and 20-year UK government bonds, 1985 to 1989. He finds spreads do not Granger cause changes in short rate, evidence of actual spreads being excessively volatile and the VAR restrictions are rejected. Further, plots of actual and theoretical spreads indicate some comovement, but divergence between the two is apparent. Taylor attributes this statistical rejection to economically significant departures from the model. He argues that this rejection could be due to a time varying term premium, which he models using a GARCH model, but fails to find

---

their spreads vary quite substantially over the periods considered.

[29] For Belgium, Canada, Germany, UK and US.

[30] Canada, France, Germay, Italy, Japan, UK and US.

[31] Only for the US is it significantly different from zero, and significantly different from one for three of these countries.

evidence in support of this. However, he does find some encouraging evidence for the Market Segmentation Hypothesis of the term structure. Taylor states that these results suggest that the UK policy of repurchasing government debt has had the effect of inverting the yield curve over the sample he studies, more than the effect government policy had on expected future rates. That is, his findings imply that the MSH rather that the EH is better able to describe the term structure over this period.

Turning our attention now to the UK studies that use data at the short end. As discussed earlier because "short-term instruments in the UK have, until recently, been limited to three months duration, investigation of the hypothesis at the short end of the term structure is not possible" Mills (1991). Thus until the mid 1990s it was not possible to replicate the early US studies that use rates from the entire maturity spectrum. Hurn et al (1995) use monthly data on the Interbank market rates[32] for maturities of 1, 3, 6 and 12 months over 1975 to 1991. They find cointegration, with cointegrating vector (1,-1) in almost all cases. The regression coefficients of the theoretical on the actual spread are not significantly different from their theoretical value in nearly all cases. The correlations range between 0.97 to 0.99, standard deviation ratios between 0.72 to 1.11 and the VAR restrictions can not be rejected. This study finds strong support for the EH, more so than previous studies. The authors suggest that the VAR restrictions may have been rejected by previous studies that use coupon-paying bond data, because to find equivalent yields on pure discount bonds these studies need to employ approximations. Whereas their LIBOR data does not require such approximations.

Also examining data at the short end is Rossi (1996) who tests the EH using two different monthly datasets over 1982 to 1995. He uses London Interbank middle market rates for 1-, 3-, 6- and 12-month maturities, to compute implied forward rates i.e. n-month maturity n-months ahead; and fitted gilt yields (zero coupon yields) 6, 12, 18 and 24 months to maturity. He examines the EH using both the spreads (from the

---

[32]London Interbank Offer Rates, LIBOR.

gilt yields) as in CS (1991) and forward rates (from the Interbank rates) as in Fama (1984) to predict future short rates[33]. Some evidence of spread having predictive power for both future short rates and changes in the long rate is found.

Using higher frequency weekly data Cuthbertson (1996) also considers LIBOR data for maturities of 1-week, 1-, 3-, 6- and 12-months from 1981 to 1992. He finds evidence of cointegration with cointegrating vector (1,-1) in almost all cases[34], except when 6- and 12-month rates are involved. Even though there is a lack of full support for the EH using cointegration, Cuthbertson finds forecasts from the restricted VECM[35] to be marginally more accurate than those from the unrestricted VECM. In the regression of the theoretical spread on the actual, the coefficients are not significantly different from zero; the spreads Granger cause future short rates and there is a strong correlation between actual and theoretical spreads. However, the variance ratios range between 0.55 and 2.70, and the VAR restrictions are rejected. Unlike Cuthbertson (1996), Hurn et al (1995) do not reject the VAR restrictions, even though both use LIBOR data at the short end. Hurn et al argue that this rejection may be due to Cuthbertson using higher frequency weekly data, which is more volatile compared to their monthly data.

Cuthbertson et al (1996) also use weekly data, examining the EH with liquid short term Certificate of Deposit market rates during 1975 to 1992 for 1-, 3-, 6-, 9- and 12-month maturities. They suggest this weekly data on pure discount bonds allows them to mimic agents forecasting procedures with greater accuracy than monthly or quarterly data. They find cointegration amongst the yields, but reject the hypothesis that the cointegrating vector is (1,-1). In the single equation regressions, the coefficients are

---

[33]Rossi examines the EH using both the spread as detailed in CS (1991) and summarised above, and forward rates as used by Fama (1984). The latter, which considers the predictive power of the forward rates for future short rates, is an alternative approach used by the empirical term structure literature, which we do not examine in this thesis. Here they consider the predictive power of the forward rates, rather than the spread as in the CS approach, for future short rates.

[34]Further support at the short end using the cointegration method is presented in Cuthbertson et al (1998).

[35]The unrestricted VECM does not have the spread restrictions imposed. But the restricted VECM does, such that the cointegrating vector is of the form (1,-1).

not significantly different from unity and the spread is found to Granger cause changes in the short rates in nearly all cases. There is strong correlation between actual and theoretical spreads, the standard deviation ratios (range between 0.56 and 0.88) indicate excess volatility of the actual spreads but do not significantly differ from their theoretical value. However, the VAR restrictions are rejected in almost all cases. Both Cuthbertson and Cuthbertson et al, although formally reject the VAR restrictions, find evidence to suggest that these rejections of the EH are not economically significant.

Some attribute the failure of the EH to the term premium being time varying and not constant, see Froot (1989) and Tzavalis and Wickens (1997, 1998). Cuthbertson et al (2003) investigate this claim for the UK using long rates[36]. They suggest that a time-varying (but stationary) term premium can result in a "over reaction hypothesis", where movements in the actual spread are more volatile than the expected changes in the future short rates. That is, actual spread movements are more volatile than those predicted by the EH. For $n = 2$ to 10 years they find strong support for EH with constant term premium, in particular they can not reject the VAR restrictions, which is in sharp contrast to US findings. But the correlation, variance ratio and test of the VAR restrictions indicate rejection of the EH for $n = 15, 20, 25$ i.e. at the very long end, which is consistent with the term premium being time varying for these maturities. However, the authors conclude that these time-varying term premium effects are stronger for the US.

In general, these studies of the short end of the UK term structure are more support-ive of the EH than earlier UK studies that use lower frequency data at longer maturities. As demonstrated by Cuthbertson et al (2003) a time-varying term premium may ex-plain the results found from tests of the EH for UK data at the very long end. But

---

[36]They use monthly spot rates constructed by the Bank of England from coupon bonds for maturities of 2,3,...,10 years and 15, 20 and 25 years over 1976 to 1999. To test the EH while allowing for a time varying term premium Cuthbertson et al use a trivariate VAR composed of the spread, the change in the short rate and the excess holding period return. The excess holding period return Is used as a proxy for the time-varying term premium.

the findings from studies using data from the short end suggest that the EH with a constant term premium provides a satisfactory description of the UK term structure.

Other Findings for the EH

As seen from the discussion above, although the early studies that focus on the US term structure find a lack of support for the EH[37], more favourable results are presented for the UK, this is also true of other countries. Engsted and Tanggaard (1994, 1995) using data over 1976 to 1991 find Danish yields to be cointegrated. When analysing the predictive power of the spread for future rates, their results show that during the regime of monetary targeting (target money supply) over 1976 to 1985 support for the EH is found. But during 1985 to 1991 when there was an interest rate targeting (target short rates) regime the spread appears to lose this predictive power. These findings are in line with Mankiw and Miron (1986), and further supported by Engsted (1996). Christiansen et al (2003) examine Danish data over 1993 to 2002, to find that since the ERM currency crisis of 1992 the Danish term structure has been segmented. Such that at the short end the rates are cointegrated and the EH holds, but the EH is rejected at the long end of the term structure.

Similarly, Cuthbertson et al (2000b) find during 1976 to 1993 German money markets conform closely to the EH. Like the studies above, they attribute this success of the EH to the German rates being sufficiently volatile under money supply targeting, which results in more variability of rates at the shorter end. But given the credible anti-inflation policy of the Bundesbank the rates are not highly volatile. They argue that variability in expected changes in short rates is required by econometric tests of the EH, however very large changes may increase the perceived riskiness of holding the asset and thus invalidate the EH with a constant term premium.

Other studies include Dominguez and Novales (2000) who use Eurodeposits rates

---

[37]However, more recently Engsted and Tanggaard (1994) present some encouraging results for the US using the cointegration method for the period 1952 to 1987.

over 1978 to 1998 to find the rates are cointegrated and the cointegrating vector is of the form (1,-1) as implied by the EH, and further the forward rate is found to have predictive power for future interest rates. Cuthbertson and Bredin (2000) use Irish data at the short end during 1984 to 1997, to find support using cointegration and the VAR methodology. Further support for the EH is provided by Lange (1999) for Canada, Boero and Torricelli (2002) for Germany, Konstantinou (2005) for the Polish interbank market, Lopes and Monteiro (2007) for Portugal using short rates, Koukouritakis and Michelis (2008) for the 10 newest EU countries. For Australia, Robinson (1998) cannot reject the VAR restrictions, and Fang and Lee (2003) find the rates are cointegrated. Drakos (2002) for the US, Germany and the UK using Eurocurrency market rates at the short end over 1988 to 1998, Ghazali and Low (2002) for Malaysia, Cooray (2003) for Sri Lanka, and Musti and D'Ecclesia (2008) for Italy all show the EH holds using cointegration analysis.

It is clear from the empirical literature that the support found for the EH is sensitive to the country considered, testing method employed, the time period examined and the maturity of the data. Below we discuss the stylised facts that have emerged from the empirical investigations of the EH.

Stylised Facts to Emerge from the Literature and Why the EH may be Rejected

Some stylised facts emerge from the literature[38]:

- The coefficient in the regression testing the predictive power of the spread for the future change in the long rate is often reported to be negative, e.g. CS (1991) and Hardouvelis (1994).

- Many reject the VAR cross-equation parameter restrictions, e.g. MacDonald and

---

[38]As noted in Cuthbertson et al (1996).

Speight (1991), CS (1991), Taylor (1992), Cuthbertson (1996) and Cuthbertson et al (1996).

- More support for the EH is found when interest rates are volatile, e.g. Mankiw and Miron (1986), Engsted and Tanggaard (1994, 1995), Engsted (1996) and Cuthbertson et al (2000b).

The literature offers several possible reasons why the EH is rejected, these include noise traders, time-varying term premium and over reaction of the long rate. First, noise traders follow fads purchasing assets that are fashionable, so their decisions are not solely theory informed. Hence, the prevailing long rate is not just a combination of current and future short rates as suggested by the EH, but of stochastic noise too. Secondly, the term premium is assumed constant through time in the EH model[39], however the term premium may in fact be time varying, e.g. Fama (1984), Mankiw (1986), Campbell (1987), Froot (1989), Longstaff (1990), Tzavalis and Wickens (1997, 1998), Tzavalis (2003) and Cuthbertson et al (2003).

Under the over-reaction hypothesis the current long rate does move in the direction predicted by the EH, but its movements are sluggish in comparison to those of the current short rates. Hence, the long rate under reacts to the current short rate and over reacts to future short rates. CS (1987, 1991) state that although the actual and theoretical spreads move together, this overreaction results in the EH being statistically rejected, see also Hardouvelis (1994).

Cuthbertson (1996) and Cuthbertson et al (1996) suggest that the VAR restrictions may be rejected for the following reasons first, the VAR coefficients may be biased if the econometrician does not include in the information set all variables influencing traders' perceptions. Thus, when forecasting future rates all available information is not being used optimally, thereby leading to the restrictions being rejected. Second, if the VAR

---

[39]Where the Pure EH assumes the term premium is zero.

is used to forecast then we expect agents to use "minute by minute observations" of the spread and changes in the short rate, in which case forecasts derived from even weekly data may be insufficient to mimic such behaviour. However, the formal statistical rejection of VAR cross-equation restrictions does not necessarily imply rejection of the EH from an economic perspective, if there is comovement between the actual and theoretical spreads, CS (1991).

Other Term Structure Studies and Recent Developments

Examination of the EH is merely one section of a vast term structure literature. Since this thesis is in part concerned with the EH, we concentrate our discussion in this area. However, in this section we touch on some of the other areas investigated by the term structure literature, these include the predictive power of the term structure for economic activity and regime shifts.

Many authors have examined the relationship between the term structure of interest rates and future economic activity to find a strong correlation, where the term structure possesses a leading indicator property. Two explanations are provided for this leading indicator property of the term structure. The first describes how investors try to smooth their consumption through the business cycle. That is, when a recession is anticipated long-term assets will be bought for maturity during the downturn, hence smoothing income. As demand for long term assets increase their price increases, inducing a decrease in their yields and resulting in a flattening of the yield curve, this flattening is observed to precede a recession. However, empirically this explanation is rejected. The second is that the leading indicator property of the term structure exists because of the monetary authorities, where through money market intervention the central bank can control the short-term rate. As discussed above, the EH of the term structure states that the long-term rate is given by an average of the short term rates, plus a term premium. Thus the central bank through manipulation of the short term rate,

can influence the long term rate and hence future activity. For instance through the tightening of monetary policy the short rate increases, in turn increasing the long term rate and preventing the economy from overheating. See Harvey (1989), Estrella and Hardouvelis (1991), Jorion and Mishkin (1991), Plosser and Rouwenhurst (1994), Estrella and Mishkin (1997), Bernard and Gerlach (1998), Brooks and Tsolacos (2001), Hamilton and Kim (2002), McMillan (2002), Stock and Watson (2003), Duarte et al (2005) and Gomez-Biscarri (2009).

The failure of the EH may be attributed to a regime change, as is for the US during 1979-1982, Hamilton (1988). Sola and Driffill (1994) find over pre-regime change data for the US their results are consistent with the EH. However, only when a VAR with regime switching is used, is the EH found to be consistent with the data when the regime change period is included in the sample. Further evidence provided in Kugler (1996) and Dillen (1997). Jardet (2004) tries to detect a structural break in the correlation between the interest rate spread and future activity in 1984 for the US. He finds that the term structure had greater explanatory power prior to the break in 1984, which coincides with a policy regime change of inflation targeting. See also Brooks and Rew (2002), Ang and Bekaert (2002), Bansal and Zhou (2002), Clarida et al (2006) and Dai et al (2007).

Developments in the literature include using very short term data, new testing methods for the EH and macro-finance models of the term structure. Longstaff (2000) examines the term structure at the very short end using maturities ranging from one day to 3 months, to find new support for the EH, see also Brown et al (2008). Bekaert and Hodrick (2001)[40] test the EH by the conventionally used Wald tests together with the newly developed Lagrange multiplier test. Sarno, Thornton and Valente (2007) and Della Corte, Sarno and Thornton (2008) also implement this new testing procedure. A growing branch of the term structure literature is that which incorporates core macro-

---

[40] Sarno et al (2007) state that this LM test was originally proposed by Campbell and Shiller (1987), but was made operational by Bekaert and Hodrick (2001).

economic variables, for instance as discussed in Brooks and Skinner (2000), Ang and Piazzesi (2003), Carriero et al (2006), Ang et al (2006), Dewachter and Lyrio (2006), Diebold et al (2006) and Rudebusch and Wu (2008).

Shea (1992) raises the questions (i) if the EH can at times be valid and at other times be invalid? (ii) if some of the testable implications are more easily accepted than others? and (iii) whether the theory describes only portions of the yield curve? these questions are as relevant now as they were then. From this review of the empirical literature that tests the EH, it can be seen that (i) the EH describes the data better under some regimes than others e.g. Mankiw and Miron (1986), Engsted and Tanggaard (1995), Cuthbertson et al (2000b), (ii) the testing method used is important with more support being found for the EH using cointegration analysis and the VAR methodology than under the single equation tests. And (iii) in the UK stronger support for the EH is found at the short end, this is also true for Denmark see Christiansen et al (2003). Although the evidence in favour of the EH is quite mixed using statistical criteria, many report strong comovement between actual spread and that predicted by the EH, in this case it is difficult to reject a hypothesis that is clearly economically significant.

2.3 Modelling the UK Term Structure

Earlier we specifically discussed the EH of the term structure, its implications and how they can be tested using several methods including single equations and a bivariate VAR model. In this section we discuss how yields can be modelled more generally using a range of statistical and theory informed time series models, and show how these models are nested within a VAR framework. The statistical models include the Autoregressive (AR) and Vector Autoregressive Model in Differences (VARD) models. And the theory informed models, that embed the cointegration implied by the EH both explicitly in a Vector Error Correction Model (VECM) and implicitly using spreads

in a VAR in Transformed Interest Rates (VART) Model[41]. It is this last model that the EH literature uses. This section provides the theoretical framework of these above mentioned models, which will be used in the subsequent chapters to model the UK term structure.

2.3.1 Autoregressive (AR) Model

If the yields are difference stationary, then change in the n-period yield $\Delta r_t^{(n)}$ can be modelled by a pth order AR model

$$\Delta r_t^{(n)} = \beta + \sum_{i=1}^{p} \Psi_i \Delta r_{t-i}^{(n)} + u_t \tag{2.23}$$

where the white noise process $u_t \sim i.i.dN(0, \boldsymbol{\sigma}^2)$, $E(u_t) = 0$, $E(u_t^2) = \boldsymbol{\sigma}^2$ and $E(u_t u_{t-s}) = 0$ for all $s \neq 0$, i.e. the unobservable errors u_t are independently and identically distributed random variables that are homoskedastic and serially uncorrelated. The intercept is given by β and the coefficients Ψ_i describe the influence past changes in $\Delta r^{(n)}$ have on $\Delta r_t^{(n)}$.

From this standard statistical AR model, we explore two different modelling approaches. The first is an 'economic' approach, where theory informed term structure models can be developed as given by the VART and VECM models; the second is a 'statistical' route where the above AR model can be extended to a VARD model. Further, we demonstrate how these models are nested within each other. A bivariate framework is used to discuss each model, however this can easily be extended to a multivariate one.

---

[41]Plots of the yield data used in this thesis show that none of them exhibit a trend, so the trend terms in each of the model descriptions are not included.

2.3.2 VAR in Transformed Interest Rates (VART) Model

If we are concerned with simultaneously modelling the short m-period yields $r_t^{(m)}$ and the n- and m-period spread $s_t^{(n,m)}$ for $n \neq m$. Then by assuming the yields are difference stationary and that there exists a cointegrating relationship between n- and m-period yields, such that spreads are stationary, there exists a Wold representation which can be approximated by a Vector Autoregression (VAR) model of order p. From the following Vector Moving Average model, bold face is used to represent a vector

$$
\begin{pmatrix} s_t^{(n,m)} \\ \Delta r_t^{(m)} \end{pmatrix} = \begin{pmatrix} r_t^{(n)} - r_t^{(m)} \\ r_t^{(m)} - r_{t-1}^{(m)} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \mathbf{A}(\mathbf{L}) \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix} \tag{2.24}
$$

α_1 is the mean spread and α_2 is the mean change in the m-period rate. $\mathbf{L}$ is the lag operator, ξ_{1t} and ξ_{2t} are mean zero, stationary innovations with the covariance matrix $\mathbf{\Sigma}$ [42]. The lag filter $\mathbf{A}(\mathbf{L})$ directly and the covariance indirectly accounts for the interdependent determination of the change in the m-period rate and the spread between the n- and m-period rate. Thus the model is able to incorporate the impact of shocks to $\Delta r_t^{(m)}$ and $s_t^{(n,m)}$ directly, as well as the feedbacks that may occur from $\Delta r_t^{(m)}$ to $s_t^{(n,m)}$ and vice versa.

From this fundamental MA representation in which the invertibility condition holds and assuming that the lag filter $\mathbf{A}^{-1}(\mathbf{L})$ can be approximated by a pth order lag poly-

---

[42] The Wold Decomposition Theorem justifies the use of this representation, in that it states that "every weakly stationary process and a stochastic process can be written as a sum of a deterministic process and a stochastic process with a Moving Average representation, in which the process is explained soley in terms of past and present random innovations in the vector of variables", Lee and Shields (2000). In this case $r_t^{(m)}$ is difference stationary and the spread $s_t^{(n,m)}$ is stationary assuming that cointegration exists between $r_t^{(n)}$ and $r_t^{(m)}$.

nomial i.e. $\mathbf{A}^{-1}(\mathbf{L}) = \mathbf{B}_0 + \mathbf{B}_1\mathbf{L} + \mathbf{B}_2\mathbf{L}^2 + .. + \mathbf{B}_p\mathbf{L}^p$, where $\mathbf{B}_0 = \mathbf{I}_2$, then

$$
\begin{pmatrix} s_t^{(n,m)} \\ \Delta r_t^{(m)} \end{pmatrix} = \beta + \mathbf{B}_1 \begin{pmatrix} s_{t-1}^{(n,m)} \\ \Delta r_{t-1}^{(m)} \end{pmatrix} + ... + \mathbf{B}_p \begin{pmatrix} s_{t-p}^{(n,m)} \\ \Delta r_{t-p}^{(m)} \end{pmatrix} + \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix} \tag{2.25}
$$

where $\beta = \mathbf{A}^{-1}(1)\alpha$, now $\Delta r_t^{(m)}$ and $s_t^{(n,m)}$ are described by past values of $\Delta r_t^{(m)}$ and $s_t^{(n,m)}$. Equation (2.25) describes what will henceforth be called the VAR in Transformed Interest Rates (VART) model. This VART is the same as the bivariate VAR (unrestricted) given by equation (2.20) used in the Campbell-Shiller VAR methodology to test the EH, above just provides a more detailed statistical description.

2.3.3 Vector Error Correction Model (VECM)

The above VART model is composed of $\Delta r_t^{(m)}$ and $s_t^{(n,m)}$ which are both stationary and can be written in levels, see Appendix 3, as follows

$$
\begin{pmatrix} r_t^{(n)} \\ r_t^{(m)} \end{pmatrix} = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \begin{pmatrix} r_{t-1}^{(n)} \\ r_{t-1}^{(m)} \end{pmatrix} + ... + \boldsymbol{\Phi}_{p+1} \begin{pmatrix} r_{t-p-1}^{(n)} \\ r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{2.26}
$$

From equation (2.26) it is assumed that the VAR model only contains endogenous I(1) variables, in a general form

$$
\mathbf{z}_t = \boldsymbol{\mu} + \sum_{i=1}^{p+1} \boldsymbol{\Phi}_i \mathbf{z}_{t-i} + \boldsymbol{\varepsilon}_t \tag{2.27}
$$

where $\mathbf{z}_t$ is a $q \times 1$ vector of variables, $\boldsymbol{\mu}$ is a $q \times 1$ vector of intercepts and $\boldsymbol{\varepsilon}_t$ is a $q \times 1$ vector of white noise errors as previously defined, $\boldsymbol{\varepsilon}_t \sim i.i.d.N(0, \boldsymbol{\Sigma})$. Where all the

roots of the determinantal equation fall on and/or lie outside the unit root circle

$$\left| \mathbf{I}_k - \boldsymbol{\Phi}_1 \boldsymbol{\lambda} + \boldsymbol{\Phi}_2 \boldsymbol{\lambda}^2 + ... + \boldsymbol{\Phi}_p \boldsymbol{\lambda}^p \right| = 0$$

Manipulation of this levels form, given by equation (2.26), yields the following VECM representation

$$\begin{pmatrix} \Delta r_t^{(n)} \\ \Delta r_t^{(m)} \end{pmatrix} = \boldsymbol{\mu} + \sum_{i=1}^{p} \boldsymbol{\Gamma}_i \begin{pmatrix} \Delta r_{t-i}^{(n)} \\ \Delta r_{t-i}^{(m)} \end{pmatrix} - \boldsymbol{\Pi} \begin{pmatrix} r_{t-1}^{(n)} \\ r_{t-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{2.28}$$

where the difference operator is given by $\Delta = (1 - L)$, and noting that

$$\boldsymbol{\Gamma}_i = - \sum_{j=i+1}^{p+1} \boldsymbol{\Phi}_j \quad i = 1 \text{ to } p \qquad \text{and} \qquad \boldsymbol{\Pi} = \mathbf{I}_2 - \sum_{i=1}^{p+1} \boldsymbol{\Phi}_i \tag{2.29}$$

Thus the VART model given by equation (2.25) can be written as a VECM. The VECM allows us to capture the cointegrating relations that exist between the yields and the error correction term $\left(r_{t-p-1}^{(n)} - r_{t-p-1}^{(m)} \right)$ ensures that the long and short yields do not deviate greatly from each other in the long-run. From equation (2.28) the general VECM representation is

$$\Delta \mathbf{z}_t = \boldsymbol{\mu} + \sum_{i=1}^{p} \boldsymbol{\Gamma}_i \Delta \mathbf{z}_{t-i} - \boldsymbol{\Pi} \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t \tag{2.30}$$

with $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}_i$ defined by (2.29). Where $\mathbf{z}_t$ is a $q \times 1$ vector of stochastic I(1) variables in this case the yields, $\boldsymbol{\mu}$ is a $q \times 1$ vector of constants, $\boldsymbol{\varepsilon}_t$ is a $q \times 1$ vector of white

noise errors where $\boldsymbol{\varepsilon}_t \sim i.i.d.N(0, \boldsymbol{\Sigma})$, $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Pi}$ are $q \times q$ matrices of coefficients. The short-run coefficients contained in $\boldsymbol{\Gamma}_i$ relates $\mathbf{z}_t$ to its past values. The rank of $\boldsymbol{\Pi}$ given by r determines the number of stationary linear combinations of $\mathbf{z}_t$, where the existence of cointegration amongst the yields implies that the long-run matrix $\boldsymbol{\Pi}$ has reduced rank $r < q$. Thus, under cointegration there exists r cointegrating vectors among the q yields contained in the $q \times r$ matrix $\boldsymbol{\beta}$, such that $\boldsymbol{\Pi} = \boldsymbol{\alpha}'\boldsymbol{\beta}$. The long-run matrix of coefficients $\boldsymbol{\beta}'$ is the matrix of cointegrating vectors. The $(q \times r)$ loading matrix $\boldsymbol{\alpha}$ holds the error-correction coefficients (also known as adjustment or feedback coefficients), which determine how quickly deviations of the long-run relationships from equilibrium feedback to the system $\mathbf{z}_t$, and give the weights with which each cointegrating vector enters each $\Delta\mathbf{z}_t$ equation.

If $\mathbf{z}_t$ is composed of I(0) variables then the short- and long-run effects can not be separated and shocks to the system have no long-run effects, with the variables reverting to their unconditional mean. If however, $\mathbf{z}_t$ is composed of I(1) variables that are not cointegrated, then all shocks have a persistent effect. In this case, we have $\mathbf{z}_t$ that is composed of I(1) variables that theory does suggest are cointegrated. So even though $\mathbf{z}_t$ is non-stationary the error correction terms $\boldsymbol{\beta}'\mathbf{z}_{t-1}$ give the stationary relationships that exist amongst these non-stationary variables. So shocks to the system may have persistent effects on the variables, but do not have a persistent effect on the equilibrium relations, since any effect eventually dies.

When identifying the cointegrating vectors, r^2 *exactly identifying* restrictions are imposed on the $\boldsymbol{\beta}$ matrix, that is r restrictions on each of the r cointegrating relations. However, these r^2 restrictions are not sufficient to uniquely identify the economically meaningful $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Hence further *over-identifying* restrictions are necessary to test the relationships proposed by economic theory. These restrictions can be tested by comparing the likelihood of the VECM subject to the exactly identifying restrictions

to that of the VECM subject to a full set of restrictions[43].

As previously seen if the EH is valid then there should exist $r = q - 1$ cointegrating vectors, where "the cointegration space should be spanned by the columns of the matrix such that the bivariate interest rate spreads are stationary." Christiansen et al (2003, pp. 9). In the next chapter we test the hypothesis that the cointegrating relationship between the n- and m-period yields has cointegrating vector $(1, -1)$, as implied by the EH such that spreads are stationary. If indeed a long-run relationship of this form exists, then this justifies the estimation of the VART model which is based on this assumption holding. More specifically, we impose and test two sets of over-identifying restrictions, in the first set each row corresponds to the cointegrating vector $\left(1, -1, c^{(n,m)}\right)$ with the term premium $c^{(n,m)}$ free from restriction. The second set tests the PEH which assumes the term premia are zero, such that the cointegrating vector corresponds to $(1, -1, 0)$.

2.3.4 VAR in Differences (VARD) Model

If however, from equation (2.28) $\mathbf{z}_t$ is I(1) and not cointegrated then Π would be a null matrix of rank 0, i.e. $\Pi = 0$, where no long-run relationships are found to hold amongst the yields in the form of a stationary spread. Then a VAR in Differences model of order p, VARD(p), is appropriate to describe the system as given by equation (2.31) for the bivariate case and equation (2.32) for the general case

$$\begin{pmatrix} \Delta r_t^{(n)} \\ \Delta r_t^{(m)} \end{pmatrix} = \mathbf{a}_o + \sum_{i=1}^{p} \boldsymbol{\theta}_i \begin{pmatrix} \Delta r_{t-i}^{(n)} \\ \Delta r_{t-i}^{(m)} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \tag{2.31}$$

$$\Delta \mathbf{z}_t = \mathbf{a}_o + \sum_{i=1}^{p} \boldsymbol{\theta}_i \Delta \mathbf{z}_{t-i} + \mathbf{e}_t \tag{2.32}$$

---

[43]The full set comprises exact and over-identifying restrictions.

This models simultaneously the change in the n- and m-period yields at time t where $n \neq m$. Here $\mathbf{z}_t$ is a $q \times 1$ vector containing the n- and m-period yields. Where $\mathbf{a}_o$ is a $q \times 1$ vector of intercepts and $\mathbf{e}_t$ is a $q \times 1$ vector of white noise errors, with $\mathbf{e}_t \sim i.i.d.N(0, \mathbf{\Sigma})$. No exogenous variables are considered in this model.

The VARD(p) model can be transformed to an AR(p) specification by imposing restrictions on θ_i. If $n = 1, 3, 6$ and 12, and $j = 1, 3, 6$ and 12 then for each $\Delta r_t^{(n)}$ equation in the VARD(p) model set $\theta_i = 0$ for all $\Delta r_{t-i}^{(j)}$ where $j \neq n$, then the change in the n-month yield depends only on its past values.

This section describes four different types of models that can be used to model the term structure of interest rates, both under a bivariate setup and where possible using a general form in terms of $\mathbf{z}_t$ [44]. We demonstrate how the simple AR model can be extended in two ways, first by continuing down the atheoretic path to a VARD model, or secondly through the incorporation of economic theory to a VART model. Further, it can be shown that the equivalent representation of this VART model is a VECM. Where the VECM explicitly defines and tests the long-run relationships that exist between the yields and the VART model assumes that these long-run relations hold enabling the spread terms to be modelled directly. Lastly, if in the VECM cointegration is not found then a VARD specification is appropriate. It can be seen that both the AR and VARD models are nested within the VECM, and the Bivariate VARD and AR models are nested within the Multivariate VARD model.

Although these models can be viewed individually as either statistical or theory based, it is evident some specifications are nested within others. This implies that we can move from one specification to another by imposing restrictions, where these restrictions can then be tested.

---

[44] In the next chapter $n = 12, 6, 3$ and $m = 1$, so the multivariate models contain all yields with $\mathbf{z}_t = \left(r_t^{(12)}, r_t^{(6)}, r_t^{(3)}, r_t^{(1)} \right)'$. The bivariate models have $q = 2$, with four yields this gives rise to six possible bivariate combinations of $\mathbf{z}_t$.

2.4 Recent Innovations In and the Use of Forecasting in Financial Economics

The first part of this chapter deals with the term structure of interest rates, its theories and a review of the empirical literature, together with a general overview of how the term structure can be modelled using a range of time series models. We now turn our attention to the second part of this chapter which deals with forecasting. The area of economic forecasting is vast, so we focus our discussion on the areas of density forecasting and decision-based forecast evaluation. Specifically, we describe how point and density forecasts are generated. Followed by a review of the literature on forecasting interest rates, forecast evaluation and the use of decision-based forecast evaluation in the context of portfolio choice. We also provide a short discussion of statistical forecast evaluation methods and combining density forecasts, but these techniques are not applied in this thesis so their discussion is kept brief.

A forecast makes a statement about the future and is of value because it is needed in decision making. Forecasts of economic and financial variables are used by governments to inform decisions about monetary and fiscal policy; private firms to inform decisions regarding investment and production; financial risk managers, investors and speculators are interested in forecasts of asset returns including exchange rates, interest rates and stock returns to both price assets and to evaluate risk when allocating assets.

Hence it is important to consider the context in which the forecasts will be used, the decisions they will inform and thus the best way to determine their accuracy. Other considerations include the type of forecast to generate; what data the information set should contain and the modelling approach to be used. There are many methods of forecasting including leading indicator which rely on a stable relationship between the leading variables and the variables being led, consumer and business surveys and time series models describing historical trends in data, Clements and Hendry (2004).

There are generally three types of forecasts (1) a point forecast which is a single value, (2) an interval forecast which states the probability of the future value falling within a given interval and (3) a density forecast which gives the entire probability distribution for the future value. Although point forecasts convey predictions in a simple manner, as Diebold (2004) highlights all time series are subject to random and unpredictable shocks, so even very accurate forecasts will make some error. In which case point forecasts fail to communicate the uncertainty surrounding the forecast. Interval forecasts however, give with a known probability determined by the forecaster, a range of values in which the realised value of the variable is expected to fall, e.g. there is a 90% chance that inflation will be between 1.5 and 4.5%. Here information about the forecast uncertainty is presented, with larger intervals suggesting greater uncertainty about the forecast[45].

A density forecast however, provides the entire probability distribution of the future value of the variable. With this entire density we have information like the distribution, mean and variance of the density. Density forecasts provide a clear and transparent means to convey the uncertainty surrounding computed forecasts, conveying more information than point and interval forecasts[46]. These forecasts can be used to give the likelihood of a particular event being realised as given by the estimated model, e.g. the probability that inflation will equal 2.5%. Moreover, density forecasting allows the different forms of forecast uncertainty to be examined.

The uncertainty surrounding a forecast "reflects the dispersion of possible outcomes relative to the forecast being made" Ericsson (2004). There are two types, first is predictable uncertainty "what we know that we don't know" this includes stochastic

---

[45]From the interval forecast, the point forecast can be deduced by looking to the midpoint of the interval. However, the interval forecast does not have to be symmetric about the point forecast, in which case the point forecast is not necessarily mid of the interval, but in most cases it is reasonable to assume it is, Diebold (2004).

[46]From the density an interval forecast for any confidence level can be constructed. See GLPS (2006) for a discussion of probability forecasting compared to interval forecasting.

(future) and parameter uncertainty. The second is unpredictable uncertainty "what we don't know that we don't know" including future structural changes in the economy and misspecification of the model, Clements and Hendry (2004) and Ericsson (2004). The first type can be accommodated, the second is unpredictable otherwise it would already be incorporated into the model. Stochastic uncertainty is that associated with the effect of unobserved future shocks on forecasts. Parameter uncertainty reflects that parameters are unknown and estimates are used to produce forecasts[47]. Model uncertainty arises when the true data generating process is unknown and alternative model specifications are used to generate forecasts. By computing the predictable forecast uncertainty the expected range of possible outcomes can be deduced. Below we discuss the methods by which we account for stochastic and parameter uncertainty when computing density forecasts.

Although density forecasts convey the most information, point forecasts are more widely used with confidence intervals presenting the forecast uncertainty. Reasons for this offered by the literature include density forecasts requiring computer-intensive techniques and point forecasts being easier to comprehend and utilise. However, for most decision making point forecasts are insufficient and density forecasts are needed, GLPS (2006). Furthermore, the probability of joint events can not be inferred from forecast intervals, but requires the entire joint forecast density function, which we will discuss in more detail below. Although density forecasts analytically can be complicated, we demonstrate below how this can be circumvented by computing forecasts using stochastic simulation methods.

We have thus far described the different types of forecasts and the corresponding forecast uncertainties. In the sections that follow we first describe how point forecasts

---

[47]The cumulative effect of future uncertainty tends to grow with the forecast horizon, which results in interval and density forecasts that widen with the horizon. In the case of parameter uncertainty, the parameter estimates are subject to sampling variability, such that the parameter uncertainty decreases as the sample size grows, see Diebold (2004).

are generated, then how density forecasts are computed using simulation methods when accounting for both stochastic and parameter uncertainty. Followed by how forecasts are evaluated, focusing our attention to decision-based evaluation methods.

2.4.1 Point Forecasts

Point forecasting is commonly discussed and used, so here we only briefly describe how point forecasts are generated. This is then followed by a more detailed discussion of the less frequently used density forecasting. A Vector Autoregression (VAR) of order p is

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=1}^{p} \mathbf{B}_i \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t \tag{2.33}$$

where $\mathbf{x}_t$ is a $(q \times 1)$ vector of variables, $\mathbf{B}_i$ is a $(q \times q)$ matrix of parameters, $\boldsymbol{\mu}$ is a $(q \times 1)$ vector of intercepts and $\boldsymbol{\epsilon}_t$ is a $(q \times 1)$ vector containing $i.i.d.$ serially uncorrelated errors with zero means and a positive definite covariance matrix $\boldsymbol{\Sigma}$. The vector $\mathbf{x}_t$ is often considered stationary in which case each of the term structure models discussed above[48], with the exception of the VECM, can be summarised by this VAR(p). However, equation (2.33) could accommodate I(1) variables subject to restrictions on B_i, therefore the VECM too can be summarised by this VAR(p). Hence the exact composition of $\mathbf{x}_t$ will depend upon the chosen model.

For we are concerned with forecasting $\mathbf{x}_t$, following the description given in Garratt and Lee (2009, GL), then $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{qt})'$ which is $(q \times 1)$ and includes at least the variables of interest and $\mathbf{X}_T = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)'$ is a $(q \times T)$ vector containing observations 1 to T of these q variables.

For $\mathbf{x}_t$ the h-step ahead forecast is $\hat{\mathbf{x}}_{T+h}$, where T is the last observation of the sample and the forecast error[49] is given by $\mathbf{e}_{T+H}^f = \mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h}$. An optimal forecast

---

[48]In Section 2.3, "Modelling the UK Term Structure".

[49]Note that '$\mathbf{e}^f$' denotes the forecast error, which is different to the error term '$\mathbf{e}$' in the VARD model.

of $\mathbf{x}_{T+h}$ is often defined as that which minimises the expected squared forecast error[50] $E\left[\left(\mathbf{e}_{T+H}^{f}\right)^{2}\right]$. The conditional expectation of $\mathbf{x}_{T+h}$ on all past information known at T is an optimal forecast

$$\hat{\mathbf{x}}_{T+h} = E_{T}\left(\mathbf{x}_{T+h} \mid \mathbf{x}_{1}, \mathbf{x}_{2}, ..., \mathbf{x}_{T}\right) = E_{T}\left(\mathbf{x}_{T+h} \mid \mathbf{X}_{T}\right) \quad (2.34)$$

For a VAR(p) the optimal h-step ahead forecast is

$$\begin{aligned}
\hat{\mathbf{x}}_{T+h} &= E_{T}\left(\hat{\boldsymbol{\mu}} + \sum_{i=1}^{p} \hat{\mathbf{B}}_{i} \hat{\mathbf{x}}_{T+h-i} + \boldsymbol{\epsilon}_{T+h} \mid \mathbf{X}_{T}\right) \\
\hat{\mathbf{x}}_{T+h} &= \hat{\boldsymbol{\mu}} + \sum_{i=1}^{p} \hat{\mathbf{B}}_{i} \hat{\mathbf{x}}_{T+h-i}
\end{aligned} \quad (2.35)$$

From equation (2.33), the maximum likelihood estimates of the model parameters are denoted $\widehat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\mu}}, \hat{\mathbf{B}}_{i}, \hat{\boldsymbol{\Sigma}}\right)$, for $i = 1$ to p. The model is iterated forward to produce the point estimates of the h-step ahead forecasts, conditional on the observed data $\mathbf{X}_{T}$ and the estimated parameters $\widehat{\boldsymbol{\theta}}$, for $h = 1, 2, ..., H, ...$ Using the initial values of the variables $\mathbf{x}_{T}, \mathbf{x}_{T-1}, ..., \mathbf{x}_{T-p+1}$ these forecasts are produced recursively.

To illustrate the chain rule of forecasting, writing this VAR(p) in companion form

$$\mathbf{z}_{t} = \mathbf{A}\mathbf{z}_{t-1} + \boldsymbol{\epsilon}_{t} \quad (2.36)$$

where $\mathbf{z}_{t} = \left(\mathbf{x}_{t}, ..., \mathbf{x}_{t-p+1}\right)'$ is ($qp \times 1$), $\mathbf{A}$ is a ($qp \times qp$) matrix of coefficients, $\mathbf{Z}_{T} = \left(\mathbf{z}_{1}, \mathbf{z}_{2}, ..., \mathbf{z}_{T}\right)'$ and $\boldsymbol{\epsilon}$ is ($qp \times 1$). The chain rule of forecasting is used to obtain future

---

[50] Here we adopt the standard statistical definition and define the loss function as being the expected squared forecast error. However, the loss function can take a number of forms, e.g. quadratic cost function, see Pesaran and Skouras (2004).

values of $\mathbf{z}_T$. Using equations (2.34) and (2.36) the one-step, two-step and h-step ahead forecasts are

$$
\begin{aligned}
\widehat{\mathbf{z}}_{T+1} &= E_T\left(\mathbf{z}_{T+1}|\mathbf{Z}_T\right) \\
&= E_T\left(\mathbf{A}\mathbf{z}_T+\boldsymbol{\epsilon}_{T+1}|\mathbf{Z}_T\right) = \mathbf{A}\mathbf{z}_T \\
\widehat{\mathbf{z}}_{T+2} &= E_T\left(\mathbf{z}_{T+2}|\mathbf{Z}_T\right) \\
&= E_T\left(\mathbf{A}\mathbf{z}_{T+1}+\boldsymbol{\epsilon}_{T+1}|\mathbf{Z}_T\right) = \mathbf{A}^2\mathbf{z}_T \\
\widehat{\mathbf{z}}_{T+h} &= E_T\left(\mathbf{z}_{T+h}|\mathbf{Z}_T\right) = \mathbf{A}^h\mathbf{z}_T
\end{aligned}
\tag{2.37}
$$

These optimal h-step ahead forecasts are computed recursively. This method of forecasting provides a point forecast of the variable. Given the forecast is an expectation based on the information set it is subject to error, so often an interval forecast is presented to convey the uncertainty about the forecast, e.g. an interval in which we are 95 or 99% confident that the forecast will lie.

2.4.2 Density Forecasts

We will now describe how density forecasts of variables of interest are generated using stochastic simulation techniques. The estimation procedure is discussed first by considering how the density forecasts are calculated for given values of the parameters and then by taking into account parameter uncertainty[51]. This method is not so frequently discussed or used, so we try to provide some detail and intuition. It should be noted that this description demonstrates one way of generating density forecasts from a specific model. However, density forecasts can be generated in different ways, see Britton et al (1998) and Tay and Wallis (2004) for a discussion on how the Bank of England and the Survey of Professional Forecasters respectively, generate theirs.

---

[51] See GLPS (2006, Chapter 7) for a discussion on model uncertainty.

Since forecasts of the q variables in $\mathbf{x}_t$ are required, the conditional probability density function $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ is of interest. This predictive density function gives the probability density function of the forecast values of the q variables over the horizon $T+1$ to $T + H$, where $\mathbf{X}_{T+1,H} = \left(\mathbf{x}_{T+1}, \mathbf{x}_{T+2}, ..., \mathbf{x}_{T+H}\right)'$ conditional on the observed values of the q variables from 1 to T. That is to say, the probability of observing $\mathbf{X}_{T+1,H}$ given that $\mathbf{X}_T$ has already been observed. The form the density function $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ takes is determined by the types of uncertainty surrounding the forecasts, as well as the way in which the function is characterised and estimated. The forecasts are influenced by various uncertainties including stochastic, parameter and model uncertainty. In Chapters 4 and 5 we consider both parameter and stochastic uncertainty, so we concentrate our explanation of computing density forecasts subject to these two uncertainties only.

A fully Bayesian approach can be taken to estimate the density function, see for instance Kandel and Stambaugh (1996), Barberis (2000) and Abhyankar et al (2005). This involves the construction of a posterior distribution and the use of priors for the parameters. Alternatively Garratt, Lee, Pesaran and Shin (2003 and 2006, GLPS) and Garratt and Lee (2009) take a classical stance on the Bayesian approach to estimating the density function. They use approximations of certain probabilities of interest, thereby avoiding the need for priors. Here we use this alternative approach, a detailed discussion of which is provided in the empirical chapters.

The approach taken to estimate the density function of the forecast values when parameter uncertainty is ignored and when it is incorporated, is as follows. From equation (2.33), the maximum likelihood (ML) estimates of the model parameters are $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\mu}}, \widehat{\mathbf{B}}_i, \widehat{\boldsymbol{\Sigma}}\right).$ In the absence of parameter uncertainty it is assumed that there is no uncertainty about the model parameters and they are fixed at the estimated values.

With density forecasting, first consider stochastic uncertainty only ignoring parameter uncertainty, the forecast values of the variables $\mathbf{x}_{T+h}$ can be computed using stochas-

tic simulations, this provides an estimate of the predictive density $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$ from

$$\mathbf{x}_{T+h}^{(\widetilde{r})} = \widehat{\boldsymbol{\mu}} + \sum_{i=1}^{p} \widehat{\mathbf{B}}_i \widehat{\mathbf{x}}_{T+h-i} + \boldsymbol{\epsilon}_{T+h}^{(\widetilde{r})} \tag{2.38}$$

where $\mathbf{x}_{T+h}$ is the h-step ahead forecast. Further, let $\widetilde{R}$ denote the total number of replications of the above simulation, $\widetilde{r} = 1$ to $\widetilde{R}$ and gives the $\widetilde{r}^{th}$ replication. For current and past values of $\mathbf{x}$, the actual values are used such that $\mathbf{x}_{T+h-i}^{(\widetilde{r})} = \mathbf{x}_{T+h-i}$, e.g. $\mathbf{x}_T^{(\widetilde{r})} = \mathbf{x}_T, \mathbf{x}_{T-1}^{(\widetilde{r})} = \mathbf{x}_{T-1}$... for each replication. Note that density forecasts require forecasts of the errors $\boldsymbol{\epsilon}_{T+h}$ too.

To generate forecasts in the presence of parameter uncertainty the Monte Carlo procedure is used, this provides an estimate of the predictive density $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$. First, the (in-sample) past values of $\mathbf{x}_t$ are simulated $\widetilde{H}$ times, i.e. simulate $\widetilde{H}$ 'histories' of $\mathbf{x}_t$, $t = 1, 2, ..., T$, denoted $\mathbf{x}_t^{(\widetilde{h})}$, $\widetilde{h} = 1, 2, ..., \widetilde{H}$ [52]. Where

$$\mathbf{x}_t^{(\widetilde{h})} = \widehat{\boldsymbol{\mu}} + \sum_{i=1}^{p} \widehat{\mathbf{B}}_i \widehat{\mathbf{x}}_{t-i}^{(\widetilde{h})} + \boldsymbol{\epsilon}_t^{(\widetilde{h})} \tag{2.39}$$

the actual realised values of $\mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-p}$ are used for initial values, together with the estimated model parameters $\widehat{\boldsymbol{\theta}}$ obtained using the actual observed data.

With the $\widetilde{H}$ simulated histories for $\mathbf{x}_t$ i.e. $\mathbf{x}_1^{(\widetilde{h})}, \mathbf{x}_2^{(\widetilde{h})}, ..., \mathbf{x}_T^{(\widetilde{h})}$ such that for each past value of $\mathbf{x}$ there are $\widetilde{H}$ possible values, it is now possible to estimate the VAR(p) model given by equation (2.33) $\widetilde{H}$ times, yielding $\widetilde{H}$ sets of ML parameter estimates $\widehat{\boldsymbol{\mu}}^{(\widetilde{h})}, \widehat{\mathbf{B}}_i^{(\widetilde{h})}, \boldsymbol{\epsilon}_t^{(\widetilde{h})}$ and $\boldsymbol{\Sigma}^{(\widetilde{h})}$, one set of estimates for each Monte Carlo replication, where $i = 1, 2, ..., p$.

For each Monte Carlo replication, compute h-step ahead point forecasts of $\mathbf{x}_T$, where $\widetilde{R}$ replications of these forecasts are generated, i.e. for each of the $\widetilde{H}$ generated histories

---

[52] Note that $'\widetilde{r}'$ refers to the number of 'futures' generated in the simulation, whereas $'r'$ refers to the asset return. Equally $'\widetilde{h}'$ refers to the 'histories' generated and $'h'$ refers to the step ahead forecasts.

simulate $\widetilde{R}$ futures

$$\mathbf{x}_{T+h}^{(\widetilde{h},\widetilde{r})} = \widehat{\boldsymbol{\mu}}^{(\widetilde{h})} + \sum_{i=1}^{p} \widehat{\mathbf{B}}_{i}^{(\widetilde{h})} \widehat{\mathbf{x}}_{T+h-i}^{(\widetilde{h},\widetilde{r})} + \boldsymbol{\epsilon}_{T+h}^{(\widetilde{h},\widetilde{r})} \tag{2.40}$$

for $h = 1, 2, ..., H$; $\widetilde{r} = 1, 2, ..., \widetilde{R}$ and $\widetilde{h} = 1, 2, ..., \widetilde{H}$. Note that h refers to the horizon, $\widetilde{h}$ to the number of histories generated and $S = \widetilde{H} \times \widetilde{R}$ gives the total number of simulations.

The error terms $\boldsymbol{\epsilon}_{T+h}^{(\widetilde{r})}, \boldsymbol{\epsilon}_{t}^{(\widetilde{h})}$ and $\boldsymbol{\epsilon}_{T+h}^{(\widetilde{h},\widetilde{r})}$'s can be drawn using either parametric or non-parametric methods[53]. Here we utilise parametric methods, where the errors are assumed to be $i.i.d.N(0, \boldsymbol{\Sigma})$ serially uncorrelated white noise errors, details of the exact procedure is provided in Appendix 7.

These simulations provide an estimate of the predictive densities $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$ when parameter uncertainty is ignored and $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ when it is considered. The predictive density gives every possible outcome from T to $T+H$ described as a density. That is, at every step ahead we do not just have a point forecast, but an entire density of all possible outcomes.

The above demonstrates how density forecasts can be computed through simulation methods. These techniques allow probabilities of single, joint and dynamic events of interest to be computed, that point and interval forecasting methods can not. For simplicity we abstract from parameter uncertainty. From above $\mathbf{x}_t$ contains q variables of interest, e.g. z_t and y_t, the inflation rate and output growth respectively. The conditional density function of z_{T+1} is $f(z_{T+1} \mid z_T)$, where an estimate of this density in T is $\widehat{f}_T(z_{T+1} \mid z_T) = P(z_{T+1} \mid z_T)$. Density forecasting considers $\widehat{f}_T(z_{T+1} \mid z_T)$ for all possible values of z_{T+1}, with the probability density function

$$\widehat{F}_T(z) = \int_{-\infty}^{z} \widehat{f}_T(z_{T+1} \mid z_T) \; dz_{T+1} \tag{2.41}$$

---

[53]See GLPS (2006, pp. 166-168) for further details.

for all possible z. This provides an entire density of all possible outcomes for z_{T+1}. Whereas a probability event forecast gives the probability of a particular event e.g. $P(z_{T+1} < a \mid z_T)$, inflation is below target, given by $\widehat{F}_T(a)$. Similarly, forecasts for single events concerning y_t can be computed. However, if we are concerned with the probability of joint events, e.g. $P(z_{T+1} < a \; and \; y_{T+1} > b \mid z_T, \; y_T)$, inflation is below target and output growth is positive, computing this probability requires the joint density function $\widehat{f}_T(z_{T+1}, y_{T+1} \mid z_T, \; y_T)$ because the two events are dependent. Solving this analytically[54] could be complex, but can be easily done using simulations[55], where the simulations account for this interdependence.

Further, dynamic events concerning multi-step ahead forecasts can be derived, for instance $P(z_{T+1} < a, \; z_{T+2} < a \mid z_T)$, in two successive periods inflation is below target. Again this probability requires a joint density, since we have two variables z_{T+1} and z_{T+2} that are dynamically interdependent i.e. forecast of z_{T+2} is dependent on the forecast of z_{T+1}. Furthermore, we may be concerned with dynamic joint events, e.g. $P(z_{T+1,H} < a \; and \; y_{T+1,H} > b \mid z_T, \; y_T)$. It can be seen that density forecasting allows the probability of events to be computed, where this event may concern the values of a single variable or a set of variables, measured at a particular point in time or over a length of time, GLPS (2006). Although analytically these events may be difficult to compute[56], this can easily be circumvented using simulation methods, as demonstrated above.

In general, density forecasting can be applied to a number of situations. The Survey of Professional Forecasters is the longest running series of macroeconomic density fore-

---

[54]Would require $\int_{-\infty}^{a} \int_{b}^{\infty} \widehat{f}_T(z_{T+1}, y_{T+1}) \, dz_{T+1} dy_{T+1}$ to be solved.

[55]For instance simulate 1000 values of z_{T+1} and y_{T+1}, from this count the number of times out of the 1000 futures that $z_{T+1} < a \; and \; y_{T+1} > b$. If this occurs say 300 times, then the probability of this joint event is 0.3.

[56]GLPS (2006) argue that evaluating probabilty event forecasts can be complicated because of the form of the functions, difficulties in choosing appropriate limits of integration or because the event being forecast is complex.

casts dating back to 1968[57], see Tay and Wallis (2004). Another well known example is the Bank of England's fan charts of its inflation and GDP growth forecasts, as seen earlier in Figure (2-1), the inflation fan chart depicts the Bank's predicted probability distribution of inflation, presenting the predictable forecast uncertainty. Further, Garratt, Lee, Pesaran and Shin (2006, GLPS, Chapter 11) apply probability forecasting techniques to their estimated macroeconomic model of the UK. Specifically, they consider the events of the inflation rate being in a target range and the economy going into recession, illustrating that probability event forecasts provide a means of conveying forecast uncertainty[58].

Density forecasts are more common in finance and risk management. Where in finance density forecasts of asset and portfolio returns are produced, these forecasts allow the risk associated with the return to be assessed, since they present a complete characterisation of the uncertainty surrounding these returns. This is what we are concerned with in this thesis. More specifically, we generate dynamic density forecasts of asset returns contained in $\mathbf{x}_t$ that are interdependent, use these forecasts to compute expected wealth and derive optimal portfolio allocations, and then examine the economic value of the forecasts of competing models. Specific details of how we compute the density forecasts in the empirical chapters is provided in Appendix 7.

---

[57]This survey of macroeconomic forecasters was previously conducted by American Statistical Association (ASA) and the National Bureau of Economic Research (NBER), but now by the Federal Reserve Bank of Philadelphia. They request density forecasts for inflation and output growth, the respondents are asked to assign forecast probabilities to predetermined intervals into which the future value of the variable may fall, the responses are averaged such that the mean probability distribution is reported.

[58]See also GLPS Chapter 7 for a detailed discussion of probability forecasting, including accounting for future, parameter and model uncertainty.

2.4.3   Forecasting Interest Rates

Previous research that considers the ability of theory based models to forecast interest rates in comparison to atheoretic, purely statistical models like a naive random walk, primarily focus on using statistical criteria to evaluate the accuracy of the forecasts. As discussed in detail earlier in this chapter, the literature that investigates the Expectations Hypothesis considers the ability of the spread to forecast future interest rates, key papers include Campbell and Shiller (1991), Engsted and Tanggaard (1995), Engsted (1996) and Roberds et al (1996).   Also, if the forward rate has predictive power for future spot rates, see Fama (1984a, 1990), Fama and Bliss (1987), Mishkin (1988) and Rossi (1996).   As seen, the literature reports mixed findings in favour of the EH. More recently, Guidolin and Thornton (2008) compare the ability of the EH to predict Treasury yields with forecasts generated by a random walk and a three factor model. They find that the models including the EH model, do not produce forecast errors that are significantly smaller than those from a no change model[59].   They argue that this does not invalidate the EH, but suggests that "the dominant factor in the change in the short-term rate between now and h-months from now is news which is not forecastable." Further, they find that the forecasts of short rates do not vary significantly when the risk premium is assumed constant or allowed to vary, suggesting that the failure of the EH can not be attributed, as many propose, to the time-variation in the risk premium.

Fauvel, Paquet and Zimmermann (1999) provide a detailed survey of different methods used to forecast interest rates, these include univariate methods, multivariate systems, regime switching models and consensus forecasts.   In their assessment of the forecasting performance of these various models, they conclude that despite their simplicity univariate models perform well.   Further, multivariate models i.e. VAR and VECM also perform well, while integrating short-run dynamics and long-run relation-

---

[59]This evidence of the random walk being difficult to beat is further to that reported by Duffee (2002) and Carriero et al (2006).   See also Guidolin and Timmermann (2009) who report gains at longer forecast horizons from using the EH model.

ships. This conclusion is supported by Sarno, Thornton and Valente (2005) who compare the forecasting performance of a range of time series models of the US federal funds rate, to find a simple univariate model to be best and further that combining forecasts resulted only in marginal improvements to accuracy. Bidarkota (1998) who finds multivariate models that embed cointegration between the nominal interest rate and inflation to be more accurate than univariate models of real interest rates in the US. Dua et al (2008) also report a superiority of multivariate models using Indian interest rates. A similar exercise is conducted for Canada by Deaves (1996). Vereda et al (2008) find using Brazilian interest rates, that if one is concerned with forecasting in the short run, then it is suffice to use a model that incorporates past yield data only. However, the information content in macroeconomic variables is of use when concerned with forecasting in the long run.

Other studies examine interest rates forecasting in the context of survey and market based forecasts, of the Taylor rule and the economic information seen by the policy making committee of the central bank and different yield curve estimation techniques including factor models. See Hafer et al (1992), Gosnell and Kolb (1997), Soderlind et al (2005), Kim et al (2008), Ioannides (2003) and Diebold and Li (2006)[60].

2.4.4 Forecast Evaluation Techniques and Decision Making

Forecast evaluation provides a means of judging if the estimated model provides an accurate representation of reality and further how any given model compares with alternative representations, Granger (2003). In what follows, we consider how forecasts

---

[60]There is a branch of the term structure literature that models interest rates using no-arbitrage factor models (affine models) that model the level, slope and curvature of the yield curve. As stated in De Pooter, Ravazzolo and van Dijk (2007) these models explain yields using latent factors that can be extracted from a panel of different maturity yields. We do not provide a discussion of this literature here, but key studies include Dai and Singleton (2000), Duffee (2002), Ang and Piazessi (2003), DeWachter and Lyrio (2006) and Rudebusch and Wu (2008).

are evaluated using both statistical[61] and decision-based measures, with a review of studies that examine asset return predictability using decision-based techniques.

Statistical Evaluation Techniques

With quantitative forecasts, a forecast error and statistical error measures including the mean squared error (MSE) and root mean squared error (RMSE) can be calculated[62]. When the aim is to compare two or more models, measures including ratios of MSEs may be used to ascertain predictive ability, West (2006). Diebold and Mariano (1995) propose several tests to examine the null that there is no difference in the accuracy of two competing forecasts. Stekler (1991) and Clements and Hendry (1998, Chapter 3) review commonly used statistical forecast evaluation techniques for point forecasts, and Christoffersen (1998), Clements and Hendry (1998) and Clements and Taylor (2003) for interval forecast evaluation.

With more attention being paid to density forecasting, there is an increasing focus on the evaluation of density forecasts. Statistical evaluation methods for univariate forecast densities have been developed by Dawid (1984, cited in Pesaran and Skouras (2004)) and Diebold et al (1998). Where Diebold et al (1999) propose a framework for evaluating multivariate density forecasts[63]. Clements and Smith (2000) apply these techniques to density forecasts of US output growth and unemployment produced by linear and non-linear models[64]. They argue that since non-linear models may be better at capturing the higher moments, then evaluation techniques that consider the entire density of forecasts may be better able to discriminate between models, see also

---

[61]There is an extensive range of statistical measures of forecast accuracy, but since our focus in this thesis is on decision-based forecast evaluation we implement only the RMSE measure. Hence we provide only a brief review of the statistical measures used by the literature.

[62]Non-quantitative forecasts include trying to correctly predict changes in the business cycle. Forecast evaluation here may involve comparing the number of correctly predicted turns with the number of incorrect predictions.

[63]Which is a generalisation of Diebold et al's (1998) univariate procedure.

[64]They find that although non-linear models are better able to characterise cyclical features of the data, MSE type criteria indicate that they are not better than linear models at forecasting.

Clements and Smith (2002).

Tay and Wallis (2004) and Corradi and Swason (2006) provide a survey of density forecasting including model evaluation tests presented in the literature[65]. Further, Hall and Mitchell (2009) discuss recent developments in density forecasting, noting that evaluation methods either employ goodness-of-fit tests or scoring rules. The goodness-of-fit tests are used to ascertain if the "probability integral transforms of the forecast density with respect to the realisations of the variables are uniform or, via a transformation, normal", Hall and Mitchell (2009). Briefly, following the description in Pesaran and Skouras (2004), for N consecutive pairs of the actual values and density forecasts $\left\{x_{t+1}, \widehat{f}_t(x_{t+1} \mid \Omega_t),\ t = T, T+1, ..., T+N-1\right\}$, Diebold et al (1998) show under the null that the forecasts are equal to the actual value i.e. $f(x_{t+1} \mid \Omega_t) = \widehat{f}_t(x_{t+1} \mid \Omega_t)$, the N sequence of probability integral transforms

$$z_{t+1} = \int_{-\infty}^{x_{t+1}} \widehat{f}_t(u \mid \Omega_t) du \qquad t = T, T, ..., T+N-1$$

is $i.i.d.$ Uniform (0,1). Thus testing if $\{z_{t+1},\ t = T, T+1, ..., T+N-1\}$ are $i.i.d.$ U(0,1) provides a means of statistically evaluating density forecasts. Further, Hall and Mitchell state that scoring rules are specific loss functions that allocate a numerical value (score) based on the density forecast and the prevailing realised value of the variable. This provides a relative measure of forecast performance, with densities being ranked by their score. Tests developed using scoring rules test for equal predictive performance, with the null that the difference between two or more competing densities is zero. Discussions on statistical evaluation methods that incorporate factors like parameter uncertainty, tests that allow for non-quadratic loss functions, forecast errors

---

[65]See Corradi and Swanson (2006, pp. 206) for a table that summarises key specification tests and model evaluation papers. It gives details of whether the test focuses on the conditional mean compared to conditional distribution, if it is for 1-step or multi-step ahead forecasts and if it evaluates single or multiple models.

that are non-Gaussian, non-zero mean and serially correlated, are given in Diebold and Lopez (1996), Mariano (2004), McCracken and West (2004), Sarno and Valente (2004), Clements (2006) and West (2006).

Decision-Based Forecast Evaluation

This section discusses the importance and use of decision-based forecast evaluation in economics and finance. There is a growing empirical literature that examines predictability of asset returns and use decision-based forecast evaluation in an investment decision making context. Since this thesis makes a contribution in this area, this section is followed by a review of this literature.

It is important to measure the accuracy of forecasts in the context for which the forecasts are intended. This point is asserted by Pesaran and Skouras (2004), who argue that it is necessary to distinguish between whether the forecast is being evaluated by the producer or user of the forecast. Typically statistical measures like the RMSE are used, which may be sufficient for the producer. But for the user who is concerned with making a profit and maximising utility, these statistical measures ignore how the forecasts will be used and the preferences of the user.

Given that different decision environments have different associated loss functions, using an economic cost function to evaluate forecasts in a decision making environment, will reflect that the user wants to maximise expected utility[66] or the payoff subject to constraints, Granger and Machina (2006). Decision-based forecast evaluation[67] considers the economic value of the forecast to the user, as a means of evaluating forecasts by incorporating an economic cost function into the decision making. In the case of density forecasts this is where the whole distribution is used and not just the mean.

---

[66]The user is interested in the forecast that minimises expected cost or maximises expected utility. Where the expeced cost/utility is an estimate of actual cost/utilty computed using the predictive density.

[67]We sometimes refer to decision-based measures as economic value measures.

The importance of decision-based evaluation is highlighted in the early examination of exchange and interest rate forecasts by Wright et al (1986). They note that using forecast accuracy alone in evaluation may result in incorrect and costly decisions. They state that forecasts are used to examine future outcomes of alternative plans, therefore rather than viewing forecasting and decision making as two separate activities, a greater integration between them is needed to make better and more profitable decisions. Key papers by Granger and Pesaran (1996, 2000) and Pesaran and Skouras (2004) make important contributions to the link between forecasting and decision theory.

Decision-based forecast evaluation has been used extensively in the meteorology forecasting literature[68]. In finance, decision-based forecast evaluation has been applied to measure the economic significance of asset return predictability, this literature is discussed below. These decision-based measures are appropriate in finance because the objective is clear, to make profit while minimising risk. However, in economics forecast evaluation is nearly always done using statistical criteria. Pesaran and Skouras suggest this lack of use of decision-based methods in economics may be because (1) evaluation is considered from the producers perspective, (2) the method requires full specification of the decision environment, which they argue is usually absent from the formulation of economic forecasting models, (3) until recently there was little to suggest that a decision-based evaluation approach would differ from a statistical one and (4) the approach can be technically difficult.

Although, Granger and Pesaran (2000) describe how these methods can be applied in macroeconomics. For instance, the Bank of England sets the nominal interest rate using their inflation rate forecasts, where they increase interest rates if their forecast of inflation is beyond a target rate. Granger and Pesaran (1996) argue that here the Bank needs the entire predictive distribution function of the inflation rate, not just a point forecast. Together with a complete description of the gains/losses associated

---

[68]I.e. the study of atmospheric conditions, used to forecast the weather, see Pesaran and Skouras for references.

with making a correct/incorrect forecast. In this case, the cost of incorrectly forecasting inflation to exceed its target would be interest rates higher than necessary. See Clements (2004) who uses both statistical and decision-based methods to evaluate the Bank of England's density forecasts of inflation.

The statistical methods of evaluating density forecasts, as briefly discussed in the previous section, are general in that they are not conditioned on a particular decision making environment. However, Pesaran and Skouras (2004) highlight the need to assess forecasts in the relevant decision environment, suggesting that discrepancies between the actual and forecast density may be more costly in some decision environments than others[69].

In short decision-based forecast evaluation allows both point and density forecasts to be evaluated in the decision environment for which the forecasts are ultimately intended. Below we review the literature that investigates predictability of asset returns and considers decision-based forecast evaluation in investment decision making.

Asset Return Predictability and Decision-Based Forecast Evaluation

Recent evidence of predictability in asset returns has been reported by a number of studies. This overturning the long standing view held up until the 1970s in financial economics that returns are not predictable, but follow closely a random walk, Pesaran and Timmermann (1995). Those including Campbell (1987), Fama and French (1988, 1989), Kandel and Stambaugh (1996) and Ang and Bekaert (2007) show publicly available data on business cycle related, financial and macroeconomic variables including the dividend yield and term structure variables, have predictive power for stock returns.

Most of this evidence is based on studies that assess predictability and forecast accuracy using conventional statistical criteria. These criteria are usually based on point forecasts and some measure of the forecasts' error. However, Leitch and Tanner

---

[69]For instance risk managers may be concerned with extreme values of asset returns, whereas the middle of the distribution is important when inflation is of concern, Pesaran and Skouras.

(1991) argue that given firms use forecasts to increase profits it is better to evaluate forecasts by using a direct measure of profit, rather than using some unrelated statistical measure.

Asset return predictability and decision-based forecast evaluation in the context of stock returns is examined by those including Pesaran and Timmermann (1995). They investigate the robustness of US stock return predictability using a range of macroeconomic and financial variables, to find the predictive power of these variables varies through time and predictability is higher during periods of higher volatility in the markets, see also Marquering and Verbeek (2004). In their examination of the impact of regime changes[70] and predictability on optimal allocation Guidolin and Timmermann (2005) find the optimal allocation is sensitive to what the investor believes to be the underlying state. For instance, if the investor observes a high probability of being in the bear state he invests little in stocks in the short run, but allocates more to stocks in the longer run as the likelihood of moving to the normal or bull state increases.

Conventionally, estimated parameters are treated as being the true parameters and are used to determine optimal allocation. Barberis (2000) argues that with weak statistical evidence for predictability, it is neither correct to assume returns are not predictable or to ignore the sizeable uncertainty surrounding the true predictive power of the explanatory variable. Instead, parameter uncertainty should be considered when making portfolio decisions. Klein and Bawa (1976) examine the effect of parameter uncertainty on optimal portfolio allocation, to find that this additional uncertainty significantly alters how the investor allocates, see also Kandel and Stambaugh (1996). Barberis (2000) explores this further, considering how asset return predictability, the investment horizon and parameter uncertainty affect optimal portfolio choice for long horizon investors. Earlier findings of Samuelson (1969) and Merton (1969) show that

---

[70]They define three states, the high volatility bear state with large negative mean returns, the normal state where returns are close to their historical mean and the bull state which has a high mean return.

if returns are *i.i.d.* then an investor with power utility has an optimal allocation that is insensitive to the investment horizon. Barberis however, demonstrates that if returns are predictable and not *i.i.d.*, then the investment horizon may not be irrelevant. He uses US data for T-bills and the stock index[71], to find that even when parameter uncertainty is incorporated there is sufficient predictability of returns, such that investors allocate significantly more to stocks the longer their investment horizon. Further, those who ignore this estimation risk over allocate to stocks.

Boudry and Gray (2003) extend Barberis by including two additional predictor variables-term spread and the relative bill rate to predict Australian stock returns. Stating that if an asset is important for predicting asset returns then knowledge of the variable's value will cause a utility maximising investor to alter her optimal allocation. They too find there is enough predictability to encourage the risk-averse investor to allocate more to stocks at longer horizons. However, when parameter uncertainty is incorporated the investor allocates more to bonds with the horizon. They suggest that this extra uncertainty "negates the horizon effect from predictability". This latter finding is contrary to Barberis, who finds that parameter uncertainty reduces, not eliminates the positive horizon effects. Other studies include Xia (2001) who examines the impact of learning about the stock return predictability on the optimal allocation. Noting that uncertainty about the model's predictive parameter introduces dynamic learning, through which the uncertainty about predictability affects the allocation. Avramov (2002) argues that although support may be found for predictability there is still uncertainty about the " "correct" regression specification", he finds return predictability is weakened when model uncertainty is incorporated. Further, Brooks and Persand (2003) assess the accuracy of a range of statistical models of key financial time series, to find sensitivity of this accuracy to the evaluation method used.

---

[71] No predictability is defined as the investor assuming that stock returns are *i.i.d.*, and predictability as him believing that a single lagged dividend yield term has predictive power for stock returns, bond returns are assumed constant. Predictability has the effect of making stocks look less risky and parameter uncertainty makes them look more so.

As well as stock return predictability, several papers examine the predictability of exchange rates and interest rates. West, Edison and Cho (1993) model exchange rate volatility using several alternative models for the conditional variance, and examine the out-of-sample performance of these models using both statistical and economic criteria. Abhyankar, Sarno and Valente (2005, henceforth ASV) consider for an investor allocating between domestic and foreign bonds, if the assumption of predictability alters the optimal allocation and if there is economic value of exchange rate forecasts from a monetary fundamentals model. Here forecast accuracy is determined by the utility-based value to an investor who uses this model to optimally allocate her wealth. ASV note the findings of Meese and Rogoff (1983) and the subsequent exchange rate forecasting literature, that models based on monetary fundamentals can not beat the random walk or no change models of the exchange rate when generating out-of-sample forecasts[72]. ASV find predictability significantly alters the optimal allocation and evidence of economic value to exchange rate predictability.

Further evidence is provided by Garratt and Lee (2009, GL) who also incorporate model uncertainty in their study of exchange rate predictability. They consider four different exchange rate models, together with a weighted average of the four models' density forecasts computed using Bayesian model averaging. Using statistical criteria they find that a simple random walk model outperforms the theory based exchange rate models at all investment horizons, but the model average performs best. This is largely consistent with the earlier findings of Meese and Rogoff. However like ASV, GL find evidence of economic value to exchange rate predictability with the theory informed models outperforming the random walk and the model average in terms of generating a higher end-of-period terminal utility.

Della Corte, Sarno and Thornton (2008, DST) assess the validity of the Expectations Hypothesis of the term structure of very short-term US repo rates[73], and more

---

[72]Where these findings are based on statistical measures of forecast accuracy.

[73]They implement Bekaert and Hodrick (2001) Lagrange Multiplier tests, a new testing method

importantly they measure the economic value of departures from the EH. In that, is there a gain in using the unconstrained VAR over the constrained VAR, where the VAR is constrained by the restrictions implied by the EH, see Campbell and Shiller (1991). Like many, as discussed in the first part of this chapter, they reject the EH under statistical criteria. However, they find favourable evidence for the EH when using an economic value criteria. Since the gains from using the unconstrained VAR are small, which they suggest implies that from an economic perspective, statistical rejections of the EH in the repo market are insignificant.

The results reported by ASV, DST, GL and others who use decision-based forecast evaluation techniques, illustrate that the conclusion drawn of how well theory informed models perform in comparison to atheoretic models from a forecasting perspective, is sensitive to the evaluation criterion used. To be exact, under statistical measures atheoretic models like the random walk are difficult to beat, as reported in the exchange rate and interest rate forecasting literature. But under economic value methods, encouraging evidence in favour of predictability, as captured by theory informed models, is found in an investment decision making context.

In summary, the studies described in this section demonstrate the importance of predictability, parameter and model uncertainty in asset allocation, generating density forecasts to capture the risk and return of the asset, and the use of decision-based forecast evaluation.

Forecast Combination

As well as evaluating the accuracy of individual forecasts, researchers may be concerned with comparing the accuracy of alternative forecasts. Usually several forecasts for the same variable are available, which could reflect the differences in modelling techniques used or the information available to the forecaster. In this case, how should the

---

which provides a more powerful test of the EH, see DST for further details.

information contained in these forecasts be exploited?[74] and further should a single best performing forecast be identified or a combination of the forecasts be used? Where a combined forecast may be more robust to misspecification biases and measurement errors in the individual forecasts, Timmermann (2006).

Early work includes Bates and Granger (1969, cited in Newbold and Harvey (2004)), Granger and Ramanathan (1984) and Clemen (1986, 1989). Evidence of gains in accuracy from combining forecasts is presented by Stock and Watson (2004) using output growth data; Kapetanios, Labhard and Price (2008) who discuss the Bank of England's 'suite' of purely statistical models of inflation and output growth[75], and Guidolin and Timmermann (2009) who forecast US short-term rates. Further, Diebold and Lopez (1996), De Menezes, Bunn and Taylor (2000) and Timmermann (2006) review alternative combining methods. Eklund and Karlsson (2007) and GLPS (2006, pp. 158-159) discuss the Bayesian Model Averaging approach. Hall and Mitchell (2005 and 2007) consider forecast combination in the context of density forecasting, with an application to the UK inflation fan charts[76].

2.5   Conclusion

This chapter provides a review of the term structure and financial economics forecasting literature relevant to the empirical chapters that follow. Particularly, in the empirical chapters that follow, we first model the UK term structure and test the Expectations Hypothesis of the term structure. Then we use a range of time series models to forecast and assess the predictability of asset returns in an investment decision making context.

---

[74] Newbold and Harvey (2004) argue that one of these models may be chosen, but this ignores the possibility that the abandoned forecasts may contain useful information about the future that the chosen forecast does not.

[75] The suite is one of the components of the Monetary Policy Committee's forecast process. The authors discuss how these forecasts are combined.

[76] See Hall and Mitchell (2009) for most recent developments in the area of density forecasting and evaluation.

The first part of this chapter discusses the importance of the term structure. We begin with the UK term structure and how interest rates post the adoption of inflation targeting in 1992 and central bank independence in 1997, are comparatively less volatile than before these reforms. This is followed by a summary of term structure theories, including the main theory, the Expectations Hypothesis and the several methods by which the EH is assessed. Together with a review of the empirical term structure literature, focusing on studies that test the EH. Mixed support is found for the EH, with this support being sensitive to the country, frequency and maturity of the data, and testing method considered. We describe how interest rates can be modelled more generally using a range of atheoretic and theory informed time series models, all of which can be summarised by a standard Vector Autoregressive modelling framework.

The second part of this chapter is concerned with forecasting. We discuss the types of forecasts, i.e. point, interval and density, and describe how point and density forecasts are computed. Density forecasts provide the probability distribution of all possible future outcomes, thus conveying the uncertainty about the forecast. They allow the probability of events of interest to be computed, which can not be derived from point or interval forecasts. We briefly review the interest rate forecasting literature, who primarily use statistical methods to evaluate the accuracy of forecasts. A comparatively new area of research is the use of decision-based forecast evaluation, which argues that given forecasts are used to inform decisions then forecasts should be evaluated in the context of these decisions. We discuss how forecasts are evaluated under both statistical and decision-based methods. Further, we review how decision-based methods have been used by recent studies to examine the predictability of asset returns and the economic value of this predictability.

Chapter 3

Time Series Overview of the UK Term Structure 1997 to 2004

Abstract

This chapter investigates the time series properties of the UK Term Structure over 1997 to 2004, through a range of statistical and theory informed models using weekly data for 1-, 3-, 6- and 12-month yields. The models include an Autoregressive, Vector Autoregressive in Differences, VAR in Transformed Interest Rates and a Vector Error-Correction model. This exercise demonstrates the importance of economic theory in explaining the term structure, as the theory informed models are found to have greater explanatory power than the atheoretic ones.

Further, we test the Expectations Hypothesis (EH) of the term structure using both cointegration analysis and the VAR methodology. We find support for the EH in the form of stationary spreads and yields sharing a common stochastic trend, such that over-identifying restrictions on the cointegrating vectors, as implied by the EH cannot be rejected. These results suggest that movements in the UK money market spot rates are consistent with the Expectations Hypothesis.

Keywords: short term rates, term structure models and Expectations Hypothesis.

3.1 Introduction

This chapter uses a comprehensive set of statistical and theory informed models to model the UK term structure of interest rates. There are two main aims of this investigation, first to test the Expectations Hypothesis (EH) of the term structure using the cointegration and VAR methods proposed by the literature, and second to make a comparison between the statistical and theory based models to see which is best able to capture the term structure at the short end.

The term structure of interest rates is defined as the relationship between the term to maturity and the interest rate, the yield curve plots this relationship. The term structure embeds the market's expectation of future interest rates and because it embeds this important information much research has been devoted to modelling the term structure, as discussed in Chapter 2. Shiller (1979), Campbell and Shiller (1987, 1991 henceforth CS) reignited an interest into evaluating the theories of the term structure, the EH with expectations being formed rationally appearing to be the leading hypothesis. One branch of this literature examines if the EH can explain the slope of the yield curve.

Being able to explain the term structure is important for the transmission mechanism of monetary policy. If there exists a relationship between short and long interest rates then by manipulating the short rate, the long rate and therefore real economic activity can be influenced. Further, the yield spread may contain information on future short rates, inflation and economic activity. Furthermore, being able to accurately model the term structure is important for those concerned with forecasting interest rates.

Fundamental to the EH is that the long rate is given by a weighted sum of the expected short rates over the maturity of the long bond, where expectations of future short rates are formed rationally, in that actual short rates differ from expected only by a random error. The implications of this view is that the spread between the long and

short rate, is given by a weighted sum of the expected changes in the short rate over the life of the long bond. Hence, (1) what determines the long rate is the expectations of the future short rate at time t and (2) the difference between current long and short rate is the expectation of a change in future short rates. Two stylised facts implied by the EH include; the ability of spread to predict future changes in the short rate, and the ability of the spread to predict future changes in the long rate. Both of these can be tested through single equation regressions. Chapter 2 provides a detailed discussion of the EH, its implications, methods of assessment and empirical studies, as such we keep the discussion in this chapter brief.

Cointegration methods provide another means by which the EH can be tested, since "cointegration between the short- and long-term interest rates suggests that over the long run interest rates move in tandem with each other and therefore can be used as evidence of the expectations hypothesis" Ghazali and Low (2002). As Thornton (2004) states a lack of cointegration amongst the yields is strong evidence against the EH. Since if the rates are I(1) but not cointegrated, then this implies that there is no long-run relationship amongst the rates, thus the EH can not hold.

The VAR methodology[1] proposed by CS tests the EH by computing the theoretical spread, given by forecasts of future changes in the short rate as suggested by the EH, and testing the null that the actual and theoretical spread are equal. This null can be tested by imposing non-linear restrictions implied by the EH, on a bivariate VAR containing the n- and m-period spread and the change in m-period yield. However, CS note that even small deviations from the null may lead to the VAR restrictions and hence the EH being rejected. So they propose assessing the economic significance of the EH by calculating the theoretical spread (not via imposing restrictions on the VAR) and comparing it to the actual spread through time series plots, standard deviation ratios and correlation coefficients. If the EH holds then there should be a high degree

---

[1] In this chapter we interchangeably use VAR approach and VAR method/methodology, both of which refer to the same method established by Campbell and Shiller (1987, 1991).

of comovement between the spreads and the two statistics should equal unity.

The findings of the empirical literature that tests the EH are sensitive to the country considered, testing method employed, time period examined and maturity of the data. Tests of the EH using US data generate mixed results Fama (1984a), Mankiw and Summers (1984), CS (1991), Shea (1992), Evans and Lewis (1994), Bekaert and Hodrick (2001) find that the EH is rejected for the US. Mankiw and Miron (1986) suggest that this rejection is a "consequence of the commitment of the Federal Reserve to stabilise interest rates resulting in random walk behaviour of the short rate". They argue that it is more likely support for the EH will be found under a policy of monetary targeting, than one of interest rate smoothing. This is because if a policy of interest rate smoothing results in the short rates exhibiting random walk behaviour, then we would expect there to be no change in the short rates, then in contrast to the EH the spread does not have predictive power for future short rates. Kugler (1988) finds support for Mankiw and Miron (1986) using data for Switzerland, Germany and the US. He finds under money supply targeting support for the EH is found in terms of spread having predictive power for the short rate, but this is not true when there is a policy of interest rate stabilisation. Further, support for the US is found by Engsted and Tanggard (1994b) and Longstaff (2000).

Mixed results for the UK are also observed with Mills (1991) and Taylor (1992) rejecting the EH. But Mills (1992) and MacDonald and Speight (1988, 1991) find support for the EH. Studies by Cuthbertson (1996) and Cuthbertson et al (1996) use weekly UK data for the interbank market and certificate of deposit rates respectively at the short end of the term structure. They find evidence of cointegration, however, the restrictions that the cointegrating vector is $(1, -1)$ is rejected in the certificate of deposits paper, but can not be rejected in most cases in the interbank rates paper. They find that actual and theoretical spreads move together, with the standard deviation ratios and correlation coefficients being quiet close to unity. They attribute these

results partly to using higher frequency data than earlier studies and to using pure discount bonds[2]. Cuthbertson et al (1998) use UK and German weekly data at the short end from 1981 to 1992. The hypothesis of the cointegrating vector being $(1, -1)$ in the bilateral combinations of interest rates could not be rejected. But the joint null that the set of $q - 1$ spreads form a basis for the cointegrating space is rejected for the UK. Although Cuthbertson (1996) finds support for the EH using the LIBOR rates, he rejects the VAR restrictions. However, Hurn et al (1995) who also use LIBOR rates, can not reject the VAR restrictions and find more support using the cointegration method too.

The importance of interest rate volatility for the EH is highlighted by Cuthbertson et al (2000b), further to the above mentioned early studies by Mankiw and Miron (1986) and Kugler (1988). Cuthbertson et al find that the German money markets conform closely to the EH during 1976 to 1993. They attribute this success of the EH, to the German rates being sufficiently volatile under money supply targeting. But not too volatile given the credible anti-inflation policy of the Bundesbank. In this case large interest rate changes are likely to only be observed in the event of pre-announced policy changes, due to real factors like oil price rises. Given that these events are sporadic and mostly predictable the EH will still be valid. They argue that the EH may not hold if there is a policy of interest rate smoothing, or if the rates are highly volatile-leading to a time-varying term premia and invalidating the EH with a constant term premium. Stating that econometric tests of the EH require there to be variability in expected changes in short rates. But very large changes may have the effect of increasing the perceived riskiness of holding the asset and thus invalidating the EH because of a time-varying term premium.

Further evidence is presented in Christiansen et al (2003), who state that previous studies of the Danish term structure, including Engsted and Tanggaard (1994a and

---

[2]Thus avoiding the need to make approximations as previous studies had.

1995) find some support of the EH. But generally the EH is rejected in times when the Danish rates have been highly volatile, due to the central banks policy of money supply targeting or in its attempt to support the currency during the ERM currency crisis in the early 1990s, during which large increases in the Danish rates were observed. Generally, through cointegration, the single equation method and the VAR approach, they find strong support for the EH at the short end during 1993 to 2002. In other countries favourable evidence for the EH is found by Hardouvelis (1994), Dominguez and Novales (2000) for Eurodeposits rates, Cuthbertson and Bredin (2000) using Irish data and Koukouritakis and Michelis (2008) for the 10 newest EU countries.

Although the theory appears to be intuitive the evidence in support of the EH is somewhat inconclusive, with the added complication of there being several assessment and evaluation methods of this term structure model. However, what is evident from the literature is that the results and conclusions drawn hinge upon the evaluation technique employed. The EH does provide a descriptive framework for the term structure of interest rates, but there appears to be excess volatility that the model does not explain. The literature offers two possible explanations for this; noise traders and time-varying term premium. Noise traders are those whose decisions are influenced by fads, such that decisions made are not theory informed. Hence the prevailing long rate is not just a combination of current and future short rates as suggested by the EH, but of stochastic noise too. Secondly, the EH assumes the term premium is constant through time[3], but it could be time varying. Developments in the literature since CS include new testing methods of the EH, modelling time-varying term premia, regime changes and macro-finance models of the term structure[4]

In this chapter we model the UK term structure at the short end and test the EH using weekly data for 1-, 3-, 6- and 12-month zero-coupon bonds over the period 1997

---

[3]The Pure Expectations Hypothesis (PEH), assumes that the term premium is zero. Whereas the EH assumes that it is constant for a given pair of maturities.

[4]See Bekaert and Hodrick (2001), Clarida et al (2006), Rudebusch and Wu (2008), Sarno et al (2007) and Della Corte et al (2008). See Chapter 2 for more detail.

to 2004. More specifically, we model the term structure using four models that incorporate varying degrees of economic theory. At one end of the spectrum the atheoretic Autoregressive (AR) model considers each yield in isolation, followed by the Vector Autoregressive in Differences (VARD) model which allows for some interaction amongst them. At the other end, the theory informed VAR in Transformed Interest Rates (VART) and Vector Error Correction (VECM) models embed economic theory that implies a long-run relationship exists between the yields, and model the dynamic relationship of the yields along the curve. The models are considered under both bivariate and multivariate specifications where appropriate. We make a formal comparison of these models to ascertain if a statistical or a theory informed model is best placed to explain the term structure. Further, we examine how well the EH is able to capture the UK money market using the cointegration method and the VAR approach.

Although the EH has been previously examined using UK data at the short end, as detailed in Chapter 2, these studies utilise pre-1997 data. As such the contributions of this chapter are empirical, we employ a more recent UK dataset, post-1997 to test the EH and examine if more support for the EH can be found over our sample in comparison to earlier studies. Further, we test the significance of economic theory for explaining the term structure.

We find the in-sample properties of the VECM and VART models suggest that they have greater explanatory power for the term structure compared to the statistical AR and VARD models. Further, three cointegrating relationships amongst the four yields are found, with the restrictions that the cointegrating vector between each pair of yields is $(1, -1)$, as suggested by the EH, can not be rejected. Under the VAR approach, although the Wald tests indicate a statistical rejection of the EH in almost all cases, a divergence between the actual and theoretical spread is not apparent. We find evidence to suggest the EH provides a good approximation of the long-run dynamics of the UK yield curve at short end.

The UK since 1992 has adopted a policy of inflation targeting. Further, in 1997 the Bank of England was granted independence and made responsible for monetary policy to ensure the objective of price stability is achieved. These changes were made with the intention of promoting transparency and giving credibility to the anti-inflation policy[5]. Interest rates since 1992 and particularly after 1997 have become considerably less volatile compared to the levels observed in the 1970s and 80s, as discussed in Chapter 2. This reduction in the volatility could be in part due to the above changes in the monetary policy regime. But as appears more likely now, were largely due to the stable economic climate of the "nice" decade observed until recently, Hall and Henry (2000) and King (2008)[6].

We offer this reduction in the volatility of interest rates as a potential explanation for the strong support we find for the EH in comparison to earlier UK studies that use pre-1997 data. We suggest that the interest rates observed during our sample, 1997 to 2004 are sufficiently volatile for the EH to hold. But not too volatile, as observed in previous decades, as to invalidate the EH with a constant term premium.

The organisation of this chapter is as follows, the EH is discussed in Section 3.2, with a description of the term structure models to be estimated in Section 3.3. An overview of the data, together with the results of the estimated models, a formal comparison of these models and an analysis of the dynamic properties in Section 3.4. Results from the VAR approach of the EH in Section 3.5 and to finally conclude in Section 3.6.

3.2 Expectations Hypothesis of the Term Structure

Recall from Chapter 2 the EH in a linearized form, Campbell and Shiller (1991), states that the return on a n-period zero-coupon bond $r_t^{(n)}$ should equal the return on a rolling

---

[5]See Chapter 2 for more detail.

[6]Mervyn King, the Governor of the Bank of England uses the term "nice" to describe the steady growth and low inflation observed in the UK since 1997 until the recent slow down. Nice standing for "non-inflationary consistent expansion".

investment in a sequence of k m-period bonds, plus a term/liquidity premium $c^{(n,m)}$ that may vary with maturity, i.e. with n and m, but is time-invariant. Where $r_t^{(m)}$ is the return from a m-period bond and the integer $k = n/m$ and $n > m$. This for spot yields on zero coupon bond is the 'fundamental' term structure equation

$$r_t^{(n)} = \frac{1}{k} \left[\sum_{i=1}^{k} E_t \left(r_{t+(i-1)m}^{(m)} \right) \right] + c^{(n,m)} \tag{3.1}$$

The expectations operator conditional on information available at t is denoted by E_t. Thus in equation (3.1), the long rate is given by an average of the expected future short rates plus a term premium. Further, the EH can also be formulated as

$$s_t^{(n,m)} = \sum_{i=1}^{k-1} \left(1 - \frac{i}{k} \right) E_t \left(\Delta^m r_{t+im}^{(m)} \right) + c^{(n,m)} \tag{3.2}$$

this describes the spread by expected future changes in the short rate[7]. That is, aside from the constant premium, the spread is a reflection of the expected change in the short term rates over the life of the long bond. Hence a spread arises if the short rate is expected to change or if there is a term premium. If the yields contain a stochastic trend, then from above, the spreads should be stationary[8] if the EH holds. The validity of the EH can be tested as follows[9]:

1. If the yields share a common stochastic trend, then in a set of q non-stationary yields we should find $(q-1)$ cointegrating vectors, as implied by stationary bi-variate spreads.

---

[7] Where the long-short yield spread is $s_t^{(n,m)} = r_t^{(n)} - r_t^{(m)}$ and the m period change in the one-period rate is $\Delta^m r_t = r_t - r_{t-m}$.

[8] Assuming that $r_t^{(n)}$ is I(1) and $c^{(n,m)}$ is I(0), the right-hand side of equation (3.2) is stationary and this implies that the left-hand side of equation (3.2) which gives the spread between the n- and m-period rate should also be stationary.

[9] See Drakos (2002) and Christiansen et al (2003).

2. Each of the 3-, 6- and 12-month yields are cointegrated with the 1-month yield such that the cointegrating vector is of the form $\left(1, -1, c^{(n,m)}\right)'$, with the liquidity/term premium free from restriction.

3. The Pure Expectations Hypothesis states that the liquidity premium is zero, this can be tested through the imposition of the restrictions that the premia are zero, such that the cointegrating vector is now $(1, -1, 0)'$.

Here we conduct these three tests use the Johansen maximum likelihood estimation procedure, Johansen (1988, 1991) and Johansen and Juselius (1990). We test the EH using cointegration techniques in a multivariate framework, rather than a bivariate as in many studies. As highlighted in Drakos (2002, pp .42) "this is likely to produce informational efficiency gains since one would expect term structure innovations to diffuse across the yield curve and not be confined to a pair of maturities."

3.2.1 VAR Methodology

Another test of the EH conducted in the literature[10] examines if the spread has predictive power for future interest rate changes[11]. This predictive power is conventionally examined using a single equation framework, however as explained in Cuthbertson et al (2000a) and Christiansen et al (2003), this method has some drawbacks[12]. Given the shortcomings of these singles equation tests, we test the EH using the cointegration method described above and the VAR approach. The VAR approach proposed by CS

---

[10]See CS, Cuthbertson et al (1996, 2000b) and Christiansen et al (2003).

[11]According to the EH with constant term premium, the spread between a n- and m-period rate should equal (1) the weighted average of future changes in the m-period rate over n periods and (2) the m-period change in the n-period rate. If the EH holds then the spread is a reflection of the markets expectation of, ignoring the term premium, (1) changes in the short rate over the life of the long bond and (2) changes in the long rate over the life of the short bond. See Chapter 2.

[12]When n is large relative to m in (1) or when m is large in (2) then the "degree of time-overlap becomes large and hence the error-terms become long moving average processes. This has the unfortunate implication that the Hansen (1982) and Newey and West (1987) corrections become very unreliable." i.e. the GMM corrected covariance matrix. Further, the sample size is substantially reduced with large n and m, such that the "order of the MA-error becomes extremely high, making " the regression for (1) "more or less dubious". Christiansen et al (2003, all quotes taken from pp. 6).

provides an alternative method to examine the predictive power of the spread for future changes in the short rate. As stated in Christiansen et al (2003) the VAR model is not plagued by the problem of time-overlapping variables (does not contain them) or with the sample size being significantly reduced as n increases, since with this approach a well-specified system can be achieved with a limited number of lags. Moreover, inference based on the VAR method is generally more reliable than that derived from the single equation method, Hodrick (1992, cited in Christiansen et al (2003)).

From Cuthbertson et al (1996 and 2000a), if $\mathbf{z}_t = \left(s_t^{(n,m)}, \Delta r_t^{(m)} \right)'$ is a vector of stationary variables, then there exists a bivariate Wold representation which can be approximated by a Vector Autoregression (VAR) of order p, which as a VAR(1) in companion form is

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \boldsymbol{\epsilon}_t \tag{3.3}$$

where $\mathbf{z}_t = \left(s_t^{(n,m)}, \Delta r_t^{(m)}, ..., s_{t-p+1}^{(n,m)}, \Delta r_{t-p+1}^{(m)} \right)'$ is $(2p \times 1)$, $\mathbf{A}$ is a $(2p \times 2p)$ matrix of coefficients and $\boldsymbol{\epsilon}$ is $(2p \times 1)$. Further, the $(2p \times 1)$ selection vectors $\mathbf{e1} = (1, 0, ..., 0)'$ and $\mathbf{e2} = (0, 1, 0, ..., 0)$ i.e. with unity in the first and second rows of $\mathbf{e1}$ and $\mathbf{e2}$ respectively and zero elsewhere, such that $s_t^{(n,m)} = \mathbf{e1}'\mathbf{z}_t$ and $\Delta r_t^{(m)} = \mathbf{e2}'\mathbf{z}_t$. The chain rule of forecasting is used to obtain future values of $\mathbf{z}_t$. Testing the EH requires forecasts of changes in the short rate obtained by projecting $\mathbf{z}_{t+i}$ on the restricted information set $\mathbf{D}_t$, where $\mathbf{D}_t \subset \boldsymbol{\Omega}_t$ gives

$$E_t \left(\mathbf{z}_{t+i} | \mathbf{D}_t \right) = \mathbf{A}^i \mathbf{z}_t \tag{3.4}$$

$$E_t \left(\Delta r_{t+i}^{(m)} | \mathbf{D}_t \right) = \mathbf{e2}' \mathbf{A}^i \mathbf{z}_t \tag{3.5}$$

Cuthbertson et al (1996) refer to this as the weakly rational expectations prediction of future $\Delta r_t^{(m)}$, because the limited information set $\mathbf{D}_t$ is used as selected by the econometrician and not the full information set $\boldsymbol{\Omega}_t$. Further, they show that the

m-period change in the short rate can be written as the sum of m one-period changes

$$\Delta^m r_{t+im}^{(m)} = \sum_{j=q}^{im} \Delta r_{t+j}^{(m)} \tag{3.6}$$

where $q = m(i-1) + 1$, from equations (3.5) and (3.6)

$$E_t \left(\Delta^m r_{t+im}^{(m)} \right) = \mathbf{e2}' \sum_{j=q}^{im} \mathbf{A}^j \mathbf{z}_t \tag{3.7}$$

substituting equation (3.7) into (3.2) gives the theoretical spread

$$s_t^{(n,m)*} = \mathbf{e2}' \sum_{i=1}^{k-1} (1 - i/k) \sum_{j=q}^{im} \mathbf{A}^j \mathbf{z}_t \tag{3.8}$$

which is a function of the estimated VAR parameters contained in the matrix $\mathbf{A}$. From CS (1987, 1991) the VAR non-linear restrictions as implied by the EH are

$$\mathbf{e1}' - \mathbf{e2}'\mathbf{A} \left[I - (m/n)(I - \mathbf{A}^n)(I - \mathbf{A}^m)^{-1} \right] (I - \mathbf{A})^{-1} = 0 \tag{3.9}$$

where the above VAR methodology can be used as follows to compute the theoretical spread $s_t^{(n,m)*}$

$$s_t^{(n,m)*} = \mathbf{e2}'\mathbf{A} \left[I - (m/n)(I - \mathbf{A}^n)(I - \mathbf{A}^m)^{-1} \right] (I - \mathbf{A})^{-1} \mathbf{z}_t \tag{3.10}$$

The EH states that the theoretical spread $s_t^{(n,m)*}$, is given by the weighted sum

of the optimal forecasts of the changes in the short rates. If the EH holds then the theoretical spread should equal the actual. A formal test of the VAR parameter restrictions given in equation (3.9) provides a test of the hypothesis $s_t^{(n,m)*} = s_t^{(n,m)}$, however as highlighted in Cuthbertson (1996) and Cuthbertson et al (2000a), CS note that even small deviations from this null may lead to a rejection of the EH under these formal tests of the VAR restrictions.

CS hence propose an alternative to imposing the above non-linear VAR restrictions. They suggest computing the theoretical spread by using the VAR estimates of the $\mathbf{A}$ matrix[13] to generate forecasts of future changes in the short rate as in equation (3.7), these are then substituted into equation (3.8), this theoretical spread is then compared to the actual. The null that the EH holds implies:

1. The restrictions under (3.9) hold, such that the spread is an optimal predictor of future changes in the short rates and the information beyond that contained in the spread at t, should not help predict these future changes.

2. Plots of $s_t^{(n,m)*}$ and $s_t^{(n,m)}$ over time should move together.

3. More formally, the degree of comovement between $s_t^{(n,m)*}$ and $s_t^{(n,m)}$ can be measured by the standard deviation ratio $SDR = \sigma\left(s_t^{(n,m)*}\right)/\sigma\left(s_t^{(n,m)}\right)$ and the correlation coefficient $Corr\left(s_t^{(n,m)*}, s_t^{(n,m)}\right)$, which should both equal unity under the EH[14].

By comparing the theoretical and actual spreads as CS suggest using plots, the SDR

---

[13] See Appendix 2.

[14] From Cuthbertson et al (1996, footnote 10) the standard error (SE) of the SDR and the correlation coefficient are non-linear functions of the matrix $\mathbf{A}$ from the estimated VAR, and can be computed using:

$$SE = f(\gamma)' \mathbf{\Psi} f(\gamma)$$

where the statistic for which the SE is being computed is denoted $f(\gamma)$, and $\mathbf{\Psi}$ is the GMM variance-covariance matrix of the parameters γ. In the case of the SE of the SDR, $f(\gamma)$ is a matrix of the SE of the estimated coefficients/parameters γ. For the SE of the correlation coefficient, $f(\gamma)$ is a correlation matrix of the parameters. GMM estimation is used to compute the variance-covariance matrix of the variables in the VAR system, corrected for heteroskedasticity and autocorrelation.

and the correlation between them, the extent to which the predictions made by the EH are close to actual observations can be established. The existence of a high degree of comovement suggests that the actual spread does largely reflect the markets perception of future changes in the short rate.

3.3 Models of the Term Structure

In the previous section we discussed the EH and the two methods by which we will test this hypothesis of the term structure. However, in this investigation we are also concerned with modelling the term structure generally. In this section we describe a set of atheoretic and theory informed models including an Autoregressive Model (AR), Vector Autoregressive Model in Differences (VARD), VAR in Transformed Interest Rates Model (VART) and a Vector Error Correction Model (VECM)[15]. Each model is discussed in Chapter 2, so here we only recall the key equations.

3.3.1 Autoregressive (AR) Model

The change in the n-month yield $\Delta r_t^{(n)}$ can be modelled by a pth order AR model, if the yields are difference stationary, for $n = 1, 3, 6, 12$

$$\Delta r_t^{(n)} = \beta + \sum_{i=1}^{p} \Psi_i \Delta r_{t-i}^{(n)} + u_t \tag{3.11}$$

where the white noise process $u_t \sim i.i.dN(0, \boldsymbol{\sigma^2})$, $E(u_t) = 0$, $E(u_t^2) = \boldsymbol{\sigma^2}$ and $E(u_t u_{t-s}) = 0$ for all $s \neq 0$ [16], β is the intercept and Ψ_i are the coefficients that capture the influence of past changes on $\Delta r_t^{(n)}$.

---

[15]None of the yields exhibit a trend, so the trend terms are excluded from the models.

[16]The unobservable errors u_t are independently and identically distributed random variables that are homoskedastic and serially uncorrelated

Two different modelling approaches can be explored from this standard statistical AR model. The first is an 'economic' approach, where theory informed term structure models can be developed as given by the VART and VECM models; the second is a 'statistical' route where the above AR model can be extended to a VARD model. Further, we demonstrate how the models described are nested within each other. A bivariate framework is used to discuss each model, however this can be extended to a multivariate one.

3.3.2 VAR in Transformed Interest Rates (VART) Model

If yields are difference stationary and there exists a cointegrating relationship between n- and m-period yields, such that spreads are stationary, then there exists a bivariate Wold representation which can be approximated by the following VAR(p) model. Bold face is used to represent a vector.

$$
\begin{pmatrix} s_t^{(n,m)} \\ \Delta r_t^{(m)} \end{pmatrix} = \beta + \mathbf{B}_1 \begin{pmatrix} s_{t-1}^{(n,m)} \\ \Delta r_{t-1}^{(m)} \end{pmatrix} + ... + \mathbf{B}_p \begin{pmatrix} s_{t-p}^{(n,m)} \\ \Delta r_{t-p}^{(m)} \end{pmatrix} + \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix} \tag{3.12}
$$

Equation (3.12) describes the VAR in Transformed Interest Rates (VART) model which simultaneously models the short m-period yields and the n- and m-period spread. Here $\Delta r_t^{(m)}$ and $s_t^{(n,m)}$ are described by their past values, for $n = 3, 6, 12$ and $m = 1, 3, 6$, where $n \neq m$.

3.3.3 Vector Error Correction Model (VECM)

The above VART model is composed of $\Delta r_t^{(m)}$ and $s_t^{(n,m)}$, which are both stationary and can be written in levels, see Chapter 2 and Appendix 3, manipulation of this levels form yields the following VECM representation

$$\Delta \mathbf{z}_t = \boldsymbol{\mu} + \sum_{i=1}^{p} \boldsymbol{\Gamma}_i \Delta \mathbf{z}_{t-i} - \boldsymbol{\Pi} \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t \qquad (3.13)$$

where $\mathbf{z}_t$ is a $q \times 1$ vector of stochastic I(1) variables in this case the yields r_t, $\boldsymbol{\mu}$ is a $q \times 1$ vector of constants, $\boldsymbol{\varepsilon}_t$ is a $q \times 1$ vector of white noise errors where $\boldsymbol{\varepsilon}_t \sim i.i.d.N(0, \boldsymbol{\Sigma})$, $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Pi}$ are $q \times q$ matrices of coefficients. The rank of $\boldsymbol{\Pi}$ given by r determines the number of stationary linear combinations of $\mathbf{z}_t$. The long-run matrix $\boldsymbol{\Pi}$ has reduced rank $r < q$ if the yields are cointegrated. Thus, under cointegration there exists r cointegrating vectors among the q yields contained in the $q \times r$ matrix $\boldsymbol{\beta}$, such that $\boldsymbol{\Pi} = \boldsymbol{\alpha}' \boldsymbol{\beta}$. The long-run matrix of coefficients $\boldsymbol{\beta}'$ contains the cointegrating vectors and $\boldsymbol{\alpha}$ is a $q \times r$ matrix of error-correction coefficients. Under cointegration even though $\mathbf{z}_t$ is non-stationary, the error correction terms $\boldsymbol{\beta}' \mathbf{z}_{t-1}$ give the stationary relationships that exist amongst these non-stationary variables.

On the $\boldsymbol{\beta}$ matrix, r^2 *exactly identifying* restrictions are imposed, since these are not sufficient to impose the complete structure suggested by the EH, the VECM is then estimated subject to further *over-identifying* restrictions. We can test the hypothesis that the cointegrating relationship between the n- and m-period yields has cointegrating vector $(1, -1)$, as implied by the EH such that spreads are stationary, by comparing the likelihood of the VECM subject to the exactly identifying restrictions to that of the VECM subject to a full set of restrictions[17]. If indeed a long-run relationship of this form exists, then this justifies the estimation of the VART model which is based on this assumption holding.

As previously noted, if the EH is valid then there should exist $r = q-1$ cointegrating vectors, whereby "the cointegration space should be spanned by the columns of the matrix such that the bivariate interest rate spreads are stationary." Christiansen et al (2003, pp. 9). With $\mathbf{z}_t = \left(r_t^{(12)}, r_t^{(6)}, r_t^{(3)}, r_t^{(1)} \right)'$ we test for three cointegrating relations

---

[17]The full set comprises exact and over-identifying restrictions.

between 12-, 6- and 3-month each with the 1-month[18]. With the restricted matrix of the cointegration vectors

$$\beta' = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \tag{3.14}$$

this is the first set of over-identifying restrictions, where each row corresponds to the cointegrating vector $\left(1, -1, c^{(n,m)}\right)$ as implied by the EH and the term premium $c^{(n,m)}$ is free from restriction. The second set of over-identifying restrictions test the Pure EH (PEH), which assumes the liquidity premia are zero, where the cointegrating vector corresponds to $(1, -1, 0)$. The PEH could comply with our short term interest rate data for the money markets, since the maturities of the assets are close together such that we do not expect to observe a large premium.

As shown the VART model given by equation (3.12) can be written as a VECM, where the VECM captures the cointegrating relations that exist between the yields and the error correction term ensures that the long and short yields do not deviate greatly from each other in the long-run.

3.3.4 VAR in Differences (VARD) Model

From equation (3.13) if $\mathbf{z}_t$ is I(1) and not cointegrated then $\Pi = 0$, where no long-run relationships are found to hold amongst the yields in the form of a stationary spread. In this case a VAR in Differences (VARD) model of order p is appropriate to describe the system

$$\Delta\mathbf{z}_t = \mathbf{a}_o + \sum_{i=1}^{p} \boldsymbol{\theta}_i \Delta\mathbf{z}_{t-i} + \mathbf{e}_t \tag{3.15}$$

---

[18]Confirmation of these three cointegrating relationships exisiting, will provide support for the existence of a cointegrating relation between all other possible pairwise combinations of these yields.

this simultaneously models the change in the n- and m-period yields at time t, where $n = 3, 6, 12$, $m = 1, 3, 6$ and $n \neq m$. Here $\mathbf{z}_t$ is a $q \times 1$ vector containing the n- and m-period yields, $\mathbf{a}_o$ is a $q \times 1$ vector of intercepts and $\mathbf{e}_t$ is a $q \times 1$ vector of white noise errors, with $\mathbf{e}_t \sim i.i.d. N(0, \boldsymbol{\Sigma})$. No exogenous variables are considered in this model.

This VARD(p) model can be transformed to an AR(p) specification by imposing restrictions on θ_i. If $n = 1, 3, 6$ and 12, and $j = 1, 3, 6$ and 12, by setting $\theta_i = 0$ for all $\Delta r_{t-i}^{(j)}$ where $j \neq n$ in each $\Delta r_t^{(n)}$ equation in the VARD(p) model, results in the change in the n-month yield depending only on its past values.

The above models are described for the bivariate case and where possible in a general form in terms of $\mathbf{z}_t$. The vector of variables $\mathbf{z}_t$ in the multivariate cases contain all yields with $q = 4$ and $\mathbf{z}_t = \left(r_t^{(12)}, r_t^{(6)}, r_t^{(3)}, r_t^{(1)} \right)'$. For the bivariate models $q = 2$, so the four yields give rise to six possible bivariate combinations of $\mathbf{z}_t$.

This section describes four different models that can be used to explain the term structure of interest rates. As seen a simple AR model can be extended in two ways, first by continuing down the atheoretic path to a VARD model, or secondly through the incorporation of economic theory to a VART model. The VECM and VART representations are equivalent, where the VECM explicitly tests for cointegrating relationships amongst the yields and the VART model assumes these relationships hold and uses spreads directly. Further, it can be seen that both the AR and VARD models are nested within the VECM, and the Bivariate VARD and AR models are nested within the Multivariate VARD model. Although these models can be described as either statistical or theory based, some specifications are nested within others, this allows the unrestricted models to be compared to the restricted, as will be done in the next section.

3.4 Modelling the UK Term Structure

3.4.1 Data

Zero coupon bonds or risk free discount bonds are used in the construction of yield curves and in empirical studies of the term structure, because it is desirable for the instruments to differ only in their term to maturity. Here we use a weekly dataset, this allows for a closer approximation of the information available to traders in comparison to monthly or quarterly observations. UK data for maturities of 1, 3, 6, 12 months over the period 1997 week 10 to 2005 week 18 is employed, specifically Wednesday observations of the nominal government spot rates, giving a total of 427 observations for each maturity, all yields are continuously compounded and annualised.

The data is official Bank of England (BoE) data on the Government liability curve. We use the reported daily data for the nominal spot rates curve at the short end, from which we select the Wednesday observations. This nominal zero coupon yields data is calculated using gilt prices and General Collateral (GC) repo rates. From the BoE data notes these n-month nominal government spot interest rates refer to those applicable today on a n-month risk-free nominal loan and by definition this (the nominal government spot rate) is the yield to maturity of a nominal zero coupon bond[19].

Figure 3-1 illustrates the time series characteristics of the yields. In general, all the series appear to decline until 2003, after which an upward trend is apparent. At the beginning of the sample the yields are around 6%, increasing quite steadily throughout 1997 and the first half of 1998, before a sharp decline was experienced during the latter quarter of 1998. This decline continued into the middle of 1999 with yields averaging 4.8%, increasing until the first quarter of 2000 to levels just shy of 6.4% for the 6- and 12-month yields. After which the yields declined gradually until the second quarter of

---

[19]Here we use estimated yield curve data, official data estimated by the Bank of England, see Data Appendix for further details.

2003, with the lowest yields of the sample period considered being observed at 3.16% and 3.17% for the 1- and 3-month, and 3.13% and 3.11% for the 6- and 12-month yields, respectively. This decline was followed by an upturn to levels ranging 4.26% to 4.65% for the maturities considered.

The first differences of the yields, $r_t^{(n)} - r_{t-1}^{(n)}$, are graphed in Figure 3-2. Each of the series clearly exhibit mean reverting behaviour with frequent crossings of the mean, characteristics consistent with stationary processes.

The EH suggests that there exists a long-run relationship between $r_t^{(n)}$ and $r_t^{(m)}$ such that the prevailing spread is stationary. Given the four maturities considered; namely the 1-, 3-, 6- and 12-month yields[20] six yields spreads can be defined. If these six spreads follow an I(0) process then this is support for the EH. From Figure 3-3 the spreads fluctuate about the mean with reasonably frequent crossings of their respective means. Also apparent are the persistent deviations from the mean, suggesting significant differences between the pair of yields remain. However, mean reversion is evident and a more formal method of assessment is required. Table 3.1 provides the descriptive statistics for the yields and spreads considered over the sample period.

3.4.2 Unit Roots

Interest rates are usually highly persistent, whereby it takes a stationary series a long time to revert back to its mean. Hence, statistically the null of a unit root in short-term interest rates often can not be rejected. However, from an economic stance it is difficult to "defend non-stationarity of interest rates" Verbeek (2000, pp. 263), Drakos (2002) and Christiansen et al (2003). As such, empirical investigations differ in their treatment of interest rates depending on the assumptions made regarding the order of integration.

---

[20]Recall that the long n-period bond yield is denoted $r_t^{(n)}$, the short m-period bond yield by $r_t^{(m)}$ and the spread between them $s_t^{(n,m)} = r_t^{(n)} - r_t^{(m)}$. With $n > m$ and $n, m = 12, 6, 3, 1$. Note that the 9-month rate was excluded from the empirical analysis because we want $k = n/m$ to take an integer value.

The Augmented Dickey-Fuller (ADF), Phillips Perron (PP) and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) unit root tests are employed to determine the order of integration of the yields and the spreads[21]. The three unit root tests are performed over the sample period 1997 week 10 to 2005 week 18, results of which are summarised in Table 3.2 and Table 3.3. For each of the yields the null of a unit root can not be rejected under the ADF and PP tests at the 5% level of significance, equivalently the null of stationarity is rejected under the KPSS test at the 1% level. For the differenced yield series, the null of a unit root is rejected and the null of stationarity can not be rejected at the 1% level. Hence the 1-, 3-, 6- and 12-month rates are integrated of order one, in that they are difference stationary.

The spreads between the (n, m) month maturities for $(3, 1)$, $(6, 1)$, $(12, 1)$ and $(6, 3)$ are found to be stationary by all three unit root tests. However, when testing the $(12, 3)$ and $(12, 6)$ spreads, the null of a unit root could not be rejected at the 10% level under the ADF or PP test, implying that these spreads are integrated of order one. But under the KPSS test the null of stationarity could not be rejected. The failure to reject the null of a unit root may be due to the sensitivity of the tests to the persistent deviations from the mean. But it can be seen that the statistics produced by each of the ADF and PP tests are close to their respective critical values, such that they fall just short of rejecting the null of non-stationarity. Equally, from Figure 3-3 it can be seen that mean reversion is observed by all the spreads.

---

[21]Phillips and Perron (1987) propose the PP test as an alternative to the ADF unit root test, in which the t-statistics of the Dickey-Fuller regressions are adjusted to take account of possible autocorrelation in the errors, as opposed to the inclusion of additional lags in the regressions with the aim of achieving an error term absent of autocorrelation as in the ADF test. Kwiatkowski, Phillips, Schmidt and Shin (1992) suggest that the issue of low power unit root tests can be circumvented and proposed the KPSS test with the null hypothesis of stationarity and the alternative of a unit root. This is contrary to the ADF and PP tests, which test the null of a unit root against the alternative of a stationary process.

3.4.3 Estimation of the Term Structure Models

In this section we present the estimation results, compare the estimated statistical models to the theory informed ones and evaluate their dynamic properties. We estimate the following previously discussed models for the term structure:

- AR(p) model: for $\Delta r_t^{(n)}$ where $n = 1, 3, 6, 12$.

- VART(p) model: the bivariate model denoted BVART$^{(n,m)}(p)$ for each of the six bivariate (n, m) combinations. The multivariate MVART(p) model simultaneously models $s_t^{(12,1)}, s_t^{(6,1)}, s_t^{(3,1)}$ and $\Delta r_t^{(1)}$.

- VECM(p) model: for the multivariate case where $\mathbf{z}_t = \left(r_t^{(12)}, r_t^{(6)}, r_t^{(3)}, r_t^{(1)} \right)'$.

- VARD(p): the multivariate model MVARD(p) for $\mathbf{z}_t = \left(r_t^{(12)}, r_t^{(6)}, r_t^{(3)}, r_t^{(1)} \right)'$ and six bivariate models BVARD$^{(n,m)}(p)$ for each pairwise (n, m) combination.

The models are estimated over the period 1997 week 10 to 2004 week 18, using 374 observations[22]. Residual serial correlation of order 9, normality of residuals and equality of error-variances are all tested for in each regression[23], these diagnostics are presented with the estimates.

The estimated models are all of the same order p, in that they all possess the same lag structure, see Appendix 4. With the same lag structure the models only differ by the restrictions imposed on them, e.g. if $\mathbf{\Pi} = 0$ in the VECM this yields the VARD model of the same order, this allows for tests of the restrictions to be conducted. A sufficiently high order AR and unrestricted VAR model is run over the sample period, the chosen order p is that which maximises the Akaike Information Criterion (AIC), the Schwarz Bayesian Criterion (SBC), and the order for which the null that the restrictions are

---

[22]The remaining 53 observations are reserved for the forecasting exercises in Chapters 4 and 5, in which the dataset is extended to include observations up to 2007 week 19.

[23]With the exception of the AR regressions, VAR models of various specifications are estimated, so the diagnostic tests are conducted on the whole system rather than the individual regressions.

correct can not be rejected, as given by the Likelihood Ratio Test (LRT). An AR(12) is run for each $\Delta r_t^{(n)}$, the three criteria suggest an order of 9 for the AR models of the 1- and 3-month yields, and order 6 for the 6- and 12-month yields. Further, an unrestricted multivariate VAR(12) in differences is run to determine the order of the underlying VAR, the AIC and LRT suggest an order of 9, and the SBC indicates 1. We choose $p = 9$ as the order for all the estimated models, because with a large sample the implications of choosing a higher order is less damaging than that of choosing a lower one.

After the order of the underlying VAR has been determined the VECM requires that the number of cointegrating relations are tested for using the maximum eigenvalue, trace, SBC and AIC statistics. The outcome of these tests are often 'supplemented' with the number implied by economic theory, as results can be inconclusive and may not coincide with theory[24]. Maximum likelihood is used to estimate the model subject to the restrictions imposed, exact and over, where these restrictions are tested using the χ^2 with degrees of freedom corresponding to the number of over-identifying restrictions.

Results

Table 3.4 describes the results for the simplest model estimated for the four yields; namely the AR(9) model. The coefficients are jointly found to be significantly different from zero at the 1% level for the $\Delta r_t^{(1)}, \Delta r_t^{(3)}$ and $\Delta r_t^{(6)}$ equations and at the 10% for the $\Delta r_t^{(12)}$ equation, with explanatory power of the AR regressions between 1.7% and 13.2%.

From Tables 3.5 and 3.6 the simple AR framework is extended to the multivariate VARD model, under which the explanatory power for all but the $\Delta r_t^{(12)}$ equation increases and the coefficients are jointly significant for all equations at the 1% level except

---

[24]With r cointegrating vectors amongst the q variables, r^2 restrictions i.e. r on each of the r cointegrating vectors are imposed to exactly identify the long-run relationships. Once these exactly identifying restrictions have been imposed, then further over-identifying restrictions can be imposed to test the economic theory from which these restrictions have come.

for the $\Delta r_t^{(12)}$ equation. Tables 3.7 to 3.13 show that under the bivariate specification estimated for each pairwise combination of $\Delta r_t^{(n)}$, the coefficients are jointly significant at the 1% level for all of the equations except for those of $\Delta r_t^{(12)}$, for which the null that the coefficients are jointly no different from zero could not be rejected.

Moving to the theory informed models, a VECM of order 9 is estimated, using the Johansen (1988) procedure, containing restricted intercepts and no trend coefficients, Tables 3.14 and 3.15. The cointegration rank statistics presented in these tables show that under the Trace and Maximum Eigenvalue Statistics the null that $r \leqslant 2$ can not be rejected at the 5% level of significance. The AIC, SBC and the Hannan-Quinn Criteria (HQC) are maximised when $r = 3$, $r = 0$ and $r = 2$ respectively. Since three cointegrating relations are expected amongst the four yields we impose $r = 3$, followed by nine exactly identifying restrictions, Table 3.16. The estimates of the Reduced Form Error Specification with the exactly identifying restrictions imposed are presented in Tables 3.6 and 3.18. For the $\Delta r_t^{(1)}$, $\Delta r_t^{(3)}$ and $\Delta r_t^{(6)}$ equations the estimated coefficients are found to be jointly significant at the 1% level, but are not significant for the $\Delta r_t^{(12)}$ equation. The error correction coefficients $\boldsymbol{\alpha}$, for each error correction term denoted $ect1_{t-1}$, $ect2_{t-1}$ and $ect3_{t-1}$ corresponding to the three long-run relationships between the yields in $\boldsymbol{\beta}'\mathbf{z}_{t-1}$, are all found to be significant suggesting that the short-run behaviour of each Δr_t is significantly affected by all three long-run relationships. The $\overline{R}^2$ ranges from 3 to 30%.

The VECM model is then estimated subject to two sets of over-identifying restrictions, Table 3.19. The first set tests for cointegration of each of the 3-, 6- and 12-month yields with the 1-month, such that the cointegrating vector corresponds to $(1, -1, c^{(n,m)})$. The second set tests the hypothesis that the liquidity premia are zero, so now the cointegrating vector corresponds to $(1, -1, 0)$. The log likelihood of the VECM subject to exact identifying restrictions is 9645.3, and 9642.1 and 9639.1 subject to the first and second set of over-identifying restrictions respectively. A Likelihood

Ratio Test of these over-identifying restrictions yields $\chi^2[3] = 6.55$ for the first set and $\chi^2[6] = 12.53$ for the second[25], both of which can not be rejected at the 5% level of significance. This implies that the cointegrating relations amongst the yields are of the form (1,-1) as suggested by the EH and further the liquidity premia are not significantly different from zero. In which case we find evidence in support of the Pure EH.

Confirmation of these long-run relations amongst the yields justifies the estimation of the MVART(9). From Tables 3.20 and 3.21, the coefficients of $s_t^{(12,1)}, s_t^{(6,1)}, s_t^{(3,1)}$ and $\Delta r_t^{(1)}$ are all found to be jointly significant at the 1% level of significance. With $\overline{R}^2$ for the spread equations ranging from 83% to 93%, and that of the $\Delta r_t^{(1)}$ equation being slightly lower than that obtained under the VECM specification. Tables 3.22 to 3.28 show the BVART(9) model estimates for the six spread and yield combinations. We observe a high explanatory power for the spread equations, and the explanatory power for the $\Delta r_t^{(m)}$ equations are of the same magnitude as that observed under other specifications.

Inspecting the diagnostics reveals the null that the residuals are normal is rejected for almost all the estimated equations. There is also evidence of heteroskedasticity in almost all equations apart from those of the VECM[26] and further of serial correlation under the BVARD and MVARD models. The problem of residuals being non-normal does not undermine the parameter estimates of the models, since at the estimation stage the Least Squares method is used which does not rely on the normality of errors. The presence of heteroskedasticity and non-normal errors is not surprising here given that we are using financial data. In our assumption of homoskedastic and normal residuals, we follow the literature who also utilise such data. Given this evidence, the regressions are estimated with Newey-West heteroskedastic and autocorrelation corrected errors, and it is these corrected standard errors that are reported in the tables.

---

[25] The degrees of freedom are given by the total restrictions (in this case 12) minus the exactly identifying restrictions (here 9).

[26] Generally OLS estimates in the presence of heteroskedasticity have standard errors that are too low, this overestimates t-statistics and hence increases the chance of incorrectly rejecting the null.

From the models estimated it is apparent that the theory informed models have greater explanatory power than the statistical AR and VARD models. This holds whether it be the VECM which explicitly defines and tests the long-run relations or the VART models that are based on the assumption that these long-run relations hold. Additionally, the multivariate forms of the respective models are better able to describe the yields compared to the bivariate equivalents.

The VART specification has greater explanatory power than the VECM model, even though they are theoretically equivalent. This is because the variables being modelled and the estimated equations under each are not the same. The VECM has equations of $\Delta r_t^{(n)}$, whereas the VART has equations of $s_t^{(n,1)}$ and $\Delta r_t^{(1)}$. The spreads exhibit less volatility than the changes in yields, allowing the movements of the spreads to be captured and modelled with greater accuracy.

3.4.4 Comparison of Models

From the estimations, the theory informed VECM and MVART models exhibit greater explanatory power over the statistical AR and VARD models. A formal statistical comparison between the models can be made using Wald tests. We can move from an unrestricted specification to a restricted one by imposing q restrictions, the Wald tests then test the null that the restrictions imposed are correct. Here we compare: (i) MVARD against AR ($q = 27$), (ii) MVARD against BVARD ($q = 18$), (iii) MVART against AR ($q = 27$), (iv) MVART against BVART ($q = 18$) and (v) VECM against MVARD ($q = 3$), where the *unrestricted against the restricted* model is tested.

For the change in the 1-month yield $\Delta r_t^{(1)}$, there is an equation in each model with it as the dependent variable. So in cases (i) to (v) we can move from the unrestricted model to the restricted by imposing q restrictions and directly testing these restrictions. However, only a subset of the above tests are conducted for $\Delta r_t^{(3)}, \Delta r_t^{(6)}$ and $\Delta r_t^{(12)}$ because some models have $\Delta r_t^{(n)}$ and others $s_t^{(n,m)}$ as their dependent variable;

so hypotheses involving the VART models are not tested. Table 3.29 summarises the results of the Wald tests conducted. In almost all cases the restrictions are rejected. Whereby the restrictions under the AR and BVARD models are rejected in favour of the MVARD model, the restrictions of the AR and BVART models are rejected for the MVART model and the restrictions of the MVARD model are rejected for the VECM. This with the exception of $\Delta r_t^{(12)}$, where the restrictions under the BVARD can not be rejected.

In short, the restrictions under the statistical VARD and AR models are rejected in favour of both multivariate models of the same specification and the theory informed VECM and VART models. This implying that not only are all the yields important in explaining $\Delta r_t^{(m)}$, but so too are the long-run relations between the yields.

3.4.5 Dynamic Properties

The dynamic properties of the yields are examined using impulse responses and persistence profiles. We focus on the AR and VECM models in order to gain an insight into the dynamic behaviour of the yields, under a statistical compared to a theory informed specification. The generalised impulse response of a one standard error variable specific shock in a particular equation is considered under both models. The generalised impulse response of a one standard error variable specific shock on the r cointegrating relations, together with a persistence profile which is the time profile of the effect of system wide shocks on the cointegrating relations, are considered under the VECM specification only. These dynamics are considered over horizons of up to 150 weeks.

Consider first the AR models that describe $\Delta r_t^{(n)}$ by its past values, Figure 3-4 shows the generalised impulse response of a one standard error $\Delta r_t^{(n)}$ variable shock to the $\Delta r_t^{(n)}$ AR equation, for $n = 1, 3, 6$ and 12. All four yields respond to a shock in their respective equations similarly, with an initial positive response. These shocks have no long-run consequence for $\Delta r^{(n)}$ under the AR models. Figure 3-4 also shows

113

the responses in levels, considering the impact of a shock to $r_t^{(n)}$ on $r_t^{(n)}$. The 12-month rate is most responsive to a shock in its AR equation followed by the 6-, 3- and then the 1-month rate. These rates converge to a constant above their pre-shock level within 9 months of the initial shock, this constant is of similar magnitude for each yield.

Under the VECM[27] framework the yields are modelled in a system which captures any feedbacks that exist. This allowing us to not only observe the impact of a $r_t^{(n)}$ shock on $r_t^{(n)}$, but also the effect of this shock on the other yields in the system and equally the response of $r_t^{(n)}$ to a shock in the other yields. These interactions and feedbacks are not accommodated in the univariate AR framework. Figure 3-5 shows the generalised impulse responses of a variable specific shock to different variables in the cointegrating VAR, i.e. the response of all yields to a shock in the $r_t^{(12)}, r_t^{(6)}, r_t^{(3)}$ and $r_t^{(1)}$ equations[28]. All yields respond positively to a one unit shock in the 12- and the 6-month yields and then converge to a constant above their pre-shock level within 150 weeks. The 12-month yield responds the strongest and adjusts the quickest. The 12-month yield is the least responsive to a one unit shock to the 3-month yield and as before all the yields converge to a higher equilibrium above zero. Each rate however, does not respond as strongly to a unit shock in the 1-month rate and their convergence to zero is comparatively speedier. What is clear from these impulse responses is that feedbacks from $r_t^{(n)}$ to $r_t^{(m)}$ exist, where $n \neq m$, with the dynamics being more complex than those under the AR model.

Figure 3-6 considers the generalised impulse response to variable specific shocks on the r cointegrating relations. These shocks are imposed on the cointegrating vectors so their impact will be transitory. $CV1, CV2$ and $CV3$ refer to the three cointegrating relations $\left(r_t^{(12)} - r_t^{(1)}\right), \left(r_t^{(6)} - r_t^{(1)}\right)$ and $\left(r_t^{(3)} - r_t^{(1)}\right)$ respectively, which reflects the

---

[27]The VECM with the exactly identifyig restrictions $\left(1, \beta, c^{(n,m)}\right)$ imposed on the cointegrating vectors.

[28] The impact of a shock to $r_t^{(n)}$ on $r_t^{(n)}$ in each case, is comparable to the univariate generalised impulse response. Here the VECM also incorporates the impact the shock has on the other yields.

gap between $r_t^{(n)}$ and $r_t^{(m)}$ given by the spread $s_t^{(n,m)}$. The impact on $\left(r_t^{(12)} - r_t^{(1)}\right)$ from a shock to $r_t^{(12)}$ has the effect of increasing the gap between the two yields as the response is positive. This gap then narrows to the point where the two yields are equal 20 weeks after the shock. The spread $s_t^{(12,1)}$ is then negative for 60 weeks before turning positive and converging to zero within the 150 week horizon. This reaction to a 12-month yield shock is mirrored by the other cointegrating vectors $\left(r_t^{(6)} - r_t^{(1)}\right)$ and $\left(r_t^{(3)} - r_t^{(1)}\right)$, where $s_t^{(6,1)}$ and $s_t^{(3,1)}$ are not as responsive to the shock and quicker to adjust. The spreads' respond to a $r_t^{(6)}$ and $r_t^{(3)}$ shock in a similar fashion. The initial response of the spreads to a unit shock to the 1-month rate is negative, implying that $r_t^{(n)} < r_t^{(1)}$, before zero spreads at around 60 weeks, a slightly positive spread which is then followed by convergence. In general, a shock in the 12-month yield has the greatest affect on the long-run relationships and a shock to the 1-month has the least. Further, the $s_t^{(12,1)}$ is the most responsive to a given shock and the $s_t^{(3,1)}$ the least.

Figure 3-7 gives the Persistence Profiles, which map the time profile of the effect of system wide shocks on the long-run relationships, i.e. the response of the spreads to a simultaneous shock to all the yields. All three long-run relationships converge to their respective equilibria, with the initial response of the cointegrating vectors to a system wide shock being positive. This initial response halves within 3 months and completely dies within two years. The $s_t^{(3,1)}$ is the least responsive to the shocks and adjusts to its long-run equilibrium the quickest. We would expect this spread to have the quickest speed of convergence, since it is composed of the shortest maturities. Whereas $s_t^{(12,1)}$ is the most responsive to the shocks and the slowest to adjust, which again is not surprising as it contains the 12-month yield which is the longest maturity considered.

3.5 Results from the VAR Methodology

The VAR methodology is applied to our weekly dataset for each of the six pairs of (n,m)[29]. We test the hypothesis $s_t^{(n,m)*} = s_t^{(n,m)}$ as implied by the EH, first by estimating the regression $s_t^{(n,m)*} = \alpha + \beta s_t^{(n,m)}$ and conducting Wald tests[30] of (i) $H_0 : \beta = 1$ and (ii) $H_0 : \alpha = 0, \beta = 1$, where rejection of the null suggests that the spread is not an optimal predictor of future short rate changes. Second, by examining the time series plots of $s_t^{(n,m)*}$ and $s_t^{(n,m)}$, and third by computing the SDRs and correlation coefficients which should both equal unity under the EH. The first test provides a statistical test of the EH, whereas the second and third tests examine the economic validity of the EH.

From Table 3.30 the estimated values for the coefficients α and β are close to zero and one respectively. Though, both the null that $\beta = 1$ and the joint null that $\alpha = 0$ and $\beta = 1$ are rejected for all pairs of (n,m) except for $(3,1)$. This implies that the spread between the 3- and 1-month is an optimal predictor of future changes in the 1-month rate, and according to the Wald test only for this pair does the EH hold. However, the accuracy and reliability of Wald tests have been criticised[31].

The SDRs are very close to unity, with excess volatility being apparent in all but the $(3,1)$ case i.e. SDR<1, the null that the SDR is equal to unity is rejected for all (n,m). There is strong correlation between the actual and theoretical spread for all maturities and the null that the correlation coefficient is unity, is rejected only for $(6,3)$ at the 5% level. Further Figure 3-8 shows the high degree of comovement between the actual and theoretical spreads, with the series moving in unison for all pairs.

As highlighted in Cuthbertson et al (2000b, footnote 10) there exists a trade-off when testing the null that an estimated coefficient is equal to its theoretical value.

---

[29] Where $n = 3, 6, 12$ and $m = 1, 3, 6$ and $\mathbf{z}_t = \left(s_t^{(n,m)}, \Delta r_t^{(m)} \right)'$ in equation (2.20).

[30] It should be noted that a Wald test of the VAR cross-equation restrictions, see CS, is different to the Wald tests conducted here, however both provide a test of the hypothesis $s_t^{(n,m)*} = s_t^{(n,m)}$.

[31] Bekaert et al. (1997) and Bekaert and Hodrick (2000) provide monte carlo evidence to show that the null is grossly over-rejected when asymtotic critical values are used, because the Wald test suffers from severe size distortions.

Where an estimated coefficient that is numerically close to its theoretical value, but has a small standard error rejects the null. Whereas, an estimated coefficient that is further from its theoretical value, but has a large standard error results in non-rejection of the null. Here our results are consistent with this, in that the estimated values are numerically close to what the EH suggests for both the regression of $s_t^{(n,m)*}$ on $s_t^{(n,m)}$ and the SDRs. However, small standard errors are observed for both, indicating a statistical rejection of the EH in the regression of s^* on s.

Rejection of the VAR restrictions is common in the literature and some explanations as to why the tests under the VAR method indicate a rejection are provided, see Cuthbertson et al (1996). First, the VAR method requires an explicit information set chosen by the econometrician known to both him and agents, however biased VAR coefficients may be estimated if the econometrician mistakenly fails to incorporate variables influencing traders' perceptions, resulting in a rejection of the restrictions. The second point raised is even if agents use the VAR method to forecast, given that new information becomes available moment by moment one would expect this continuous stream of information to be used, in this respect even data as frequent as weekly appears inadequate.

However as Cuthbertson (1996, pp. 589) states these informational inefficiencies "may not be numerically important enough to cause a large discrepancy between forecast movements in future rates and the optimal predictor of these movements namely, the actual spread". CS emphasise that although upon statistical grounds the VAR restrictions may be rejected, this does not necessarily imply rejection of the EH on economic grounds, especially when it is apparent that the actual and theoretical spreads exhibit a high degree of comovement. In short, the EH is economically significant if the theoretical spread can explain the majority of the variation in the actual spread. Which it clearly does here. Although the Wald tests largely indicate a statistical rejection of the EH, large deviations between forecasts of changes in future short rates given by

$s^{(n,m)*}$, from $s^{(n,m)}$ are not apparent. The time series plots of $s_t^{(n,m)*}$ and $s_t^{(n,m)}$, the high values of the correlation coefficients and the point estimates of the SDRs all indicate that the limited information set used here is able to predict the direction of change of the future short rates. Hence in this study the EH is "economically significant". Thus with some support also being provided by the Wald test, we find evidence in favour of the EH.

3.6 Conclusion

This investigation models the UK term structure for zero-coupon 1-, 3-, 6- and 12-month bonds, using a weekly dataset over the period 1997 to 2004, through a comprehensive set of statistical and theory based models. There are two main aims of this study, first to test the EH of the term structure using both cointegration and the VAR approach, second to ascertain if a statistical or a theory informed model is best placed to explain the UK term structure.

All the yields are difference stationary and the (n, m) spreads for $(3, 1), (6, 1), (12, 1)$ and $(6, 3)$ are stationary. Although all the spreads exhibit mean reverting behaviour, persistent deviations from their respective means are observed. The ADF and PP unit root tests appear to be sensitive to this persistence, as the null of a unit root in the $(12, 3)$ and $(12, 6)$ spreads could not be rejected at the 10% level. However, three cointegrating relationships amongst the four yields are found, implying that a long-run relationship exists between the n- and m-month yields and that their spread is stationary. Furthermore, the restrictions that the cointegrating vector between each pair of yields is $(1, -1)$ as suggested by the EH could not be rejected, together with evidence to suggest that the Pure EH is applicable to this dataset.

Four models that incorporate varying degrees of economic theory are estimated. Specifically, an Autoregressive Model, Vector Autoregressive Model in Differences, VAR

in Transformed Interest Rates Model and a Vector Error Correction Model, each of order 9 so that a test of the model restrictions could be made. The theory informed VECM model explicitly defines and tests the long-run relations. Whereas the equivalent VART model assumes that these long-run relations hold and embeds the cointegration relations implied by the EH. The results from the estimated models suggest that these theory informed models have greater explanatory power for the yields than the atheoretic AR and VARD models. We also find that the restrictions under the AR, VARD and the bivariate specifications are rejected in favour of the multivariate theory informed VECM and VART models, when Wald tests of the model restrictions are conducted. This providing additional evidence of the superiority of these theory based models in capturing the term structure.

The dynamic properties of the yields under the AR model and the VECM are examined through impulse responses and persistence profiles, all the plots show convergence indicating that the systems are stable. The dynamics are found to be far more complex under the VECM, since it captures the feedbacks and interactions that exist between the yields. Whereas the univariate framework ignores such interactions and feedbacks by abstracting from economic theory and the influence of other variables.

The results from the VAR approach of the EH indicate high comovement between the theoretical spread given by forecasts of future changes in short rates, and the actual observed spread. We find both the standard deviation ratios and correlation coefficients to be close to unity. Even though the Wald test indicates a statistical rejection of the hypothesis that $s_t^{(n,m)*} = s_t^{(n,m)}$ in all cases except the spread between the 3- and 1-month yields, a divergence between the actual and theoretical spread as measured by the time series plots, SDRs and correlation coefficients is not apparent.

The VAR method makes the strong assumption that the limited information set selected by the econometrician captures all variables influencing traders' perceptions, the literature suggest this may be a reason why the VAR restrictions are rejected.

Although here we do not directly impose and test the VAR cross-equation restrictions, we do derive the theoretical spread and test the hypothesis $s_t^{(n,m)*} = s_t^{(n,m)}$ from the unrestricted VAR which also makes this assumption. The cointegration test of the EH is not subject to these restrictive assumptions, because here we test for a long-run relationship between the yields that translates into stationary spreads as implied by the EH. We find clear support for the EH using the cointegration method.

The results presented here are more supportive of the EH than that of previous studies. This may be due to the fact that we use comparatively high frequency weekly data for short maturities than earlier studies. In the context of the UK studies, we find more support using the cointegration method, compared to Cuthbertson (1996) and Cuthbertson et al (1996, 1998) who also use weekly data at the short end. In particular, unlike Cuthbertson et al (1998) we find the joint null that the set of $q - 1$ spreads form a basis for the cointegrating space can not be rejected. Our SDRs are closer to unity and correlation coefficients of the same magnitude. These UK studies use pre-1997 interest rate data during which UK interest rates were more volatile. Taking the example of Cuthbertson et al (1996) they use data from October 1975 to October 1992, over their sample the interest rates fluctuate between 4 and 18%, compared to our post-1997 data which varies between 3 and 7%. In 1997 the UK had a change of government who granted independence to the central bank to add credibility to its anti-inflation policy. This marked the beginning of an era of comparatively lower, less volatile interest rates for the UK.

We argue that the additional support we find for the EH is due to our UK data 1997 to 2004 being comparatively less volatile. With the evidence presented here further validating earlier findings, that variability in expected changes in short rates is required, suffice to conduct tests of the EH, assuming a constant term premium. But rates that are not highly volatile and unpredictable as to deem the EH with a constant term premium invalid.

Our findings imply that the volatility of the interest rates is important for the EH, and adds to those reported by Mankiw and Miron (1986), Kugler (1988), Cuthbertson et al (2000b) and Christiansen et al (2003) for the US, Germany and Denmark. That is, the EH may not hold if there is a policy of interest rate smoothing or if interest rates are highly volatile and unpredictable. We argue that the support found here for the EH is a combination of the Bank of England's credible anti-inflation policy and the benign economic conditions of the "nice decade" observed until recently.

In this chapter we model the term structure and analyse the in-sample properties of the various models. We find support for the EH and for modelling the term structure using a theory informed as opposed to a purely statistical model. Given these findings, in the next chapter we take the theory informed multivariate VART model that embeds the cointegration relations implied by the EH, and use it to generate forecasts of the short term yields. These forecasts are then evaluated in an investment decision making context.

To conclude, in this examination we find evidence in favour of the much discussed and tested EH of the term structure. The results of the various tests of the EH are not unanimously in favour, but on the whole supportive of the EH. Overall we may deduce that the EH has significant economic content and provides a good representation of this short term UK data. This exercise demonstrates the importance of economic theory in explaining the term structure, as the theory informed models are found to have greater explanatory power than the purely statistical based ones.

Figure 3-1: Yields 1997 to 2005
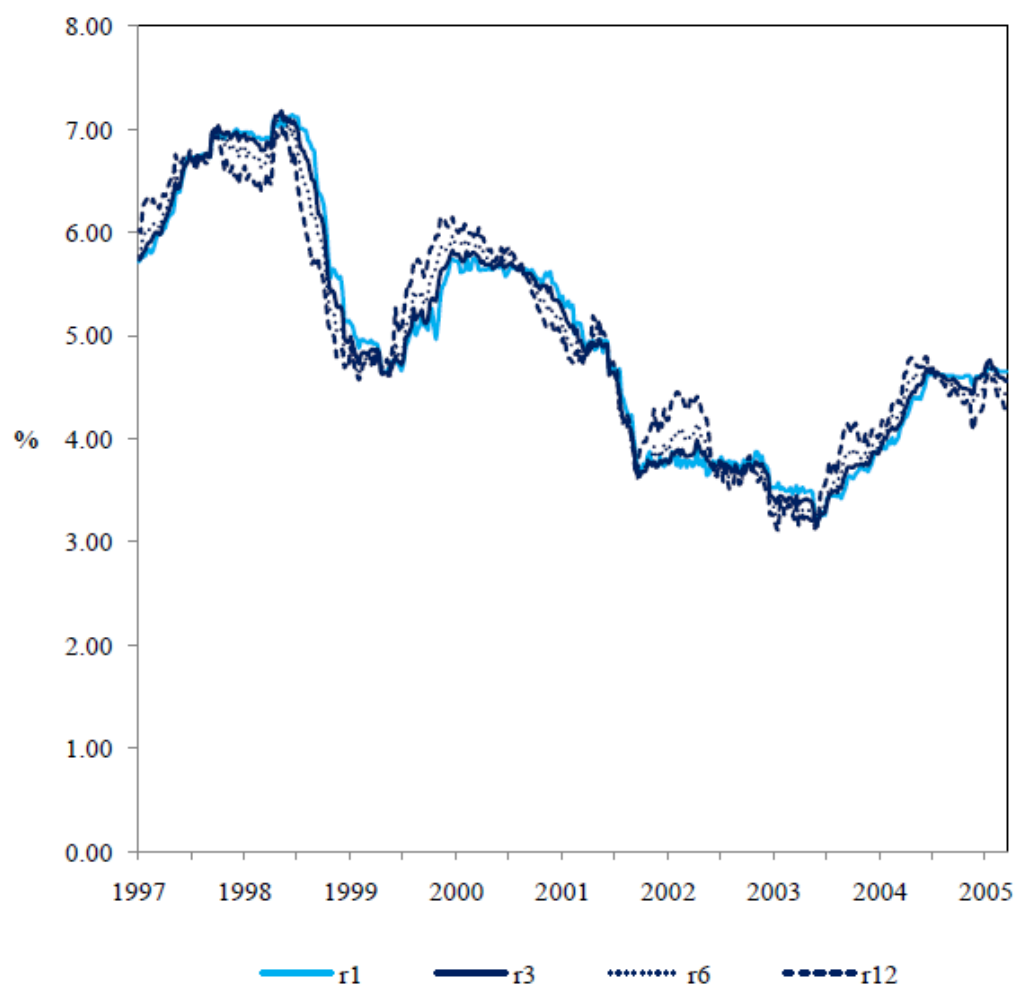
Figure 3-2: Change in Yields
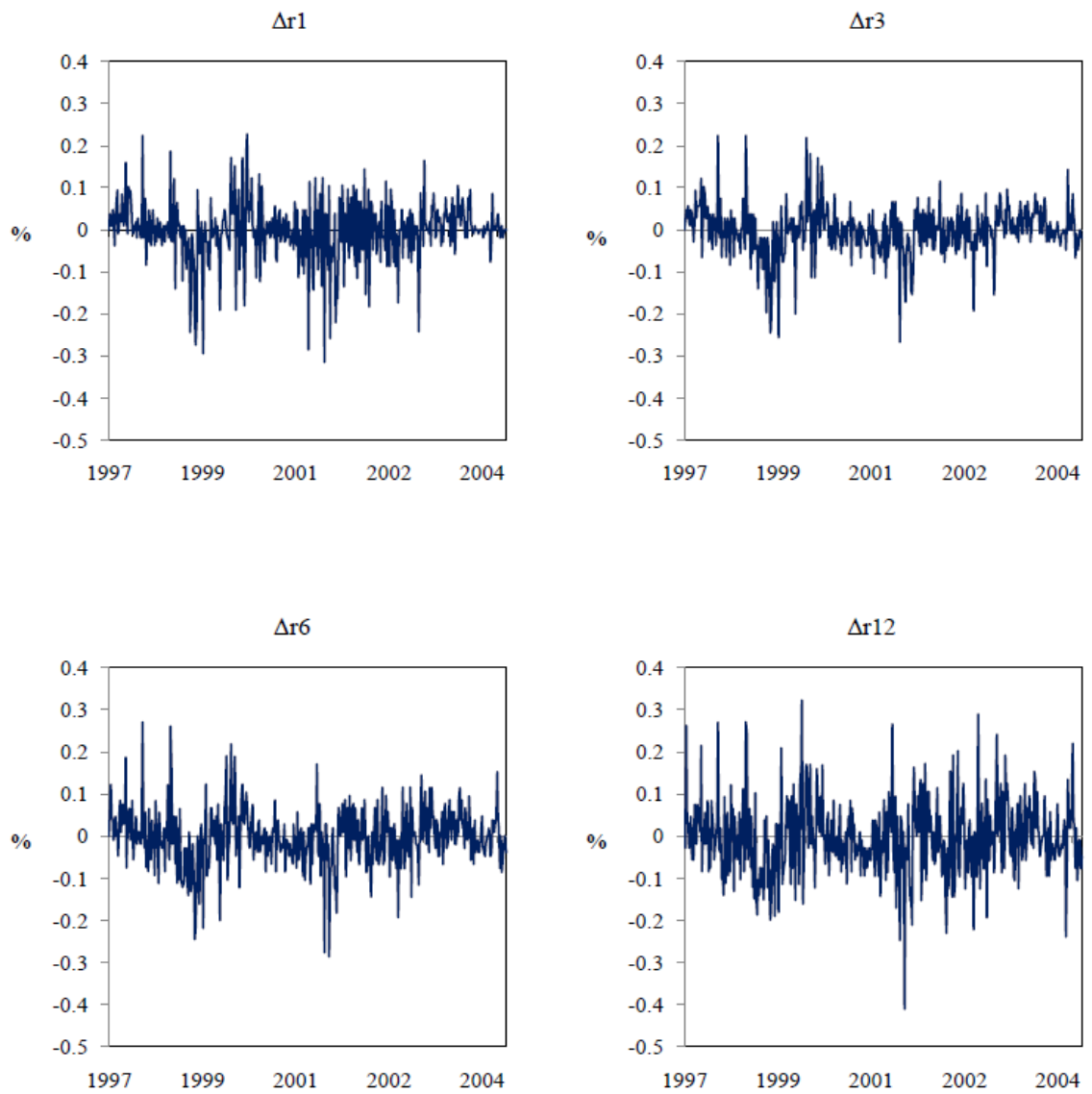
Figure 3-3: Spread between n-month and m-month Yields (snm)
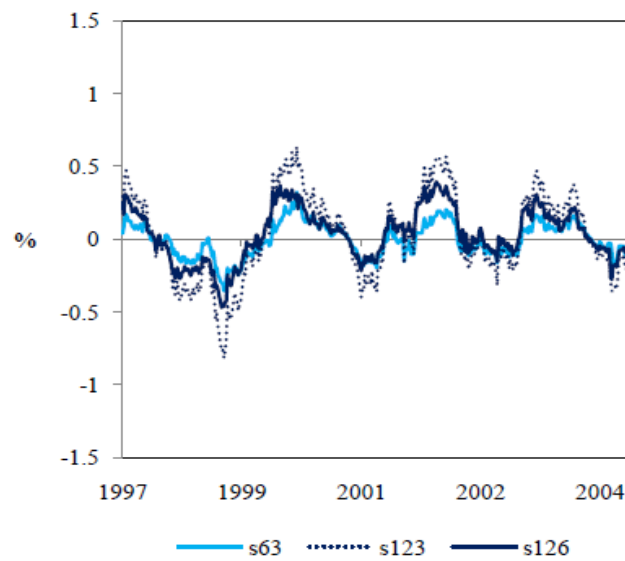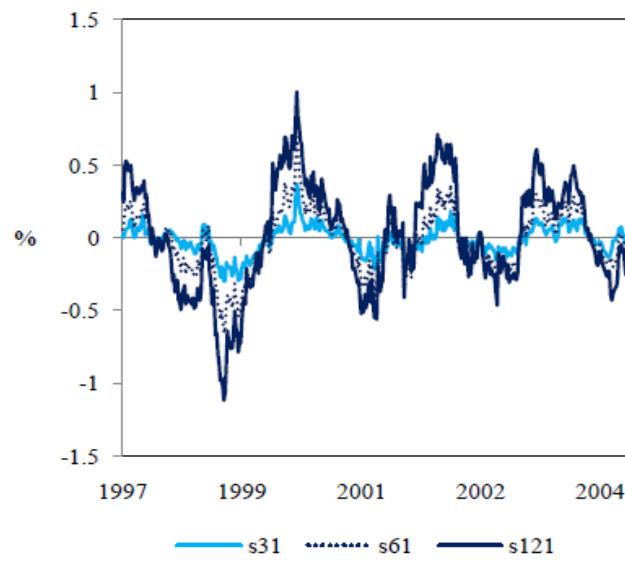
Figure 3-4: Generalised Impulse Response to one S.E. shock in AR(p) Equation

shock in AR(p) equation

Horizon (weeks)

- - - - Δr12   ········ Δr6   —— Δr3   —— Δr1

response in levels

Horizon (weeks)

•••••r12   ········r6   ——r3   ——r1
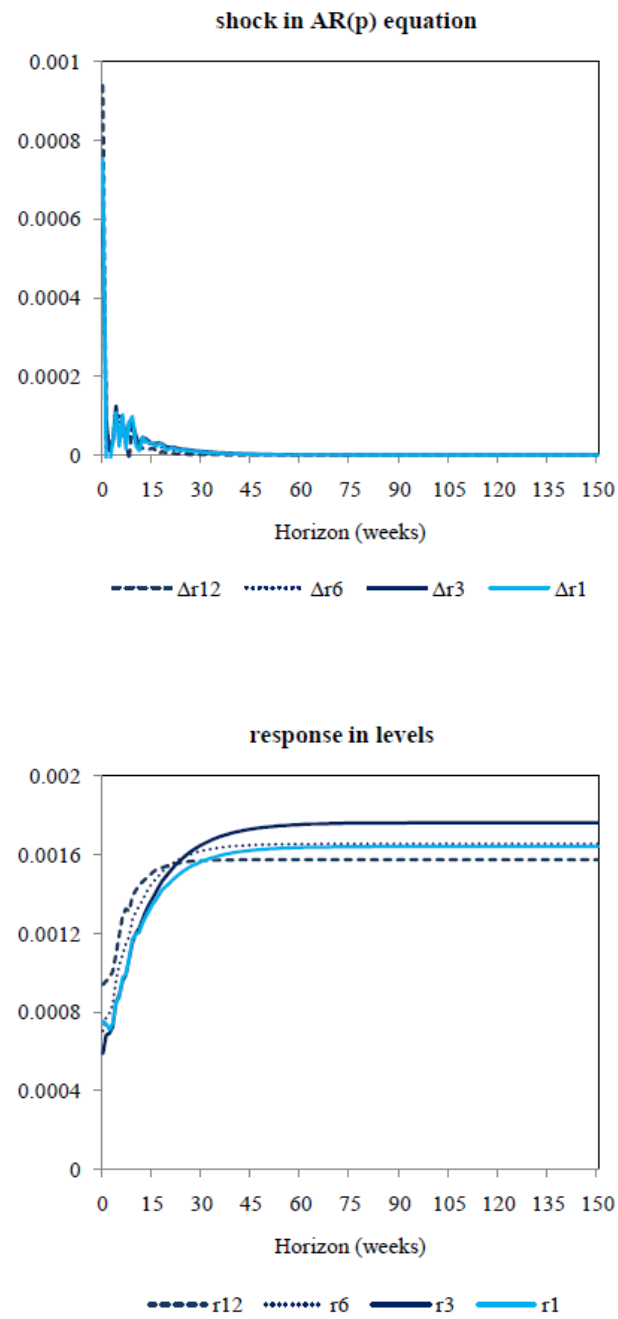
Figure 3-5: Generalised Impulse Response to one S.E. shock in n-month Yields (rn)

126
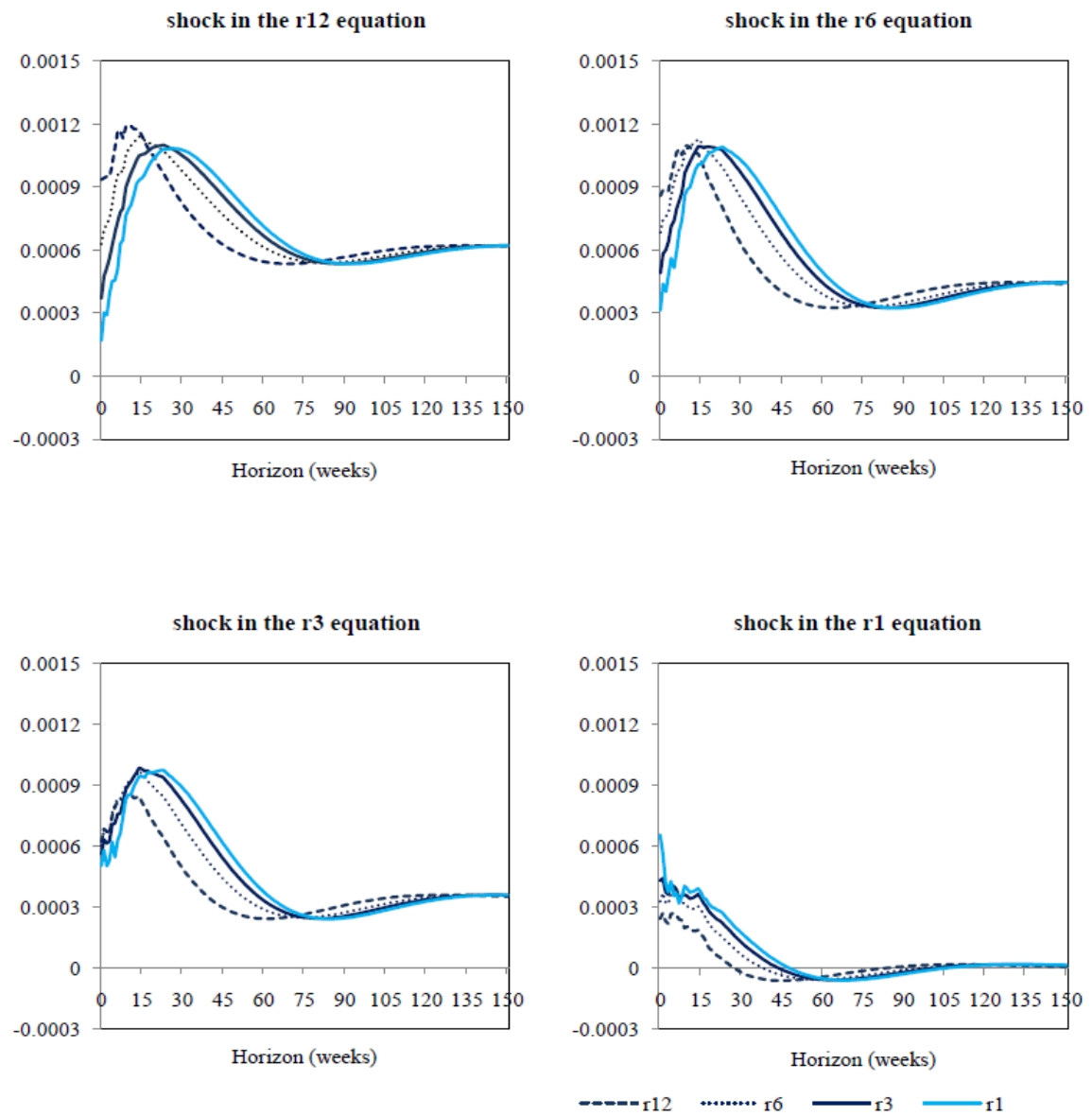
Figure 3-6: Generalised Impulse Response of Cointegrating Vectors (CV) to one S.E. shock in n-month Yields (rn)
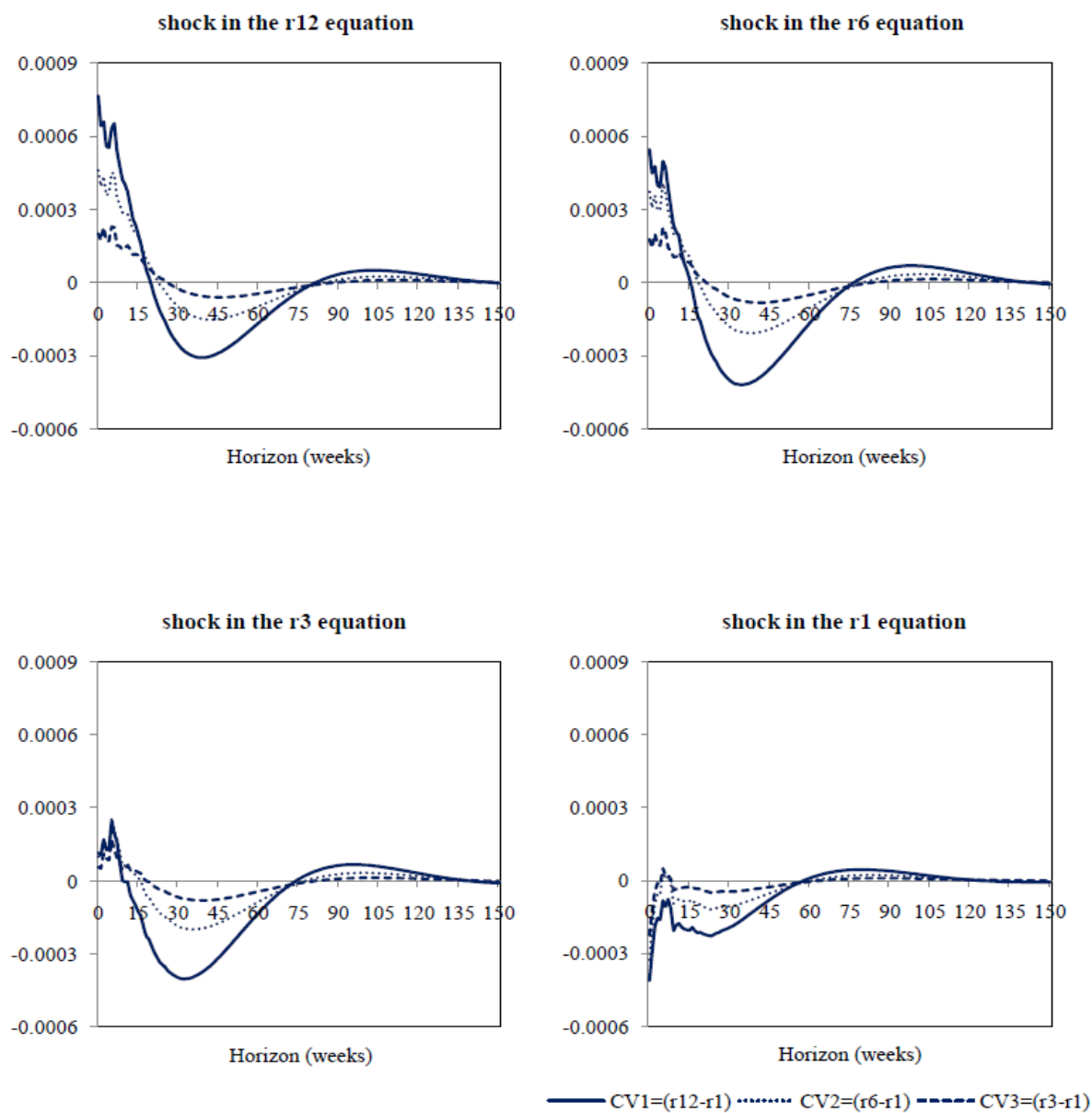
Figure 3-7: Persistence Profile of the effect of a System-Wide Shock on the Cointegrating
Vectors
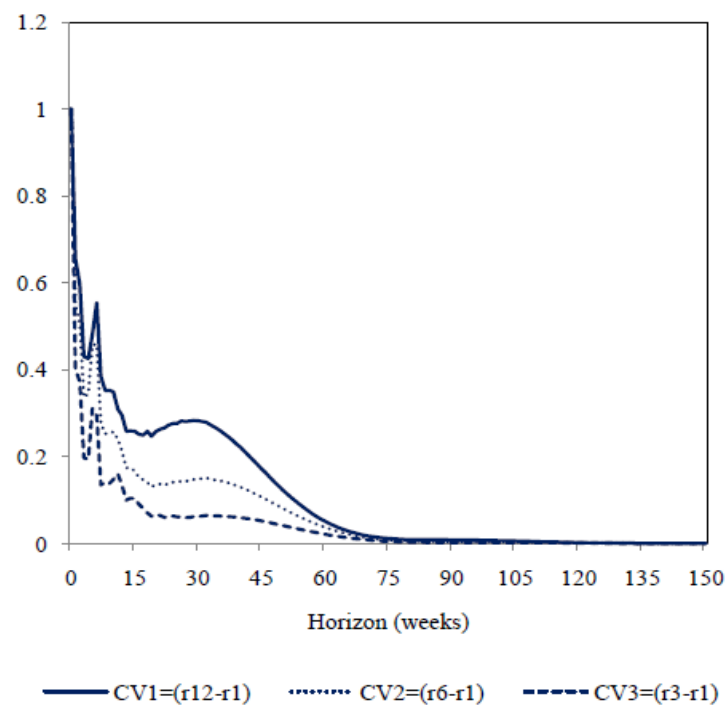
Figure 3-8: Actual (s) and Theoretical (s*) Spreads
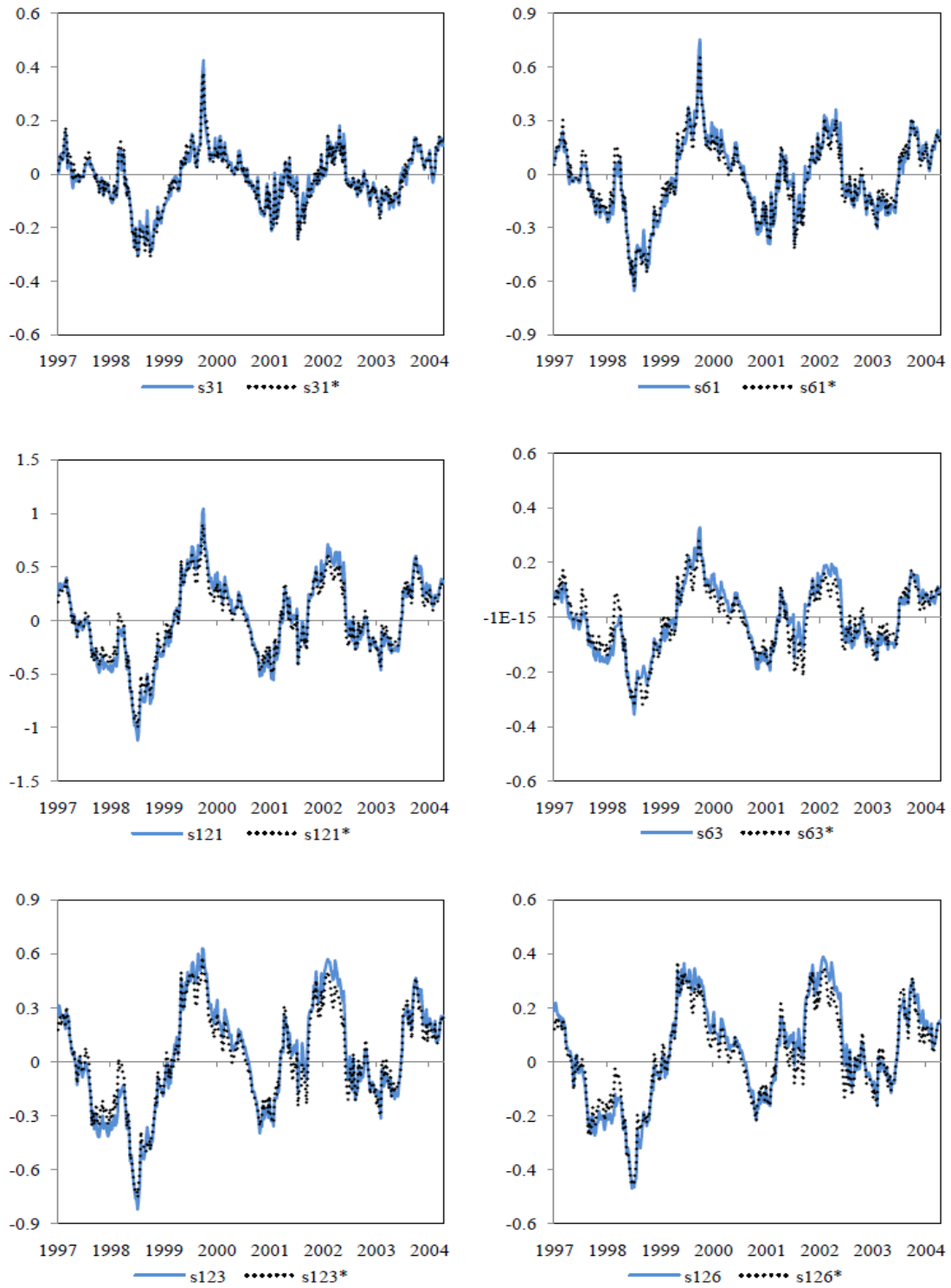
Table 3.1: Descriptive Statistics for Yields and Spreads

Variable	Mean	St Dev	Minimum	Maximum
r_t^1	4.984	1.136	3.160	7.167
r_t^3	4.969	1.129	3.169	7.176
r_t^6	4.964	1.111	3.131	7.148
r_t^{12}	4.995	1.064	3.111	7.018
$s_t^{3,1}$	−0.015	0.099	−0.299	0.370
$s_t^{6,1}$	−0.020	0.209	−0.656	0.692
$s_t^{12,1}$	0.011	0.373	−1.118	1.004
$s_t^{6,3}$	−0.006	0.118	−0.357	0.322
$s_t^{12,3}$	0.025	0.292	−0.819	0.633
$s_t^{12,6}$	0.031	0.180	−0.470	0.393

Notes: Descriptive Statistics of the annualised 1-, 3-, 6- and 12-month yields (%) and their respective spreads over the sample period 1997 week 10 to 2005 week 18.

Table 3.2: Unit Root Tests of Yields

Variable	ADF t-statistic (lag length)	PP adj t-statistic (bandwidth)	KPSS LM statistic (bandwidth)
r_t^1	−0.787 (0)	−1.027 (11)	1.884*** (16)
Δr_t^1	−19.610*** (0)	−20.777*** (11)	0.189 (11)
r_t^3	−1.246 (4)	−1.089 (13)	1.871*** (16)
Δr_t^3	−6.499*** (3)	−18.211*** (12)	0.155 (13)
r_t^6	−1.299 (4)	−1.141 (13)	1.877*** (16)
Δr_t^6	−7.241*** (3)	−18.864*** (12)	0.123 (13)
r_t^{12}	−1.022* (0)	−1.262 (11)	1.901*** (16)
Δr_t^{12}	−19.256*** (0)	−19.836*** (10)	0.088 (11)

Notes for this table are on the next page.

131

Table 3.3: Unit Root Tests of Spreads

Variable $s_t^{(n,m)}$	ADF t-statistic (lag length)	PP adj t-statistic (bandwidth)	KPSS LM statistic (bandwidth)
$(3,1)$	-3.959^{***} (1)	-4.384^{***} (6)	0.106 (16)
$(6,1)$	-2.974^{**} (1)	-3.202^{**} (7)	0.117 (16)
$(12,1)$	-2.700^{*} (0)	-2.667^{*} (7)	0.165 (16)
$(6,3)$	-2.824^{*} (0)	-2.767^{*} (6)	0.132 (16)
$(12,3)$	-2.433 (0)	-2.487 (7)	0.190 (16)
$(12,6)$	-2.476 (0)	-2.422 (6)	0.229 (16)

Critical Values

	ADF Test	PP Test	KPSS Test
1% level	-3.445	-3.445	0.739
5% level	-2.868	-2.868	0.463
10% level	-2.570	-2.570	0.347

Notes: The ADF test statistics are computed using ADF regressions with an intercept and 'L' lagged first differences of the dependent variable, applied to both the levels and first differences. The order of augmentation in the Dickey-Fuller regressions are chosen using the Schwarz Information Criterion, with maximum lag length of 20. The bandwidth for both the PP and KPSS test is selected using the Newey-West (1994) method based on the Bartlett Kernel. The PP test statistics are calculated with an intercept only in the underlying DF regressions, for both the levels and first differences of the variables. The statistics for each test were also computed using regressions with an intercept and linear time trend, but the results are not significantly different from those found above so are not reported. Null Rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 3.4: AR(p) Model Estimates

Equation	Δr_t^{12}	Δr_t^6	Δr_t^3	Δr_t^1
Δr_{t-1}^n	0.021 (0.051)	0.101* (0.052)	0.161* (0.058)	−0.018 (0.075)
Δr_{t-2}^n	0.014 (0.044)	0.016 (0.041)	−0.010 (0.040)	−0.036 (0.052)
Δr_{t-3}^n	0.038 (0.050)	0.055 (0.050)	0.0534 (0.053)	0.040 (0.053)
Δr_{t-4}^n	0.089 (0.060)	0.165*** (0.053)	0.196*** (0.062)	0.143*** (0.054)
Δr_{t-5}^n	0.100* (0.058)	0.075 (0.060)	−0.002 (0.061)	0.039 (0.071)
Δr_{t-6}^n	0.094* (0.051)	0.047 (0.049)	0.106** (0.051)	0.147** (0.067)
Δr_{t-7}^n	0.026 (0.046)	0.040 (0.053)	−0.007 (0.050)	0.016 (0.062)
Δr_{t-8}^n	−0.032 (0.063)	0.001 (0.061)	0.090 (0.059)	0.010* (0.058)
Δr_{t-9}^n	0.056 (0.059)	0.075 (0.054)	0.079 (0.054)	0.114 (0.054)
$inpt$	−0.00003 (0.00005)	−0.00002 (0.00004)	−0.00002 (0.00003)	−0.00002 (0.00004)
$\overline{R}^2$	0.017	0.071	0.132	0.061
$\widehat{\sigma}$	0.0009	0.0007	0.0006	0.0008
$F[9, 354]$	1.70*	4.10***	7.13***	3.63***
$eq^n LL$	2032.01	2136.67	2203.36	2111.42
χ_N^2	21.99***	50.16***	135.82***	90.41***
$\chi_H^2[9]$	8.38	4.26	10.97	16.94**
$\chi_{SC}^2[9]$	5.94	8.70	8.04	6.60

Notes: Standard errors in parenthesis. The AR(9) regressions are estimated for each yield over 1997 week 10 to 2004 week 18 (364 observations), with yields expressed as changes. The regressions are estimated with Newey-West heteroskedastic and autocorrelation corrected errors, reported above. The $\overline{R}^2$, standard error of the regression ($\widehat{\sigma}$), F-statistic to test the joint significance of the estimated coefficients and the log likelihood of the equation (LL) are presented, together with the chi-squared statistics for Breusch-Pagan Serial Correlation test (SC) for the null of no serial correlation at lag 9, the Jarque-Bera Test for Normality (N) for the null that the residuals are normal, Breusch-Pagan-Godfrey test for Heteroskedasticity (H) for the null that the residuals are homoskedastic. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 3.5: Estimation of MVARD(p) Model

Equation	Δr_t^{12}	Δr_t^6	Δr_t^3	Δr_t^1
Δr_{t-1}^{12}	0.225 (0.209)	0.461*** (0.158)	0.355** (0.146)	0.243 (0.154)
Δr_{t-2}^{12}	0.237 (0.221)	0.238 (0.165)	0.105 (0.136)	0.042 (0.156)
Δr_{t-3}^{12}	−0.256 (0.225)	−0.181 (0.165)	−0.150 (0.141)	0.069 (0.167)
Δr_{t-4}^{12}	0.214 (0.237)	0.183 (0.164)	0.102 (0.130)	0.204 (0.169)
Δr_{t-5}^{12}	0.429 (0.273)	0.453*** (0.186)	0.436*** (0.146)	0.477*** (0.159)
Δr_{t-6}^{12}	0.256 (0.243)	0.140 (0.166)	0.018 (0.139)	0.052 (0.172)
Δr_{t-7}^{12}	−0.082 (0.218)	−0.187 (0.156)	−0.202* (0.120)	0.039 (0.131)
Δr_{t-8}^{12}	0.094 (0.205)	−0.071 (0.140)	−0.094 (0.108)	−0.00 (0.135)
Δr_{t-9}^{12}	−0.198 (0.211)	−0.145 (0.157)	−0.160 (0.123)	−0.261* (0.149)
Δr_{t-1}^{6}	−0.801 (0.495)	−1.075*** (0.370)	−0.696** (0.339)	−0.68* (0.353)
Δr_{t-2}^{6}	−0.613 (0.558)	−0.478 (0.400)	−0.123 (0.325)	−0.330 (0.382)
Δr_{t-3}^{6}	0.719 (0.599)	0.635 (0.434)	0.552 (0.361)	−0.174 (0.410)
Δr_{t-4}^{6}	−0.614 (0.591)	−0.446 (0.399)	−0.268 (0.311)	−0.739* (0.409)
Δr_{t-5}^{6}	−0.726 (0.649)	−0.699* (0.437)	−0.731** (0.329)	−1.07*** (0.382)
Δr_{t-6}^{6}	−0.380 (0.547)	−0.209 (0.374)	−0.058 (0.318)	−0.443 (0.408)
Δr_{t-7}^{6}	0.302 (0.567)	0.578 (0.395)	0.575* (0.293)	0.026 (0.326)
Δr_{t-8}^{6}	−0.388 (0.503)	−0.003 (0.355)	−0.022 (0.270)	−0.397 (0.327)
Δr_{t-9}^{6}	0.736 (0.469)	0.573 (0.352)	0.537* (0.289)	0.692** (0.346)
Δr_{t-1}^{3}	1.022** (0.460)	0.982*** (0.335)	0.672** (0.307)	1.174*** (0.337)
Δr_{t-2}^{3}	0.597 (0.541)	0.382 (0.366)	0.115 (0.285)	0.655* (0.348)
Δr_{t-3}^{3}	−0.594 (0.559)	−0.612 (0.395)	−0.459 (0.318)	0.521 (0.387)
Δr_{t-4}^{3}	0.811 (0.541)	0.561 (0.368)	0.403 (0.312)	1.090*** (0.412)
Δr_{t-5}^{3}	0.530 (0.581)	0.335 (0.402)	0.315 (0.320)	0.721* (0.424)
Δr_{t-6}^{3}	0.286 (0.502)	0.165 (0.345)	0.228 (0.284)	0.938** (0.400)
Δr_{t-7}^{3}	−0.302 (0.544)	−0.573 (0.393)	−0.464 (0.292)	0.184 (0.314)
Δr_{t-8}^{3}	0.443 (0.519)	0.254 (0.371)	0.354 (0.272)	0.910*** (0.302)
Δr_{t-9}^{3}	−0.515 (0.459)	−0.357 (0.320)	−0.217 (0.250)	−0.183 (0.304)

Table 3.6: Estimation of MVARD(p) Model (continued)

Δr^1_{t-1}	-0.297^* (0.157)	-0.188 (0.114)	-0.157 (0.099)	-0.570^{***} (0.113)
Δr^1_{t-2}	-0.237 (0.196)	-0.133 (0.133)	-0.112 (0.109)	-0.480^{***} (0.120)
Δr^1_{t-3}	0.101 (0.186)	0.174 (0.126)	0.105 (0.110)	-0.357^{**} (0.161)
Δr^1_{t-4}	-0.254 (0.189)	-0.113 (0.127)	-0.067 (0.107)	-0.380^{***} (0.144)
Δr^1_{t-5}	-0.161 (0.209)	0.009 (0.144)	0.018 (0.116)	-0.240 (0.159)
Δr^1_{t-6}	-0.096 (0.195)	0.007 (0.135)	-0.055 (0.106)	-0.357^{**} (0.147)
Δr^1_{t-7}	0.052 (0.198)	0.154 (0.148)	0.063 (0.115)	-0.232^* (0.129)
Δr^1_{t-8}	-0.135 (0.208)	-0.079 (0.147)	-0.130 (0.103)	-0.345^{***} (0.104)
Δr^1_{t-9}	-0.034 (0.180)	-0.039 (0.121)	-0.049 (0.09)	-0.046 (0.115)
$inpt$	-0.00003 (0.00005)	0.00001 (0.00004)	-0.00001 (0.00003)	-0.00002 (0.00003)
$\overline{R^2}$	0.007	0.106	0.192	0.244
$\widehat{\sigma}$	0.0009	0.0007	0.0006	0.0007
$F[36, 327]$	1.07	2.20^{***}	3.40^{***}	4.26^{***}
$eq^n LL$	2044.6	2158.0	2230.9	2159.6
$system\ LL$	9638.86			
$\chi^2_N[8]$	40.96^{***}			
$\chi^2_H[720]$	799.14^{**}			
$\chi^2_{SC}[16]$	25.57^*			

Notes: Standard errors in parenthesis. The MVARD(9) is estimated over 1997 week 10 to 2004 week 18 (364 observations). The regressions are estimated with Newey-West heteroskedastic and autocorrelation corrected errors. The $\overline{R}^2$, standard error of the regression ($\widehat{\sigma}$), F-statistic to test the joint significance of the estimated coefficients and the log likelihood of the equation (LL) are presented, together with the model diagnostic tests which are all carried out on the VAR residuals. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. Chi-squared statistics presented for: (N) the VAR Residual Normality Test (orthogonalization: residual correlation (Doornik-Hansen) this test statistic is not sensitive to the ordering or the scale of the variables) for the null that the residuals are multivariate normal; (H) the VAR Residual Heteroskedasticity Test (no cross terms, but the conclusion was the same when cross terms were included) for the null that the residuals are homoskedastic, and (SC) the VAR Residual Serial Correlation LM Test for the null of no serial correlation at lag 9. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 3.7: Estimation of BVARD(p) Model: n=3, m=1

$Equation$	Δr_t^3	Δr_t^1
Δr_{t-1}^3	0.291*** (0.091)	0.547*** (0.121)
Δr_{t-2}^3	0.120 (0.073)	0.205** (0.085)
Δr_{t-3}^3	0.116 (0.081)	0.321*** (0.108)
Δr_{t-4}^3	0.280*** (0.094)	0.347*** (0.119)
Δr_{t-5}^3	−0.014 (0.090)	−0.119 (0.125)
Δr_{t-6}^3	0.193** (0.077)	0.334*** (0.122)
Δr_{t-7}^3	0.117 (0.085)	0.277*** (0.100)
Δr_{t-8}^3	0.190** (0.095)	0.285** (0.114)
Δr_{t-9}^3	0.250*** (0.094)	0.318*** (0.119)
Δr_{t-1}^1	−0.154** (0.067)	−0.437*** (0.090)
Δr_{t-2}^1	−0.142 (0.078)	−0.295*** (0.076)
Δr_{t-3}^1	−0.078 (0.074)	−0.281*** (0.107)
Δr_{t-4}^1	−0.082 (0.072)	−0.149* (0.089)
Δr_{t-5}^1	0.006 (0.072)	−0.034 (0.093)
Δr_{t-6}^1	−0.109 (0.070)	−0.162 (0.101)
Δr_{t-7}^1	−0.121* (0.071)	−0.231** (0.094)
Δr_{t-8}^1	−0.096 (0.071)	−0.149* (0.082)
Δr_{t-9}^1	−0.147** (0.061)	−0.105 (0.082)
$inpt$	−0.00002 (0.00003)	−0.00002 (0.00003)

Table 3.8: Estimation of BVARD(p) Model: n=6, m=1

$Equation$	Δr_t^6	Δr_t^1
Δr_{t-1}^6	0.093 (0.058)	0.260*** (0.072)
Δr_{t-2}^6	0.041 (0.055)	0.068 (0.050)
Δr_{t-3}^3	0.071 (0.052)	0.190*** (0.063)
Δr_{t-4}^6	0.152** (0.067)	0.137** (0.068)
Δr_{t-5}^6	0.072 (0.068)	−0.084 (0.076)
Δr_{t-6}^6	0.069 (0.063)	0.121 (0.074)
Δr_{t-7}^6	0.052 (0.061)	0.181*** (0.059)
Δr_{t-8}^6	0.012 (0.079)	0.063 (0.075)
Δr_{t-9}^6	0.141** (0.063)	0.191*** (0.071)
Δr_{t-1}^1	0.022 (0.050)	−0.193*** (0.071)
Δr_{t-2}^1	−0.056 (0.067)	−0.151** (0.060)
Δr_{t-3}^1	−0.018 (0.059)	−0.129 (0.068)
Δr_{t-4}^1	0.023 (0.068)	0.036 (0.061)
Δr_{t-5}^1	0.013 (0.072)	−0.0001 (0.074)
Δr_{t-6}^1	−0.022 (0.060)	0.025 (0.075)
Δr_{t-7}^1	−0.039 (0.058)	−0.090 (0.068)
Δr_{t-8}^1	−0.003 (0.060)	0.021 (0.060)
Δr_{t-9}^1	−0.100* (0.056)	0.039 (0.056)
$inpt$	−0.00002 (0.00004)	−0.00002 (0.00004)

Table 3.9: Estimation of BVARD(p) Model: n=12, m=1

Equation	Δr_t^{12}	Δr_t^1
Δr_{t-1}^{12}	0.011 (0.052)	0.146*** (0.050)
Δr_{t-2}^{12}	0.020 (0.052)	0.031 (0.034)
Δr_{t-3}^{12}	0.042 (0.050)	0.133*** (0.042)
Δr_{t-4}^{12}	0.076 (0.065)	0.064 (0.048)
Δr_{t-5}^{12}	0.107* (0.060)	0.001 (0.050)
Δr_{t-6}^{12}	0.110* (0.059)	0.0446 (0.054)
Δr_{t-7}^{12}	0.029 (0.050)	0.118*** (0.037)
Δr_{t-8}^{12}	-0.027 (0.071)	0.004 (0.049)
Δr_{t-9}^{12}	0.089 (0.066)	0.086 (0.050)
Δr_{t-1}^{1}	0.049 (0.058)	-0.105 (0.074)
Δr_{t-2}^{1}	-0.054 (0.075)	-0.108** (0.053)
Δr_{t-3}^{1}	0.001 (0.081)	-0.051 (0.055)
Δr_{t-4}^{1}	0.032 (0.081)	0.083 (0.056)
Δr_{t-5}^{1}	-0.016 (0.085)	0.013 (0.072)
Δr_{t-6}^{1}	-0.038 (0.068)	0.096 (0.069)
Δr_{t-7}^{1}	-0.022 (0.073)	-0.016 (0.062)
Δr_{t-8}^{1}	-0.006 (0.068)	0.090 (0.055)
Δr_{t-9}^{1}	-0.087 (0.076)	0.111 (0.057)
$inpt$	-0.00003 (0.00005)	-0.00001 (0.00004)

138

Table 3.10: Estimation of BVARD(p) Model: n=6, m=3

$Equation$	Δr_t^6	Δr_t^3
Δr_{t-1}^6	-0.029 (0.095)	0.116 (0.089)
Δr_{t-2}^6	0.084 (0.111)	0.156^* (0.092)
Δr_{t-3}^6	0.161 (0.103)	0.186^{**} (0.088)
Δr_{t-4}^6	0.050 (0.126)	0.016 (0.105)
Δr_{t-5}^6	0.133 (0.137)	0.059 (0.125)
Δr_{t-6}^6	0.092 (0.115)	0.046 (0.103)
Δr_{t-7}^6	0.139 (0.100)	0.166^{**} (0.077)
Δr_{t-8}^6	-0.089 (0.145)	-0.090 (0.116)
Δr_{t-9}^6	0.254^{**} (0.117)	0.195^{**} (0.088)
Δr_{t-1}^3	0.200^* (0.117)	0.040 (0.110)
Δr_{t-2}^3	-0.113 (0.137)	-0.192^* (0.114)
Δr_{t-3}^3	-0.140 (0.132)	-0.143 (0.100)
Δr_{t-4}^3	0.149 (0.151)	0.173 (0.129)
Δr_{t-5}^3	-0.078 (0.167)	-0.057 (0.147)
Δr_{t-6}^3	-0.034 (0.129)	0.060 (0.117)
Δr_{t-7}^3	-0.155 (0.130)	-0.172^* (0.101)
Δr_{t-8}^3	0.142 (0.151)	0.195 (0.128)
Δr_{t-9}^3	-0.235^* (0.137)	-0.092 (0.101)
$inpt$	-0.00002 (0.00004)	-0.00002 (0.00003)

Table 3.11: Estimation of BVARD(p) Model: n=12, m=3

$Equation$	Δr_t^{12}	Δr_t^3
Δr_{t-1}^{12}	-0.053 (0.067)	0.084^* (0.043)
Δr_{t-2}^{12}	0.031 (0.073)	0.077^* (0.040)
Δr_{t-3}^{12}	0.048 (0.068)	0.077^* (0.038)
Δr_{t-4}^{12}	0.036 (0.080)	0.012 (0.053)
Δr_{t-5}^{12}	0.120 (0.078)	0.078 (0.061)
Δr_{t-6}^{12}	0.130^* (0.071)	0.017 (0.051)
Δr_{t-7}^{12}	0.049 (0.061)	0.051 (0.036)
Δr_{t-8}^{12}	-0.034 (0.089)	-0.054 (0.054)
Δr_{t-9}^{12}	0.108 (0.091)	0.060 (0.046)
Δr_{t-1}^{3}	0.192^* (0.111)	0.069 (0.076)
Δr_{t-2}^{3}	-0.076 (0.107)	-0.111^* (0.062)
Δr_{t-3}^{3}	-0.038 (0.126)	-0.019 (0.063)
Δr_{t-4}^{3}	0.119 (0.119)	0.173^{**} (0.081)
Δr_{t-5}^{3}	-0.069 (0.148)	-0.066 (0.094)
Δr_{t-6}^{3}	-0.076 (0.099)	0.101 (0.070)
Δr_{t-7}^{3}	-0.048 (0.123)	-0.038 (0.067)
Δr_{t-8}^{3}	0.020 (0.112)	0.163^{**} (0.077)
Δr_{t-9}^{3}	-0.118 (0.132)	0.051 (0.075)
$inpt$	-0.00003 (0.00005)	-0.00001 (0.00003)

Table 3.12: Estimation of BVARD(p) Model: n=12, m=6

$Equation$	Δr_t^{12}	Δr_t^6
Δr_{t-1}^{12}	-0.103 (0.128)	0.055 (0.090)
Δr_{t-2}^{12}	0.105 (0.124)	0.117 (0.088)
Δr_{t-3}^{12}	0.030 (0.132)	0.041 (0.091)
Δr_{t-4}^{12}	-0.015 (0.132)	-0.044 (0.109)
Δr_{t-5}^{12}	0.219 (0.163)	0.201 (0.133)
Δr_{t-6}^{12}	0.217 (0.132)	0.085 (0.105)
Δr_{t-7}^{12}	0.047 (0.116)	-0.001 (0.091)
Δr_{t-8}^{12}	-0.017 (0.138)	-0.118 (0.107)
Δr_{t-9}^{12}	0.078 (0.164)	0.043 (0.118)
Δr_{t-1}^{6}	0.183 (0.171)	0.036 (0.127)
Δr_{t-2}^{6}	-0.142 (0.152)	-0.122 (0.112)
Δr_{t-3}^{6}	0.010 (0.181)	0.015 (0.130)
Δr_{t-4}^{6}	0.152 (0.165)	0.217 (0.136)
Δr_{t-5}^{6}	-0.187 (0.236)	-0.161 (0.180)
Δr_{t-6}^{6}	-0.167 (0.163)	-0.030 (0.128)
Δr_{t-7}^{6}	-0.005 (0.179)	0.059 (0.136)
Δr_{t-8}^{6}	-0.023 (0.163)	0.154 (0.125)
Δr_{t-9}^{6}	-0.017 (0.208)	0.044 (0.154)
$inpt$	-0.00003 (0.00005)	-0.00002 (0.00004)

Table 3.13: BVARD(p) Model Diagnostics

$Model$	$\Delta r_t^3, \Delta r_t^1$	$\Delta r_t^6, \Delta r_t^1$	$\Delta r_t^{12}, \Delta r_t^1$	$\Delta r_t^6, \Delta r_t^3$	$\Delta r_t^{12}, \Delta r_t^3$	$\Delta r_t^{12}, \Delta r_t^6$
$\overline{R^2}$	0.144	0.060	0.001	0.078	0.01	0.006
	0.205	0.168	0.134	0.153	0.153	0.071
$\widehat{\sigma}$	0.0006	0.0007	0.0009	0.0007	0.0009	0.0009
	0.0007	0.0007	0.0007	0.0006	0.0006	0.0007
$F[18, 345]$	4.41***	2.28***	1.022	2.72***	1.21	1.11
	6.22**	5.07***	4.133***	4.66***	4.66***	2.55***
$eq^n LL$	2210.7	2139.1	2033.8	2142.8	2035.5	2034.6
	2146.5	2138.1	2130.9	2212.7	2212.7	2141.4
$system\ LL$	4519.1	4329.1	4181.0	4626.5	4353.8	4509.7
$\chi_N^2[4]$	37.85***	25.81***	24.24***	24.25***	35.73***	44.41
$\chi_H^2[108]$	127.99*	141.76**	149.15***	133.67**	128.19**	125.96
$\chi_{SC}^2[4]$	7.73	8.26*	7.16	10.25**	8.49*	7.15

Notes: Standard errors in parenthesis. A BVARD(9) is estimated over 1997 week 10 to 2004 week 18 (364 observations) for each of the six pairs of yields, the first number refers to the statistic obtained for the Δr_t^n equation and the second for the Δr_t^m equation. The regressions are estimated with Newey-West heteroskedastic and autocorrelation corrected errors. The $\overline{R}^2$, standard error of the regression ($\widehat{\sigma}$), F-statistic to test the joint significance of the estimated coefficients and the log likelihood of the equation (LL) are presented, together with the model diagnostic tests which are all carried out on the VAR residuals. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. Chi-squared statistics presented for: (N) the VAR Residual Normality Test (orthogonalization: residual correlation (Doornik-Hansen) this test statistic is not sensitive to the ordering or the scale of the variables) for the null that the residuals are multivariate normal; (H) the VAR Residual Heteroskedasticity Test (no cross terms, but the conclusion was the same when cross terms were included) for the null that the residuals are homoskedastic, and (SC) the VAR Residual Serial Correlation LM Test for the null of no serial correlation at lag 9. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 3.14: Cointegration Tests

H_0	H_1	λ_{trace} Stat	95% cv	λ_{max} Stat	95% cv
$r = 0$	$r = 1$	72.32**	53.48	32.47**	28.27
$r \leqslant 1$	$r = 2$	39.85**	34.87	25.20**	22.04
$r \leqslant 2$	$r = 3$	14.65	20.18	11.62	15.87
$r \leqslant 3$	$r = 4$	3.03	9.16	3.03	9.16

Table 3.15: Number of Cointegrating Relations using the Model Selection Criteria

Rank	Max LL	AIC	SBC	HQC
$r = 0$	9610.7	9466.7	9186.1	9355.2
$r = 1$	9626.9	9474.9	9178.8	9357.2
$r = 2$	9639.5	9481.5	9173.7	9359.2
$r = 3$	9645.3	9483.3	9167.7	9357.9
$r = 4$	9646.9	9482.9	9163.3	9355.8

Notes: The underlying VAR model is of order 9 and contains restricted intercepts and no trend coefficients. The statistics refer to Johansen's log-likelihood based trace and maximum eigenvalue statistics together with the Akaike Information Criteria (AIC), Schwarz Bayesian Criteria (SBC) and the Hannan-Quinn Criteria (HQC). All of which have been computed using 364 observations over the period 1997 week 10 to 2004 week 18, to test for the number of cointegration vectors. ** Null rejected at 5%

Table 3.16: Imposing Identifying Restrictions

Yield	Vector 1	Vector 2	Vector 3
12-month	1	0	0
6-month	0	1	0
3-month	0	0	1
1-month	-0.92221 (0.06505)	-0.97597 (0.032201)	-0.99432 (0.01176)
Intercept	-0.00402 (0.00330)	-0.00103 (0.00164)	-0.00173 (0.00060)

Notes: Maximum Likelihood estimates subject to the exactly identifying restrictions, where the standard errors are given in the parentheses.

Table 3.17: Reduced Form Error Correction Specification

Equation	Δr_t^{12}	Δr_t^6	Δr_t^3	Δr_t^1
$ect1_{t-1}$	0.495*** (0.148)	0.386*** (0.115)	0.236** (0.100)	0.256** (0.103)
$ect2_{t-1}$	−1.502*** (0.459)	−1.145*** (0.354)	−0.760** (0.300)	−1.050*** (0.329)
$ect3_{t-1}$	1.441*** (0.524)	1.160*** (0.399)	0.976*** (0.337)	1.658*** (0.384)
Δr_{t-1}^{12}	−0.182 (0209)	0141 (0152)	0.144 (0.140)	−0.028 (0.157)
Δr_{t-2}^{12}	−0.148 (0.200)	−0.60 (0.148)	−0.089 (0.129)	−0.224 (0.172)
Δr_{t-3}^{12}	−0.625** (0.262)	−0.469** (0.193)	−0.344** (0.160)	−0.205 (0.186)
Δr_{t-4}^{12}	−0.136 (0.250)	−0.095 (0.195)	−0.089 (0.160)	−0.059 (0.188)
Δr_{t-5}^{12}	0.108 (0.268)	0.201 (0.198)	0.267 (0.168)	0.247 (0.172)
Δr_{t-6}^{12}	−0.033 (0.235)	−0.073 (0.165)	−0.103 (0.140)	−0.096 (0.154)
Δr_{t-7}^{12}	−0.332 (0.205)	−0.366** (0.144)	−0.298** (0.122)	−0.085 (0.134)
Δr_{t-8}^{12}	−0.121 (0.223)	−0233 (0.161)	−0.171 (0.134)	−0.032 (0.162)
Δr_{t-9}^{12}	−0.344 (0.226)	−0.250 (0.163)	−0.202* (0.122)	−0.273* (0.50)
Δr_{t-1}^6	0.367 (0.487)	−0.154 (0.362)	−0005 (0335)	0.405 (0.389)
Δr_{t-2}^6	0.466 (0.527)	0.368 (0.35)	0.519 (0.340)	0.716 (0.459)
Δr_{t-3}^6	1.712** (0.667)	1.422*** (0.515)	1.169*** (0.431)	0.855* (0487)
Δr_{t-4}^6	0.312 (0.645)	0.305 (0.495)	0.330 (0.393)	0.237 (0.457)
Δr_{t-5}^6	0.111 (0.675)	−0.029 (0.500)	−0.208 (0.419)	−0.221 (0.432)
Δr_{t-6}^6	0.366 (0.556)	0.361 (0.404)	0.343 (0.349)	0.174 (0.369)
Δr_{t-7}^6	0.933 (0.538)	1.048*** (0.380)	0.894*** (0.320)	0.528 (0.336)
Δr_{t-8}^6	0.149 (0.540)	0.382 (0.401)	0.237 (0.333)	−0.130 (0.389)
Δr_{t-9}^6	1.087** (0.513)	0.829** (0.382)	0.681** (0.301)	−0.858** (0.358)
Δr_{t-1}^3	−0.101 (0.498)	0.017 (0.395)	−0.247 (0.369)	−0.469 (0.449)
Δr_{t-2}^3	−0.408 (0.593)	−0.486 (0.449)	−0.726* (0.373)	−0.880* (0.480)
Δr_{t-3}^3	−1.483** (0.635)	−1.388*** (0.499)	−1.237*** (0.418)	−0.929* (0.91)
Δr_{t-4}^3	−0.006 (0.590)	0.166 (0.470)	−0.327 (0.397)	−0.242 (0.458)
Δr_{t-5}^3	−0.185 (0.636)	−0.299 (0.481)	−0.319 (0.411)	−0.434 (0.471)
Δr_{t-6}^3	−0.330 (0.521)	−0.377 (0.396)	−0.305 (0.340)	−0.012 (0.385)

144

Table 3.18: Reduced Form Error Correction Estimates (continued)

Δr^3_{t-7}	-0.807 (0.545)	-1.011^{**} (0.408)	-0.888^{***} (0.337)	-0.578^{*} (0.336)
Δr^3_{t-8}	0.019 (0.537)	-0.129 (0.406)	0.005 (0.322)	0.400 (0.338)
Δr^3_{t-9}	-0.777 (0.486)	-0.590^{*} (0.350)	-0.433 (0.273)	-0.535^{*} (0.320)
Δr^1_{t-1}	0.036 (0.202)	0.135 (0.174)	0.228 (0.160)	0.181 (0.194)
Δr^1_{t-2}	0.054 (0.231)	0.151 (0.180)	0.230 (0.152)	0.193 (0.174)
Δr^1_{t-3}	0.350 (0.226)	0.418^{**} (0.178)	0.407^{***} (0.157)	0.248 (0.192)
Δr^1_{t-4}	-0.033 (0.201)	0.103 (0.162)	0.198 (0.139)	0.141 (0.163)
Δr^1_{t-5}	0.020 (0.226)	0.193 (0.170)	0.248^{*} (0.143)	0.208 (0.176)
Δr^1_{t-6}	0.052 (0.197)	0.161 (0.150)	0.139 (0.127)	0.013 (0.149)
Δr^1_{t-7}	0.171 (0.200)	0.276^{*} (0.155)	0.215^{*} (0.127)	0.054 (0.130)
Δr^1_{t-8}	-0.035 (0.209)	0.025 (0.154)	-0.013 (0.115)	-0.156 (0.11)
Δr^1_{t-9}	0.024 (0.179)	0.020 (0.122)	0.019 (0.097)	0.069 (0117)
$\overline{R^2}$	0.032	0.133	0.221	0.298
$\widehat{\sigma}$	0.0009	0.0007	0.0006	0.0007
$F\,stat$	1.31	2.47^{***}	3.71^{***}	5.05^{***}
$eq^n LL$	2044.3	2158.5	2232.2	2174.4
$system\ LL$	9645.9			
$\chi^2_N [8]$	73.91^{***}			
$\chi^2_H [720]$	725.89			
$\chi^2_{SC} [16]$	21.09			

The three error correction terms are given by:

$$ect1_t \;=\; r_t^{12} \underset{(0.06505)}{-\,0.92221 r_t^1} \underset{(0.00330)}{-\,0.00402}$$

$$ect2_t \;=\; r_t^{6} \underset{(0.032201)}{-\,0.97597\, r_t^1} \underset{(0.00164)}{-\,0.00103}$$

$$ect3_t \;=\; r_t^{3} \underset{(0.01176)}{-\,0.99432\, r_t^1} \underset{(0.00060)}{-\,0.17300}$$

Notes: Standard errors in parenthesis. The VECM(9) contains restricted intercepts and no trend coefficients, estimated over 1997 week 10 to 2004 week 18 with Newey-West heteroskedastic and autocorrelation corrected errors. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. Chi-squared statistics presented for: (N) the VAR Residual Normality Test; (H) the VAR Residual Heteroskedasticity Test and (SC) the VAR Residual Serial Correlation LM Test for the null of no serial correlation at lag 9. The VECM is estimated with cointegrating vectors subject to the exactly identifying restrictions, the coefficients of the error correction terms contained in the matrix α are given in the first three rows of the table. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 3.19: Testing the EH: Imposing Over-Identifying Restrictions

(a) Restriction Set 1

Yield	Vector 1	Vector 2	Vector 3
12-month	1	0	0
6-month	0	1	0
3-month	0	0	1
1-month	−1	−1	−1
Intercept	−0.00015 (0.00091)	0.00015 (0.00041)	0.00010 (0.00014)

(b) Restriction Set 2

Yield	Vector 1	Vector 2	Vector 3
12-month	1	0	0
6-month	0	1	0
3-month	0	0	1
1-month	−1	−1	−1
Intercept	0	0	0

Restriction Set	LR Statistics
1	6.5506
2	12.5334

Notes: The first set of over-identifying restrictions test for cointegration of each of the 3-, 6- and 12-month yields with the 1-month, such that the cointegrating vector corresponds to $(1, -1, c)$. The second set tests the hypothesis that the liquidity premia are zero, such that the cointegrating vector corresponds to $(1, -1, 0)$, as under the PEH. The LR test statistic is distributed as χ^2, the 1% and 5% critical values are 11.345 and 7.815 respectively for three degrees of freedom and for six degrees of freedom the 1% and 5% critical values are 16.812 and 12.592 respectively. ** Null rejected at 5%.

Table 3.20: Estimation of MVART(p) Model

Equation	$s_t^{12,1}$	$s_t^{6,1}$	$s_t^{3,1}$	Δr_t^1
$s_{t-1}^{12,1}$	1.054*** (0.216)	0.289** (0.144)	0.159 (0.099)	0.172 (0.151)
$s_{t-2}^{12,1}$	0.214 (0.259)	−0.028 (0.174)	−0.054 (0.109)	−0.153 (0.184)
$s_{t-3}^{12,1}$	−0.489* (0.249)	−0.435** (0.176)	−0.285*** (0.110)	0.089 (0.137)
$s_{t-4}^{12,1}$	0.337 (0.264)	0.232 (0.186)	0.115 (0.112)	0.056 (0.192)
$s_{t-5}^{12,1}$	−0.055 (0.280)	−0.009 (0.198)	0.047 (0.124)	0.294 (0.245)
$s_{t-6}^{12,1}$	0.209 (0.263)	0.075 (0.181)	−0.020 (0.118)	−0.325* (0.178)
$s_{t-7}^{12,1}$	−0.329 (0.263)	−0.329* (0.183)	−0.225* (0.119)	0.055 (0.179)
$s_{t-8}^{12,1}$	0.216 (0.283)	0.124 (0.199)	0.106 (0.130)	0.051 (0.172)
$s_{t-9}^{12,1}$	0.013 (0.208)	0.156 (0.143)	0.131* (0.094)	−0.068 (0.150)
$s_{t-1}^{6,1}$	−0.407 (0.514)	0.392 (0.337)	−0.120 (0.234)	−0.540 (0.357)
$s_{t-2}^{6,1}$	−0.186 (0.582)	0.257 (0.398)	0.247 (0.268)	0.250 (0.449)
$s_{t-3}^{6,1}$	1.102* (0.595)	0.942** (0.422)	0.543** (0.266)	−0.021 (0.362)
$s_{t-4}^{6,1}$	−0.771 (0.621)	−0.520 (0.441)	−0.245 (0.278)	−0.379 (0.464)
$s_{t-5}^{6,1}$	0.234 (0.664)	0.123 (0.465)	−0.068 (0.289)	−0.471 (0.538)
$s_{t-6}^{6,1}$	−0.159 (0.614)	−0.030 (0.426)	0.122 (0.276)	0.385 (0.434)
$s_{t-7}^{6,1}$	0.235 (0.61)	0.373 (0.411)	0.231 (0.256)	0.236 (0.402)
$s_{t-8}^{6,1}$	−0.242 (0.659)	−0.099 (0.454)	−0.068 (0.298)	−0.633 (0.443)
$s_{t-9}^{6,1}$	−0.068 (0.485)	−0.383 (0.338)	−0.337 (0.232)	0.369 (0.357)
$s_{t-1}^{3,1}$	0.081 (0.467)	−0.060 (0.283)	0.525*** (0.184)	1.126*** (0.338)
$s_{t-2}^{3,1}$	0.089 (0.538)	−0.125 (0.360)	−0.095 (0.242)	−0.391 (0.393)
$s_{t-3}^{3,1}$	−1.040* (0.570)	−0.883** (0.406)	−0.489* (0.255)	0.061 (0.358)
$s_{t-4}^{3,1}$	0.786 (0.606)	0.559 (0.419)	0.251 (0.266)	0.512 (0.424)
$s_{t-5}^{3,1}$	0.034 (0.607)	0.052 (0.443)	0.183 (0.288)	−0.145 (0.478)
$s_{t-6}^{3,1}$	−0.546 (0.546)	−0.468 (0.381)	−0.371 (0.256)	0.380 (0.413)
$s_{t-7}^{3,1}$	0.080 (0.514)	−0.089 (0.344)	−0.037 (0.206)	−0.4512 (0.347)

Table 3.21: Estimation of MVART(p) Model (continued)

$s_{t-8}^{3,1}$	-0.074 (0.600)	-0.025 (0.388)	-0.024 (0.260)	0.909^{**} (0.441)
$s_{t-9}^{3,1}$	0.191 (0.453)	0.402 (0.325)	0.358 (0.220)	-0.543^{*} (0.319)
Δr_{t-1}^{1}	0.034 (0.114)	0.053 (0.083)	0.032 (0.054)	0.089 (0.085)
Δr_{t-2}^{1}	0.167^{*} (0.094)	0.171^{**} (0.069)	0.127^{***} (0.047)	-0.175^{***} (0.061)
Δr_{t-3}^{1}	-0.016 (0.096)	0.015 (0.065)	0.024 (0.038)	-0.023 (0.056)
Δr_{t-4}^{1}	0.065 (0.104)	0.080 (0.068)	0.041 (0.042)	0.066 (0.073)
Δr_{t-5}^{1}	0.255 (0.155)	0.263^{**} (0.104)	0.184^{***} (0.059)	-0.176^{**} (0.078)
Δr_{t-6}^{1}	-0.018 (0.091)	-0.003 (0.070)	-0.005 (0.047)	0.088 (0.067)
Δr_{t-7}^{1}	0.045 (0.93)	0.028 (0.064)	-0.001 (0.040)	-0.066 (0.067)
Δr_{t-8}^{1}	-0.069 (0.092)	-0.034 (0.059)	-0.021 (0.035)	0.090 (0.071)
Δr_{t-9}^{1}	-0.111 (0.069)	-0.086^{*} (0.046)	-0.044 (0.030)	0.040 (0.052)
$inpt$	-0.0001 (0.00007)	-0.00009^{*} (0.00005)	-0.00004 (0.00003)	0.000005 (0.00004)
$\overline{R^2}$	0.934	0.901	0.829	0.273
$\widehat{\sigma}$	0.001	0.0007	0.0004	0.0007
$F[36,327]$	144.42^{***}	93.27^{***}	50.02^{***}	4.80^{***}
$eq^{n}LL$	2028.0	2163.5	2337.9	2172.6
$system\ LL$	9659.1			
$\chi_{N}^{2}[8]$	35.03^{***}			
$\chi_{H}^{2}[720]$	910.12^{***}			
$\chi_{SC}^{2}[16]$	18.48			

Notes: Standard errors in parenthesis. A MVART(9) is estimated over 1997 week 10 to 2004 week 18 (364 observations). The regressions are estimated with Newey-West heteroskedastic and autocorrelation corrected errors. The $\overline{R}^{2}$, standard error of the regression ($\widehat{\sigma}$), F-statistic to test the joint significance of the estimated coefficients and the log likelihood of the equation (LL) are presented, together with the model diagnostic tests which are all carried out on the VAR residuals. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. Chi-squared statistics presented for: (N) the VAR Residual Normality Test; (H) the VAR Residual Heteroskedasticity Test, and (SC) the VAR Residual Serial Correlation LM Test for the null of no serial correlation at lag 9. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 3.22: Estimation of BVART(p) Model: n=3, m=1

$Equation$	$s_t^{3,1}$	Δr_t^1
$s_{t-1}^{3,1}$	0.733*** (0.075)	0.570*** (0.122)
$s_{t-2}^{3,1}$	0.155* (0.090)	−0.322** (0.132)
$s_{t-3}^{3,1}$	−0.131 (0.084)	0.145 (0.122)
$s_{t-4}^{3,1}$	0.122 (0.076)	0.070 (0.137)
$s_{t-5}^{3,1}$	0.163 (0.119)	−0.437** (0.175)
$s_{t-6}^{3,1}$	−0.234** (0.108)	0.401*** (0.135)
$s_{t-7}^{3,1}$	−0.036 (0.107)	−0.011 (0.145)
$s_{t-8}^{3,1}$	0.054 (0.103)	0.028 (0.149)
$s_{t-9}^{3,1}$	0.038 (0.063)	−0.137 (0.108)
Δr_{t-1}^1	0.052 (0.050)	0.067 (0.080)
Δr_{t-2}^1	0.095 (0.045)	−0.144** (0.062)
Δr_{t-3}^1	0.028 (0.036)	−0.018 (0.057)
Δr_{t-4}^1	0.034 (0.042)	0.119* (0.069)
Δr_{t-5}^1	0.171*** (0.061)	−0.197** (0.083)
Δr_{t-6}^1	−0.048 (0.050)	0.085 (0.074)
Δr_{t-7}^1	−0.015 (0.040)	−0.032 (0.066)
Δr_{t-8}^1	−0.009 (0.035)	0.075 (0.072)
Δr_{t-9}^1	−0.067** (0.028)	0.050 (0.049)
$inpt$	−0.00001 (0.00003)	0.00001 (0.00004)

Table 3.23: Estimation of BVART(p) Model: n=6, m=1

Equation	$s_t^{6,1}$	Δr_t^1
$s_{t-1}^{6,1}$	0.827*** (0.063)	0.266*** (0.069)
$s_{t-2}^{6,1}$	0.131 (0.088)	-0.179** (0.077)
$s_{t-3}^{6,1}$	-0.096 (0.070)	0.131* (0.077)
$s_{t-4}^{6,1}$	0.124* (0.070)	-0.027 (0.084)
$s_{t-5}^{6,1}$	0.139 (0.112)	-0.219** (0.105)
$s_{t-6}^{6,1}$	-0.199* (0.105)	0.178** (0.090)
$s_{t-7}^{6,1}$	-0.083 (0.096)	0.074 (0.085)
$s_{t-8}^{6,1}$	0.072 (0.097)	-0.113 (0.089)
$s_{t-9}^{6,1}$	0.018 (0.060)	0.000 (0.072)
Δr_{t-1}^1	0.075 (0.072)	0.026 (0.073)
Δr_{t-2}^1	0.094 (0.059)	-0.114** (0.057)
Δr_{t-3}^1	0.024 (0.060)	0.017 (0.058)
Δr_{t-4}^1	0.036 (0.067)	0.118* (0.063)
Δr_{t-5}^1	0.197* (0.098)	-0.117* (0.076)
Δr_{t-6}^1	-0.062 (0.062)	0.093 (0.068)
Δr_{t-7}^1	-0.041 (0.064)	0.034 (0.065)
Δr_{t-8}^1	-0.044 (0.053)	0.044 (0.069)
Δr_{t-9}^1	-0.143*** (0.046)	0.093* (0.05)
$inpt$	-0.00001 (0.00004)	-0.00002 (0.0004)

Table 3.24: Estimation of BVART(p) Model: n=12, m=1

$Equation$	$s_t^{12,1}$	Δr_t^1
$s_{t-1}^{12,1}$	0.863*** (0.060)	0.146*** (0.048)
$s_{t-2}^{12,1}$	0.121 (0.081)	−0.107** (0.047)
$s_{t-3}^{12,1}$	−0.080 (0.075)	0.108** (0.052)
$s_{t-4}^{12,1}$	0.101 (0.079)	−0.063 (0.059)
$s_{t-5}^{12,1}$	0.092 (0.103)	−0.065 (0.070)
$s_{t-6}^{12,1}$	−0.041 (0.090)	0.039 (0.064)
$s_{t-7}^{12,1}$	−0.156** (0.073)	0.077 (0.056)
$s_{t-8}^{12,1}$	0.054 (0.086)	−0.112* (0.057)
$s_{t-9}^{12,1}$	0.017 (0.063)	0.026 (0.048)
Δr_{t-1}^1	0.040 (0.094)	0.005 (0.073)
Δr_{t-2}^1	0.067 (0.081)	−0.102* (0.054)
Δr_{t-3}^1	−0.013 (0.088)	0.047 (0.059)
Δr_{t-4}^1	−0.013 (0.095)	0.109* (0.061)
Δr_{t-5}^1	0.101 (0.128)	−0.019 (0.069)
Δr_{t-6}^1	−0.040 (0.077)	0.106* (0.064)
Δr_{t-7}^1	−0.070 (0.086)	0.062 (0.062)
Δr_{t-8}^1	−0.103 (0.080)	0.065 (0.066)
Δr_{t-9}^1	−0.186*** (0.068)	0.119** (0.057)
$inpt$	−0.00001 (0.00005)	−0.00003 (0.00004)

Table 3.25: Estimation of BVART(p) Model: n=6, m=3

$Equation$	$s_t^{6,3}$	Δr_t^3
$s_{t-1}^{6,3}$	0.845*** (0.053)	0.149* (0.086)
$s_{t-2}^{6,3}$	0.067 (0.070)	0.046 (0.101)
$s_{t-3}^{6,3}$	0.048 (0.063)	0.055 (0.126)
$s_{t-4}^{6,3}$	0.061 (0.065)	−0.157 (0.131)
$s_{t-5}^{6,3}$	0.038 (0.068)	0.030 (0.133)
$s_{t-6}^{6,3}$	−0.031 (0.067)	−0.020 (0.121)
$s_{t-7}^{6,3}$	−0.077 (0.065)	0.113 (0.120)
$s_{t-8}^{6,3}$	0.021 (0.075)	−0.256** (0.122)
$s_{t-9}^{6,3}$	−0.004 (0.056)	0.139 (0.114)
Δr_{t-1}^3	0.021 (0.033)	0.123** (0.058)
Δr_{t-2}^3	0.020 (0.024)	−0.049 (0.040)
Δr_{t-3}^3	−0.016 (0.033)	0.018 (0.051)
Δr_{t-4}^3	0.020 (0.034)	0.164*** (0.059)
Δr_{t-5}^3	0.061 (0.049)	−0.020 (0.060)
Δr_{t-6}^3	−0.037 (0.029)	0.084* (0.050)
Δr_{t-7}^3	0.002 (0.037)	−0.027 (0.055)
Δr_{t-8}^3	−0.042 (0.031)	0.080 (0.056)
Δr_{t-9}^3	−0.074** (0.029)	0.082 (0.059)
$inpt$	−0.000004 (0.00002)	−0.00002 (0.00003)

Table 3.26: Estimation of BVART(p) Model: n=12, m=3

$Equation$	$s_t^{12,3}$	Δr_t^3
$s_{t-1}^{12,3}$	0.861^{***} (0.055)	0.094^{**} (0.042)
$s_{t-2}^{12,3}$	0.089 (0.061)	-0.003 (0.046)
$s_{t-3}^{12,3}$	0.022 (0.071)	0.007 (0.053)
$s_{t-4}^{12,3}$	0.050 (0.074)	-0.066 (0.061)
$s_{t-5}^{12,3}$	0.017 (0.072)	0.064 (0.069)
$s_{t-6}^{12,3}$	0.069 (0.061)	-0.061 (0.061)
$s_{t-7}^{12,3}$	-0.117^{*} (0.053)	0.033 (0.058)
$s_{t-8}^{12,3}$	0.016 (0.066)	-0.105^{*} (0.058)
$s_{t-9}^{12,3}$	-0.020 (0.057)	0.074 (0.054)
Δr_{t-1}^3	-0.009 (0.067)	0.123^{**} (0.059)
Δr_{t-2}^3	0.002 (0.054)	-0.052 (0.037)
Δr_{t-3}^3	-0.040 (0.069)	0.033 (0.054)
Δr_{t-4}^3	-0.015 (0.073)	0.166^{***} (0.057)
Δr_{t-5}^3	0.046 (0.10)	-0.009 (0.060)
Δr_{t-6}^3	-0.052 (0.059)	0.098^{*} (0.049)
Δr_{t-7}^3	0.0005 (0.074)	-0.005 (0.053)
Δr_{t-8}^3	-0.112^{*} (0.068)	0.090^{*} (0.057)
Δr_{t-9}^3	-0.114^{*} (0.067)	0.091^{*} (0.059)
$inpt$	-0.00001 (0.00004)	-0.00003 (0.00003)

Table 3.27: Estimation of BVART(p) Model: n=12, m=6

Equation	$s_t^{12,6}$	Δr_t^6
$s_{t-1}^{12,6}$	0.844*** (0.063)	0.083 (0.085)
$s_{t-2}^{12,6}$	0.140** (0.062)	0.067 (0.106)
$s_{t-3}^{12,6}$	0.004 (0.069)	−0.072 (0.107)
$s_{t-4}^{12,6}$	0.036 (0.072)	−0.090 (0.130)
$s_{t-5}^{12,6}$	−0.009 (0.069)	0.247* (0.149)
$s_{t-6}^{12,6}$	0.111* (0.067)	−0.111 (0.128)
$s_{t-7}^{12,6}$	−0.084 (0.055)	−0.087 (0.126)
$s_{t-8}^{12,6}$	0.046 (0.071)	−0.118 (0.138)
$s_{t-9}^{12,6}$	−0.099* (0.057)	0.137 (0.111)
Δr_{t-1}^6	−0.008 (0.033)	0.068 (0.057)
Δr_{t-2}^6	−0.029 (0.030)	−0.023 (0.040)
Δr_{t-3}^6	−0.012 (0.036)	0.036 (0.061)
Δr_{t-4}^6	−0.030 (0.035)	0.156*** (0.056)
Δr_{t-5}^6	−0.006 (0.048)	0.023 (0.071)
Δr_{t-6}^6	−0.002 (0.032)	0.038 (0.053)
Δr_{t-7}^6	−0.011 (0.034)	0.041 (0.064)
Δr_{t-8}^6	−0.072** (0.035)	0.020 (0.058)
Δr_{t-9}^6	−0.015 (0.037)	0.077 (0.055)
$inpt$	−0.00001 (0.00002)	−0.00005 (0.00004)

Table 3.28: BVART(p) Model Diagnostics

Model	$s_t^{3,1}, \Delta r_t^1$	$s_t^{6,1}, \Delta r_t^1$	$s_t^{12,1}, \Delta r_t^1$	$s_t^{6,3}, \Delta r_t^3$	$s_t^{12,3}, \Delta r_t^3$	$s_t^{12,6}, \Delta r_t^6$
$\overline{R}^2$	0.812	0.894	0.932	0.924	0.944	0.950
	0.236	0.186	0.154	0.160	0.168	0.084
$\widehat{\sigma}$	0.0004	0.0007	0.001	0.0003	0.0007	0.0004
	0.0007	0.0007	0.0007	0.0006	0.0006	0.0007
$F[18, 345]$	88.54***	171.21***	276.26***	246.14***	344.95***	387.66***
	7.24***	5.628***	4.69***	4.84***	5.08***	2.86***
$eq^n LL$	2311.1	2140.5	2011.2	2412.3	2140.4	2338.3
	2153.7	2142.2	2135.2	2214.0	2215.8	2143.9
$system\ LL$	4526.3	4333.0	4184.8	4629.5	4344.2	4513.8
$\chi^2_N[4]$	43.88***	38.81***	33.08***	44.91***	43.67***	38.99***
$\chi^2_H[108]$	137.52**	166.56***	151.40***	156.15***	142.61**	139.58**
$\chi^2_{SC}[4]$	4.52	6.75	4.23	7.20	5.52	2.27

Notes: Standard errors in parenthesis. A BVART(9) is estimated over 1997 week 10 to 2004 week 18 (364 observations) for each of the six pair of yields. The first number refers to the statistic obtained for the $s_t^{n,m}$ equation and the second for the Δr_t^m equation. The regressions are estimated with Newey-West heteroskedastic and autocorrelation corrected errors. The $\overline{R}^2$, standard error of the regression ($\widehat{\sigma}$), F-statistic to test the joint significance of the estimated coefficients and the log likelihood of the equation (LL) are presented, together with the model diagnostic tests which are all carried out on the VAR residuals. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. Chi-squared statistics presented for: (N) the VAR Residual Normality Test; (H) the VAR Residual Heteroskedasticity Test, and (SC) the VAR Residual Serial Correlation LM Test for the null of no serial correlation at lag 9. Null rejected at *** 1%, ** 5%, * 10% level of significance.

Table 3.29: Wald Test of Model Restrictions

Dependent Variable	Δr_t^1	Δr_t^3	Δr_t^6	Δr_t^{12}
MVARD *vs* AR (q=27)	163.87*** [0.00]	86.39*** [0.00]	69.43*** [0.00]	39.97* [0.05]
MVARD *vs* BVARD (q=18) $(\Delta r_t^3, \Delta r_t^1)$	61.45*** [0.00]	47.64*** [0.00]	...	...
$(\Delta r_t^6, \Delta r_t^1)$	88.24*** [0.00]	...	51.25*** [0.00]	...
$(\Delta r_t^{12}, \Delta r_t^1)$	99.09*** [0.00]	...	...	26.36* [0.09]
$(\Delta r_t^6, \Delta r_t^3)$	...	67.37*** [0.00]	39.71*** [0.00]	...
$(\Delta r_t^{12}, \Delta r_t^3)$	...	60.60*** [0.00]	...	19.72 [0.35]
$(\Delta r_t^{12}, \Delta r_t^6)$	...	...	44.71*** [0.00]	30.42** [0.03]
MVART *vs* AR (q=27)	174.87*** [0.00]	...	...	...
MVART *vs* BVART (q=18) $\left(s_t^{3,1}, \Delta r_t^1\right)$	50.94*** [0.00]	...	...	...
$\left(s_t^{6,1}, \Delta r_t^1\right)$	89.70*** [0.00]			
$\left(s_t^{12,1}, \Delta r_t^1\right)$	101.53*** [0.00]			
VECM *vs* MVARD (q=3)	25.55*** [0.00]	12.49** [0.01]	11.59** [0.01]	12.96*** [0.00]

Notes: p-value in $[.]$. In the comparison of *unrestricted vs restricted* models q restrictions are imposed on the unrestricted model. Above are the $\chi^2(q)$ statistics for the Wald test of these restrictions, testing the null that the restrictions imposed on the model are correct. Null rejected at *** 1% level, ** 5% level, * 10% level of significance. When a comparison between the multivariate and bivariate specification is made for a given model, the z_t for the bivariate model is given in $(.,.)$. In some cases it is not possible to make all comparisons for each Δr_t^n, which are indicated by '...'.

Table 3.30: Testing the EH using Weakly Rational Expectations

(a) Regression $s_t^* = \alpha + \beta s_t + e_t$

Spread (n,m)	α (s.e)	β (s.e)	$H_0 : \beta = 1$ [p − value]	$H_0 : \alpha = 0, \beta = 1$ [p − value]
$(12,6)$	0.00008** (0.00004)	0.8802*** (0.0198)	36.61 [0.00]	44.97 [0.00]
$(12,3)$	−0.00010* (0.00005)	0.8743*** (0.0159)	62.66 [0.00]	71.60 [0.00]
$(12,1)$	−0.00006 (0.00005)	0.8638*** (0.0113)	144.55 [0.00]	144.90 [0.00]
$(6,3)$	0.00005 (0.00004)	0.9030*** (0.0327)	8.81 [0.0032]	11.22 [0.00]
$(6,1)$	−0.00002 (0.00004)	0.9496*** (0.0169)	8.90 [0.003]	8.92 [0.01]
$(3,1)$	0.00000 (0.00002)	1.0131*** (0.0168)	0.60 [0.4381]	0.61 [0.74]

(b) Standard Deviation Ratios and Correlation Coefficients

Spread (n,m)	$SDR = \boldsymbol{\sigma}(\mathbf{s}_t^{(n,m)*})/\boldsymbol{\sigma}(\mathbf{s}_t^{(n,m)})$ (s.e)	$\mathbf{Corr}(\mathbf{s}_t^{(n,m)*}, \mathbf{s}_t^{(n,m)})$ (s.e)
$(12,6)$	0.9001 (0.0001)	0.9780 (0.0247)
$(12,3)$	0.8902 (0.0002)	0.9822 (0.0339)
$(12,1)$	0.8727 (0.0004)	0.9898 (0.0474)
$(6,3)$	0.9599 (0.0001)	0.9408 (0.0189)
$(6,1)$	0.9672 (0.0001)	0.9818 (0.0252)
$(3,1)$	1.0264 (0.0002)	0.9870 (0.0174)

Notes: The standard errors are given in $(.)$ and the p-values are given in $[.]$. Table (a) presents the estimated coefficients for the regression $s_t^* = \alpha + \beta s_t + e_t$, together with the $\chi^2(q)$ statistic for the Wald test of two sets of restrictions, to test (i) H_0: $\beta=1$ and (ii) H_0: $\alpha=0, \beta=1$, where $q=1$ and 2 respectively. The regression is estimated with Newey-West heteroskedastic and autocorrelation corrected errors. The theoretical spread s_t^* is computed using a VAR(9) for each bivariate (n,m) combination over the period 1997 week 10 to 2004 week 18. Table (b) presents the Standard Deviation Ratios (SDR) and the correlation coefficients (Corr) between the actual and theoretical spreads, both of which take a value of unity under the EH. The null that the SDR=1 is rejected for all (n,m), the null that the correlation=1 is rejected only for (6,3) at the 5% level . Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Chapter 4

Decision-Based Forecast Evaluation of Interest Rate Predictability

Abstract

This chapter illustrates the importance of density forecasting in portfolio decision making involving bonds of different maturities. The forecast performance of an atheoretic and a theory informed model of bond returns is evaluated. The decision making environment is fully described for an investor seeking to optimally allocate his portfolio between long and short Treasury Bills, over investment horizons of up to two years. Using weekly data over 1997 to 2007 we examine the impact of parameter uncertainty and predictability in returns on the investor's allocation. We describe how the forecasts are computed and used in this context. Both statistical and decision-based criteria are used to assess the out-of-sample forecasting performance of the models. Our results show sensitivity to the evaluation criterion used. In the context of investment decision making under an economic value criterion, we find some potential gain for the investor from assuming predictability.

Keywords: density forecasting, interest rate predictability, parameter uncertainty and decision-based forecast evaluation.

4.1 Introduction

Conventionally forecast accuracy is assessed using statistical measures, which are usually based on point forecasts and some measure of the forecast errors, but these measures convey little information about the value of the forecast. Leitch and Tanner (1991) argue that given economists believe firms use forecasts to increase profits, then it would be more appropriate to evaluate forecast accuracy using profitability. Granger and Pesaran (2000) and Pesaran and Skouras (2004) formalise this view, asserting that forecasts are ultimately intended to assist in decision making and hence should be evaluated in the decision making context for which they are intended.

Decision-based forecast evaluation[1] is becoming increasingly popular in research. Whereby measures like profit, wealth and utility are used as opposed to forecast errors, to judge forecasts and compare the accuracy with which competing models make projections. Recent research investigates predictability in asset returns and decision-based forecast evaluation in the context of investment decision making. This includes Pesaran and Timmermann (1995), Xia (2001), Avramov (2002), Brooks and Persand (2003), Boudry and Gray (2003) and Marquering and Verbeek (2004), who consider stock return predictability, the effects of parameter and model uncertainty on the optimal allocation and the economic value of this predictability. Further, West et al (1993) consider the economic value of predictability in exchange rate volatility[2].

However, little attention has been paid to decision-based forecast evaluation of interest rates and determining if there is economic value to interest rate predictability. This is what we seek to address here. Previous research that compare the ability of theory based models to forecast interest rates to atheoretic models like a naive random walk, primarily focus on using statistical criteria to evaluate the accuracy of the fore-

---

[1]We may also refer to economic value measures, these are the same as decision-based measures.

[2]Note that we only briefly mention studies key to our investigation, since a detailed review of the literature is provided by Chapter 2.

casts[3]. In this chapter we consider first, how the allocation decisions of the investor are influenced by parameter uncertainty and predictability. Second, if a utility maximising investor gains, in terms of higher wealth, from using a theory informed model to forecast interest rates as opposed to a random walk model, in that we assess if there is economic value to interest rate predictability.

When considering interest rate predictability, we turn to the dominant theory of the term structure, the Expectations Hypothesis (EH), which links interest rates of different maturities together. The explanatory power of the EH has be examined extensively using various testing methods and datasets[4]. In Chapter 3 we model interest rates using a set of statistical and theory informed models, to find past changes in the yields and spreads to have explanatory power. Hence, in this chapter we use the previously estimated Multivariate VAR in Transformed Interest Rates (MVART) model, which embeds the cointegrating relations between the yields as implied by the EH, to capture predictability. Such that, if the investor believes yields are predictable, he uses the MVART model to forecast and inform his allocation decisions. We assess the forecasting ability of this theory based model using both statistical and decision-based measures.

The importance of parameter uncertainty in asset allocation is demonstrated by Klein and Bawa (1976), and Kandel and Stambaugh (1996). Further, the impact of parameter uncertainty, asset return predictability and the investment horizon on optimal portfolio choice is examined in the key paper by Barberis (2000). He draws on the early findings of Samuelson (1969) and Merton (1969), who show that if returns are *i.i.d.* then an investor with power utility has an optimal allocation that is insensitive to the investment horizon, i.e. allocation in the long-run will be the same as in the short-run. However, Barberis finds if returns are predictable, and not *i.i.d.*, then horizon effects may in fact be observed. Further, even with parameter uncertainty there is

---

[3]See Fama (1990), Hafer et al (1992), Gosnell and Kolb (1997), Fauvel et al (1999), Diebold and Li (2006), De Pooter et al (2007) and Guidolin and Timmermann (2008).

[4]As seen in Chapter 2 key studies include Campbell and Shiller (1991), Taylor (1992), Cuthbertson (1996), Cuthbertson et al (1996, 2003), Longstaff (2000) and Sarno et al (2007).

sufficient predictability of returns, such that investors allocate significantly more to stocks the longer their investment horizon and that those who ignore this estimation risk over allocate to stocks[5]. He highlights the importance of parameter uncertainty, since the standard errors of the estimated coefficients may suggest, that the true forecasting ability of the explanatory variable is lower than the actual coefficient estimate implies.

Key studies that evaluate the predictive power of theory informed models under a decision-based criteria include Abhyankar, Sarno and Valente (2005, henceforth ASV), who find exchange rate predictability to significantly alter the optimal allocation and evidence of economic value to exchange rate predictability. Whereby, the realised terminal wealth of an investor over a 10-year horizon who uses a monetary model to forecast the exchange rate, is higher than that of the investor who assumes no predictability. Further evidence is provided by Garratt and Lee (2009, GL) who also incorporate model uncertainty in their investigation[6]. Della Corte, Sarno and Thornton (2008, DST) assess the validity of the EH and examine the economic value of departures from the EH[7]. Under statistical criteria, like many in the literature as discussed in Chapter 2, they reject the EH. But under economic value criteria they find support for the EH.

The findings of ASV, DST and GL illustrate that the conclusion of how well theory informed models perform compared to atheoretic models from a forecasting perspective, is sensitive to the evaluation criterion used. To be precise, under statistical measures atheoretic models like the random walk are difficult to beat. But under decision-

---

[5]Barberis models returns as being $i.i.d.$ under the assumption of no predictability, compared to a VAR model when modelling stock returns. Under $i.i.d.$ returns the mean and variance evolve linearly. He discusses how the variance of the cumulative returns grow slower/faster than linearly with the investment horizon when predictability/parameter uncertainty is considered. He demonstrates how the allocations differ under the two models because the variances evolving linearly under the $i.i.d.$ and less than linearly under the VAR. This is examined in more detail in the Chapter 5.

[6]They use four different models: the Efficient Market Hypothesis, the Monetary Fundamental, the Purchasing Power Parity and the Autoregressive Models to predict the exchange rate, together with a weighted average of the four models' density forecasts computed using Bayesian Model Averaging.

[7]In that, is there any gain in using the unconstrained VAR over the constrained VAR, where the VAR is constrained by the restrictions implied by the EH, see Campbell and Shiller (1991).

based methods encouraging evidence in favour of predictability, as captured by theory informed models, is found.

The contributions of this chapter are empirical. The studies mentioned above that examine asset return predictability focus their attention on stock returns and exchange rates. We consider asset return predictability and the economic value of this predictability in the context of interest rates, and to my knowledge DST is the only other to consider bond return predictability using a decision-based criteria. Our work differs from DST in several ways; they focus their attention on testing the EH and seeing if there is economic value of departures from the EH. We however, use an unrestricted VAR model, since we do not seek to test the EH in this chapter. Here we are concerned with firstly, our models' ability to forecast out-of-sample and provide a statistical and decision-based assessment and secondly, we examine the effects of parameter uncertainty and predictability on optimal allocation, neither of which DST consider.

We focus our attention on the importance of predictability and parameter uncertainty in asset allocation, generating density forecasts to capture the risk as well as the return of the asset and consider the economic value of these forecasts to the investor. We compare, not just the optimal allocation of an investor who assumes no predictability against one who believes returns are predictable, but see if there are significant gains to an investor in terms of wealth from assuming predictability. Our work is based on the asset allocation framework used in Barberis, ASV and GL.

In brief, we compute the optimal portfolio allocation for a buy-and-hold investor with power utility over terminal wealth, using weekly data for the UK during 1997 week 10 to 2007 week 19, for two assets the 1-month and the n-month T-bill for $n = 3, 6, 12$ months, over investment horizons of up to 2 years. The assets considered here provide risk-free returns and only differ in their maturity. We consider two models that make opposing assumptions regarding return predictability. If the investor believes returns

162

are not predictable, he uses a random walk with drift model to forecast returns and inform his allocation decisions. If however, he believes that returns are predictable he uses the MVART model. Using these two models, we examine the impact predictability and parameter uncertainty have on how the investor optimally allocates his portfolio. Here we do not take into account transactions costs.

Two types of uncertainty are considered here future and parameter, where future uncertainty is that surrounding forecasts which is the result of unobserved future shocks. Parameter uncertainty for a given model, is that "concerned with the robustness of forecasts to the choice of parameter values" GLPS (2006, pp. 153). Initially parameter uncertainty is ignored, such that the investor assumes that there is no uncertainty surrounding the parameter values and takes them to be fixed at their estimated values. However, potentially insignificant standard errors may imply that the true forecasting ability of past changes in the yields and the spreads between them may be weaker than that suggested by the coefficients estimates. Hence, by accommodating parameter uncertainty an improvement in the portfolio decision may be observed.

Both statistical and decision-based criteria are used to evaluate the out-of-sample forecasting performance of the models, to ascertain if indeed there is economic value to assuming that bond returns are predictable and given by the MVART model. Our results suggest that the investor allocates differently when he assumes predictability, to an investor who assumes that returns are not predictable. Further, the results under the statistical and decision-based criteria do not entirely coincide, with the evidence in favour of predictability being different under each criteria. We do not find economic value of predictability at all the portfolio combinations and investment horizons considered. However, under the economic value measure the random walk model is not superior to the MVART by the margin implied by the root mean squared errors. In fact under the economic value measure some evidence in favour of predictability is found, suggesting that the results are sensitive to the assessment criteria used.

Although we recognise that it may be unrealistic to assume that an investor will have a portfolio consisting only of T-bills, the aim here is not to propose and test realistic allocation strategies. But the objective is to use this illustrative strategy to ascertain the effect of predictability and parameter uncertainty on allocation decisions and further, determine if there is economic value to interest rate predictability.

As highlighted in ASV, the measure of economic value based on utility, that we also use here, is one way to define economic value but not the only way. The aim here is to present an alternative assessment criteria i.e. decision-based, that takes into consideration the uncertainty about the forecast and the investor's feelings about risk by computing utility, which the investor seeks to maximise. As opposed to solely using a statistical measure which does not take these factors into account. Under this decision-based criteria, we use the end-of-period terminal utility to establish if there are gains to the investor from assuming that interest rates are explained by the MVART model.

The setup of this paper is as follows, Section 4.2 provides details of how we model the interest rates, the investment decision and the framework used to evaluate the economic value of predictability, when parameter uncertainty is both ignored and accounted for. Section 4.3 describes the dataset, the estimated models and provides a statistical evaluation of the forecasting performance of each model. In Sections 4.4 and 4.5 we analyse the effects of predictability and parameter uncertainty on the optimal allocations, and judge the models' forecasting performance by comparing the realised end-of-period wealth generated under each, respectively. Section 4.6 concludes.

4.2 Optimal Allocation, Parameter Uncertainty and Predictability

We examine how a utility-maximising investor allocates his portfolio between 1-month and 3-, 6-, or 12-month T-bills, that is between a selection of short run risk-free bonds. We consider if there are gains in utility for an investor who believes returns are predictable and employs a theory informed model to forecast interest rates, in comparison to one who believes returns are not predictable. Here we describe the model estimated when we consider predictability and that when bond return predictability is ignored. Further, we introduce how we measure the economic value of interest rates under parameter uncertainty and predictability.

4.2.1 Modelling the Interest Rate

As described in Chapter 2 and 3, the EH states that the return on a n-period zero coupon bond should equal the return on a rolling investment in a sequence of k m-period bonds[8], plus a time invariant term premium/liquidity premium $c^{(n,m)}$

$$r_t^{(n)} = \frac{1}{k} \left[\sum_{i=1}^{k} E_t \left(r_{t+(i-1)m}^{(m)} \right) \right] + c^{(n,m)} \tag{4.1}$$

That is, the long rate is given by an average of the expected future short rates plus a liquidity premium. Further recall

$$s_t^{(n,m)} = \sum_{i=1}^{k-1} \left(\frac{k-i}{k} \right) E_t \left(\Delta r_{t+im}^{(m)} \right) + c^{(n,m)} \tag{4.2}$$

---

[8]Where the integer $k = n/m$ and $n > m$.

which describes the spread by expected future changes in the short rate. That is, aside from the constant premium, the spread is a reflection of the expected change in the short term rates over the life of the long bond.

If the above yields contain a stochastic trend, then if the EH holds the spreads should be stationary. That is to say, if the yields share a common stochastic trend, then we should find $(q-1)$ cointegrating vectors, as implied by stationary bivariate spreads, in a set of q non-stationary yields. This is one method by which the validity of the EH is tested using a VECM framework, as discussed in the previous chapter.

Assuming that yields are difference stationary and that there exists a cointegrating relationship between n- and m-period yields, such that spreads are stationary, then there exists a Wold representation which can be approximated by the following VAR(p) model

$$\mathbf{x}_t = \boldsymbol{\mu} + \mathbf{B}_1 \mathbf{x}_{t-1} + \mathbf{B}_2 \mathbf{x}_{t-2} + ... + \mathbf{B}_p \mathbf{x}_{t-p} + \boldsymbol{\epsilon}_t \tag{4.3}$$

where in this multivariate case $q = 4$ with $\mathbf{x}_t = \left(s_t^{(12,1)}, s_t^{(6,1)}, s_t^{(3,1)}, \Delta r_t^{(1)} \right)'$. From the previous chapter this VAR(p) is denoted the Multivariate VAR in Transformed Interest Rates (MVART) model. It embeds the cointegrating relations between the yields, describing the change in the m-period rate and the spread between the n- and m-period yields using past changes and spreads.

The previous chapter models the UK term structure using a set of statistical and theory informed models including an Autoregressive, Vector Autoregressive in Differences, VAR in Transformed Interest Rates and a Vector Error-Correction model. These are estimated under both a bivariate and a multivariate specification. In brief, evidence in favour of the EH is found in the form of stationary spreads, yields sharing a common stochastic trend and the over-identifying restrictions on the cointegrating vectors implied by the EH can not be rejected. The in-sample properties of the theory informed VECM and MVART models suggest that they have greater explanatory power for the

term structure, in comparison to the statistical based models. Further, the restrictions under the bivariate models are rejected in favour of the multivariate models.

The EH can be further tested by imposing restrictions on the MVART model[9]. Here we do not impose such a stringent structure. But given we find evidence of cointegration amongst the yields, we use the MVART model that embeds the cointegration as implied by the EH to explain and capture the UK term structure and in turn forecast the yields. As such, if the investor believes bill returns are predictable, he employs the MVART model to forecast future returns.

In contrast, if however the investor believes that returns are not predictable, the random walk with drift (RW) model is used to forecast returns. Where in the VAR $\mathbf{x}_t = \left(\Delta r_t^{(12)}, \Delta r_t^{(6)}, \Delta r_t^{(3)}, \Delta r_t^{(1)} \right)'$ and there are no predictor variables. Hence $\mathbf{B}_i = 0$ and $\Delta r_t^{(n)} = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t$, i.e. the returns are given by a random walk with drift.

So under the assumption of no predictability in returns the investor uses the RW model to forecast returns. Under the assumption of predictability he uses the MVART model to forecast returns. By modelling T-bill returns in these two ways allows us to examine whether it is beneficial to the investor, in terms of wealth gains, to assume that returns are predictable as opposed to assuming they are not. Both the RW and the MVART models are estimated when parameter uncertainty, which is the uncertainty about the true values of the model's parameters, is both ignored and accounted for[10].

4.2.2   Investment Strategies

At time T, the investor has to choose how to allocate his wealth between investing in m-month or n-month T-bills. Where $r_t^{(n)}$ and $r_t^{(m)}$ are the annualised, continuously compounded nominal zero coupon yields on a m-month and n-month bill respectively.

---

[9]Such that the spreads are determined in accordance to the EH. That is, the restrictions impose the EH structure on the VAR, see Campbell and Shiller (1991).

[10]We differentiate between when the model is estimated subject to stochastic uncertainty only and when it is estimated subject to stochastic and parameter uncertainty, by denoting them as RW and MVART, and RWPU and MVARTPU respectively.

The investor wishes to make the investment over the period T to $T + H$ i.e. over H periods, assuming a buy-and-hold strategy where once the investment is made it remains untouched until the end of the investment horizon in $T + H$.

If $H = n$, then the investment horizon is equal to the maturity of the longest asset considered. So the investor allocates the proportion ω of his initial wealth in a sequence of rolling investments in m-month bills and $(1 - \omega)$ in a single n-month bill. However, if $H > n$ then the investor allocates ω in a sequence of m-month bills and $(1 - \omega)$ in a sequence of n-month bills. That is, he invests ω in a sequence of $s = H/m$ rolling investments in short m-period bills, and $(1 - \omega)$ in a sequence of $l = H/n$ in long n-period bills. Assuming that initial wealth is $W_T = 1$, then the cumulative return at the end of the investment period W_{T+H}, is

for $H > n$

$$
\begin{aligned}
W_{T+H} &= \omega \prod_{i=1}^{s} \left(1 + R_{T+(i-1)m}^{(m)}\right)^{\frac{1}{s}} + (1 - \omega) \prod_{i=1}^{l} \left(1 + R_{T+(i-1)n}^{(n)}\right)^{\frac{1}{l}} \\
&= \omega \left\{ \exp\left(\frac{1}{s} \sum_{i=1}^{s} r_{T+(i-1)m}^{(m)}\right) \right\} + (1 - \omega) \left\{ \exp\left(\frac{1}{l} \sum_{i=1}^{l} r_{T+(i-1)n}^{(n)}\right) \right\} \quad (4.4)
\end{aligned}
$$

for $H = n$

$$
W_{T+H} = \omega \left\{ \exp\left(\frac{1}{k} \sum_{i=1}^{k} r_{T+(i-1)m}^{(m)}\right) \right\} + (1 - \omega) \exp\left(r_T^{(n)}\right) \quad (4.5)
$$

where $\frac{1}{s} \sum_{i=1}^{s} \left(r_{T+(i-1)m}^{(m)}\right)$ and $\frac{1}{l} \sum_{i=1}^{l} \left(r_{T+(i-1)n}^{(n)}\right)$ give the cumulative returns from a sequence of rolling investments in m-period and n-period T-bills over a H period horizon respectively[11], see Appendix 5. It can be seen that if $H = n$, then the return

---

[11] This data, from the BoE, is continuously compounded and annualised such that the monthly return for an n-month bill can be obtained by dividing by 12. Since we want to compare the expected return under each investment the annualised returns need to be scaled and are done so by the $\frac{1}{s}$ and $\frac{1}{l}$ terms.

from the investment in the long n-period bill is known with certainty[12], such that $E_T\left(r_T^{(n)}\right) = r_T^{(n)}$. From here on we will not explicitly state a separate equation for the case where $H = n$, as setting $H = n$ in equation (4.4) will yield the same result.

Into this decision making process risk aversion can be incorporated, using the end-of-horizon wealth W_{T+H} from the standard constant relative risk aversion (CRRA) power utility function[13], the utility is

$$v\left(W\right) = \frac{W^{1-A}}{1-A} \tag{4.6}$$

where A is the coefficient of risk aversion. The investor faces the following optimisation problem in T

$$\max_\omega E_T\left(v\left(W_{T+H}\left(\omega\right)\right) \mid \Omega_T\right) \tag{4.7}$$

where the investor computes the expectation above conditional upon the information set available at T. That is, the investor maximises the expected utility with respect to the proportion of the portfolio allocated to the investment in the m-month bills, e.g. $\omega = 1$ suggests all is invested in the m-month, equally $\omega = 0$ suggests all in the n-month bills.

Assessment of the above strategy requires expectations of $v\left(W_{T+H}\left(\omega\right)\right)$ to be formed based on the information available to the investor at T. Due to the non-linear nature of W_{T+H}, the point forecasts of $r_{T+(i-1)m}^{(m)}$ and $r_{T+(i-1)n}^{(n)}$ over the H period investment horizon are insufficient to evaluate $E_T\left(v\left(W_{T+H}\left(\omega\right)\right) \mid \Omega_T\right)$. Since point forecasts do not convey the future uncertainty surrounding the returns, if the risk about the future returns is of concern then the entire distribution surrounding each forecast must be considered, see Appendix 6. Here we incorporate risk in two ways; firstly for each step ahead forecast of the returns calculated we compute a distribution of values which

---

[12]In this sense the long n-period return is riskless. But the return from the rolling investments in m-period bills are risky, since future short rates are unknown. However, when $H > n$ then the cumulative returns from both the n- and m-period investments will contain unknown future returns.

[13]See Campbell and Viceira (2003, pp. 24 and 42) for details on the properties of the CRRA utility function.

accounts for the uncertainty surrounding the projections; secondly we consider the investors feelings about risk through the calculation of expected utility rather than expected wealth.

The entire joint probability distribution of the forecast values of $r^{(m)}_{T+(i-1)m}$ and $r^{(n)}_{T+(i-1)n}$, where $i = 1$ to s, or l respectively, is considered. From which the expected utility can be calculated for each possible proportion, i.e. $\omega = 0$ to 1, increasing by 0.01 each time. The optimal allocation is that which yields the maximum expected utility from all possible allocations.

Fundamental to this optimisation problem is the distribution the investor employs to evaluate this expectation. The distribution used depends upon whether the investor assumes predictability in bill returns or not. To ascertain the influence of predictability on allocation decisions a comparison between the allocations of an investor who ignores predictability, to that of one who takes it into account can be made, this is discussed in greater detail below.

The above investment strategy will be explored with $n = 3, 6$ and 12-month T-bill rates and $m = 1$-month rate. Where each of the three pairwise combinations of n and m will be examined. So over the investment horizon H, the investor will consider how to optimally allocate when faced with the following three portfolio choices:

(1) 1-month vs 3-month under $H = 3, 6, 12, 18, 24$ months

(2) 1-month vs 6-month under $H = 6, 12, 18, 24$ months

(3) 1-month vs 12-month under $H = 12, 24$ months

In the case where $H = n$ there is uncertainty surrounding the future values of the short rates only and for $H > n$ there is uncertainty about the future long rates too. Each of (1), (2) and (3) are considered for the levels of risk aversion $A = 2, 5, 10$, with $A = 10$ being the highest level. We consider how (i) the assumptions regarding predictability, (ii) whether the investor incorporates parameter uncertainty or not, (iii)

his level of risk aversion and (iv) the length of the investment horizon, affect the way in which he allocates his portfolio. This is examined under the three different portfolio combinations detailed above[14].

4.2.3 The Probability Density Function of the Forecast Values

The approach taken to estimate the density function when parameter uncertainty is ignored and when it is incorporated will now be discussed. The form the density function $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ takes is determined by the types of uncertainty surrounding the forecasts, as well as the way in which the function is characterised and estimated. The estimation methods adopted by those including Kandel and Stambaugh (1996), Barberis and ASV estimate the density function using a fully Bayesian approach. This involves the construction of a posterior distribution and the use of priors for the parameters. An alternative approach takes a classical stance on the Bayesian approach to estimating the density function, see Garratt, Lee, Pesaran and Shin (2003 and 2006, GLPS) and GL, where the need for priors is avoided since approximations of certain probabilities of interest are made. Here we use the second approach.

In order to evaluate each investment decision over the investment horizon, the investor needs the probability density function of the forecast values of the m- and n-month rates. Following the description in GL, then $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{qt})'$ is a $q \times 1$ vector of q variables (that includes at least the variables of interest i.e. $r_t^{(m)}$ and $r_t^{(n)}$), and $\mathbf{X}_T = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)'$ is a $q \times T$ vector containing the observations 1 to T of the q variables. Since forecasts of the variables are required, the conditional probability density function $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ is of interest. This predictive density function gives the

---

[14]We assume here that there are 4 weeks in a month and 13 weeks in 3 months. From equation (4.4) it would appear that the number of weeks under each of the three investment choices do not match, e.g. if $H = 12$ months it appears that a rolling investment in twelve 1-month bills matures after 48 weeks and not 52 weeks like in a rolling investment in 3-, 6- or 12-month bills. This assumption is made for ease of notation, in practice interest is accured daily and the length of time each investment is held for are equivalent.

probability density function of the forecast values of the q variables, over the horizon $T + 1$ to $T + H$, where $\mathbf{X}_{T+1,H} = (\mathbf{x}_{T+1}, \mathbf{x}_{T+2}, ..., \mathbf{x}_{T+H})'$ conditional on the observed values of the q variables from 1 to T. That is to say, the probability of observing $\mathbf{X}_{T+1,H}$ given that $\mathbf{X}_T$ has already been observed.

The investment problem that the investor is faced with depends on whether he considers the uncertainty surrounding the parameters. In the case where the investor ignores parameter uncertainty and is only concerned with the uncertainty about the future values, he calculates the expectation over the distribution of returns conditional on the fixed parameter values $\widehat{\boldsymbol{\theta}}$, such that the predictive density is $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$. So the investor's problem to solve ignoring parameter uncertainty is

$$\max_{\omega} \left\{ E_T \upsilon \left(W_{T+H}\left(\omega\right)\right) = \int \upsilon \left(W_{T+H}\left(\omega\right)\right) . P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right) d\mathbf{X}_{T+1,H} \right\} \quad (4.8)$$

However, if the investor incorporates parameter uncertainty then the predictive density for the returns is given by $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$, which is conditional on the observed data only

$$P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right) = \int P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right) P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right) d\boldsymbol{\theta} \quad (4.9)$$

The posterior probability of $\boldsymbol{\theta}$, denoted $P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right)$ gives the uncertainty surrounding the parameters given the observed data. So here the investor acknowledges that $\boldsymbol{\theta}$ has a distribution conditional on $\mathbf{X}_T$. Now the investment problem is

$$\max_{\omega} \left\{ E_T \upsilon \left(W_{T+H}\left(\omega\right)\right) = \int \upsilon \left(W_{T+H}\left(\omega\right)\right) . P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right) d\mathbf{X}_{T+1,H} \right\} \quad (4.10)$$

the posterior density $P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right)$, equation (4.9), is proportionate to the product $P\left(\boldsymbol{\theta}\right) . P\left(\mathbf{X}_T \mid \boldsymbol{\theta}\right)$,

i.e. of the prior on $\boldsymbol{\theta}$ and the likelihood function[15].

GLPS and GL suggest that the predictive density $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ can be estimated using Monte Carlo integration techniques if meaningful priors exist. However, in the circumstance where meaningful priors are difficult to obtain, they propose the use of approximations to the key probabilities needed to estimate the predictive density. They make the following assumption for the posterior probability of $\boldsymbol{\theta}$

$$\boldsymbol{\theta} \mid \mathbf{X}_T \overset{\omega}{\sim} N\left(\widehat{\boldsymbol{\theta}}_T, T^{-1}\widehat{\mathbf{V}}_{\boldsymbol{\theta}}\right) \tag{4.11}$$

where $\widehat{\boldsymbol{\theta}}_T$ is the maximum likelihood estimate of the true parameter value of $\boldsymbol{\theta}$ and $T^{-1}\widehat{\mathbf{V}}_{\boldsymbol{\theta}}$ is the asymptotic covariance matrix of the estimated parameters $\widehat{\boldsymbol{\theta}}_T$.

The forecasts are influenced by various uncertainties including stochastic, parameter and model uncertainty. In this exercise we consider both the uncertainty associated with the model (stochastic) and that surrounding the estimated model parameters (parameter). Although we abstract from model uncertainty, we do however model interest rates under two different assumptions. First assuming returns are not predictable as given by the RW model and second that they are predictable as given by the MVART model.

For these two models, using stochastic simulation techniques an estimate of the probability density function of the forecasts can be obtained. Where these simulations provide an estimate of the predictive densities $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$ in the case where parameter uncertainty is ignored and $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ when it is considered. It is then possible to evaluate $E_T\left(\upsilon\left(W_{T+H}\right) \mid \Omega_T\right)$ for a range of portfolio weights ω. In practice $\upsilon\left(W_{T+H}\left(\omega\right)\right)$ is computed $\widetilde{R}$ times for each value of ω, then the mean across these $\widetilde{R}$

---

[15]Using Bayes rule, if (i) $P\left(\theta \mid X_T\right) = P\left(\theta \cap X_T\right)/P\left(X_T\right)$ and (ii) $P\left(X_T \mid \theta\right) = P\left(X_T \cap \theta\right)/P\left(\theta\right)$, rearrange (ii) and substitute into (i) $P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right) = \frac{P(\boldsymbol{\theta}).P(\mathbf{X}_T|\boldsymbol{\theta})}{P(\mathbf{X}_T)}$. If $P\left(\mathbf{X}_T\right)$ is the likelihood of observing our dataset and is equal to some fixed likelihood, then $P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right) \propto P\left(\boldsymbol{\theta}\right).P\left(\mathbf{X}_T \mid \boldsymbol{\theta}\right)$, i.e.they are equal subject to this scalar.

replications is calculated, from which the investor chooses the weight ω that maximises the expected utility $E_T v \left(W_{T+H} \left(\omega \right) \right)$. Here ω takes values 0, 0.01,...,0.99,1, where $\omega = 0$ suggests that the investor should allocate all to n-month bills. Equally $\omega = 1$ suggests that all should be allocated to 1-month bills. Since the weight is between 0 and 1 we do not allow for short selling. Details of the estimation procedure, how the computations are carried out and the method by which the errors are calculated[16] are provided in Appendix 7.

So here we consider four possibilities for the distribution of future returns, given by when the investor assumes no predictability compared to predictability, both when ignoring and then incorporating parameter uncertainty. From this how the optimal allocations differ under the assumptions of predictability and parameter uncertainty, when each of the four distributions are used to forecast returns can be observed.

4.3   Modelling the UK Treasury Bill Rates

4.3.1   Data

We employ data for UK Treasury Bills of maturities 1, 3, 6, 12 months over the period 1997 week 10 to 2007 week 19. Specifically, Wednesday observations of the nominal government spot rates, giving a total of 532 observations for each maturity, all yields are continuously compounded and annualised[17].

The data is official Bank of England (BoE) data on the Government liability curve, we use the reported daily data for the nominal spot rates curve at the short end, from

---

[16]The errors can be drawn using either parametric or non-parametric methods, see GLPS (2006, pp. 166-168). Here parametric methods are utilised, where the errors are assumed to be $i.i.d.N \left(0, \mathbf{\Sigma} \right)$ serially uncorrelated white noise errors.

[17]We use estimated yield curve data, official data estimated by the BoE because actual T-bill data is unavailable during some periods of our sample. However, when plotting the constructed data against the T-bill data for the same maturity when it is available, little difference between the two is observed. So we are satisfied that the data used here is a fair reflection of what the investor would get should he want to undertake an investment in T-bills.

which we select the Wednesday observations. This nominal zero coupon yields data was calculated using gilt prices and General Collateral (GC) repos rates. Zero coupon bonds or risk free discount bonds are used in the construction of yield curves and in empirical studies of the term structure, because it is desirable for the instruments to differ only in their term to maturity. From the BoE data notes these n-month nominal government spot interest rates refer to those applicable today, on a n-month risk-free nominal loan and by definition this (the nominal government spot rate) is the yield to maturity of a nominal zero coupon bond[18].

We assume here a buy-and-hold investor, i.e. the investor holds the bond to maturity[19]. In order to assess the various investment strategies the investor requires the holding period returns for the bonds. Since the investor receives zero coupon payments, the yields used here reflect the total return from holding this asset. In which case the zero coupon yields that we use are equivalent to the holding period returns. So here what we refer to as returns/yields denoted $r_t^{(n)}$, are the holding period returns.

From the summary of this data provided in Chapter 3, the 1-, 3-, 6-, and 12-month T-bill yields in general appear to decline until 2003, after which an upward trend is apparent, Figure 4-1. With average yields of 4.98%, 4.97%, 4.96% and 4.99% for each rate respectively. Further, the ADF, PP and KPSS unit root tests are employed to determine the order of integration of each T-bill rate and the spreads between them over the entire sample. The results indicate that the yields are difference stationary, and the (n, m) rate spreads between the $(3, 1)$, $(6, 1)$ and the $(12, 1)$ rates are stationary.

The two models are each estimated over the period 1997 week 10 to 2004 week 18 (374 observations) and then recursively at weekly intervals through to 1997 week 10 to 2005 week 18 (427 observations), giving 54 recursions in total. For each recursion we generate h-step ahead out-of-sample forecasts[20] for $h = 1, 2, ..., H, ...$ and the investment

---

[18]Further details of the data is provided in the previous chapter and in the Data Appendix.

[19]ASV, Barberis and GL also assume a buy-and-hold investor.

[20]We denote the investment horizon H in months since the T-bills are denoted as n-months to

horizons including $H = 3, 6, 12, 18$ and 24 months. So for the first recursion, we forecast over the period 2004 week 19 to 2006 week 18 and for the last recursion 2005 week 19 to 2007 week 19. For each recursion the investor will use his generated forecasts to determine how to allocate his portfolio optimally[21]. For each of the three portfolio choices i.e. 1-month vs 3-month, 1-month vs 6-month and 1-month vs 12-month, under each A and H, we will have 54 allocation decisions with which to compare the allocations and utility gains under each model, both without and with parameter uncertainty.

4.3.2 Estimation

Estimates of the RW model are given in Table 4.1. From the previous chapter, the MVART model was estimated of order 9 the results are recalled in Tables 4.2 and 4.3[22]. Comparing the two models, a gain in explanatory power for the 1-month return is observed when assuming returns are predictable, all coefficients are jointly significant at the 1% level under the MVART model. The diagnostics show evidence of serial correlation in the RW model, in contrast to the MVART model. The nulls that the residuals are homoskedastic and normal are rejected under both models, which is unsurprising given that we are using financial data. These results do indicate gains in terms of explanatory power and having a model free of serial correlation when predictability is assumed.

4.3.3 Statistical Evaluation of the Forecasting Performance

A statistical evaluation of the out-of-sample forecast performance of the two models can be made using the root mean squared error (RMSE). Table 4.4 gives the RMSEs of the

---

maturity. However, the data has a weekly frequency, so when we refer to the 'h-step' ahead forecasts each 'step' is a week.

[21]In order to evaluate the investment decision the investor does not require all of the h-step ahead forecasts generated, only the forecasts $r^{(m)}_{T+(i-1)m}$ and $r^{(n)}_{T+(i-1)n}$ for $i = 1$ to s, or l respectively.

[22]Estimates of each model for the first recursion only are provided, to give an overall impression of the in-sample predictability. At the forecasting stage the models are estimated recursively.

1-, 3-, 6- and 12-month return forecasts, for forecast horizons $H = 1, 3, 6, 12, 18$ and 24 months for each model, both ignoring and incorporating parameter uncertainty. Table 4.5 reports the ratio of the RMSEs for each model to the benchmark model, which is taken to be the RW model. A value of the ratio less than one indicates that the RMSE of the model is lower than that of the benchmark.

The RMSEs of the bill return forecasts indicate that only at $H = 1$ for the 1-month return does the MVART model beat the benchmark. The RW and RWPU models, that make the strong assumption of returns not being predictable, outperform the theory informed models at each horizon for the 3-, 6- and 12-month returns under this criteria. Comparing the RW models without and with parameter uncertainty the RMSEs are virtually the same, unsurprising since we only estimate μ; whereas small differences are observed between MVART and MVARTPU. In general, the differences observed in the RMSEs amongst the models are small. These results broadly correspond to those found in the interest rate and exchange rate forecasting literature, which in general find sophisticated theory informed models are outperformed by a simple random walk.

From the ratios, it is apparent that not only do the RW and RWPU models outperform the MVART and MVARTPU models. But the ratio of the MVART models to the benchmark increases with H, suggesting their forecasting ability deteriorates relative to the RW model with the investment horizon. Generally, the RMSEs increase up until $H = 6$ before decreasing, implying that they are non-monotonic. Whereby they do not increase with H, but instead oscillate in relative value. This suggests that both models are better at forecasting over the longer horizon, than they are over the shorter. Although this statistical evaluation provides an indication of the forecasting performance of the models, a clear indication of how these models perform in an investment decision making context, in terms of the economic value of the models' forecasts, is not provided[23].

---

[23]When we refer to the 'RW models' and the 'MVART models' this includes without and with

4.4   Predictability & Parameter Uncertainty Effects

We now examine the implications for the optimal allocation when the investor assumes either that returns are not predictable or predictable, in both cases parameter uncertainty is ignored and accounted for.   In the case where parameters are assumed fixed the maximisation problem is given by equation (4.8) and under parameter uncertainty by equation (4.10).

The models are estimated first over 1997 week 10 to 2004 week 18, the optimal weights are calculated from the forecasts generated from each estimated model, this is then repeated moving forward by one week re-calculating the expected utility to find the optimal weight for this new augmented sample.   This is repeated for each recursion, such that we have results for 54 recursions over the total evaluation period 2004 week 19 to 2007 week 19.   Figures 4-7 to 4-9 and the results in Table 4.6, are based on the optimal allocation averaged over the 54 recursions for a particular A, H, model and portfolio combination.

Figures 4-2 to 4-6 show the expected utility for each recursion, from a rolling invest- ment in the 1-month bill given by $E(U1)$ and n-month given by $E(Un)$ for $n = 3, 6, 12$, $A = 2$ and a particular H.   When the investment horizon $H = n$ then this is the actual utility gained.   We only present the plots for $A = 2$ because a significant difference in the allocations between this and that for $A = 5$ and 10 is not observed, where $A = 10$ is the highest level of risk aversion.   These plots show that the computed expected utilities differ under each model for a particular H.   Although generally the expected utilities under each model are of the same magnitude and they change over the recur- sions in a similar way for $H = 3$, differences between the expected utilities calculated under the RW models and the MVART models (both without and with parameter un- certainty) emerge at $H = 6, 12, 18, 24$ months, these differences will be reflected in how

---

parameter uncertainty for each model.   When analysing the results later we compare RW with MVART first ignoring parameter uncertainty and then when they both incorporate it.

the investor ultimately allocates.

Figures 4-7 to 4-9 provide an illustration of the link between the expected utilities computed by the investor and the optimal allocation. They depict the optimal allocation to the 1-month, $100\omega\%$, given the difference in $E(U)$ between the 'all in the n-month' and 'all in the 1-month' investments, for each portfolio combination, for $A = 2$ under the MVART model ignoring parameter uncertainty. It can be seen that when the difference is positive, where we expect to gain a higher utility from investing 'all in the n-month' than 'all in the 1-month', the investor allocates his entire initial wealth in the n-month and zero to the 1-month, and vice versa when the difference is negative[24].

The impact of the various effects upon the allocations is summarised in Table 4.6. Here under each portfolio combination, for a given A, H and model, the table gives as a percentage the number of times out of the 54 recursions the investor allocates everything in the 1-month bill, i.e. $\omega = 1$. We present the results like this for two reasons, first nearly all the allocation results suggest an optimal weight of $\omega = 0$ or 1. This implying that the investor, given the aim is to maximise expected utility, invests everything either in the 1-month or n-month, depending on which yields the higher expected utility and not a mix of the two bills considered. Secondly, using these percentages we can see how the allocations differ under varying degrees of risk aversion, investment horizons, assumptions of predictability and the inclusion of parameter uncertainty.

When considering how the allocations change with the level of risk aversion, they vary by 0 to 8% amongst the three values of A considered[25]. In the majority of cases the difference in allocations is very small. This is arguably not surprising given that the assets considered here belong to the same asset class, only differing in their term to maturity and are positively correlated. So even though the investor may be highly risk

---

[24]Again because a significant difference in allocations between the different values of A is not observed, we only present the plots for $A = 2$. Further, we only show the plots under the MVART model, to provide the reader with a general impression of the link between the expected utilities the investor calculates and how he eventually allocates between the two bills.

[25]This exercise was also carried out with extreme degrees of risk aversion i.e. $A = 20, 50, 100$, the results were not significantly altered.

179

averse, the opportunity for the investor to diversify out the risk here is small because the risk is non-diversifiable.

Now comparing the allocations for different investment horizons for a particular portfolio combination, 1-month vs n-month and particular model. Under the 1-month vs 3-month assuming no predictability as H increases, the allocation to the 1-month increases by up to 10%; assuming predictability the increase is bigger of up to 61%. Under 1-month vs 6-month, assuming no predictability the allocation to the 1-month increases by up to 10%; assuming predictability by up to 24%. Under 1-month vs 12-month, assuming no predictability the allocation increases by 4% ignoring parameter uncertainty, but decreases by up to 14% with parameter uncertainty; under predictability the allocation is unchanged. In general, the allocation to the 1-month increases with the investment horizon, where the increases are larger under predictability. Suggesting that if the investor assumes predictability, then the allocations are more sensitive to the investment horizon.

Here both the RW and MVART models' variances will evolve in the same way i.e. faster than linearly[26] with H, suggesting that the bills appear riskier in the longer run than the shorter horizons for both models. Since it is not possible to rank the variances under each model a priori, we later discuss how the actual variances of the forecasts evolve with H here using our computed RMSEs under each model.

We now examine the effect of predictability ignoring parameter uncertainty. When comparing the RW with the MVART model the investor moves from assuming no predictability to predictability of returns. Under predictability if the investor is better able to capture and explain the yields i.e. in-sample $\overline{R}^2$ is higher, then assuming that the relationship remains stable over the forecast horizon e.g. no structural breaks,

---

[26]Appendix 8 discusses the mean and variance of returns when the returns are modelled as non-stationary, in comparison to when they are treated as stationary, and further examine how the variance behaves over T to $T + H$ under both. The yields are found to be and thus modelled as non-stationary under both the RW and MVART models. So the discussion of how the variances of the returns and cumulative returns will evolve under a RW model in Appendix 8 is appropriate here.

we would expect the MVART model to produce more accurate forecasts. Thus the variance of the forecasts under the MVART would be lower than that under the RW, hence the asset looks less risky under the MVART.

Moving from the assumption that returns are not predictable as under the RW model, to that they are as under the MVART model, there is a gain (in-sample) in explanatory power as can been seen by the $\overline{R}^2$ for $\Delta r_t^{(1)}$, which increases from 0% to 27%. Although only $\Delta r_t^{(1)}$ is directly comparable under both models, from the previous chapter where we modelled $\Delta r_t^{(n)}$ for $n = 1, 3, 6$ and 12 months, past changes and spreads were found to have explanatory power. So it is reasonable to expect all the n-month bills to gain from moving from the RW to the MVART model in terms of explanatory power. Such that it applies to all the returns that as they become more predictable, then they become more attractive to the investor.

However, this gain in in-sample predictability is not translated into an out-of-sample gain, as the RMSEs of the MVART model are higher than those of the RW. However, GL note "...as shown in Clements and Hendry (2005), using RMSE as a criterion penalises models for including variables with low associated t-values even if the model is misspecified by their exclusion", so the poor performance of the MVART model according to the RMSE criterion, may be largely due to the fact that it is heavily parameterised in comparison to the RW model.

In this exercise we consider two different models for forecasting T-bill returns, the random walk model which assumes no variables are able to explain the change in the n-month T-bill return and the MVART model which on the opposite end of the spectrum assumes that past changes and spreads have explanatory power. So if the investor believes that returns are not predictable, then he will use the random walk model to forecast interest rates. Conversely, if he believes they are predictable he will use the MVART model. Ultimately, how he allocates his portfolio is conditional on which model he believes to be a correct representation of reality.

The difference in allocations between the RW and MVART models varies from 8 to 58%. Under 1-month vs 3-month, for all H the MVART models allocate more to the 1-month. However, under the 1-month vs 6-month and the 1-month vs 12-month for each H, it is the RW models that allocate more to the 1-month. For each portfolio combination (1-month vs n-month, for $n = 3, 6$ and 12-month bills) both of the assets will go from being not predictable as determined by the random walk, to being predictable as described by the MVART, so both will gain in terms of predictability. How the allocation differs under the RW to that under the MVART, will depend on which of the two bills gains more from the assumption of predictability.

Generally, large differences are observed between the RW models and MVART models, suggesting that the investor who assumes that returns are not predictable will allocate differently to one who assumes that they are predictable. Thus the assumptions made regarding predictability are important in determining how the investor allocates.

Looking to the effect of parameter uncertainty, incorporating parameter uncertainty increases the variance of the forecast returns at all H, so the asset looks riskier relative to when it was modelled ignoring parameter uncertainty. If all the assets are affected by this additional form of uncertainty, how the allocation changes when parameter uncertainty is incorporated to when it is ignored under a given model will be determined by, for which asset the variance has increased the most.

Comparing the allocations without and with parameter uncertainty for both the RW and the MVART models in turn, allows the impact parameter uncertainty has on the allocation to be isolated. The impact on allocation varies by 0 to 21% under the RW model. In the 1-month vs 3-month portfolio combination, at all H the allocation changes by 0 to 2%; 1-month vs 6-month by 0 to 6%; 1-month vs 12-month $H = 12$ by 0 to 2% and up to 21% for $H = 24$. In most cases the allocation to the 1-month increases under parameter uncertainty, suggesting that the 1-month looks comparatively less risky when parameter uncertainty is incorporated than the n-month.

Under the MVART model parameter uncertainty has more of an impact, where the change in the allocation ranges from 0 to 17%. For each of the portfolio combinations, for all H, in the 1-month vs 3-month the allocation changes by 0 to 6%; 1-month vs 6-month at all H by 7 to 17% and no change is observed in the 1-month vs 12-month. Mostly the allocation to the 1-month decreases under parameter uncertainty, implying that under the MVART model with parameter uncertainty the 1-month looks riskier.

To help explain the optimal allocations observed, we can consider how the variances about the distribution of future predicted returns evolves over the forecast horizon[27]. Here it is reasonable to suppose that the RMSEs and the variances are closely related[28], allowing us to use the RMSEs as an indication of how the variances of the forecasts evolve. Recall Tables 4.4 and 4.5, the non-monotonic RMSEs imply that the variances of the forecasts are also non-monotonic. They increase up until $H = 6, 12$ months and then decline. Which as the following quote from Hall and Hendry (1988, pp. 256-7) highlights may not be so surprising "Hendry (1984) has demonstrated that the standard error need not increase monotonically, as there is a term in the formulae for the model standard error which reaches a maximum and which then may decline.", further Hall and Hendry mention that if this non-monotonicity (in the model standard error) is stronger than the rest of the formulae, then the total standard error will behave in this non-monotonic way.

Since the variance about the forecasts contracts and expands with H, the asset will appear more risky at some horizons than at others. Further, the variances of the different returns oscillate at different rates, otherwise the RMSE for each return would be equal. This indicates that some n-month returns have a greater variance about their distribution of forecasts than others. So at some horizons the 1-month bill will

---

[27]Given that the bill returns are non-stationary in levels, then as demonstrated in Appendix 8 the variances will grow faster than linearly with H under both the RW and MVART models.

[28]The RMSE measures the dispersion around the actual value of a variable, whereas the variance measures the dispersion about the mean of the distribution. If the distribution is unbiased then the mean of the distribution equals the actual value, in which case the RMSE equals the variance of the forecast.

appear more risky than the n-month and at others less. Under the RW models a clear ranking emerges with the 1-month having the largest variance and the 12-month the smallest, for all H. Under the MVART models the variances are not only bigger for each bill and H, but seem to oscillate more. With the 1-month looking less risky than the n-month over the shorter horizons and then more risky over the longer horizons.

This non-monotonicity combined with the fact that the variances of the returns expand and contract at different rates could provide an explanation for the optimal allocations observed here. In that, earlier we saw that when computing expectations of $\upsilon\left(W_{T+H}\left(\omega\right)\right)$, due to the non-linear nature of W_{T+H} the investor requires the variances and the covariances of the forecast returns at each step ahead as well as their means. So how the variances of the forecasts differ for each n-month bill over H and with assumptions regarding predictability and parameter uncertainty, will serve to influence expected utility because of the way it is calculated and thus the ultimate optimal allocation.

4.5 Economic Value of Predictability

It is clear from the results that the allocations are sensitive to the assumptions made regarding predictability, whether parameter uncertainty is incorporated and the investment horizon. In this investigation we also seek to ascertain if there is economic value to interest rate predictability. The RMSE provides a statistical measure of forecast accuracy, now we will assess forecast performance by considering the economic value of the forecasts to the investor. An economic evaluation of the forecast performance of each model is reported in Tables 4.7 to 4.9 under each portfolio combination for each A and H. We compute the end-of-period wealth[29] that the risk averse investor would have achieved over 2004 week 19 to 2007 week 19, had he allocated his portfolio

---

[29]We follow ASV and GL in our measure of economic value being based on wealth, as mentioned in ASV this is only one way to define economic value.

as suggested by the optimal weights of each model for a particular A and H. Where the optimal weight ω is that calculated by solving the utility maximisation problem[30]. These realised wealths are averaged over 54 recursions and then ranked in descending order, so the performance of each model can be compared.

Apart from the RW and MVART models described above, under which parameter uncertainty is both ignored and incorporated to derive the optimal allocations, we introduce two passive 'lazy' strategies. Under the lazy strategies the investor makes no attempt to model or predict the returns, but instead either invests (1) all in 1-month bills (A1) or (2) all in n-month bills (An) for $n = 3$, 6 or 12.

The top position is mostly occupied by the lazy 'all in 1-month' strategy, with the 'all in n-month' strategy coming last. However, during a large part of the forecast horizon 2004 week 19 to 2007 week 19, over which this evaluation of the models is made, the 1-month return was higher than the others, Figure 4-1. Looking to positions 2 to 5, under the 1-month vs 3-month the RW models perform well at $H = 3, 6$ and the MVART models at $H = 12, 18, 24$. However, under 1-month vs 6-month and 1-month vs 12-month the success of the RW models is apparent at all A and H.

We have considered two different forecast evaluation criteria, the RMSE which provides a statistical measure and the realised wealths which is an economic value measure of forecast accuracy. When comparing the results under the two criteria, under the 1-month vs 3-month the RW models outperform the MVART models at all H, when considering the RMSEs. Whereas the MVART models achieve a higher realised wealth at $H = 12, 18, 24$ under the economic value measure. So the conclusions drawn under each criteria do not entirely correspond. Under the 1-month vs 6-month and 1-month vs 12-month both the statistical and the economic value criteria do correspond, to find the RW models perform best at all H.

Also from the RMSE ratios Table 4.5, the performance of the MVART models rela-

---

[30]The forecasts produced by each model are used to determine the optimal weight. These weights are then combined with actual/realised returns to give the realised end-of-horizon wealth.

tive to the benchmark deteriorates as H increases. So we would expect the difference between the realised wealths of the two models to also increase as H increases to reflect this. But the differences in the realised wealths between the RW and MVART models are small. Further, the ratios of the realised wealths to the benchmark, Table 4.10, are quite constant over H.

So the RMSEs suggest that the MVART models perform much worse than the RW. But the economic value results suggest there is little difference between the two models. This implies that the forecasting errors of the MVART are larger, but when this is translated through to realised wealths a huge difference is not observed, with the realised wealths not exhibiting such an obvious difference in performance between assumptions of no predictability and predictability.

We ignore transactions costs in this exercise. But appreciate that in the comparison of an investment in a n-period long bond with a rolling investment in m-period bills, the transaction costs incurred under the two alternate investments will differ. With that of the rolling investment being higher. From the realised wealths in Tables 4.7 to 4.9, comparing the realised wealths of A1 with An for $n = 3, 6, 12$ months, compares investing in a long bond with the rolling strategy. For all A and H in this sample the strategy of investing all in a sequence of 1-month bills yields the highest realised wealth. However, the differences in the realised wealths between the two strategies are very small, so any gain from adopting a rolling strategy is likely to be eradicated with the inclusion of transactions costs.

4.6   Conclusion

Previous studies find that by using an alternative assessment criterion, that considers the economic value of the forecasts e.g. using profits or utility, in comparison to conventional statistical methods can yield favourable results for theory informed models.

These papers find evidence to support the argument made by Leitch and Tanner (1991), Granger and Pesaran (2000) and Pesaran and Skouras (2004) that forecasts should be judged in the decision making context for which they are intended. In our case, like others before us, this context is that of investment decision making.

In this investigation we examine how the optimal allocations of a utility maximising investor are affected by, the assumptions he makes regarding the predictability of returns and parameter uncertainty. If the investor believes that returns are not predictable he uses an atheoretic random walk with drift model in his decision making. Alternatively, if he believes they are predictable he uses the theory informed MVART model. Further, we evaluate the economic value of the out-of-sample forecasts of bill returns generated under these two models. The investment decision is whether to invest in 1-month or n-month bonds, this is examined in a framework that both ignores parameter uncertainty and explicitly allows for it.

The importance of the Expectations Hypothesis is well documented in the interest rate literature. Here given that we find evidence of cointegration amongst the yields from the previous chapter, we use the MVART model that embeds the cointegration relations implied by the EH to model the interest rates and then evaluate interest rate predictability in this economic value framework.

The effect of assuming predictability on the optimal allocation is considerable, where the optimal weights under predictability were in some cases greatly different to those under no predictability. The effect of parameter uncertainty is small over the investment horizon considered here, even though previous studies such as Barberis (2000) report significant parameter uncertainty effects these are prominent at horizons longer than considered here.

Under the statistical evaluation criterion, the RW models outperform the MVART models at almost all horizons when forecasting bond returns. Although evidence of misspecification at the estimation stage is found under the RW model. When

187

an economic value approach is used over the sample investigated here, we find some evidence that an investor seeking to optimally allocate his wealth between 1-month and n-month UK T-bills, is better off in terms of higher end-of-horizon wealth by assuming predictability than an investor who assumes no predictability. Notably, for $n = 3$ at $H = 12, 18, 24$. For $n = 6$ and 12 months under the assumption of predictability, the realised wealths are marginally lower than those when the investor assumes returns are not predictable.

From Clements and Hendry (2005a) as quoted in GL the RMSE criteria penalises heavily parameterised models. Here this could be exaggerating the superior performance of the RW model relative to the MVART under this criteria. Although the realised wealths imply that in some cases there are no gains from assuming predictability over no predictability, the realised wealths of the two models are of the same magnitude and their ratios are very close to one. What can be deduced from these results is that the performance of the MVART under the economic value criteria is not as poor as the RMSEs would suggest. This evidence of disparity between the results obtained under the two criterion suggests that the results are sensitive to the criterion used.

This exercise considers the allocation between short term bills, with the longest maturity being 12 months and investment horizon being 24 months. We find for a utility maximising investor it is not optimal to hold a mixed portfolio of 1-month and n-month bills. As explained earlier this is largely due to the assets considered here, risk-free T-bills, having a high positive correlation amongst them. So the opportunity for the investor to diversify here is small, because the risk he faces is non-diversifiable given these assets. One possible extension to this investigation is to consider a longer investment horizon and see if parameter uncertainty has more of an impact at these longer horizons. Another would be to expand the portfolio set to include a risky asset.

In the next chapter, we use the asset allocation framework discussed here and extend it by considering a risky asset. We compare a risky stock, where it is possible to make a

loss from investing, with a risk-free T-bill. Now the investor does have the opportunity to diversify risk, so we would expect to see optimal mixed portfolios. In short, we explore how a utility maximising investor optimally allocates his portfolio between bonds and stocks, using a range of atheoretic and theory informed models of bond and stock returns. We examine the impact of parameter uncertainty and predictability in returns on how the investor allocates and if there is economic value of predictability.

In conclusion, our results highlight the importance of evaluating the forecasts using an appropriate criterion. Here the investor is concerned with optimally allocating his portfolio, so it is necessary to incorporate the investor's feelings about risk and to consider the distribution about the predicted returns into this decision making process. In which case, the RMSE criterion seems somewhat inadequate for this purpose compared to the economic value measure. In the context of investment decision making, we find some evidence of economic value to interest rate predictability, such that the investor may gain from assuming predictability.

Figure 4-1: Nominal Spot Yields 1997 to 2007
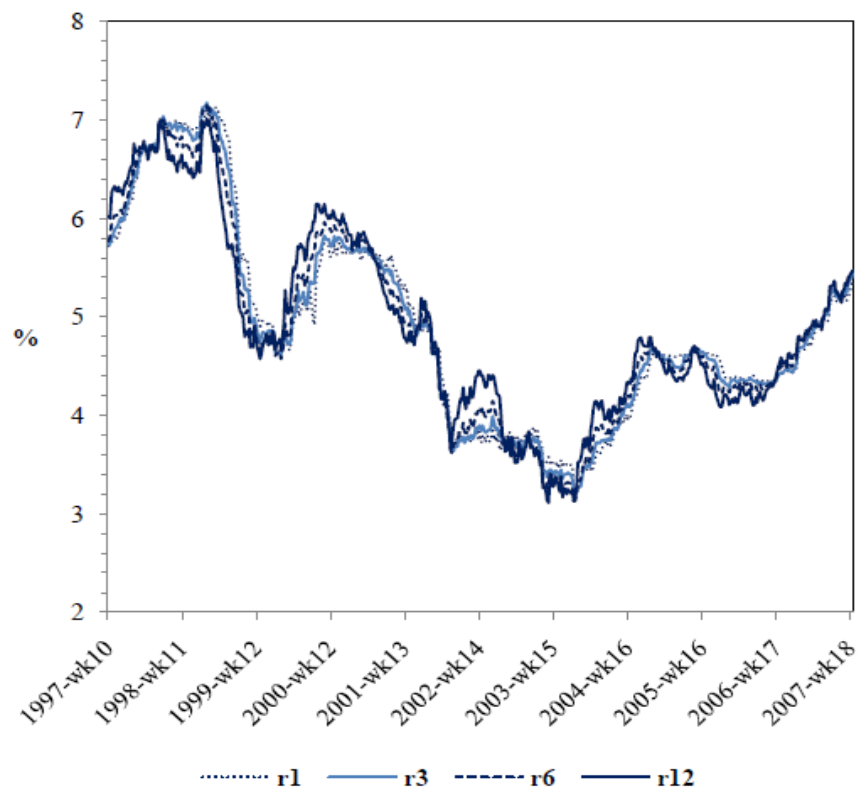
Figure 4-2: Expected Utility under RW and MVART Models, for H=3 and A=2
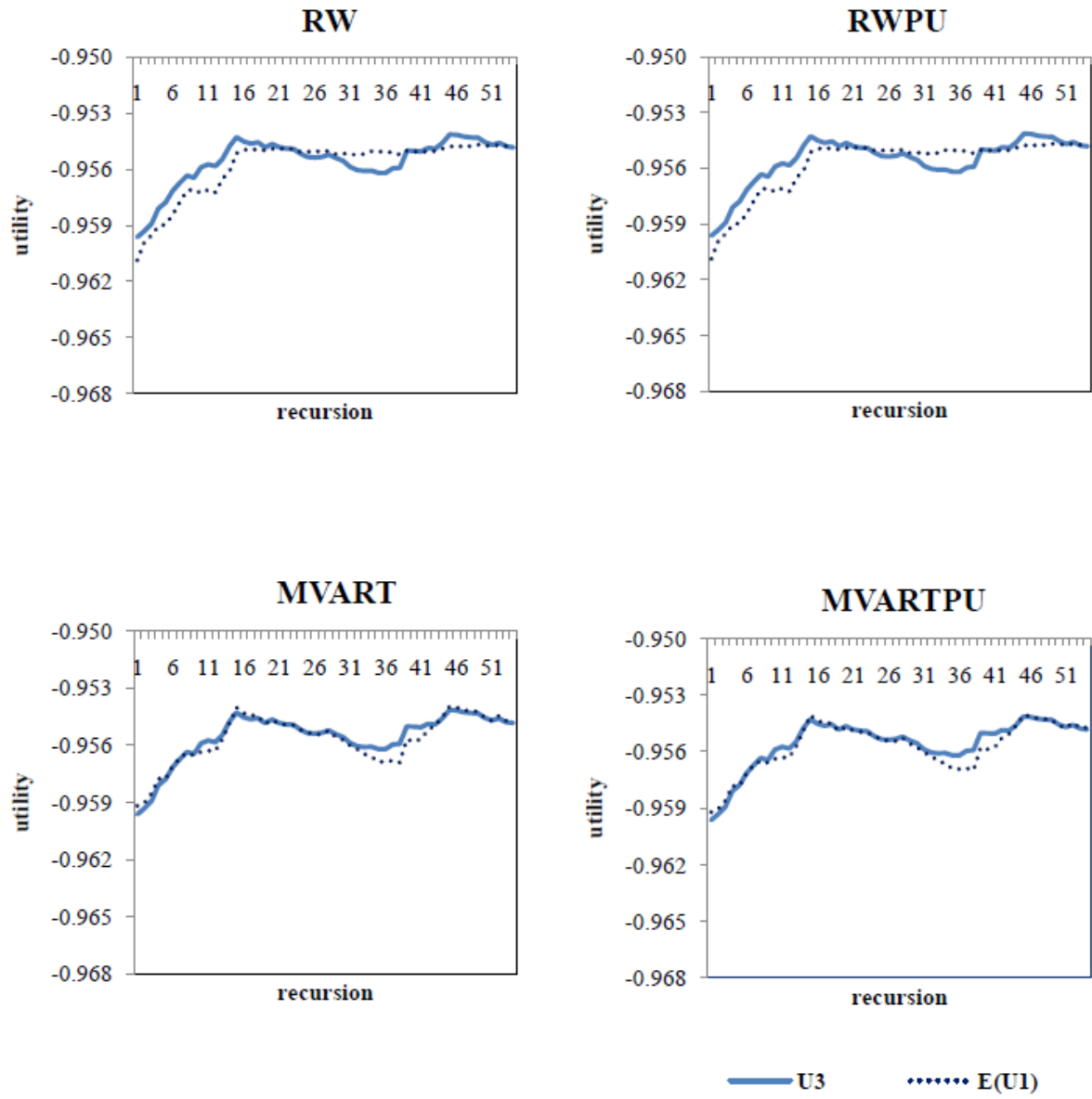
Figure 4-3: Expected Utility under RW and MVART Models, for H=6 and A=2
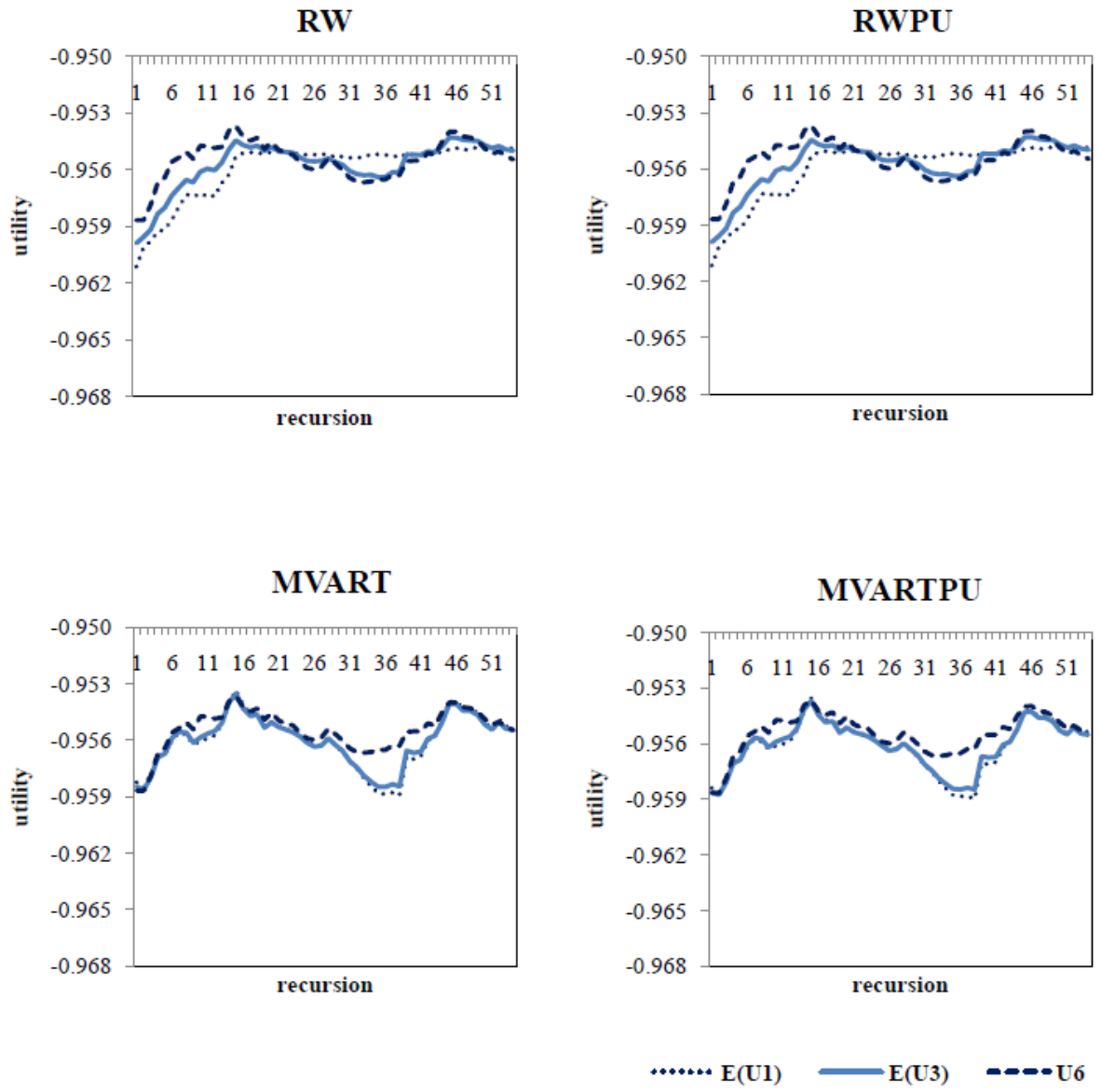
Figure 4-4: Expected Utility under RW and MVART Models, for H=12 and A=2
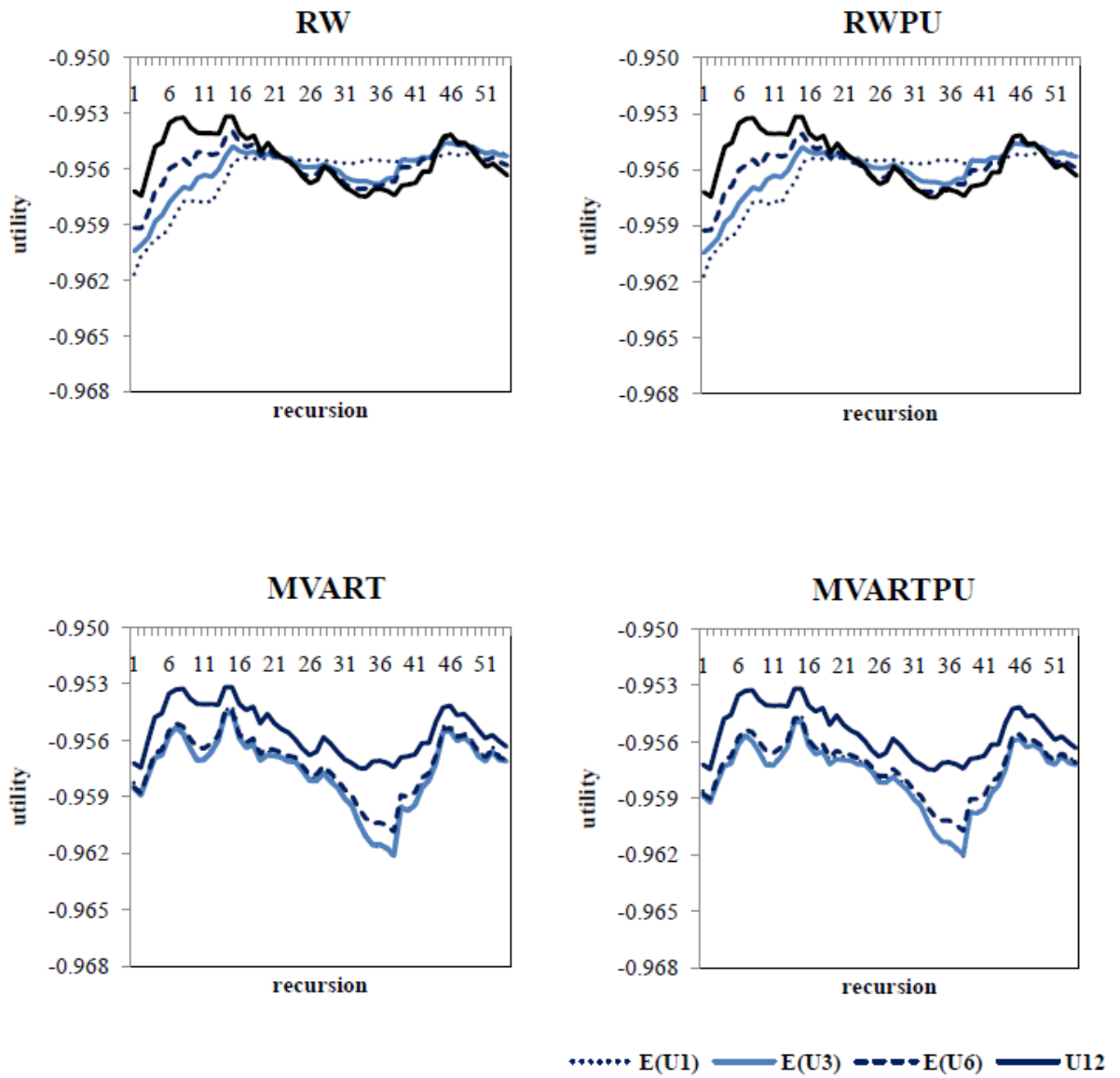
Figure 4-5: Expected Utility under RW and MVART Models, for H=18 and A=2
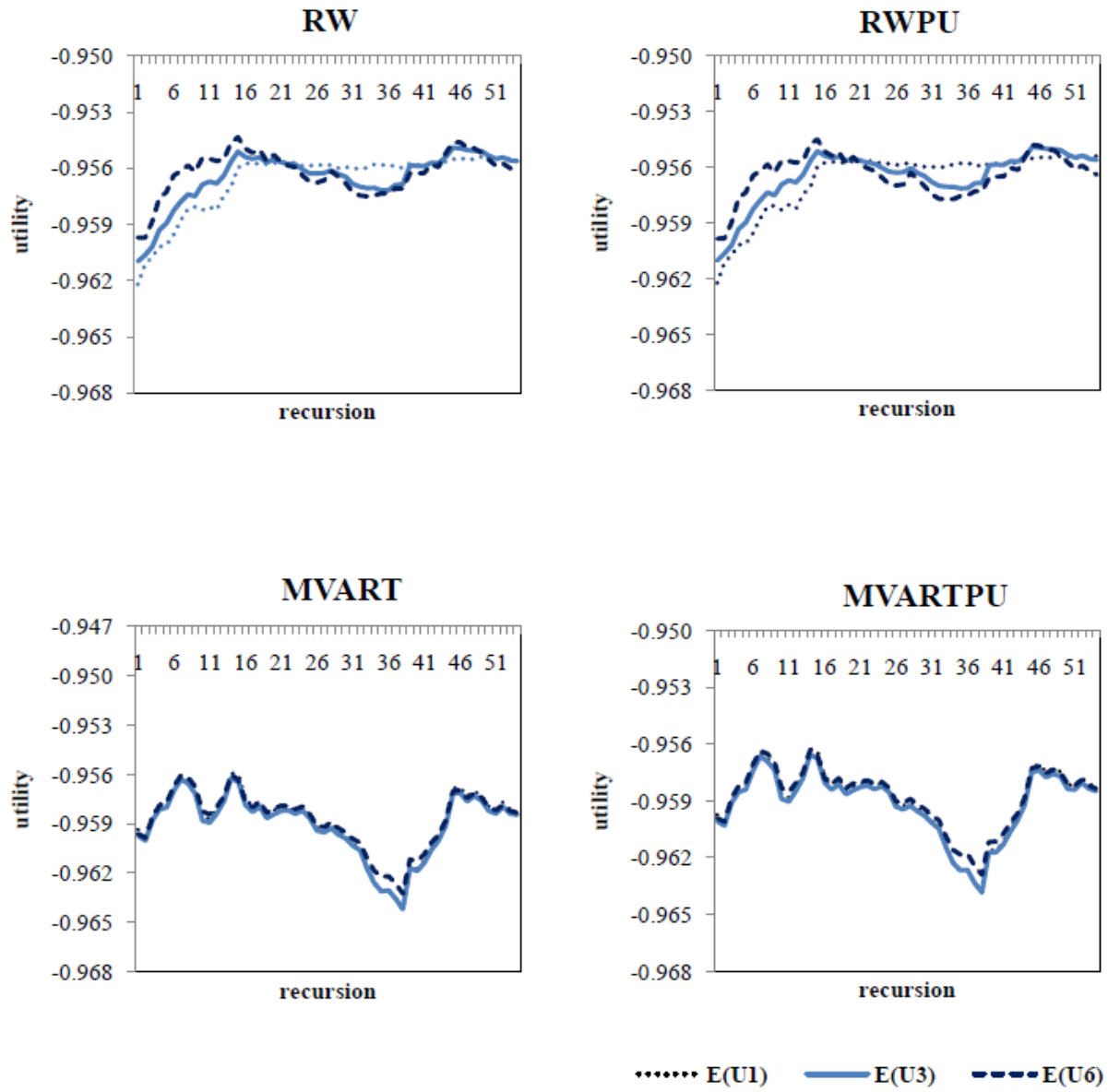
Figure 4-6: Expected Utility under RW and MVART Models, for H=24 and A=2
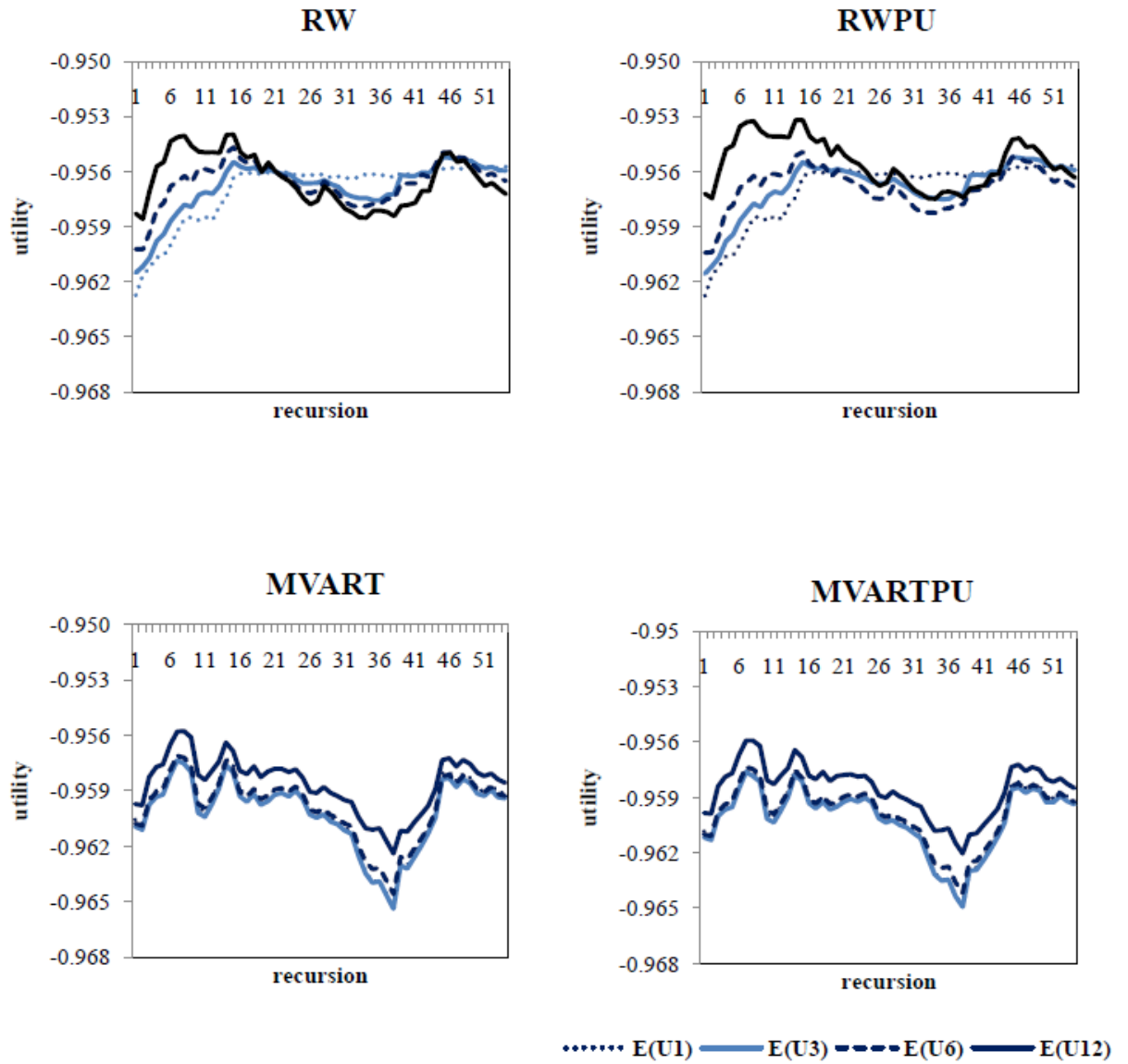
Figure 4-7: 1-month vs 3-month Allocations under the MVART Model, for A=2

196
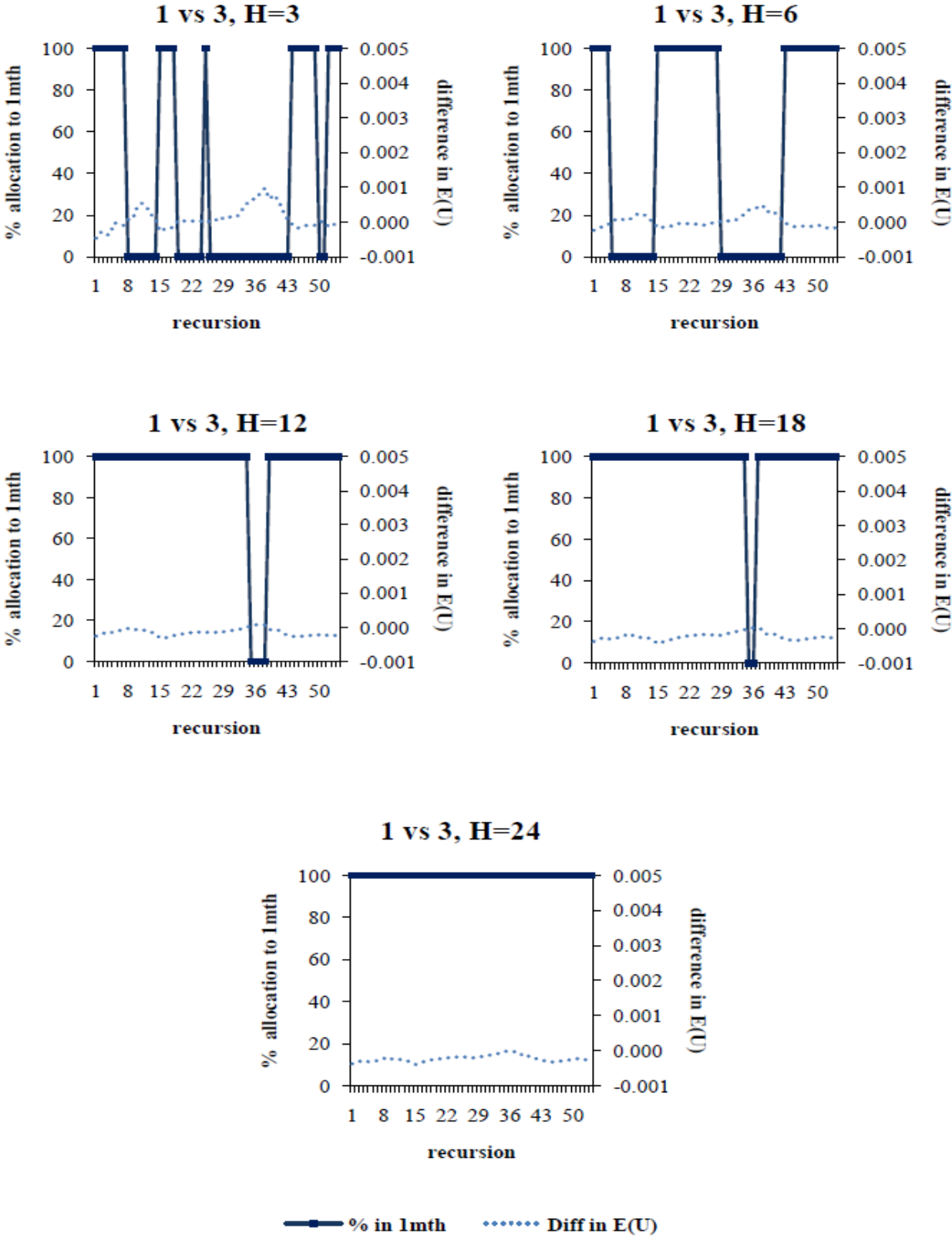
Figure 4-8: 1-month vs 6-month Allocations under the MVART Model, for A=2
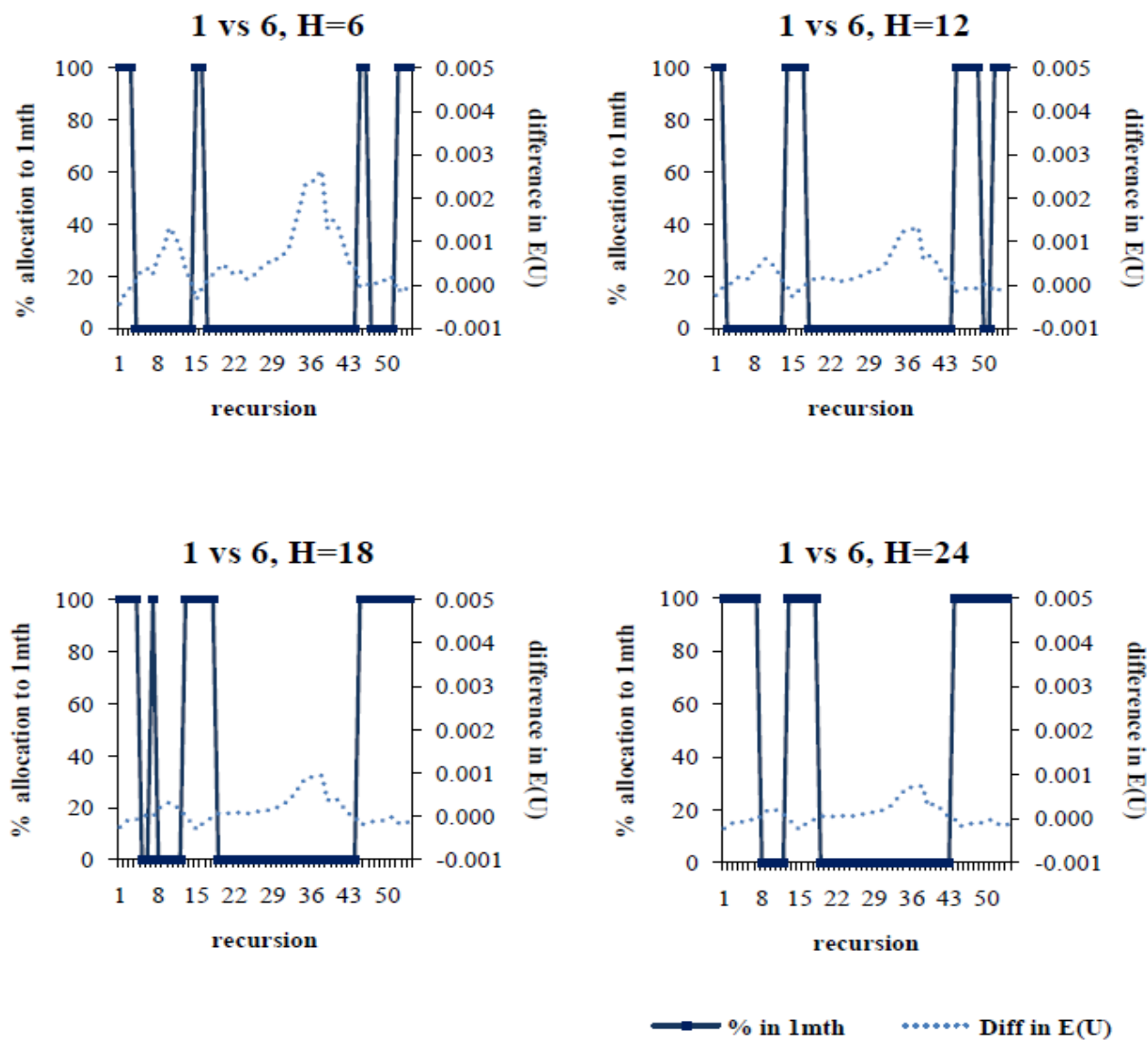
Figure 4-9: 1-month vs 12-month Allocations under the MVART Model, for A=2

1 vs 12, H=12

1 vs 12, H=24
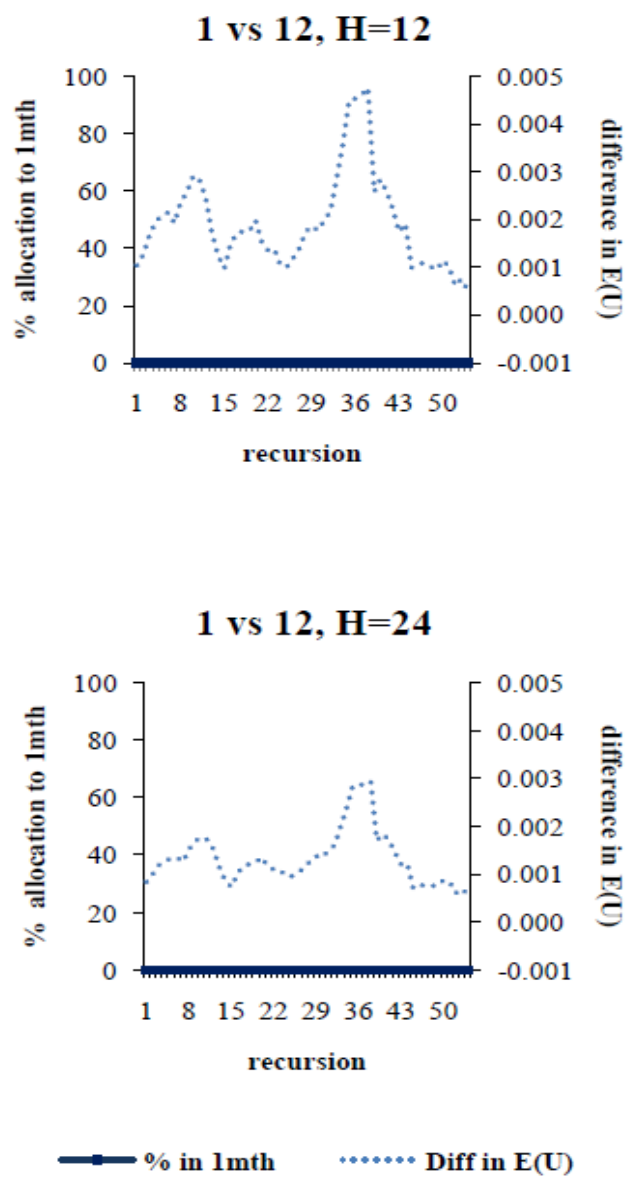
Figure 4-10: Root Mean Squared Errors of Return
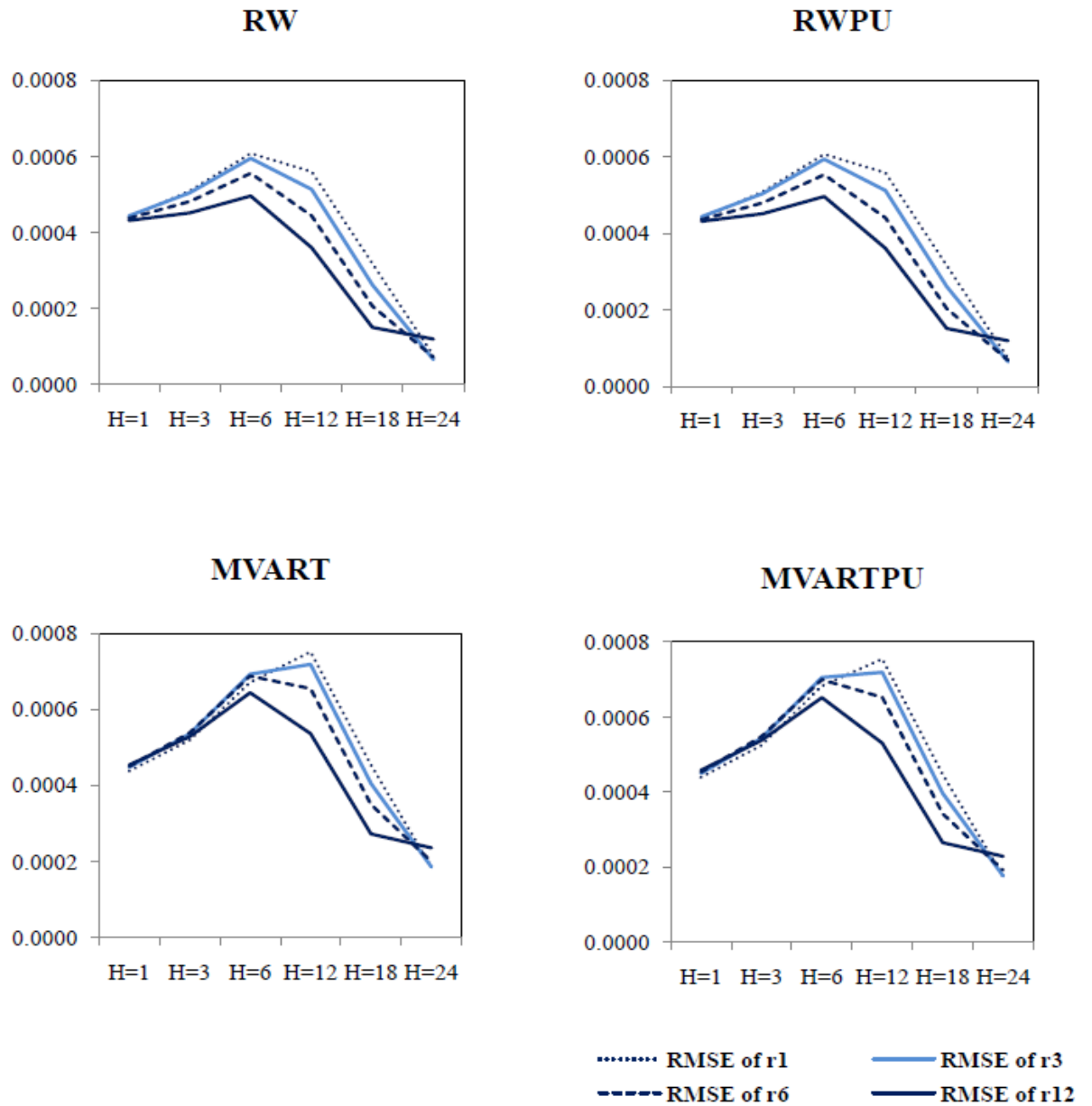
Table 4.1: Estimation of Random Walk Model

Equation	Δr_t^1	Δr_t^3	Δr_t^6	Δr_t^{12}
$\boldsymbol{\mu}$	-0.000046 (0.000041)	-0.000043 (0.000033)	-0.000042 (0.000038)	-0.000044 (0.000049)
$\overline{R^2}$	0.000	0.000	0.000	0.000
$\widehat{\sigma}$	0.0008	0.0006	0.0007	0.0009
$eq^n LL$	2137.27	2219.11	2167.24	2067.53
$\chi_N^2\,[2]$	358.23***	261.98***	76.42***	32.80***
$\chi_{SC}^2\,[1]$	0.44	20.29***	11.20***	0.78
$\chi_{SC}^2\,[2]$	0.55	21.97***	13.05***	1.47
$\chi_{SC}^2\,[6]$	20.73***	49.08***	33.34***	12.87**
$\chi_{SC}^2\,[12]$	29.45***	57.84***	40.17***	16.83
$\chi_{SC}^2\,[9]$	38.15***			
$\chi_{SC}^2\,[12]$	12.34			

Notes: Standard errors in parenthesis (.). The $\overline{R}^2$, standard error of the regression $\left(\widehat{\sigma}\right)$, log likelihood of the equation (LL) presented, together with the chi-squared statistics for Breusch-Pagan Serial Correlation test (SC) and the Jarque-Bera Test for Normality (N). The random walk with drift model assumes that $\Delta r_t^n = \mu + \epsilon_t$ for n = 1,3,6 and 12, and each is estimated over 1997 week 10 to 2004 week 18 (373 observations). Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 4.2: Estimation of MVART(p) Model

Equation	$s_t^{12,1}$	$s_t^{6,1}$	$s_t^{3,1}$	Δr_t^1
$s_{t-1}^{12,1}$	1.054*** (0.216)	0.289** (0.144)	0.159 (0.099)	0.172 (0.151)
$s_{t-2}^{12,1}$	0.214 (0.259)	−0.028 (0.174)	−0.054 (0.109)	−0.153 (0.184)
$s_{t-3}^{12,1}$	−0.489* (0.249)	−0.435** (0.176)	−0.285*** (0.110)	0.089 (0.137)
$s_{t-4}^{12,1}$	0.337 (0.264)	0.232 (0.186)	0.115 (0.112)	0.056 (0.192)
$s_{t-5}^{12,1}$	−0.055 (0.280)	−0.009 (0.198)	0.047 (0.124)	0.294 (0.245)
$s_{t-6}^{12,1}$	0.209 (0.263)	0.075 (0.181)	−0.020 (0.118)	−0.325* (0.178)
$s_{t-7}^{12,1}$	−0.329 (0.263)	−0.329* (0.183)	−0.225* (0.119)	0.055 (0.179)
$s_{t-8}^{12,1}$	0.216 (0.283)	0.124 (0.199)	0.106 (0.130)	0.051 (0.172)
$s_{t-9}^{12,1}$	0.013 (0.208)	0.156 (0.143)	0.131* (0.094)	−0.068 (0.150)
$s_{t-1}^{6,1}$	−0.407 (0.514)	0.392 (0.337)	−0.120 (0.234)	−0.540 (0.357)
$s_{t-2}^{6,1}$	−0.186 (0.582)	0.257 (0.398)	0.247 (0.268)	0.250 (0.449)
$s_{t-3}^{6,1}$	1.102* (0.595)	0.942** (0.422)	0.543** (0.266)	−0.021 (0.362)
$s_{t-4}^{6,1}$	−0.771 (0.621)	−0.520 (0.441)	−0.245 (0.278)	−0.379 (0.464)
$s_{t-5}^{6,1}$	0.234 (0.664)	0.123 (0.465)	−0.068 (0.289)	−0.471 (0.538)
$s_{t-6}^{6,1}$	−0.159 (0.614)	−0.030 (0.426)	0.122 (0.276)	0.385 (0.434)
$s_{t-7}^{6,1}$	0.235 (0.61)	0.373 (0.411)	0.231 (0.256)	0.236 (0.402)
$s_{t-8}^{6,1}$	−0.242 (0.659)	−0.099 (0.454)	−0.068 (0.298)	−0.633 (0.443)
$s_{t-9}^{6,1}$	−0.068 (0.485)	−0.383 (0.338)	−0.337 (0.232)	0.369 (0.357)
$s_{t-1}^{3,1}$	0.081 (0.467)	−0.060 (0.283)	0.525*** (0.184)	1.126*** (0.338)
$s_{t-2}^{3,1}$	0.089 (0.538)	−0.125 (0.360)	−0.095 (0.242)	−0.391 (0.393)
$s_{t-3}^{3,1}$	−1.040* (0.570)	−0.883** (0.406)	−0.489* (0.255)	0.061 (0.358)
$s_{t-4}^{3,1}$	0.786 (0.606)	0.559 (0.419)	0.251 (0.266)	0.512 (0.424)
$s_{t-5}^{3,1}$	0.034 (0.607)	0.052 (0.443)	0.183 (0.288)	−0.145 (0.478)
$s_{t-6}^{3,1}$	−0.546 (0.546)	−0.468 (0.381)	−0.371 (0.256)	0.380 (0.413)
$s_{t-7}^{3,1}$	0.080 (0.514)	−0.089 (0.344)	−0.037 (0.206)	−0.4512 (0.347)

$s_{t-8}^{3,1}$	-0.074 (0.600)	-0.025 (0.388)	-0.024 (0.260)	0.909^{**} (0.441)
$s_{t-9}^{3,1}$	0.191 (0.453)	0.402 (0.325)	0.358 (0.220)	-0.543^{*} (0.319)
Δr_{t-1}^{1}	0.034 (0.114)	0.053 (0.083)	0.032 (0.054)	0.089 (0.085)
Δr_{t-2}^{1}	0.167^{*} (0.094)	0.171^{**} (0.069)	0.127^{***} (0.047)	-0.175^{***} (0.061)
Δr_{t-3}^{1}	-0.016 (0.096)	0.015 (0.065)	0.024 (0.038)	-0.023 (0.056)
Δr_{t-4}^{1}	0.065 (0.104)	0.080 (0.068)	0.041 (0.042)	0.066 (0.073)
Δr_{t-5}^{1}	0.255 (0.155)	0.263^{**} (0.104)	0.184^{***} (0.059)	-0.176^{**} (0.078)
Δr_{t-6}^{1}	-0.018 (0.091)	-0.003 (0.070)	-0.005 (0.047)	0.088 (0.067)
Δr_{t-7}^{1}	0.045 (0.93)	0.028 (0.064)	-0.001 (0.040)	-0.066 (0.067)
Δr_{t-8}^{1}	-0.069 (0.092)	-0.034 (0.059)	-0.021 (0.035)	0.090 (0.071)
Δr_{t-9}^{1}	-0.111 (0.069)	-0.086^{*} (0.046)	-0.044 (0.030)	0.040 (0.052)
$inpt$	-0.0001 (0.00007)	-0.00009^{*} (0.00005)	-0.00004 (0.00003)	0.000005 (0.00004)
$\overline{R}^2$	0.934	0.901	0.829	0.273
$\widehat{\sigma}$	0.001	0.0007	0.0004	0.0007
$F[36, 327]$	144.42^{***}	93.27^{***}	50.02^{***}	4.80^{***}
$eq^{n}LL$	2028.0	2163.5	2337.9	2172.6
$system\ LL$	9659.1			
$\chi_N^2\,[8]$	35.03^{***}			
$\chi_H^2\,[720]$	910.12^{***}			
$\chi_{SC}^2\,[16]$	18.48			

Notes: Standard errors in parenthesis. A MVART(9) is estimated over 1997 week 10 to 2004 week 18 (364 observations). The regressions are estimated with Newey-West heteroskedastic and autocorrelation corrected errors. The $\overline{R}^2$, standard error of the regression ($\widehat{\sigma}$), F-statistic to test the joint significance of the estimated coefficients and the log likelihood of the equation (LL) are presented, together with the model diagnostic tests which are all carried out on the VAR residuals. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. Chi-squared statistics presented for: (N) the VAR Residual Normality Test; (H) the VAR Residual Heteroskedasticity Test, and (SC) the VAR Residual Serial Correlation LM Test for the null of no serial correlation at lag 9. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 4.4: Root Mean Squared Errors of Returns

(a) 1-month returns (r_t^1)

Model	$H=1$	$H=3$	$H=6$	$H=12$	$H=18$	$H=24$
RW	0.000439	0.000510	0.000608	0.000560	0.000318	0.000076
MVART	0.000438	0.000519	0.000671	0.000751	0.000454	0.000184
RWPU	0.000439	0.000509	0.000607	0.000558	0.000316	0.000074
MVARTPU	0.000440	0000525	0.000683	0.000755	0.000447	0.000175

(b) 3-month returns (r_t^3)

Model	$H=1$	$H=3$	$H=6$	$H=12$	$H=18$	$H=24$
RW	0.000444	0.000505	0.000595	0.000513	0.000262	0.000065
MVART	0.000447	0.000535	0.000692	0.000718	0.000404	0.000185
RWPU	0.000444	0.000504	0.000594	0.000512	0.000262	0.000065
MVARTPU	0.000450	0.000543	0.000706	0.000720	0.000397	0.0001762

(c) 6-month returns (r_t^6)

Model	$H=1$	$H=3$	$H=6$	$H=12$	$H=18$	$H=24$
RW	0.000437	0.00482	0.000556	0.000444	0.000206	0.000071
MVART	0.000449	0.000538	0.000687	0.000654	0.000350	0.000199
RWPU	0.000437	0.000481	0.000554	0.000441	0.000204	0.000069
MVARTPU	0.000453	0.000547	0.000699	0.000652	0.000342	0.000191

(d) 12-month returns (r_t^{12})

Model	$H=1$	$H=3$	$H=6$	$H=12$	$H=18$	$H=24$
RW	0.000431	0.000451	0.000496	0.000360	0.000150	0.000119
MVART	0.000454	0.000528	0.000643	0.000536	0.000272	0.000236
RWPU	0.000432	0.000452	0.000496	0.000361	0.000151	0.000120
MVARTPU	0.000458	0.000538	0.000652	0.000530	0.000265	0.000229

Notes: The RMSEs are computed for each model for the horizons H = 1, 3, 6, 12, 18, 24 months, and for each model as follows $\sqrt{\frac{\sum_{i=1}^{54}(r_{T+H}-\widehat{r}_{T+H})_i}{54}}$ where r_{T+H} is the actual monthly return i.e. $r_t^1, r_t^3, r_t^6, r_t^{12}$, $\widehat{r}_{T+H}$ is the forecast and the difference between the two $\left(r_{T+H} - \widehat{r}_{T+H}\right)$ is computed for each recursion i, there are 54 weekly recursions. The RW and MVART models are estimated subject to stochastic uncertainty only, the RWPU and MVARTPU models consider parameter uncertainty too.

Table 4.5: Ratio of RMSEs of Returns

(a) 1-month returns (r_t^1)

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
RW	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MVART	0.9957	1.0175	1.1026	1.3409	1.4268	2.4348
RWPU	0.9997	0.9991	0.9982	0.9966	0.9951	0.9831
MVARTPU	1.0002	1.0297	1.1239	1.3480	1.4051	2.3168

(b) 3-month returns (r_t^3)

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
RW	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MVART	1.0079	1.0605	1.1635	1.3995	1.5427	2.8367
RWPU	0.9998	0.9994	0.9989	0.9981	0.9981	0.9890
MVARTPU	1.0141	1.0765	1.1860	1.4021	1.5145	2.7034

(c) 6-month returns (r_t^6)

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
RW	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MVART	1.0287	1.1166	1.2365	1.4733	1.6996	2.7928
RWPU	0.9994	0.9983	0.9969	0.9940	0.9909	0.9700
MVARTPU	1.0372	1.1366	1.2588	1.4689	1.6629	2.6872

(d) 12-month returns (r_t^{12})

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
RW	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MVART	1.0518	1.1697	1.2972	1.4897	1.8120	1.9793
RWPU	1.0002	1.0002	1.0007	1.0025	1.0081	1.0048
MVARTPU	1.0620	1.1916	1.3144	1.4733	1.7645	1.9198

Notes: The above ratios are that of the RMSE for each model to the RMSE of the RW model, which is taken as the benchmark.

Table 4.6: Effects on Allocation

Strategy		Model			
		RW	RWPU	MVART	MVARTPU
1 vs 3, H=3	$A = 2$	35	35	39	33
	$A = 5$	33	35	39	33
	$A = 10$	33	31	38	33
1 vs 3, H=6	$A = 2$	35	35	56	54
	$A = 5$	35	35	56	54
	$A = 10$	33	35	56	52
1 vs 3, H=12	$A = 2$	41	43	93	93
	$A = 5$	41	41	93	93
	$A = 10$	35	37	93	91
1 vs 3, H=18	$A = 2$	44	44	96	93
	$A = 5$	43	4	96	93
	$A = 10$	41	39	94	93
1 vs 3, H=24	$A = 2$	44	44	100	100
	$A = 5$	44	44	100	100
	$A = 10$	43	44	100	100
1 vs 6, H=6	$A = 2$	46	46	19	9
	$A = 5$	46	46	19	9
	$A = 10$	46	46	19	7
1 vs 6, H=12	$A = 2$	48	52	26	13
	$A = 5$	48	52	26	11
	$A = 10$	46	52	24	11
1 vs 6, H=18	$A = 2$	48	54	39	26
	$A = 5$	48	54	35	26
	$A = 10$	48	52	31	24
1 vs 6, H=24	$A = 2$	52	56	44	31
	$A = 5$	52	56	43	30
	$A = 10$	48	54	41	24
1 vs 12, H=12	$A = 2$	48	50	0	0
	$A = 5$	48	48	0	0
	$A = 10$	48	48	0	0
1 vs 12, H=24	$A = 2$	52	37	0	0
	$A = 5$	52	37	0	0
	$A = 10$	52	31	0	0

Notes: Here under each portfolio combination, for a given A,H and model, the table gives as a percentage the number of times out of the 54 recursions the investor allocates everything to the 1-month bill.

Table 4.7: Realised Wealth under 1-month vs 3-month Strategy

(a) under $A = 2$

Position	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	A1 1.046746	A1 1.046877	A1 1.04654	A1 1.046072	A1 1.046091
2nd	RW 1.046740	RWPU 1.046804	MVARTPU 1.046511	MVART 1.046062	MVARTPU 1.046091
3rd	RWPU 1.046740	RW 1.046796	MVART 1.046511	MVARTPU 1.046044	MVART 1.046091
4th	MVART 1.046511	MVART 1.046774	RWPU 1.046432	RWPU 1.045910	RW 1.046010
5th	A3 1.046509	MVARTPU 1.046769	RW 1.046431	RW 1.045910	RWPU 1.046009
6th	MVARTPU 1.046505	A3 1.046649	A3 1.046252	A3 1.045755	A3 1.045967

(b) under $A = 5$

Position	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	A1 1.046746	A1 1.046877	A1 1.046544	A1 1.046072	A1 1.046091
2nd	RW 1.046740	RWPU 1.046798	MVARTPU 1.046511	MVART 1.046062	MVARTPU 1.046091
3rd	RWPU 1.046740	RW 1.046796	MVART 1.046511	MVARTPU 1.046044	MVART 1.046091
4th	MVART 1.046511	MVART 1.046774	RWPU 1.046429	RWPU 1.04591	RW 1.04601
5th	A3 1.046509	MVARTPU 1.046769	RW 1.046429	RW 1.045908	RWPU 1.046009
6th	MVARTPU 1.046504	A3 1.046649	A3 1.046252	A3 1.045755	A3 1.045967

(c) under $A = 10$

Position	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	A1 1.046746	A1 1.046877	A1 1.046544	A1 1.046072	A1 1.046091
2nd	RW 1.046739	RW 1.046796	MVART 1.046511	MVART 1.046059	MVARTPU 1.046091
3rd	RWPU 1.046739	RWPU 1.046796	MVARTPU 1.046506	MVARTPU 1.046044	MVART 1.046091
4th	MVART 1.046511	MVART 1.046774	RW 1.046423	RWPU 1.045907	RWPU 1.04601
5th	A3 1.046509	MVARTPU 1.046766	RWPU 1.046423	RW 1.045905	RW 1.046009
6th	MVARTPU 1.046504	A3 1.046649	A3 1.046252	A3 1.045755	A3 1.045967

Notes: See those for Table (4.9).

Table 4.8: Realised Wealth under 1-month vs 6-month Strategy

(a) under $A = 2$

Position	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	RW 1.047030	A1 1.046877	A1 1.046544	A1 1.046072
2nd	RWPU 1.047027	RWPU 1.04638	RWPU 1.045804	RWPU 1.045836
3rd	A1 1.046746	RW 1.046372	RW 1.045792	RW 1.045828
4th	MVARTPU 1.046634	MVARTPU 1.046079	MVART 1.045625	MVART 1.045740
5th	A6 1.046609	MVART 1.046078	MVARTPU 1.045542	MVARTPU 1.045666
6th	MVART 1.046602	A6 1.046041	A6 1.045444	A6 1.045548

(b) under $A = 5$

Position	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	RW 1.04703	A1 1.046877	A1 1.046544	A1 1.046072
2nd	RWPU 1.047029	RWPU 1.046380	RWPU 1.045804	RWPU 1.045836
3rd	A1 1.046746	RW 1.046372	RW 1.045788	RW 1.045828
4th	MVARTPU 1.046634	MVART 1.046078	MVART 1.045610	MVART 1.045733
5th	A6 1.046634	MVARTPU 1.046077	MVARTPU 1.045542	MVARTPU 1.045665
6th	MVART 1.046602	A6 1.046041	A6 1.045444	A6 1.045548

(c) under $A = 10$

Position	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	RW 1.047030	A1 1.046877	A1 1.046544	A1 1.046072
2nd	RWPU 1.047030	RWPU 1.046380	RWPU 1.045805	RWPU 1.045838
3rd	A1 1.046746	RW 1.046371	RW 1.045787	RW 1.045824
4th	MVARTPU 1.046635	MVART 1.046078	MVART 1.045590	MVART 1.045727
5th	MVART 1.046614	MVARTPU 1.046055	MVARTPU 1.045536	MVARTPU 1.045627
6th	A6 1.046609	A6 1.046041	A6 1.045444	A6 1.045548

Notes: See those for Table (4.9).

Table 4.9: Realised Wealth under 1-month vs 12-month Strategy

Position	A = 2		A = 5		A = 10	
	H = 12	H = 24	H = 12	H = 24	H = 12	H = 24
1st	RWPU 1.047033	A1 1.046072	RWPU 1.047033	A1 1.046072	RWPU 1.047029	A1 1.046072
2nd	RW 1.047027	RW 1.045806	RW 1.047026	RW 1.045805	RW 1.047025	RW 1.045805
3rd	A12 1.046566	RWPU 1.045629	A12 1.046566	RWPU 1.045621	A12 1.046566	RWPU 1.045595
4th	MVARTPU 1.046566	MVARTPU 1.044899	MVARTPU 1.046566	MVARTPU 1.044899	MVARTPU 1.046566	MVARTPU 1.044899
5th	MVART 1.046566	MVART 1.044899	MVART 1.046566	MVART 1.044899	MVART 1.046566	MVART 1.044899
6th	A1 1.046746	A12 1.044889	A1 1.046746	A12 1.044889	A1 1.046746	A12 1.044889

Notes: The realised wealths above are the end-of-investment horizon wealths that the investor would have achieved over 2004 week 19 to 2007 week 19 had he allocated according to the optimal weights for each model, A and H. These end-of-investment horizon wealths have been averaged over the 54 recursions. The realised wealth for each model are ranked in descending order, for a particular A and H. The tables above show how the two models, RW and MVART without and with parameter uncertainty perform, together with the lazy strategies in terms of their achieved realised wealths. The actual realised wealths are given below the model code. 'A1' is the 'all in 1-month' and 'An' is the 'all in n-month' lazy strategy for n = 3, 6 and 12 months.

Table 4.10: Ratios of Realised Wealth

(a) 1-month vs 3-month

Model	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
RW	1.0000	1.0000	1.0000	1.0000	1.0000
RWPU	1.0000	1.0000	1.0000	1.0000	1.0000
MVART	0.9998	1.0000	1.0001	1.0001	1.0001
MVARTPU	0.9998	1.0000	1.0001	1.0001	1.0001

(b) 1-month vs 6-month

Model	$H = 6$	$H = 12$	$H = 18$	$H = 24$
RW	1.0000	1.0000	1.0000	1.0000
RWPU	1.0000	1.0000	1.0000	1.0000
MVART	0.9996	0.9997	0.9998	0.9999
MVARTPU	0.9996	0.9997	0.9998	0.9998

(c) 1-month vs 12-month

Model	$H = 12$	$H = 24$
RW	1.0000	1.0000
RWPU	1.0000	0.9998
MVART	0.9996	0.9991
MVARTPU	0.9996	0.9991

Notes: Ratio of Realised Wealths under each model to that of the RW model, for A = 2 only because the Ratios under A = 5 and 10 were not significantly different.

Chapter 5

The Economic Value of Interest Rate and Stock Predictability

Abstract

In this chapter, we evaluate the forecast performance of a range of atheoretic and theory informed models of bond and stock returns. The decision making environment is fully described for an investor who would like to optimally allocate his portfolio between bonds and stocks, over an investment horizon of up to two years. We use a weekly dataset on UK Treasury Bill rates and the FTSE All-Share Index over the period 1997 to 2007. We examine the impact parameter uncertainty and predictability in returns have on how the investor optimally allocates his portfolio. We describe the methods by which the forecasts should be computed and used in this context. Both statistical and decision-based criteria are used to evaluate the out-of-sample forecasting performance of the models. Our results suggest that in the context of investment decision making under an economic value criterion, the investor gains from not only assuming predictability but by modelling the bond and stock returns together.

Keywords: density forecasting, decision-based forecast evaluation, interest rate and stock return models, predictability and parameter uncertainty.

5.1 Introduction

Evidence of predictability in asset returns has been reported by a number of studies including Campbell (1987), Fama and French (1988a,b and 1989), Kandel and Stambaugh (1996) and Ang and Bekaert (2007). They show variables including the dividend yield and term structure variables, have predictive power for the stock return. This overturning the long standing view held up until the 1970s in financial economics, that returns are not predictable. Most of this evidence is based on studies that assess predictability from a statistical standpoint, using measures like the significance of estimated coefficients, the explanatory power of the regressors and the RMSEs of forecasts.

However, recent research argues that conventional statistical forecast evaluation criteria, usually based on some measure of the forecasts error, may be inappropriate. Instead, it would be more appropriate to evaluate forecast accuracy using profitability, given firms use forecasts to increase profits, Leitch and Tanner (1991). Further, Granger and Pesaran (2000) and Pesaran and Skouras (2004) argue that forecasts should be evaluated in the decision making context for which they are intended. These studies advocate the use of decision-based forecast evaluation[1], where forecasts are judged in terms of their economic value to the user, rather than in terms of forecast errors.

This chapter first examines the impact of predictability in bond and stock returns, together with the effect of parameter uncertainty upon how an investor optimally allocates his portfolio. Second, we consider if there is any economic value to the investor of bond and stock return predictability.

Authors including West, Edison and Cho (1993), Pesaran and Timmermann (1995), Xia (2001), Brooks and Persand (2003), Avramov (2002), Boudry and Gray (2003), and Marquering and Verbeek (2004) have previously considered the economic value of predictability in returns within an asset allocation framework[2]. Barberis (2000) con-

---

[1]We may also refer to economic value measures, these are the same as decision-based measures.
[2]Chapter 2 provides a detailed review of the relevant literature.

211

siders how asset return predictability affects optimal portfolio choice for long horizon investors, if this allocation differs with the investment horizon and further the impact on allocation when parameter uncertainty[3] is incorporated[4]. Barberis defines no predictability as the investor assuming that stock returns are *i.i.d.* and predictability as him believing that a single lagged dividend yield term has predictive power for stock returns. In both cases bond returns are assumed constant. Predictability has the effect of making stocks look less risky and parameter uncertainty makes them look more so. Barberis demonstrates that the investment horizon may not be irrelevant if returns are predictable. Further, even with parameter uncertainty there is sufficient predictability of returns, such that investors allocate significantly more to stocks the longer the horizon and that those who ignore parameter uncertainty over allocate to stocks by a considerable amount.

Recent studies that examine the predictive power of theory informed models under a decision-based criteria for exchange rates include Abhyankar, Sarno and Valente (2005, henceforth ASV) and Garratt and Lee (2009, GL). Both find evidence of economic value to exchange rate predictability, in that the realised terminal wealth of an investor who assumes predictability is higher than that of the investor who assumes no predictability. For interest rates, Della Corte, Sarno and Thornton (2008, DST) assess the validity of the EH, to find that on the basis of statistical tests the EH is rejected, but from an economic value perspective favourable support is found.

The results reported by ASV, DST and GL illustrate that the forecasting performance of models can be significantly different depending on whether statistical or decision-based evaluation techniques are used. To re-iterate the point made in Chapter 4, under statistical measures atheoretic models like the random walk are difficult to beat.

---

[3]Earlier studies by Klein and Bawa (1976), and Kandel and Stambaugh (1996) demonstrate the importance of parameter uncertainty in asset allocation.

[4]He uses monthly US data for two assets: T-bills and the stock index to examine the potential horizon effects under buy-and-hold and dynamic optimal rebalancing strategies, in discrete time for an investor with power utility over terminal wealth.

But under economic value methods encouraging evidence in favour of predictability, as captured by theory informed models, is found. The studies described here bring to our attention several key factors including the importance of predictability and parameter uncertainty in asset allocation, generating density forecasts to capture the risk as well as the return of the asset and the economic value to the investor of these forecasts.

The contributions of this chapter are empirical. To my knowledge we are the first to model both bond and stock returns, separately and jointly, and evaluate their predictability in an asset allocation setting using economic value. Chordia et al (2005, pp. 87) argue that "A negative information shock in stocks often causes a "flight to quality" as investors substitute safe assets for risky assets". Further, "when stocks are expected to show weakness, investment funds often flow to the perceived haven of the bond market, with that shift usually going into reverse when, .., equities start to strengthen." Party (2001, cited in Chorida et al (2005))[5]. Both of these statements highlight the dynamic relationship that exists between bond and stock markets. This supports the need to model them together and try to capture these interactions, i.e. allow for the possibility that the variables of one market have explanatory power for the variables of the other.

In brief, we compute the optimal portfolio allocation for a buy-and-hold investor with power utility over terminal wealth using weekly UK data during 1997 week 10 to 2007 week 19 for two assets, the 1-month T-bill and the FTSE All-Share Index. We extend the work of Barberis by allowing for the possibility of predictability in bond returns too and further model the bond and stock returns jointly. Here under predictability the investor assumes past values of the asset returns together with key stock and term structure variables, like the dividend yield and interest rate spreads have explanatory power. We consider a set of four models that assume varying degrees of bond and stock return predictability, all under a VAR framework. We examine the

---

[5]Full reference is John Party, The Wall Street Journal, 1st August 2001, pp. C1.

impact predictability and parameter uncertainty have on how the investor optimally allocates his portfolio. Both statistical and decision-based criteria are used to evaluate the out-of-sample forecasting performance of the models, to ascertain if indeed there is economic value to bond and stock return predictability.

Our results do suggest that in the context of investment decision making under an economic value criterion, the investor allocates differently when he assumes predictability to an investor who assumes that returns are not predictable. Moreover, he gains from not only assuming predictability in both returns, but by modelling the bond and stock returns jointly.

The setup of this chapter is as follows Section 5.2 details how we model the interest rates and stocks, the investment decision and the framework used to evaluate the economic value of predictability when parameter uncertainty is both ignored and accounted for. Section 5.3 describes the dataset, the estimated models and provides a statistical evaluation of the forecasting performance of each model. In Section 5.4 we judge the models' forecasting performance by comparing the realised end-of-period wealth generated under each and Section 5.5 concludes.

5.2 Optimal Allocation, Parameter Uncertainty and Predictability

We examine how a utility-maximising investor allocates his portfolio between 1-month T-bills and the FTSE All-Share Index. That is, between the stock market and risk-free bonds. We consider if there are gains in utility for an investor, who employs a theory informed model to forecast interest rates and stock returns, in comparison to one who believes that the returns are not predictable. Here we describe the models estimated when we first ignore T-bill and stock return predictability and then when we consider predictability. Further, we introduce how we measure the economic value of interest

rates and stock returns under predictability and parameter uncertainty.

When considering the predictability in interest rates, we look to the Expectations Hypothesis (EH) of the term structure of interest rates. The EH suggests that a n-period long rate is given by a weighted average of current and future expected short m-period rates over n periods, with the addition of a time invariant term premium. Numerous tests of the EH have been carried out using various datasets and testing methods[6], with the support found for the EH being somewhat mixed.

In Chapter 3 the UK term structure was modelled by a set of statistical and theory informed models, and tests of the EH conducted. In short, support for the EH was found in the form of stationary spread; yields sharing a common stochastic trend, such that over-identifying restrictions on the cointegrating vectors as implied by the EH could not be rejected. And further support is found using Campbell and Shiller's (1991) VAR approach. Also, the in-sample properties of the theory informed VECM and MVART models suggest a greater explanatory power for the term structure, in comparison to the statistical based models.

Given the evidence of cointegration amongst the yields, we use the MVART model that embeds the cointegration implied by the EH to explain the term structure and in turn forecast the yields. As such, we proceed assuming that if the investor believes bill returns are predictable he uses the MVART model to forecast future returns. We follow previous studies including Kandel and Stambaugh (1996) and Barberis (2000), who use the dividend yield to examine stock return predictability.

5.2.1 Modelling Interest Rates and Stocks

Let r_t^s be the return on the FTSE All-Share Index in week t, $r_t^{(1)}$ be the return on a 1-month T-bill, both returns are continuously compounded monthly returns. dy_t is

---

[6]Key studies include Campbell and Shiller (1991), Taylor (1992), Cuthbertson et al (1996, 2003), Longstaff (2000) and Sarno et al (2007), see Chapter 2.

the dividend yield, the change in the 1-month T-bill rate $\Delta r_t^{(1)} = r_t^{(1)} - r_{t-1}^{(1)}$ and the spread between a n- and 1-month rate $s_t^{(n,1)} = r_t^{(n)} - r_t^{(1)}$ for $n = 3, 6, 12$. We refer to r_t^s and dy_t as the stock variables, and $\Delta r_t^{(1)}$ and $s_t^{(n,1)}$ as the bond (or term structure, TS) variables. In order to determine how the investor should optimally allocate his portfolio he requires forecasts of $r_t^{(1)}$ and r_t^s. We consider four alternative models from which the investor could derive these forecasts, generally each model can be summarised by the following VAR(p)

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=1}^{p} \mathbf{B}_i \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t \tag{5.1}$$

where $\mathbf{x}_t$ is a $(q \times 1)$ vector of variables, $\mathbf{B}_i$ is a $(q \times q)$ matrix of parameters, $\boldsymbol{\beta}$ is a $(q \times 1)$ vector of intercepts and $\boldsymbol{\epsilon}_t$ is assumed to be a $(q \times 1)$ vector containing $i.i.d$ serially uncorrelated errors with zero means and a positive definite covariance matrix $\boldsymbol{\Sigma}$. The exact composition of $\mathbf{x}_t$ will depend upon the assumption made regarding predictability, as detailed below.

The VAR framework enables one to examine how predictability affects portfolio allocation by changing the variables in the VAR. We propose four models for predicting the returns on the T-bill and stock index, each incorporating varying degrees of predictability: Barberis Non Predictability (BNP), Barberis Predictability (BP), Individual VARs (IV) and the Joint VAR (JV) model.

The Barberis Non Predictability and Predictability models are named so, since they are in the spirit of those estimated by Barberis (2000). These models assume that the risk-free T-bill rate $r^{(1)}$ is constant[7] and allow only for the possibility of predictability in stock returns. Under the assumption of no predictability as in the BNP model, there are no predictor variables in the VAR, the stock index returns are assumed to be $i.i.d.$ such that $r_t^s = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t$, i.e. a drift term plus a random error term. Hence $\mathbf{x}_t = r_t^s$ and $\mathbf{B}_i = 0$.

---

[7]The T-bill rate is assumed constant at the last value of the estimation sample, such that in the first recursion it is fixed at its 2004 week 18 value.

However, under the assumption of predictability as in the BP model, the dividend yield is included in the VAR, with $\mathbf{x}_t = (r_t^s, z_t')'$, $z_t = (z_{1,t}, ..., z_{n,t})'$ and $\mathbf{x}_t = \boldsymbol{\mu} + \mathbf{Bz}_{t-1} + \boldsymbol{\epsilon}_t$. Such that z_t is a vector containing explanatory variables for the stock index return, i.e. the dividend yield. Hence the first equation of the VAR specifies the expected stock index return as a function of the dividend yield, and the second equation specifies the stochastic evolution of the dividend yield.

Further, it is possible to relax this assumption of a constant T-bill rate and allow for predictability in both T-bill and stock returns, we do this in two ways. First, using the IV model, where the predictability of T-bill and stock returns are described separately by two VARs (IV-BOND and IV-STOCK). The form of $\mathbf{x}_t$ for the bond returns and the stock returns are given by $\mathbf{x}_t^{(1)} = \left(\Delta r_t^{(1)}, s_t^{(12,1)}, s_t^{(6,1)}, s_t^{(3,1)}\right)'$ and $\mathbf{x}_t^s = (r_t^s, dy_t)'$ respectively. Second, using the JV model, where the predictability of the bill and stock returns are modelled jointly within a single system, here $\mathbf{x}_t = \left(r_t^s, dy_t, \Delta r_t^{(1)}, s_t^{(12,1)}, s_t^{(6,1)}, s_t^{(3,1)}\right)'$.

By modelling the predictability of T-bill and stock returns in these two ways allows us to test whether it is beneficial to the investor, in terms of wealth gains, to model the two returns jointly. In that, by allowing for interactions and feedbacks to exist between the bond and stock market, will the investor who uses the JV model to generate forecasts of the T-bill rate and the return on stocks achieve a higher wealth? Each of these four models are estimated when the parameter uncertainty, which is the uncertainty about the true values of the model's parameters is both ignored and accounted for[8].

In time T the buy-and-hold investor faces the problem of how to optimally allocate his wealth over a H month investment horizon between 1-month T-bills and the FTSE All-Share Index, where these two assets yield the continuously compounded returns $r_T^{(1)}$ and r_T^s respectively.

With an initial wealth of $W_T = 1$ and ω being defined as the proportion of initial

---

[8]We differentiate between when the model is estimated subject to stochastic uncertainty only, and when it is estimated subject to stochastic and parameter uncertainty by denoting them as BNP, BP, IV, JV and BNPPU, BPPU, IVPU, JVPU respectively.

wealth allocated to bonds[9], the end-of-horizon wealth is given by

$$W_{T+H} = \omega \exp\left(\sum_{i=1}^{H} r_{T+i-1}^{(1)}\right) + (1-\omega)\exp\left(\sum_{i=1}^{H} r_{T+(i-1)}^{s}\right) \tag{5.2}$$

Further, risk aversion can be incorporated into the investor's decision making, by assuming that the utility gained from the end-of-horizon wealth follows that given by a constant relative risk-aversion (CRRA) power utility function.

$$\upsilon(W) = \frac{W^{1-A}}{1-A} \tag{5.3}$$

where A is the coefficient of risk aversion. The optimisation problem faced by the investor in T is

$$\max_{\omega} E_T\left\{\upsilon(W_{T+H}\left(\omega\right)) \mid \Omega_T\right\} \tag{5.4}$$

where the investor computes the expectation above conditional upon the information set available at T. Fundamental to this optimisation problem is the distribution the investor employs to evaluate this expectation. The distribution used depends upon whether the investor assumes predictability in bond and stock returns. To ascertain the influence of predictability on allocation decisions, a comparison between the allocations of an investor who ignores predictability, to that of one who takes it into account can be made. This will now be discussed in greater detail below.

---

[9]Under the BNP and BP models the T-bill return is assumed to be constant, such that $\sum_{i=1}^{H} r_{T+(i-1)}^{(1)} = H.\bar{r}^{(1)}$.

5.2.2 The Probability Density Function of the Forecast Values

In this section we discuss the approach taken to estimate the density function in the case where parameter uncertainty is not considered and when it is. A detailed discussion of the approach is provided in Chapter 4, here we summarise the key concepts. The form of the density $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ is determined by the types of uncertainty surrounding the forecasts, and how the function is characterised and estimated. Here we follow the method proposed by Garratt, Lee, Pesaran and Shin (2003 and 2006, GLPS) and GL, which takes a classical view of the Bayesian approach[10] to calculating the density function. This involves approximations of certain probabilities of interest, thereby avoiding the need for priors.

To evaluate each investment decision over the investment horizon, the investor needs the probability density function of the forecast values of the 1-month rate and the stock return. Following GL, $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{qt})'$ is a $q \times 1$ vector of q variables (including at least $r_t^{(1)}$ and r_t^s), and $\mathbf{X}_T = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)'$ is a $q \times T$ vector containing the observations 1 to T of the q variables. Since forecasts of the variables are required, the conditional probability density function $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ is of interest, this predictive density function gives the probability density function of $\mathbf{X}_{T+1,H} = (\mathbf{x}_{T+1}, \mathbf{x}_{T+2}, ..., \mathbf{x}_{T+H})'$ conditional on $\mathbf{X}_T$.

When the investor ignores parameter uncertainty, he calculates the expectation of the distribution of returns conditional on the fixed parameter values $\widehat{\boldsymbol{\theta}}$. So the investor's problem to solve is

$$\max_{\omega} \left\{ E_T \upsilon\left(W_{T+H}\left(\omega\right)\right) = \int \upsilon\left(W_{T+H}\left(\omega\right)\right).P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right) d\mathbf{X}_{T+1,H} \right\} \quad (5.5)$$

---

[10]Kandel and Stambaugh (1996), Barberis and ASV use a fully Bayesian approach to estimate the density function, through the construction of a posterior distribution and using priors for the parameters.

However, if the investor incorporates parameter uncertainty then the predictive density for the returns is conditional on the observed data only, given by

$$P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right) = \int P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right) P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right) d\boldsymbol{\theta} \qquad (5.6)$$

The posterior probability of $\boldsymbol{\theta}$, denoted $P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right)$ gives the uncertainty about the parameters given the observed data. Now the investor acknowledges that $\boldsymbol{\theta}$ has a distribution conditional on $\mathbf{X}_T$. So the investor's problem to solve under parameter uncertainty is

$$\max_{\omega} \left\{ E_T \upsilon\left(W_{T+H}\left(\omega\right)\right) = \int \upsilon\left(W_{T+H}\left(\omega\right)\right).P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right) d\mathbf{X}_{T+1,H} \right\} \qquad (5.7)$$

The posterior density $P\left(\boldsymbol{\theta} \mid \mathbf{X}_T\right)$ in equation (5.6) is proportionate to the prior on $\boldsymbol{\theta}$ and the likelihood function i.e. $P\left(\boldsymbol{\theta}\right).P\left(\mathbf{X}_T \mid \boldsymbol{\theta}\right)$.

GLPS and GL suggest that in the case where meaningful priors exist are difficult to obtain, approximations of key probabilities needed to estimate the predictive density $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ can be used. They assume for the posterior probability of $\boldsymbol{\theta}$

$$\boldsymbol{\theta} \mid \mathbf{X}_T \overset{\omega}{\sim} N\left(\widehat{\boldsymbol{\theta}}_T, T^{-1}\widehat{\mathbf{V}}_{\boldsymbol{\theta}}\right) \qquad (5.8)$$

where $\widehat{\boldsymbol{\theta}}_T$ is the maximum likelihood estimate of the true parameter value of $\boldsymbol{\theta}$ and $T^{-1}\widehat{\mathbf{V}}_{\boldsymbol{\theta}}$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}_T$ i.e. of the estimated parameters.

In this exercise we consider stochastic and parameter uncertainty, the uncertainty associated with the model and the estimated model parameters respectively. We appreciate that interest rates and stock returns can be modelled under various assumptions,

and thus model the two returns in four different ways: BNP, BP, IV and JV models as described above, which can all be summarised by equation (5.1).

For each of these models, through stochastic simulation techniques, an estimate of the probability density function of the forecasts can be computed. Given that these simulations provide an estimate of the predictive densities $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$ when parameter uncertainty is ignored and $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ when it is considered, it is now possible to evaluate $E_T\left(v\left(W_{T+H}\right) \mid \Omega_T\right)$ for a range of portfolio weights ω. That is, $v\left(W_{T+H}\left(\omega\right)\right)$ is computed $\widetilde{R}$ times for each value of ω. Then the mean across these $\widetilde{R}$ replications is calculated, from which the investor chooses the weight ω that maximises the expected utility $E_T v\left(W_{T+H}\left(\omega\right)\right)$. Here ω takes values 0, 0.01,...,0.99,1, where $\omega = 0$ suggests all should be allocated to bills, equally $\omega = 1$ suggests that all should be allocated to stocks. The weight is between 0 and 1, so we do not allow for short selling. Appendix 7 provides details of the estimation procedure, how the computations are carried out and the method by which the errors are calculated[11].

5.3 Modelling the UK T-Bill Rates and the FTSE All-Share Index

5.3.1 Data

In this study we use weekly observations on the continuously compounded monthly returns for both the 1-month T-bill[12] $r_t^{(1)}$ and the FTSE All-Share Index[13] r_t^s, and the dividend yield dy_t for the UK. These variables together with $\Delta r_t^{(1)}, s_t^{(3,1)}, s_t^{(6,1)}$ and

---

[11]Here we use parametric methods to draw the errors, where the errors are assumed to be $i.i.d. N\left(0, \boldsymbol{\Sigma}\right)$ serially uncorrelated white noise errors.

[12]As with Chapter 4, estimated yield curve data is used as opposed to actual T-bill data here, because data was unavailable during some periods of our sample. However, as previously mentioned, we are satisfied that the data used here is a fair reflection of what the investor would get, should he want to undertake an investment in T-bills.

[13]We use the FTSE All-Share Index since it gives a broad portfolio of stocks.

$s_t^{(12,1)}$ are used in the analysis, refer to the Data Appendix for the definitions, sources and transformations conducted. The entire sample period is from 1997 week 10 to 2007 week 19 (532 observations). Figures 5-1 and 5-2 plot the monthly stock return, the dividend yield, the monthly bill return in levels and first differences, and the three spreads over the entire sample. The monthly stock return takes an average value of 0.59% compared with 0.41% for the T-bill, with a minimum and maximum of -17.72 to 15.32% and 0.26 to 0.60% respectively over the whole sample. This corresponds to what we would expect, average returns from the stock market tend to be higher, but there is a risk of making a loss. The return from the 1-month T-bill has a general downward trend up until the end of 2004, before increasing until the end of the sample. The annual dividend yield takes an average value of 2.86%, although there are some persistent deviations, the dividend yield exhibits mean reversion. The yield difference and spreads display mean reverting behaviour which is consistent with a stationary process.

The four models are each estimated over the period 1997 week 10 to 2004 week 18 (374 observations) and then recursively at weekly intervals through to 1997 week 10 to 2005 week 18 (427 observations), giving 54 recursions in total. For each recursion we generate h-step ahead out-of-sample forecasts[14] for $h = 1, 2, ..., H, ...$ and the investment horizon $H = 3, 6, 12, 18$ and 24 months. So for the first recursion we forecast over the period 2004 week 19 to 2006 week 18 and for the last recursion 2005 week 19 to 2007 week 19. For each recursion the investor will use his generated forecasts to determine the optimal allocation of his portfolio. Hence in this exercise we will have 54 allocation decisions for each A and H, with which to compare the allocations and utility gains under each model without and with parameter uncertainty.

---

[14]We denote the investment horizon H in months since $r_t^{(1)}$ and r_t^s are monthly returns. However, the data has a weekly frequency, so when we refer to the 'h-step' ahead forecasts each 'step' is a week.

5.3.2 Estimation

Here we describe how we estimate the four models and present the estimated regression results for the first recursion[15] over 1997 week 10 to 2004 week 18. We begin by employing the ADF, PP and KPSS unit root tests to determine the order of integration of r_t^s, dy_t, $r_t^{(1)}$, $s_t^{(3,1)}$, $s_t^{(6,1)}$ and $s_t^{(12,1)}$ over the entire sample period, see Table 5.1. All three tests indicate that r_t^s and the spreads are found to be stationary in levels and $r_t^{(1)}$ is difference stationary. As for dy_t the unit root tests suggest it is non-stationary, but given the test statistics are close to their respective critical values and the series exhibits mean reversion we treat, like in previous studies, the dividend yield as stationary.

The optimal lag length for the IV and JV models is chosen by estimating a set of VAR(p) with $p = 0, 1, ..., 12$ for each model over 1997 week 10 to 2004 week 18. The optimal lag length is that which minimises the Schwarz Information lag selection criteria, as well as satisfying the diagnostic checks, in particular the model's residuals should be free of serial correlation at the 5% level. Based on this, the lag length chosen was five for the IV-STOCK model, six for the IV-BOND and JV models. Tables 5.2 to 5.8 summarise the estimates with the diagnostics of the BNP, BP, IV and JV models.

Comparing the estimated BP model to the BNP model, Table 5.2 to 5.3, there is a small gain in explanatory power by allowing for predictability in stock returns through the inclusion of a single lagged dividend yield term. Further, all coefficients in the estimated BP model are significantly different from zero. Moving from the BP to the IV-STOCK model, Table 5.3 to 5.4, allows for past values of both r^s and dy to influence current values. A substantial gain in explanatory power for stock returns is observed. All the coefficients are jointly significant, which suggests there are gains from relaxing the assumptions of no and limited predictability made under the BNP and BP models. For each equation in the IV-BOND model, Tables 5.5 and 5.6, the TS[16]

---

[15]Estimates of each model for the first recursion only are provided, to give an overall impression of the in-sample predictability. At the forecasting stage the models are estimated recursively.

[16]TS is used to denote the term structure.

variables are jointly significant. The JV model, Table 5.7 and 5.8, is a generalisation of the individual VARs, allowing for feedbacks between the two markets. In terms of explanatory power as indicated by $\overline{R}^2$, the gains from modelling the two returns together are small. However, the stock variables are jointly significant in all the TS equations, but the TS variables are jointly significant in the TS equations only. This implies that causality exists from the stock market variables to the TS variables, which provides support in favour of modelling the two markets together.

The diagnostics are satisfactory, there is indication of some serial correlation in the stock equations of the BNP and BP models, but we want to replicate those estimated in Barberis. In the IV and JV models we do not have serial correlation at the 5% level and the explanatory power of the models is quite high. Rejection of the nulls that the regression residuals are homoskedastic and normal is not surprising given that we are using financial data. But we follow the assumptions made by the literature that also utilise such data.

5.3.3 Statistical Evaluation of the Forecasting Performance

The root mean squared error (RMSE) provides a statistical evaluation of the out-of-sample forecasting performance of each model. Table 5.9 gives the RMSEs of the bond and stock return forecasts, for the forecast horizons $H = 1, 3, 6, 12, 18$ and 24 months for each model, without and with parameter uncertainty being considered. Table 5.10 reports the ratio of the RMSEs for each model to the benchmark model. A value of the ratio greater than one indicates that the RMSE of the model is lower than that of the benchmark. The benchmark taken is the BNP model which assumes $r_t^{(1)}$ is constant and $r_t^s = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t$, since it assumes no predictability a comparison can be made with the other models which assume varying degrees of predictability.

The RMSEs for forecasts of the bond returns indicate that only at $H = 1$ do the JV and JVPU models beat the benchmark. The BNP, BP, BNPPU and BPPU models

that make the strong assumption that $r_{T+H}^{(1)}$ is constant, outperform the other more theory informed models at each horizon under this criteria. However, it can be seen that the differences in the RMSEs amongst the models are small. These results broadly correspond to those found in the exchange rate forecasting literature, as summarised in ASV and GL. Which in general find sophisticated theory informed models are outperformed by a simple random walk.

With the stock returns, the RMSEs show that there is not a single model that performs consistently well over all horizons. The JV and JVPU models perform the best at $H = 1, 3$ and 12, whereas the BP and BPPU models perform well at $H = 6, 18$ and 24. These results suggest some gain in terms of forecasting performance from incorporating predictability when modelling stock returns.

When comparing the size of the RMSEs of the two returns, there is greater variance in the r_t^s forecasts than the $r_t^{(1)}$ forecasts. This is not surprising since stock returns are more volatile and thus more difficult to predict. In general, the RMSEs increase up until $H = 6$ and 12 before decreasing. This suggests that the RMSEs for both the returns are non-monotonic, i.e. they oscillate in relative value and do not just increase with H. Although the RMSEs for both the returns are non-monotonic, the rates at which the two are changing across the horizons are different. Over the shorter horizon, the rate at which the RMSEs for $r_t^{(1)}$ increase is smaller than the rate at which the RMSE for r_t^s increases. But over the longer horizon the rate at which the RMSE for $r_t^{(1)}$ decreases is greater.

This statistical evaluation provides an indication of the forecasting performance of each model. But does not provide a clear indication of how these models perform in an investment decision making context, i.e. in terms of the economic value of the gains from the models' forecasts.

5.4 Investors' Evaluation of Forecasts

We now examine the implications for optimal allocations when the returns are either *i.i.d.* or predictable, where the degree of predictability is varied and parameter uncertainty is both ignored and accounted for. In the case where parameters are assumed fixed the maximisation problem is given by equation (5.5) and under parameter uncertainty it is given by (5.7). Figures 5-3 to 5-7 give the optimal allocations to bonds, $100\omega\%$, at each investment horizon $H = 3, 6, 12, 18, 24$ months, for each model and for the levels of risk aversion $A = 2, 5$ and 10, $A = 10$ is the highest level of risk aversion. The models are estimated first over 1997 week 10 to 2004 week 18, the optimal weights are calculated from the forecasts generated from each estimated model. Then moving forward one week this is repeated, re-calculating expected wealth and utility to find the optimal weight for this new augmented sample. This is repeated for each recursion, giving results for 54 recursions over the total evaluation period 2004 week 19 to 2007 week 19. The plots are based on the optimal allocation averaged over the 54 recursions for a particular A, H and model.

Figure 5-3 gives the optimal allocation under each model, when parameter uncertainty is ignored, here allocations are conditional on the fixed parameter values estimated. A risk aversion effect is evident for all the models, where the investor allocates more to bonds at all horizons the more risk averse he is. Further, under the BP, IV and JV models the difference in the allocation to bonds under each A increases with H, with differences of up to 65% being observed for an investor with $A = 2$ compared with $A = 10$. This suggests that the allocation to bonds for a longer horizon investor greatly depends on how risk averse they are.

It can be seen that the investment horizon is also important in determining how the investor allocates. In the absence of horizon effects, the short horizon investor allocates no differently than a long horizon investor. With horizon effects there is a difference between the allocations of a short and long horizon investor, such that the 'allocation

curve' which we define as describing for a particular A how the investor allocates over H, has a slope. Further, this curve may have a positive or negative slope, if the slope is positive then the investor allocates more to bonds as H increases. Here strong horizon effects are present under all models. In general, we find as H increases under the BNP and BP models the investor allocates more to bonds for all A. This is true for $A = 5$ and 10 under the IV model, but for $A = 2$ the allocation to stocks increases with H. Equally, under the JV model for $A = 10$ the investor increases his allocation to bonds with H, for $A = 5$ he increases the allocation to stocks over the medium horizon before increasing the allocation to bonds in the longer horizon, whereas with $A = 2$ the investor increase his allocation to stocks with H.

In short, horizon effects are present. But the extent of the effect the investment horizon has on the allocation depends upon the predictability assumptions the investor makes. That is, which model he believes to be true and his level of risk aversion.

We will now try and provide an explanation for these allocation results by first considering the effects of predictability (ignoring parameter uncertainty) and then the effects of parameter uncertainty.

5.4.1   Effect of Predictability

In this exercise we consider four different models for forecasting interest rates and stock returns. The atheoretic BNP and BP models assume no predictability in regard to bond returns. Further, the BNP model assumes no variables are able to predict the stock return. However, the BP model relaxes this assumption allowing for some predictability in stock returns. On the opposite end of the spectrum, the theory informed IV and JV models not only assume predictability, but as in the case of the JV model allow for the possibility of feedbacks amongst the stock and term structure variables.

These models reflect opposing views of whether bond and stock returns are predictable, and further have a varying degree of predictability which increases as we move

from the BNP to BP to IV to JV model. If the investor assumes no predictability then
he believes in the BNP model. Conversely, if he assumes predictability he may believe
in the BP, IV or JV model depending on the extent of the predictability assumed. Ul-
timately, how the investor allocates is determined by which model he believes to be a
true depiction of reality.

From Figure 5-3 it can be seen that the BNP model allocates the most to bonds,
followed by the BP, the IV and then the JV model at each A and H. Where the JV
model allocates the most to stocks. The difference in allocation to bonds in some cases
is over 70% amongst the models, e.g. $H = 24$ and $A = 2$ the BNP model allocates 77%
more to bonds than the JV model.

Under no predictability, which is similar to assuming the stock returns follow a
random walk process, the variance of the cumulative log returns distribution $\sigma^2 \rightarrow \infty$,
i.e. the variance continues to grow with the horizon. Whereas, when the return is
modelled as a stationary process, as is the case under predictability, then $\sigma^2 \rightarrow long$
$run\ mean$ i.e. mean reversion of the variance of returns. In which case, stocks appear
less risky in the long run and are more attractive to long horizon investors, Fama and
French (1988).

Under the BNP model we find horizon effects, where the investor allocates more to
bonds as H increases. Under the assumption that log returns are independently and
identically normally distributed (assumption of normality is not necessary for this to
hold) the mean and variance of the cumulative log returns distribution grows propor-
tionally with the investment horizon[17] i.e. $H\mu$ and $H\sigma^2$. For the risk averse investor
with power utility function, although return per unit of variance is the same as H in-
creases, the higher return is coupled with higher risk in absolute terms and since the

---

[17]$r_{t,t+H} = r_{t+1} + r_{t+2} + ... + r_{t+H} \implies E(r_{t,t+H}) = E(r_{t+1}) + E(r_{t+2}) + ... + E(r_{t+H}) = H\mu$,
where each return has the same mean (identically distributed) and returns are independent in that
one return does not contain information about the other returns. Further, $var(r_{t,t+H}) = var(r_{t+1}) +
var(r_{t+2}) + ... + var(r_{t+H}) = H\sigma^2$, where the returns are uncorrelated so there is no covariance term
and all the variances are equal (identically distributed).

investor is risk averse he allocates less to stocks as H increases.

With predictability the investor recognises that rather than the returns being *i.i.d.* they may be predictable, as is the case under the BP, IV and JV models. Now returns are no longer independent, but the distribution of future returns is conditional on the current and past values of the explanatory variables. In which case the mean and the variance of the returns no longer grow linearly. Barberis highlights that under predictability the variance of cumulative log stock returns may grow slower than linearly with H, such that stocks appear comparatively less risky at longer horizons, resulting in higher allocations to stocks as H increase.

With the BP model however, we find that it is the allocation to bonds that increases with H. A possible explanation for this is that although we are now incorporating predictability the gain in terms of explanatory power for stocks returns are small, $\overline{R}^2$ increases from 0% under the BNP model to just over 2% under the BP model, so the increase in predictability is not sufficient for the investor to increase his allocation to stocks with the horizon.

The bond returns are also modelled[18] under the IV and JV models. So now both returns will be subject to future uncertainty and ultimately the optimal allocation hinges on how risky bonds look relative to stocks. With the IV model the investor allocates more to stocks at all horizons than the BNP and BP models, i.e. allocation curve shifts down for all A. This can be attributed to two factors, firstly bond returns now look relatively more risky than they did under the BNP and BP models since the return is no longer known with certainty. Secondly, stock predictability under the IV model has increased dramatically, from 2% under the BP model to nearly 70%. Both of these factors make stocks look more attractive.

Predictability increases further under the JV model, we expect an increase in the

---

[18]Note when bond returns are modelled too, the variance of cumulative log bond returns may also grow less than linearly with H. So now bond and stock returns may both be subject to these predictability effects.

allocation to the asset that has gained most from the increase in predictability. An increase in the allocation to stocks at each H in comparison to the IV model is observed. Thus stock returns appear to have gained more from modelling the returns jointly, so that they appear less risky and the investor is more willing to hold them. For $A = 2$ stock return predictability dominates as the investor increases the amount allocated to stocks as H increases. For $A = 5$ stock return predictability dominates until $H = 12$, then bond return predictability dominates such that the investor allocates more to bonds. For $A = 10$ bond return predictability dominates as the investor increases allocation to bonds with H.

Under the varying degrees of predictability that each model assumes, how the increased predictability alters the optimal allocation depends, firstly on which return (bond or stock) benefits more from the predictability effect[19]. Secondly, how risk averse the investor is. As we move from the BNP to JV model the investor allocates more to stocks at each H, so the allocation curves shifts down. This could be because the investor is able to predict stocks better as we move from the BNP to the JV model, so he is prepared to allocate more to stocks at every horizon for each A. But most evidently for $A = 2$, when moving from BNP through to JV the slope of the allocation curve changes. For the IV and JV models the investor is prepared to allocate substantially more to stocks at longer horizons, which could be attributed to σ^2 growing less than linearly combined with the investor not being very risk averse. Whereas for $A = 10$ the investor is very risk averse and increases his allocation to bonds with H.

5.4.2 Effect of Parameter Uncertainty

Figures 5-4 to 5-7 compare the allocations under each model when parameter uncertainty is ignored to that when it is considered. Incorporating parameter uncertainty

---

[19]The predictability effect results in the variance of cumulative log returns to grow less than linearly, making the asset appear less risky at longer horizons.

has the effect of increasing the variance of the distribution of cumulative returns. Further, the variance increases faster than linearly with H in the case of $i.i.d.$ returns, when this additional uncertainty is accounted for. This increase in the variance serves to make the asset seem riskier at longer horizons.

When the investor believes in the BNP model we indeed find that the allocation to stocks is reduced by 0 to 2% with parameter uncertainty, the effects are small over the horizons considered. For the BP model this additional uncertainty increases the allocation to bonds by up to 7%, with the effect of parameter uncertainty decreasing as the investor becomes more risk averse.

Under the IV and JV models the bond returns are also being modelled, such that they too are subject to parameter uncertainty. Now bonds look riskier than they did under the BNP and BP models, so the optimal allocation hinges on which asset is affected by parameter uncertainty more and hence the riskiness of bonds relative to stocks.

Parameter uncertainty under the IV model has the effect of increasing the allocation to stocks by 3 to 10% in the short to medium horizon for $A = 2$ and 5, the increase is smaller for $A = 10$, before the allocation to bonds increases in the longer horizon to levels similar to those when parameter uncertainty is ignored. Here we find that the impact of this uncertainty is different for each A, where the more risk averse the investor is, the less willing he is to hold more stocks. Allocations emerge as being non-monotonic over H, because the investor does not simply increase his allocation to stocks with the horizon, but the slope of the curve actually changes over H. Over the short to medium horizon it appears that the effect of parameter uncertainty is greater on bond returns than stock returns. That is, the variance of the cumulative stock returns is less than that of bonds, $\sigma_{rs}^2 < \sigma_{r1}^2$, making stocks look less risky and more being allocated to them. But over the longer horizon the converse seems true, such that stocks look riskier and the optimal allocation is equal to that when parameter

231

uncertainty is ignored.

The effect of parameter uncertainty is most apparent under the JV model, with allocations to stocks increasing by up to 4% for $A = 2$, and by the same margin for $A = 5$ over the short to medium horizon before the allocation to bonds increases over the longer horizon by 9 to 13%. The changes in allocation to bonds for $A = 10$ over the investment horizon are similar to those observed for $A = 5$, but of a smaller magnitude. Again allocations are non-monotonic for $A = 5$ and 10, in that after $H = 12$ the parameter uncertainty risk is less for bonds, thus making them appear more attractive.

To explain the non-monotonic allocations that arise under parameter uncertainty, we consider how the variances about the distribution of future predicted returns evolve over the forecast horizon. In this case it is reasonable to expect the RMSEs and the variances to be closely related, as in Chapter 4 we use the RMSEs as an indication of how the variances of the forecasts evolve. Recall Tables 5.9 and 5.10, the non-monotonic RMSEs imply that the variances of the forecasts are also non-monotonic[20]. This suggests that the variance about the forecasts contracts and expands with H, so under parameter uncertainty the asset will appear more risky at some horizons than at others. Further, the variances of the two returns oscillate at different rates, such that the effects of parameter uncertainty will be different at different H, so at some horizons stocks will appear more risky than bonds and at others less. This non-monotonicity combined with the fact that the variances of the two returns expand and contract at different rates could provide an explanation for the impact of parameter uncertainty observed here.

We can see that as the investor becomes more risk averse, he is less prepared to allocate more to stocks when parameter uncertainty is incorporated. Further, he is prepared to allocate more to stocks under parameter uncertainty over the short to

---

[20]Which as Hall and Hendry (1988, pp. 256-7) argue may not be so surprising, see discussion in Chapter 4.

medium horizon, but not at the longer horizons[21].

In short, predictability has the effect of making the assets appear less risky at longer H, while parameter uncertainty makes the asset look more risky. The final allocation depends on which effect dominates for that asset. Additionally, since we consider two assets-bonds and stocks, which of the two emerges as the less riskier.

The RMSE is a statistical measure of forecast accuracy, here we focus on assessing forecast performance using the economic value to an investor. An economic evaluation of the forecast performance of each model is reported in Tables 5.11 to 5.13 [22]. We compute the end-of-period wealth that the risk averse investor would have achieved over 2004 week 19 to 2007 week 19 had he allocated his portfolio as suggested by the optimal weights of each model for a particular A and H. The optimal weight ω is calculated by solving the utility maximisation problem[23]. These realised wealths are averaged over 54 recursions and then ranked in descending order so the performance of each model can be compared.

Apart from the four models described above, under which we both ignore parameter uncertainty and incorporate it to derive the optimal allocations, we also introduce three passive 'lazy' strategies. Under the lazy strategies the investor makes no attempt to model or predict the returns, but instead either invests (1) all in bonds (AB), (2) all in stocks (AS) or (3) half in bonds and half in stocks (HH). The top position is always occupied by the lazy 'all in stocks' strategy. Although it should be noted that during

---

[21]Boudry and Gray (2003, BG) extend Barberis by including two additonal predictor variables-term spread and the relative bill rate to predict Australian stock returns. Like us they too find "negative horizon effects", where the investor allocates more to bonds at longer horizons. This is contrary to Barberis, who finds that parameter uncertainty reduces not eliminates the positive horizon effects. BG argue that their model contains more predictor variables that require estimating than Barberis', which introduces a significant degree of parameter uncertainty. Thus the perceived riskiness of stocks grows faster than linearly with H and allocation to stocks decreases. Further, they state that this negative horizon effect may be intensified by the fact that the investment is buy-and-hold, whereby the consequence of inaccurately judging the level of predictability is more severe when the investor is locked-in for long horizons.

[22]Like ASV and GL our measure of economic value is based on wealth.

[23]The optimal weight is determined by the forecasts from the model. These weights are then combined with actual/realised returns to give the realised end-of-horizon wealth.

the forecast horizon 2004 week 19 to 2007 week 19 over which this evaluation of the models is made, the UK stock market was buoyant which explains the success of this strategy here. Hence during times of market growth investing 'all in bonds' would yield the lowest realised wealth. Looking to positions 2 to 10, the success of the JV models (without and with parameter uncertainty) is clear, with it occupying 2nd and 3rd place for almost all A and H. The IV models come mostly 4th and 5th, followed by the BP models and then the BNP models.

What emerges from these results is that the success of the model in terms of profitability appears to be closely related to the level of predictability the investor assumes. Whereby the more theory informed IV and JV models consistently outperform the more restricted BP models and the atheoretic BNP models. Broadly speaking these results are not sensitive to the investment horizon or level of risk aversion. This provides evidence not only in favour predictability, but of modelling the two returns jointly as under the JV models rather than separately, when we use economic value as a means to evaluate forecasts.

5.5   Conclusion

For a utility maximising investor, we compare how the optimal allocations differ under a set of atheoretic and theory informed models, and how it differs when the investor incorporates parameter uncertainty to when he ignores it. Further, we evaluate the economic value of the out-of-sample forecasts of bond and stock returns generated under each of these models. The investment decision is whether to invest in bonds or stocks, this is examined in a framework that both ignores parameter uncertainty and explicitly allows for it.

The key innovation here is that we model both returns by using the EH to model the interest rate and the dividend yield to model the stock returns, and then evalu-

ate interest rate and stock predictability in an economic value framework. Under the assumption of bond return predictability, we first model the bond and stock returns separately with their predictor variables. Then secondly model the two returns jointly with all the predictor variables. This joint modelling framework allows for the possibility of stock variables to influence the term structure variables and vice versa. Over the sample investigated here we find evidence to suggest that an investor seeking to optimally allocate his wealth between UK bonds and stocks is better off, in terms of higher end-of-horizon wealth, by assuming predictability in returns and further modelling both returns together, than an investor who assumes no predictability.

We find the effect of predictability on the optimal allocation is considerable, where the optimal weights under predictability of returns are in some cases greatly different to those under no predictability. In particular, the predictability in the bond and stock returns led to more being allocated to stocks at each horizon, and under the IV and JV models for $A = 2$ the investor increases the allocation to stocks with the horizon. These findings lend support to the predictive ability of the stock and term structure models considered here, and to modelling both returns jointly. The effect of parameter uncertainty is not large over the investment horizon considered here. Although Barberis reports significant effects of parameter uncertainty on the optimal allocation these are prominent at longer horizons, he considers horizons up to 10 years. At our comparatively shorter horizons of up to 2 years, the magnitude of the impact is of similar proportions to those reported by Barberis.

Using a statistical evaluation criterion i.e. RMSEs, the BNP and BP models outperform the models that assume predictability at almost all horizons when forecasting bond returns. However, when forecasting stocks returns there is not a single model that outperforms the others, the JV models perform well over the shorter horizons and the BNP and BP models over the longer horizons. In general, under this statistical criterion the Barberis models which assume no or limited predictability forecast well.

Conversely, when an economic value approach is used these Barberis models are the worst performing and are outperformed by the theory based models. So we observe that the results from the two differing evaluation techniques do not entirely coincide, where the model that achieves the lowest RMSE is not necessarily the one that will maximise realised wealths. It is apparent from this that models and their forecasts need to be evaluated using an appropriate criteria. Here we want to know how to optimally allocate the portfolio, so it is necessary to incorporate the investor's feelings about risk and to consider the distribution about the predicted returns, in which case the RMSE seems inadequate for this purpose.

The results show evidence of economic value to bond and stock return predictability. As we increase the degree of predictability assumed in the model, when moving from the BNP model right through to the JV model, there are increasing gains in terms of economic value to the investor. Since the end-of-horizon wealth gained by the investor who assumes bond and stock return predictability is greater, than one who assumes they are not predictable. With the investor who assumes the highest level of predictability here as given by the JV model, achieving the greatest end-of-horizon wealth.

To conclude we find further evidence to that reported by Abhyankar et al (2005), Della Corte et al (2008), and Garratt and Lee (2009) amongst others, which highlights the importance of having an evaluation criterion that reflects the purpose for which the forecasts are intended. Our results suggest that in the context of investment decision making under an economic value criterion, the investor gains from not only assuming predictability, but by modelling the bond and stock returns together.

Figure 5-1: Stock Return and Dividend Yield 1997 to 2007

Monthly return from FTSE All-Share Index

Dividend Yield
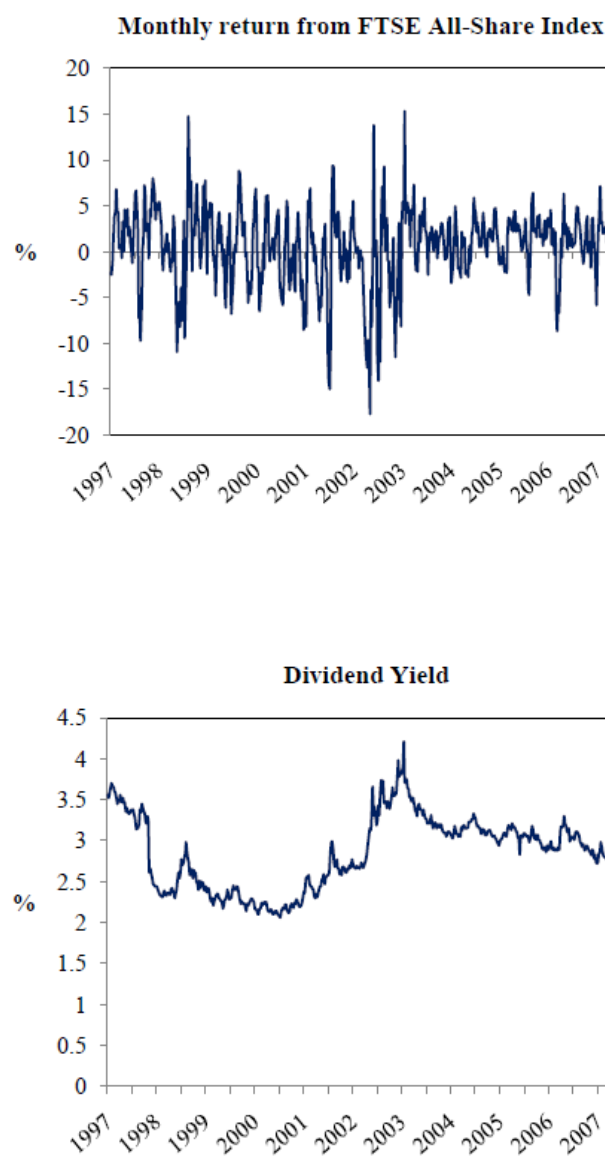
Figure 5-2: Bond Return, Changes and Spreads 1997 to 2007

Monthly return from 1-month Bill

Change in the 1-month return

Spreads

·········· s31   ——— s61   ——— s121
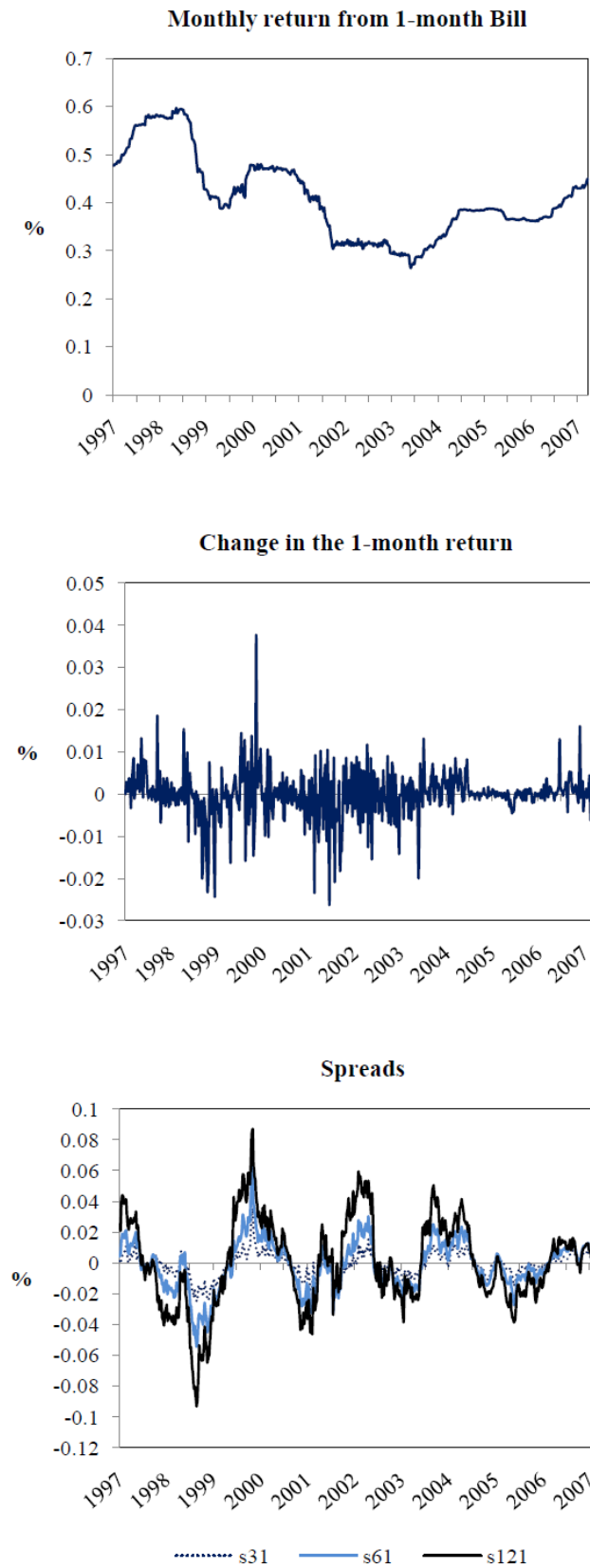
Figure 5-3: Effect of Predictability ignoring Parameter Uncertainty

Figure 5-4: Allocation under the BNP Model Without (solid line) and With (dotted line) Parameter Uncertainty

A=2

A=5

A=10

Figure 5-5: Allocation under the BP Model Without (solid line) and With (dotted line) Parameter Uncertainty

A=2

A=5

A=10
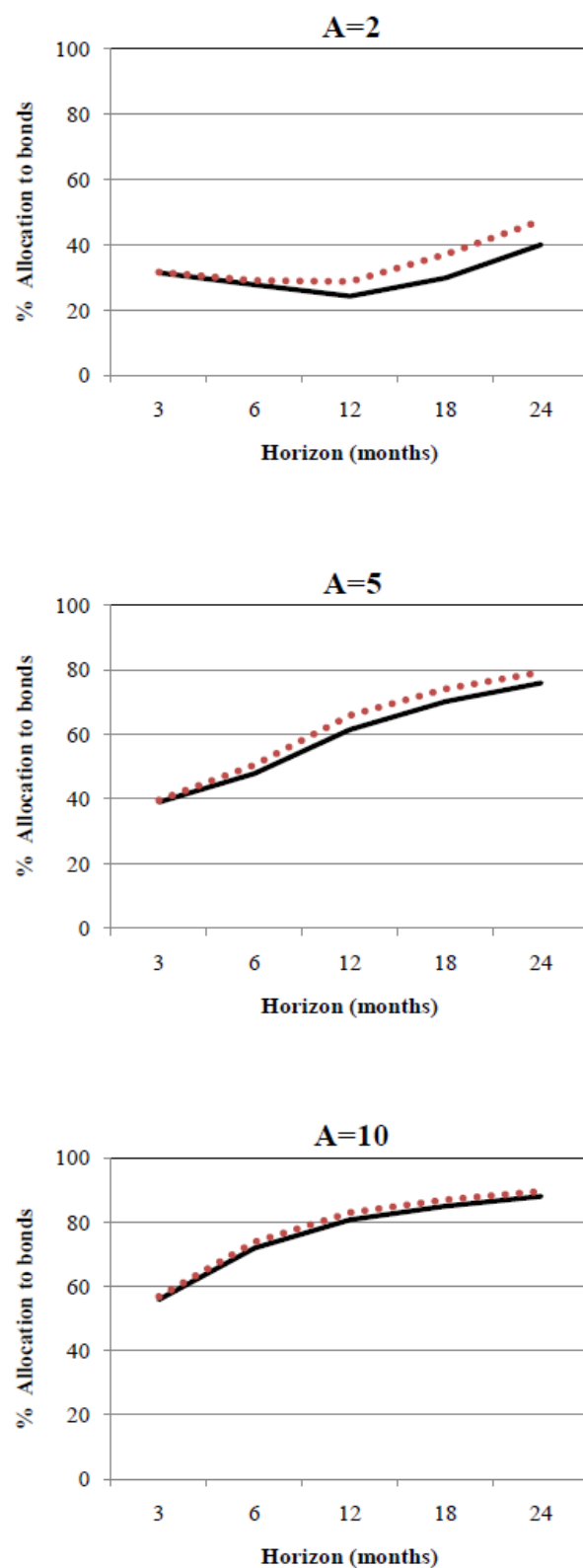
Figure 5-6: Allocation under the IV Model Without (solid line) and With (dotted line) Parameter Uncertainty
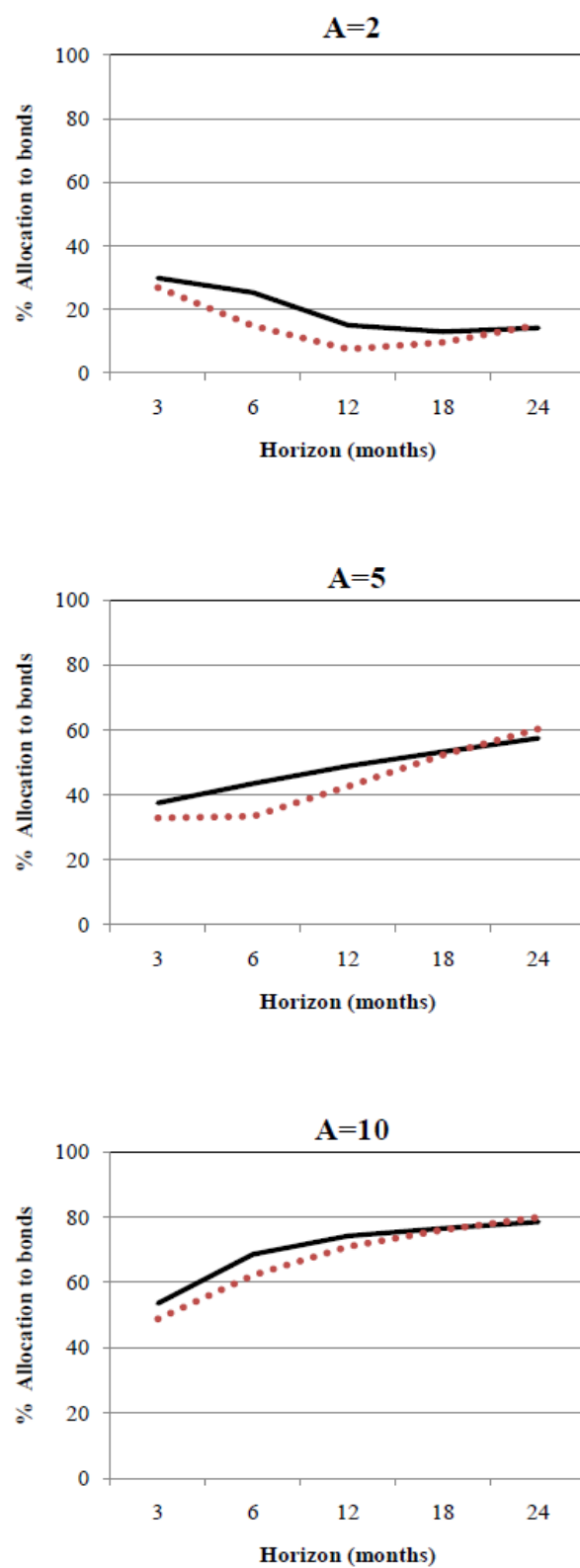
Figure 5-7: Allocation under the JV Model Without (solid line) and With (dotted line) Parameter Uncertainty

243
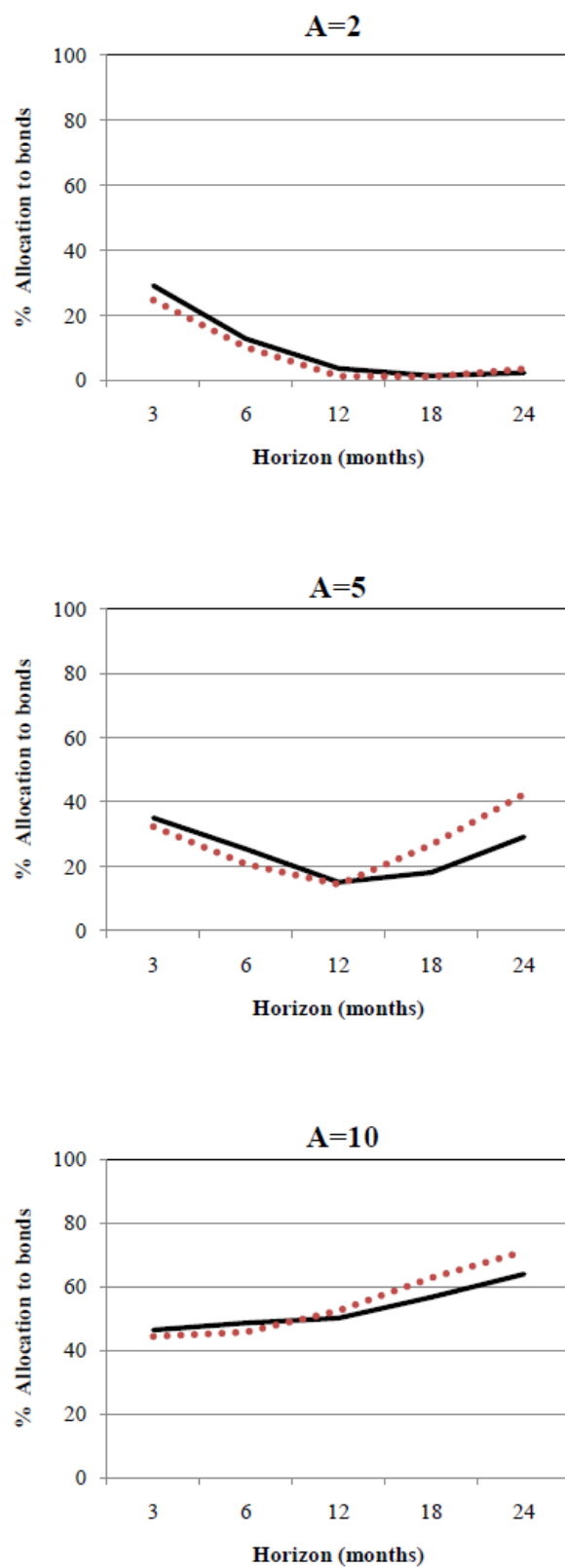
Table 5.1: Unit Root Tests

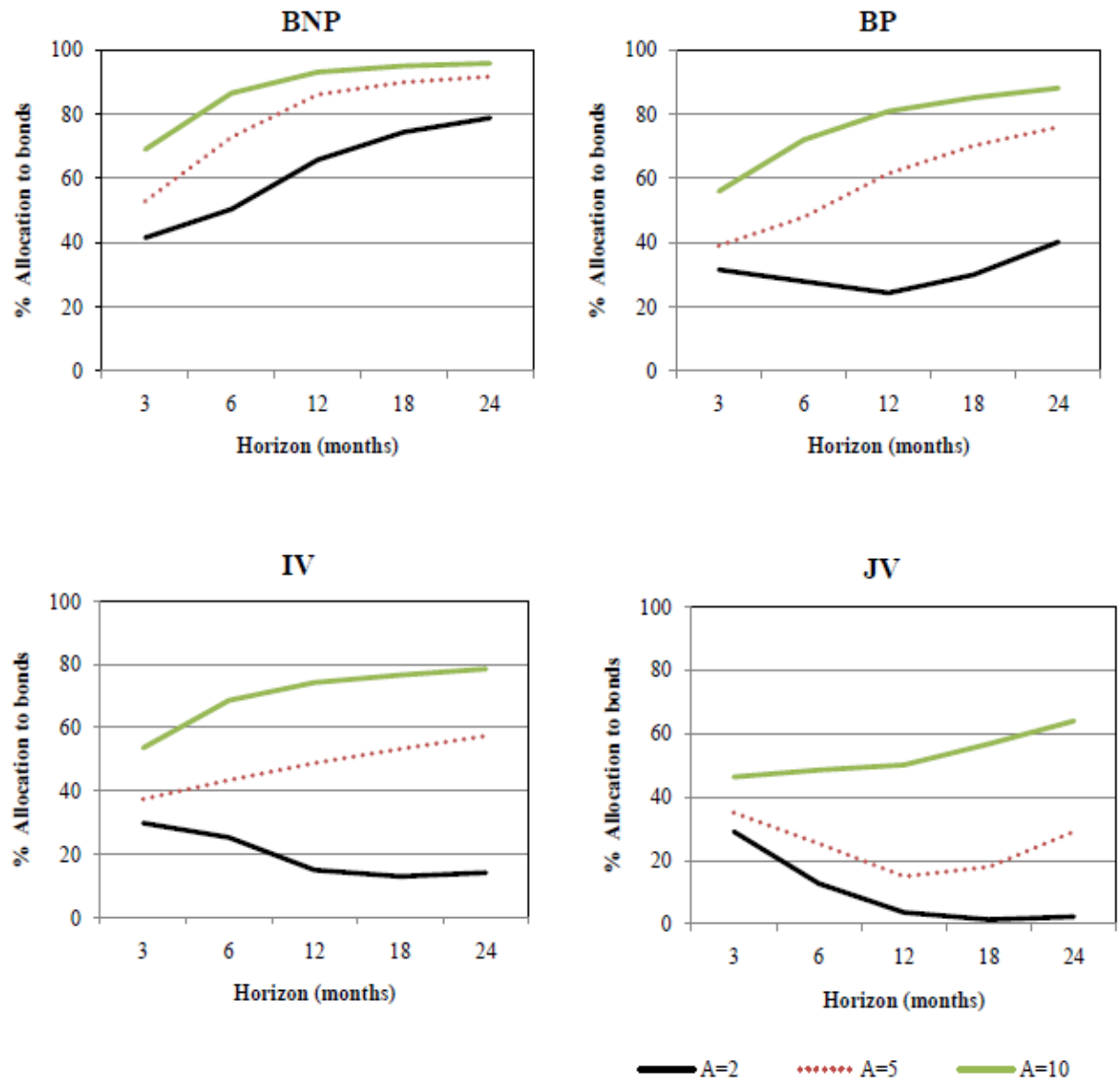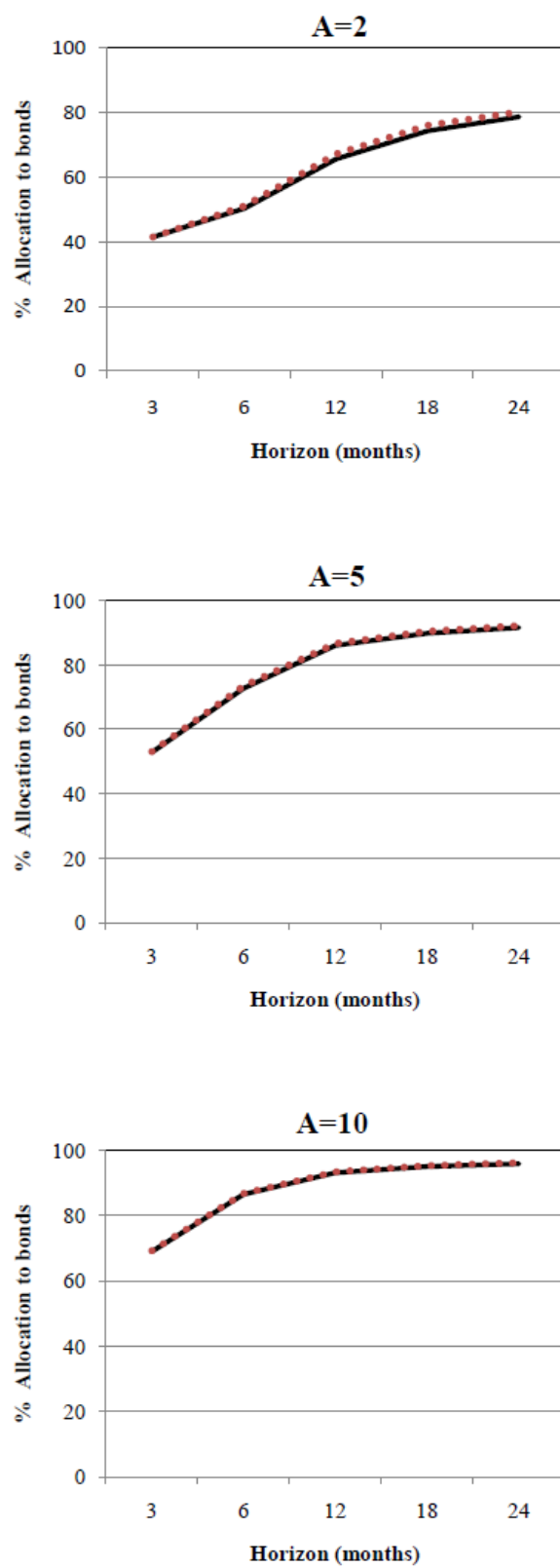| Variable	ADF t-statistic (lag length)	PP adj t-statistic (bandwidth)	KPSS LM statistic (bandwidth)
r_t^s	-5.002^{***} (0)	-9.150^{***} (22)	0.248 (9)
dy_t	-2.005 (1)	-2.107 (4)	0.799^{***} (17)
Δdy_t	-26.238^{***} (0)	-26.189^{***} (3)	0.170 (6)
r_t^1	-0.888 (0)	-1.184 (12)	1.568^{***} (18)
Δr_t^1	-21.686^{***} (0)	-23.129^{***} (12)	0.266^{***} (12)
$s_t^{12,1}$	-3.121^{**} (0)	-3.141^{**} (8)	0.114 (17)
$s_t^{6,1}$	-3.840^{***} (0)	-3.639^{***} (7)	0.128 (17)
$s_t^{3,1}$	-4.454^{***} (1)	-4.700^{***} (4)	0.141 (17)

Critical Values

	ADF Test	PP Test	KPSS Test
1% level	-3.445	-3.445	0.739
5% level	-2.868	-2.868	0.463
10% level	-2.570	-2.570	0.347

Notes: The ADF test statistics are computed using ADF regressions with an intercept and 'L' lagged first differences of the dependent variable. The order of augmentation in the Dickey-Fuller regressions are chosen using the Schwarz Information Criterion, with maximum lag length of 20. The bandwidth for both the PP and KPSS test was selected using the Newey-West (1994) method based on the Bartlett Kernel. The PP test statistics are calculated with an intercept only in the underlying DF regressions. Tests are performed on the entire sample 1997 week 10 to 2007 week 19. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 5.2: Estimation of Barberis No Predictability (BNP) Model

Equation	r_t^s
$\boldsymbol{\mu}$	0.0026 (0.0024)
$\overline{R^2}$	0.000
$\widehat{\sigma}$	0.0471
$eq^n LL$	612.42
$\chi_N^2 [2]$	40.31***
$\chi_{SC}^2 [12]$	246.35***

Notes: see those for Table (5.3).

Table 5.3: Estimation of Barberis Predictability (BP) Model

Equation	r_t^s	dy_t
$\boldsymbol{\mu}$	-0.0385*** (0.0132)	0.0004* (0.0002)
dy_{t-1}	1.4813*** (0.4666)	0.9845*** (0.0083)
$\overline{R^2}$	0.0238	0.9743
$\widehat{\sigma}$	0.0466	0.0008
$eq^n LL$	615.44	2118.61
χ_N^2	53.83***	2324.69***
$\chi_H^2 [1]$	1.22	20.11***
$\chi_{SC}^2 [12]$	240.30***	18.70*

Notes: Standard errors in parenthesis (.). The $\overline{R}^2$, standard error of the regression $(\widehat{\sigma})$, log likelihood of the equation (LL) presented, together with the chi-squared statistics for Breusch-Pagan Serial Correlation test (SC), the Jarque-Bera Test for Normality (N), Breusch-Pagan-Godfrey test for Heteroskedasticity (H). The BNP and BP models are estimated over 1997 week 10 to 2004 week 18 (364 observations), the BNP model assumes that $r_t^s = \mu + \epsilon_t$ and the BP model assumes that both r_t^s and dy_t are determined by a single lagged dividend yield term. Both assume that r_t^1 is a constant taken at the last value in the sample i.e. 2004 week 18. Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 5.4: Estimation of Individual VAR-STOCK (IV-STOCK) Model

Equation	r_t^s	dy_t
r_{t-1}^s	0.8488*** (0.0520)	−0.0021*** (0.0009)
r_{t-2}^s	0.0969* (0.0680)	0.00002 (0.0011)
r_{t-3}^s	0.0395 (0.0682)	0.0016* (0.0011)
r_{t-4}^s	−0.9627*** (0.0684)	−0.0253*** (0.0011)
r_{t-5}^s	0.6429*** (0.0800)	0.0174*** (0.0013)
dy_{t-1}	−5.2877** (2.7871)	0.7973*** (0.0464)
dy_{t-2}	−4.4200* (3.0747)	0.0130 (0.0512)
dy_{t-3}	2.4994 (3.0838)	−0.0089 (0.0514)
dy_{t-4}	25.6587*** (3.0723)	0.6554*** (0.0512)
dy_{t-5}	−18.0215*** (2.8277)	−0.4597*** (0.0471)
$inpt$	−0.0111* (0.0078)	0.00009 (0.0001)
$\overline{R^2}$	0.6835	0.9925
$\widehat{\sigma}$	0.0267	0.0004
$F\,stat$	80.48***	4862.61***
$eq^n LL$	819.78	2330.93
Excl. r^s terms	118.68***	95.82***
Excl. dy terms	4.79***	22793.15***
$\chi_N^2\,[4]$	377.63***	
$\chi_H^2\,[60]$	273.46***	
$\chi_{SC}^2\,[4]$	9.37*	

Notes: Standard errors in parenthesis (.). The $\overline{R}^2$, standard error of the regression $\left(\widehat{\sigma}\right)$, log likelihood (LL) presented with the model diagnostic tests which are all carried out on the VAR residuals. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. Chi-squared statistics presented for: (N) the VAR Residual Normality Test (orthogonalization: residual correlation (Doornik-Hansen) this test statistic is not sensitive to the ordering or the scale of the variables) for the null that the residuals are multivariate normal, (H) the VAR Residual Heteroskedasticity Test (no cross terms, but the conclusion was the same when cross terms were included), and (SC) the VAR Residual Serial Correlation LM Test. "Excl ... terms" tests the joint significance of the excluded terms, we give the F-statistic of the Wald test of these restrictions. The IV model for the stock returns is estimated of order 5, over 1997 week 10 to 2004 week 18 (364 observations). Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 5.5: Estimation of Individual VAR-BOND (IV-BOND) Model

Equation	Δr_t^1	$s_t^{12,1}$	$s_t^{6,1}$	$s_t^{3,1}$
Δr_{t-1}^1	0.0261 (0.0641)	0.0807 (0.0924)	0.1005* (0.0643)	0.0677** (0.0403)
Δr_{t-2}^1	−0.1324** (0.0642)	0.1024 (0.0926)	0.1262** (0.0644)	0.0905** (0.0404)
Δr_{t-3}^1	−0.0401 (0.0647)	0.0147 (0.0933)	0.0340 (0.0648)	0.0365 (0.0407)
Δr_{t-4}^1	0.0401 (0.0647)	0.1181 (0.0932)	0.1194** (0.0648)	0.0698** (0.0407)
Δr_{t-5}^1	−0.1556*** (0.0652)	0.2144** (0.0940)	0.2256*** (0.0654)	0.1635*** (0.0410)
Δr_{t-6}^1	0.0637 (0.0503)	−0.0425 (0.0724)	−0.0015 (0.0504)	−0.0056 (0.0316)
$s_{t-1}^{12,1}$	0.2036 (0.1694)	0.9127*** (0.2442)	0.1289 (0.1698)	0.0512 (0.1065)
$s_{t-2}^{12,1}$	−0.0399 (0.2085)	0.2616 (0.3005)	0.0256 (0.2089)	−0.0231 (0.1311)
$s_{t-3}^{12,1}$	−0.1141 (0.2099)	−0.2798 (0.3025)	−0.2414 (0.2103)	−0.1303 (0.1320)
$s_{t-4}^{12,1}$	−0.1828 (0.2102)	0.5160** (0.3029)	0.3192* (0.2106)	0.1697* (0.1321)
$s_{t-5}^{12,1}$	0.6021*** (0.2079)	−0.3520* (0.2997)	−0.1617 (0.2084)	−0.0507 (0.1307)
$s_{t-6}^{12,1}$	−0.2516* (0.1729)	0.0471 (0.2491)	−0.0724 (0.1732)	−0.1011 (0.1087)
$s_{t-1}^{6,1}$	−0.6234* (0.4141)	−0.0628 (0.5969)	0.7338** (0.4150)	0.0820 (0.2604)
$s_{t-2}^{6,1}$	0.0182 (0.5081)	−0.2724 (0.7323)	0.1476 (0.5091)	0.1774 (0.3195)
$s_{t-3}^{6,1}$	0.4275 (0.5102)	0.6366 (0.7353)	0.5034 (0.5113)	0.1914 (0.3208)
$s_{t-4}^{6,1}$	0.2740 (0.510)	−1.2991** (0.7350)	−0.7821* (0.5111)	−0.3879 (0.3206)
$s_{t-5}^{6,1}$	−1.4676*** (0.5076)	1.1729* (0.7315)	0.6392 (0.5086)	0.2476 (0.3191)
$s_{t-6}^{6,1}$	0.3932 (0.4273)	−0.1773 (0.6158)	0.0961 (0.4282)	0.2071 (0.2687)

247

Table 5.6: Estimation of Individual VAR-BOND Model (continued)

$s_{t-1}^{3,1}$	1.1372*** (0.3693)	−0.0945 (0.5322)	−0.2179 (0.3700)	0.4756** (0.2322)
$s_{t-2}^{3,1}$	−0.1628 (0.4582)	0.0026 (0.6604)	−0.1587 (0.4592)	−0.1316 (0.2881)
$s_{t-3}^{12,1}$	−0.2502 (0.4590)	−0.7137 (0.6615)	−0.5592 (0.4599)	−0.2200 (0.2886)
$s_{t-4}^{3,1}$	−0.0421 (0.4587)	1.2917** (0.6613)	0.8299** (0.4598)	0.4046* (0.2885)
$s_{t-5}^{3,1}$	0.8163** (0.4595)	−0.9184* (0.6623)	−0.5250 (0.4605)	−0.1750 (0.2889)
$s_{t-6}^{3,1}$	0.2139 (0.3898)	−0.3190 (0.5617)	−0.3820 (0.3906)	−0.3364* (0.2451)
$inpt$	−0.000001 (0.000004)	−0.000011** (0.000005)	−0.000006* (0.000004)	−0.000002 (0.000002)
$\overline{R^2}$	0.2582	0.9353	0.9015	0.8264
$\widehat{\sigma}$	0.0001	0.0001	0.0001	0.0000
$Fstat$	6.31***	221.24***	140.74***	73.69***
$eq^n LL$	3079.65	2945.51	3078.88	3249.95
Excl. Δr_t^1 terms	2.12*	1.17	1.87*	2.38**
Excl. $s_t^{12,1}$ terms	1.78	4.03***	0.78	0.60
Excl. $s_t^{6,1}$ terms	1.82	0.89	1.12	0.48
Excl. $s_t^{3,1}$ terms	1.76	1.02	0.99	1.89*
$\chi_N^2 [8]$	55.65***			
$\chi_H^2 [480]$	730.43***			
$\chi_{SC}^2 [16]$	11.53			

Notes: Standard errors in parenthesis (.). The $\overline{R}^2$, standard error of the regression $\left(\widehat{\sigma}\right)$, log likelihood (LL) presented with VAR residual diagnostic tests and the test of restrictions as detailed before. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. The IV model for the bond returns is estimated of order 6, over 1997 week 10 to 2004 week 18 (364 observations). Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 5.7: Estimation of Joint VAR (JV) Model

Equation	r_t^s	dy_t	Δr_t^1	$s_t^{12,1}$	$s_t^{6,1}$	$s_t^{3,1}$
r_{t-1}^s	0.8437*** (0.0559)	−0.0016** (0.0009)	0.00003 (0.0001)	0.00002 (0.0002)	−0.00005 (0.0001)	−0.00009 (0.0001)
r_{t-2}^s	0.0955* (0.0733)	−0.00043 (0.0012)	−0.00013 (0.0002)	0.00005 (0.0002)	0.00006 (0.0002)	0.00009 (0.0001)
r_{t-3}^s	0.0280 (0.0722)	0.0009 (0.0012)	0.00019 (0.0002)	−0.0003 (0.0002)	−0.0002 (0.0002)	−0.0001 (0.0001)
r_{t-4}^s	−0.9675*** (0.0723)	−0.0253*** (0.0012)	−0.0002* (0.0002)	0.0011*** (0.0002)	0.0007*** (0.0002)	0.0003*** (0.0001)
r_{t-5}^s	0.7132*** (0.1187)	0.0217*** (0.0019)	0.0002 (0.0003)	−0.0004 (0.0003)	−0.0004* (0.0002)	−0.0002* (0.0002)
r_{t-6}^s	−0.1010 (0.1069)	−0.0059*** (0.0017)	−0.000087 (0.0002)	−0.00006 (0.0003)	0.000043 (0.0002)	0.00007 (0.0001)
dy_{t-1}	−3.4884 (3.4040)	0.8663*** (0.0550)	0.0039 (0.0071)	0.0041 (0.0098)	0.0016 (0.0069)	0.0018 (0.0044)
dy_{t-2}	−7.4658** (4.0558)	−0.0966* (0.0655)	−0.0050 (0.0085)	0.0029 (0.0116)	0.0028 (0.0082)	−0.0009 (0.0053)
dy_{t-3}	3.2377 (3.3330)	−0.0071 (0.0539)	0.0004 (0.0070)	−0.00001 (0.0096)	−0.0004 (0.0068)	0.0006 (0.0044)
dy_{t-4}	24.7579*** (3.3222)	0.6343*** (0.0537)	−0.0042 (0.0070)	−0.0176** (0.0095)	−0.0010* (0.0067)	−0.0029 (0.0043)
dy_{t-5}	−20.4026*** (4.0628)	−0.5584*** (0.0657)	0.0068 (0.0085)	0.0055 (0.0117)	0.0049 (0.0083)	0.0023 (0.0053)
dy_{t-6}	3.8254 (3.4294)	0.1579*** (0.0554)	−0.0013 (0.0072)	0.0033 (0.0098)	0.0003 (0.0070)	−0.0011 (0.0045)
Δr_{t-1}^1	44.3657* (31.22)	−0.1888 (0.5045)	0.0345 (0.0653)	0.0490 (0.0895)	0.0810 (0.0634)	0.0611 (0.0407)
Δr_{t-2}^1	−9.8129 (31.1126)	0.2307 (0.5028)	−0.1481** (0.0651)	0.1290* (0.0892)	0.1449** (0.0632)	0.1010*** (0.0406)
Δr_{t-3}^1	−20.7582 (31.4388)	−0.1499 (0.5080)	−0.0343 (0.0657)	0.0358 (0.0901)	0.0453 (0.0638)	0.0374 (0.0410)
Δr_{t-4}^1	−0.8345 (31.2922)	0.4512 (0.5057)	0.0326 (0.0654)	0.0999 (0.0897)	0.1080** (0.0635)	0.0652* (0.0408)
Δr_{t-5}^1	5.1664 (31.4400)	0.2952 (0.5081)	−0.1546*** (0.0658)	0.1807** (0.0901)	0.2059*** (0.0638)	0.1551*** (0.0410)
Δr_{t-6}^1	5.7750 (24.2492)	0.2479 (0.3919)	0.0682* (0.0507)	−0.0697 (0.0695)	−0.0183 (0.0492)	−0.0146 (0.0316)
$s_{t-1}^{12,1}$	−45.0026 (81.7993)	0.8657 (1.3213)	0.1970 (0.1710)	0.9253*** (0.2345)	0.1357 (0.1660)	0.0610 (0.1067)
$s_{t-2}^{12,1}$	−48.8576 (101.310)	−1.5758 (1.6371)	−0.0487 (0.2119)	0.1956 (0.2905)	−0.0151 (0.2056)	−0.0512 (0.1321)
$s_{t-3}^{12,1}$	91.6225 (101.889)	1.1484 (1.6465)	−0.0599 (0.2131)	−0.2590 (0.2921)	−0.2352 (0.2068)	−0.1281 (0.1329)
$s_{t-4}^{12,1}$	13.1789 (101.991)	1.0135 (1.6481)	−0.2623 (0.2133)	0.5374** (0.2924)	0.3419** (0.2070)	0.1874* (0.1333)
$s_{t-5}^{12,1}$	−20.9624 (100.933)	−1.5036 (1.6310)	0.6723*** (0.2111)	−0.3337 (0.2894)	−0.1568 (0.2049)	−0.0592 (0.1316)
$s_{t-6}^{12,1}$	−19.0208 (84.3141)	0.7271 (1.3625)	−0.3056** (0.1763)	0.1620 (0.2417)	−0.0084 (0.1711)	−0.0770 (0.110)

Table 5.8: Estimation of Joint VAR Model (continued)

$s_{t-1}^{6,1}$	88.8499 (200.813)	−3.4790 (3.2450)	−0.6127* (0.4199)	−0.1499 (0.5758)	0.6944** (0.4076)	0.0589 (0.2619)
$s_{t-2}^{6,1}$	203.8494 (247.846)	8.8052** (4.0050)	0.0261 (0.5183)	−0.1192 (0.7106)	0.2311 (0.5030)	0.2267 (0.3233)
$s_{t-3}^{6,1}$	−238.2515 (248.877)	−5.2168* (4.0217)	0.3597 (0.5205)	0.3213 (0.7136)	0.3375 (0.5051)	0.1256 (0.3246)
$s_{t-4}^{6,1}$	−83.3880 (248.824)	−3.8379 (4.0208)	0.3672 (0.5203)	−1.0019* (0.7134)	−0.6166 (0.5050)	−0.3113 (0.3245)
$s_{t-5}^{6,1}$	82.6709 (246.359)	6.3508* (3.9810)	−1.5152*** (0.5152)	0.9559* (0.7064)	0.5057 (0.5000)	0.1922 (0.3213)
$s_{t-6}^{6,1}$	43.2428 (207.804)	−4.0780 (3.3580)	0.4916 (0.4346)	−0.4441 (0.5958)	−0.04545 (0.4218)	0.1585 (0.2710)
$s_{t-1}^{3,1}$	−15.1047 (180.825)	3.0517 (2.9220)	1.1146*** (0.3781)	0.0194 (0.5185)	−0.1664 (0.3670)	0.5014** (0.2358)
$s_{t-2}^{3,1}$	−343.980* (225.663)	−9.2975*** (3.6466)	−0.1809 (0.4719)	−0.0786 (0.6470)	−0.1952 (0.4580)	−0.1568 (0.2943)
$s_{t-3}^{3,1}$	212.502 (226.610)	4.7176* (3.6619)	−0.1935 (0.4739)	−0.2940 (0.6497)	−0.3398 (0.4599)	−0.1431 (0.2956)
$s_{t-4}^{3,1}$	120.809 (225.803)	3.9380 (3.6488)	−0.1221 (0.4722)	0.8971* (0.6474)	0.6008* (0.4583)	0.2985 (0.2945)
$s_{t-5}^{3,1}$	−70.8071 (224.02)	−6.7076** (3.620)	0.8540** (0.4685)	−0.7411 (0.6423)	−0.410 (0.4547)	−0.1196 (0.2922)
$s_{t-6}^{3,1}$	−44.6769 (189.193)	4.4755* (3.0572)	0.1246 (0.3956)	−0.0122 (0.5425)	−0.2115 (0.3840)	−0.2715 (0.2468)
$inpt$	−0.0119* (0.0085)	0.00008 (0.0001)	−0.00002 (0.00002)	0.00004* (0.00002)	0.00002 (0.00002)	0.000003 (0.00001)
$\overline{R^2}$	0.6710	0.9926	0.2611	0.9416	0.9080	0.8298
$\widehat{\sigma}$	0.0271	0.0004	0.00006	0.00008	0.00006	0.00004
$Fstat$	21.74	1360.44	4.59	165.01	101.34	50.58
$eq^n LL$	822.51	2336.48	3086.91	2971.09	3097.89	3260.17
Excl. r_t^s terms	82.35***	78.10***	1.03	4.61***	3.99***	2.64**
Excl. dy_t terms	4.40***	13821.12***	0.83	1.95*	1.20	0.31
Excl. Δr_t^1 terms	0.48	0.44	2.16**	1.18	2.9**	3.83***
Excl. $s_t^{12,1}$ terms	0.32	0.57	3.43***	19.33***	0.80	0.86
Excl. $s_t^{6,1}$ terms	0.37	1.16	3.15***	0.66	3.65***	1.17
Excl. $s_t^{3,1}$ terms	0.67	0.88	3.99***	0.52	1.32	1.47
Excl. $stock$ terms	56.68***	9301.20***	2.30**	3.48***	2.61***	1.76**
Excl. $bond$ terms	0.62	0.77	6.70***	333.43***	196.51***	89.49***
$\chi_N^2 [12]$	295.87***					
$\chi_H^2 [1512]$	2123.63***					
$\chi_{SC}^2 [36]$	24.42*					

Notes: Standard errors in parenthesis (.). The $\overline{R}^2$, standard error of the regression $(\widehat{\sigma})$, log likelihood (LL) presented with VAR residual diagnostic tests and the test of restrictions as detailed before. No roots of the characteristic polynomial lie outside the unit circle, so the VAR is stable. The JV model is estimated of order 6, over 1997 week 10 to 2004 week 18 (364 observations). Null rejected at *** 1% level, ** 5% level, * 10% level of significance.

Table 5.9: Root Mean Squared Errors of Bond and Stock Returns

(a) Bond Returns (r_t^1)

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
BNP	3.02E-06	3.43E-06	3.98E-06	3.44E-06	1.77E-06	2.56E-07
BP	3.02E-06	3.43E-06	3.98E-06	3.44E-06	1.77E-06	2.56E-07
IV	3.04E-06	3.55E-06	4.51E-06	5.03E-06	3.05E-06	1.33E-06
JV	3.00E-06	3.44E-06	4.35E-06	4.99E-06	3.15E-06	1.46E-06
BNPPU	3.02E-06	3.43E-06	3.98E-06	3.44E-06	1.77E-06	2.56E-07
BPPU	3.02E-06	3.43E-06	3.98E-06	3.44E-06	1.77E-06	2.56E-07
IVPU	3.05E-06	3.59E-06	4.61E-06	5.07E-06	3.00E-06	1.26E-06
JVPU	3.01E-06	3.47E-06	4.43E-06	4.99E-06	3.06E-06	1.34E-06

(b) Stock Returns (r_t^s)

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
BNP	0.001760	0.001599	0.002343	0.002479	0.002334	0.001431
BP	0.001607	0.001497	0.002316	0.002490	0.002316	0.001431
IV	0.001559	0.001520	0.002333	0.002482	0.002327	0.001430
JV	0.001533	0.001420	0.002368	0.002468	0.002360	0.001466
BNPPU	0.001762	0.001601	0.002340	0.002483	0.002333	0.001431
BPPU	0.001615	0.001505	0.002318	0.002485	0.002320	0.001429
IVPU	0.001504	0.001495	0.002314	0.002486	0.002327	0.001436
JVPU	0.001495	0.001411	0.002375	0.002477	0.002365	0.001470

Notes: The RMSEs are computed for each model for the horizons H = 1, 3, 6, 12, 18, 24 months as follows $\sqrt{\frac{\sum_{i=1}^{54}(r_{T+H}-\widehat{r}_{T+H})_i}{54}}$ where r_{T+H} is the actual monthly return i.e. r_t^1 and r_t^s, $\widehat{r}_{T+H}$ is the forecast and the difference between the two $\left(r_{T+H} - \widehat{r}_{T+H}\right)$ is computed for each recursion i, where there are 54 weekly recursions. The BNP, BP, IV and JV models are estimated subject to stochastic uncertainty only. The BNPPU, BPPU, IVPU and JVPU models consider parameter uncertainty too.

Table 5.10: Ratio of RMSEs of Returns

(a) Bond Returns (r_t^1)

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
BNP	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
BP	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
IV	1.0073	1.0338	1.1339	1.4635	1.7241	5.2038
JV	0.9956	1.0033	1.0927	1.4508	1.7805	5.7062
BNPPU	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
BPPU	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
IVPU	1.0111	1.0456	1.1573	1.4734	1.6958	4.9097
JVPU	0.9986	1.0125	1.1118	1.4507	1.7309	5.2440

(b) Stock Returns (r_t^s)

Model	$H = 1$	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
BNP	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
BP	0.9129	0.9361	0.9886	1.0043	0.9924	0.9999
IV	0.8859	0.9507	0.9959	1.0011	0.9970	0.9992
JV	0.8713	0.8883	1.0107	0.9953	1.0112	1.0247
BNPPU	1.0013	1.0013	0.9986	1.0016	0.9997	0.9999
BPPU	0.9174	0.9410	0.9894	1.0023	0.9942	0.9984
IVPU	0.8543	0.9349	0.9991	1.0028	0.9971	1.0031
JVPU	0.8495	0.8826	1.0136	0.9991	1.0135	1.0270

Notes: The above ratios are that of the RMSE for each model to the RMSE of the BNP model, which is taken as the benchmark.

Table 5.11: Realised Wealth under each Strategy for A=2

Position	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	AS 1.039024	AS 1.090983	AS 1.198814	AS 1.303596	AS 1.402846
2nd	JVPU 1.036934	JV 1.085841	JV 1.195839	JV 1.301244	JV 1.395425
3rd	IVPU 1.036653	JVPU 1.080824	JVPU 1.190218	JVPU 1.288895	JVPU 1.364283
4th	BPPU 1.036635	IVPU 1.078341	IV 1.179595	IV 1.279403	IV 1.359239
5th	JV 1.036284	IV 1.076682	IVPU 1.168318	IVPU 1.267401	IVPU 1.330711
6th	IV 1.036280	BP 1.075567	BP 1.166442	BP 1.242216	BPPU 1.280586
7th	BP 1.036424	BPPU 1.073885	BPPU 1.161321	BPPU 1.234480	BP 1.280391
8th	BNP 1.033901	BNP 1.06513	HH 1.122679	HH 1.186747	HH 1.248576
9th	BNPPU 1.033538	BNPPU 1.059189	BNPPU 1.106594	BNPPU 1.135350	BNPPU 1.162686
10th	HH 1.025256	HH 1.057076	BNP 1.106185	BNP 1.133146	BNP 1.158903
11th	AB 1.011487	AB 1.023170	AB 1.046544	AB 1.069898	AB 1.094306

Notes: The realised wealths above are the end-of-investment horizon wealths that the investor would have achieved over 2004 week 19 to 2007 week 19 had he allocated according to the optimal weights for each model, A and H. These have been averaged over the 54 recursions. The realised wealths for each model are ranked in descending order, for a particular A and H. The table shows how the four models, BNP, BP, IV and JV without and with parameter uncertainty perform, together with the lazy strategies in terms of their achieved realised wealths. The actual realised wealths are given below the model code. The BNP, BP, IV and JV models are estimated subject to stochastic uncertainty only. The BNPPU, BPPU, IVPU and JVPU models consider parameter uncertainty too. The lazy strategies are 'AS' is the 'all in stocks', 'AB' is the 'all in 1-month' and 'HH' is half in stocks and half in bonds.

Table 5.12: Realised Wealth under each Strategy for A=5

Position	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	AS 1.039024	AS 1.090983	AS 1.198814	AS 1.303596	AS 1.402846
2nd	JVPU 1.036578	JV 1.076580	JV 1.182284	JV 1.266296	JV 1.311270
3rd	JV 1.036177	JVPU 1.073879	JVPU 1.172839	JVPU 1.247169	JVPU 1.290901
4th	BP 1.034746	IVPU 1.067545	IV 1.130203	IVPU 1.188850	HH 1.248576
5th	IV 1.034722	IV 1.065844	IVPU 1.129411	HH 1.186747	IVPU 1.226558
6th	BPPU 1.034379	BP 1.063308	HH 1.22679	IV 1.185716	IV 1.225723
7th	IVPU 1.034076	BPPU 1.061854	BP 1.110021	BP 1.144193	BPPU 1.170952
8th	BNP 1.031437	HH 1.057076	BPPU 1.108467	BPPU 1.143424	BP 1.169210
9th	BNPPU 1.031427	BNP 1.045392	BNPPU 1.071853	BNPPU 1.096009	BNPPU 1.121266
10th	HH 1.025256	BNPPU 1.044855	BNP 1.070828	BNP 1.094976	BNP 1.119881
11th	AB 1.011487	AB 1.023170	AB 1.046544	AB 1.069898	AB 1.094306

Notes: See those for Table (5.11).

Table 5.13: Realised Wealth under each Strategy for A=10

Position	$H = 3$	$H = 6$	$H = 12$	$H = 18$	$H = 24$
1st	AS 1.039024	AS 1.090983	AS 1.198814	AS 1.303596	AS 1.402846
2nd	JVPU 1.034242	JVPU 1.062446	JV 1.127786	HH 1.186747	HH
3rd	JV 1.033848	JV 1.061033	JVPU 1.127539	JVPU 1.174776	JV 1.204549
4th	IV 1.030925	HH 1.057076	HH 1.122670	JV 1.174361	JVPU 1.200687
5th	IVPU 1.030333	IVPU 1.050169	IVPU 1.090307	IVPU 1.131481	IVPU 1.161707
6th	BP 1.030233	IV 1.047795	IV 1.088793	IV 1.127939	IV 1.160499
7th	BPPU 1.030059	BP 1.045415	BP 1.078158	BP 1.107021	BPPU 1.132265
8th	BNP 1.025752	BPPU 1.044642	BPPU 1.077299	BPPU 1.106430	BP 1.131377
9th	BNPPU 1.025671	BNP 1.034226	BNPPU 1.059303	BNPPU 1.082984	BNPPU 1.107677
10th	HH 1.025256	BNPPU 1.034002	BNP 1.058743	BNP 1.082416	BNP 1.107207
11th	AB 1.011487	AB 1.023170	AB 1.046544	AB 1.069898	AB 1.094306

Notes: See those for Table (5.11).

Chapter 6

Conclusion

In this thesis we both model and forecast the UK term structure at the short end. We model the term structure using a range of statistical and theory informed time series models. Then we generate density forecasts from a selection of these interest rate models, that embed varying degrees of economic theory and assess the predictability of the term structure in an investment decision making context. Both decision-based methods, that consider the economic value of the forecast to the user, as well as conventional statistical measures are used to determine forecast accuracy.

As discussed in Chapter 2, the importance of the term structure and its main theory, the EH is well documented. We begin by testing the EH and modelling the term structure generally in Chapter 3, Chapters 4 and 5 are concerned with forecasting the term structure. In Chapter 4, given that we find evidence of cointegration amongst the yields from Chapter 3, we use the MVART model that embeds the cointegration implied by the EH to model the interest rates, and then evaluate interest rate predictability in a decision-based framework. Chapter 5 extends this to include stock return predictability. We argue that density forecasting is appropriate here for conveying the risk and return information of each asset, where by using density forecasts of future returns the investor utilises the entire distribution about the forecast to inform his allocation decisions.

Further, decision-based forecast evaluation allows us to incorporate the investor's feeling about risk and assess the forecasts in terms of their economic value to the investor.

This chapter begins by summarising Chapters 2 to 5 in Section 6.1, Section 6.2 then details the contributions this thesis makes to the existing empirical literature. This is followed by a discussion of how this work can be extended in terms of future research in Section 6.3 and finally Section 6.4 concludes the thesis.

6.1 Summary of Chapters

Chapter 2 reviews the theories, modelling techniques, existing term structure and financial economic forecasting literature relevant to empirical Chapters 3 to 5. We first discuss the importance of the term structure and how UK interest rates have evolved over recent decades. We then summarise the main theory of the term structure, the Expectations Hypothesis (EH) and how it is assessed. The EH postulates that the long rate reflects information about the market's expectation of future short rates. The empirical literature finds mixed support for the EH, with it describing the data better under some monetary policy regimes than others, the support being sensitive to the testing method and the type of data employed. We describe how interest rates can be modelled more generally, using a range of atheoretic and theory informed time series models. The second part of Chapter 2 is concerned with forecasting, describing how point and density forecasts can be computed and briefly reviewing the interest rate forecasting literature. We discuss how forecasts are evaluated statistically and using decision-based methods, and review current studies that use decision-based methods to examine the predictability of asset returns.

Chapter 3 investigates the time series properties of the UK Term Structure over 1997 to 2004, through a set of statistical and theory based models using weekly data for 1-, 3-, 6- and 12-month yields. The models include the atheoretic Autoregressive (AR) model

which considers each yield separately, the Vector Autoregressive in Differences (VARD) model which allows for some interaction, the theory informed VAR in Transformed Interest Rates (VART) and Vector Error Correction (VECM) models which embed the long-run relationships between the yields as implied by the EH. There are two aims of this chapter, first to test the EH using cointegration analysis and the VAR methodology, and second to ascertain if a statistical or theory informed model is best placed to explain the UK term structure at the short end.

The spreads are found to be stationary, together with cointegration amongst the yields which implies that a long-run relationship exists between the n- and m-month yields and their spread is stationary. Further, the restrictions suggested by the EH, that the cointegrating vector is $(1, -1)$ can not be rejected. The term premia are insignificant, suggesting the Pure EH is applicable to this dataset. The VAR approach which compares the theoretical spread, as predicted by the EH, to the actual indicates a high comovement between actual and theoretical spreads. Although the Wald test rejects the hypothesis $s_t^{(n,m)*} = s_t^{(n,m)}$ in almost all cases, a divergence between the actual and theoretical spreads measured by time series plots, standard deviation ratios and correlation coefficients is not apparent. From the four models that incorporate varying degrees of economic theory, the in-sample properties of the VECM and VART models suggest they have greater explanatory power for the yields compared to the AR and VARD models. Also, the model restrictions of the AR, VARD and the bivariate specifications are rejected in favour of the multivariate VECM and VART models.

To conclude, Chapter 3 finds favourable evidence for the EH. Further, our findings demonstrate the importance of economic theory in explaining the term structure, with the theory informed models having greater explanatory power than the purely statistical ones. Hence, we find evidence to suggest the EH has significant economic content and provides a good representation of the UK money market.

In Chapter 4, given we find support for the EH and for modelling the term struc-

ture using a theory informed model in Chapter 3, we use the Multivariate VAR in Transformed Interest Rates (MVART) model, which embeds the cointegration relations implied by the EH, to forecasts yields. In particular, we compute the optimal portfolio allocation for a buy-and-hold investor with power utility over terminal wealth, using weekly UK data over 1997 to 2007 for two assets the 1-month and the n-month T-bill for $n = 3, 6, 12$, over investment horizons of up to 2 years. We consider two models that make opposing assumptions regarding return predictability. If the investor believes returns are not predictable, he uses a random walk with drift model to forecast returns and inform his allocation decisions. However, if he believes returns are predictable he uses the MVART model. Density forecasts of the returns are produced from both alternative models. This chapter first considers how the investor's allocation decisions are influenced by parameter uncertainty and predictability. Second, if the investor gains in terms of a higher wealth from assuming returns are predictable and using a theory informed model to forecast, in that is there economic value to interest rate predictability.

The effect of assuming returns are predictable on the optimal allocation is considerable, with the optimal weights under predictability in some cases greatly differing to those under no predictability. However, the effect of parameter uncertainty is small. Although previous studies report significant parameter uncertainty effects, these are apparent at horizons longer than those considered here. Under a statistical forecast evaluation criterion the random walk models outperform the MVART models at almost all horizons. However, under an economic value approach we find that in some cases the investor who assumes predictability is better off in terms of higher terminal wealth. In the other cases, the realised wealth under predictability is only marginally lower than that when returns are assumed not predictable. The conclusions drawn under each criterion do not entirely coincide. Under the economic value measure the random walk model is not superior to the MVART by the margin implied by the root mean squared

errors, and actually under this measure some evidence in favour of predictability is found.

In short, Chapter 4 finds the allocation decisions of an investor are influenced by the assumption made regarding predictability and to a much lesser extent by parameter uncertainty. Further, we find some evidence of economic value to interest rate predictability, where the investor may gain from assuming predictability.

Chapter 5 uses the asset allocation framework introduced in Chapter 4 and extends it by considering a risky asset. We compute the optimal portfolio allocation for a buy-and-hold investor with power utility over terminal wealth using weekly UK data over 1997 to 2007, for two assets the 1-month T-bill and the FTSE All-Share Index. We consider four models that assume varying degrees of bond and stock return predictability, where if returns are assumed predictable then key stock and term structure variables are believed to have explanatory power. Density forecasts of the returns are produced from all alternative models. This chapter first examines the effects of predictability and parameter uncertainty on how the investor optimally allocates. Secondly, if there is any economic value to the investor of bond and stock returns being predictable.

We find considerable effects of predictability on the optimal allocation, with return predictability leading to more being allocated to stocks at each horizon. This lending support to the predictive ability of the stock and term structure models considered here and to modelling both returns jointly. The effect of parameter uncertainty is not large over the investment horizon considered. Under the statistical forecast evaluation criterion, generally the models that assume no and limited predictability outperform the models that assume predictability at almost all horizons. Conversely, under the economic value approach the theory based models perform best. The results from the two evaluation techniques differ, whereby the model that achieves the lowest RMSE is not necessarily the one that will maximise realised wealths.

To conclude, Chapter 5 finds that the investor allocates differently when he assumes

260

predictability to an investor who assumes returns are not predictable. Also, we find evidence of economic value to bond and stock return predictability, with the terminal wealth gained by the investor who assumes returns are predictable being greater than one who assumes they are not. Further, the highest wealth is achieved by the investor who assumes the highest level of predictability and models both returns together.

6.2 Contributions of Thesis

This thesis makes empirical contributions to both the term structure literature and the forecasting literature. In this section we describe the contributions made by each empirical chapter.

Chapter 3 uses more recent UK data to test the EH using conventional methods; other UK studies mostly use data prior to changes in the monetary policy. In Chapter 3 we present new support for the much discussed and researched EH. Our findings are more supportive of the EH than earlier examinations, in particular previous UK studies that also use weekly data at the short end, see Cuthbertson (1996) and Cuthbertson et al (1996, 1998).

As discussed in Chapter 2, since adopting inflation targeting in 1992 and the Bank of England becoming independent in 1997, UK rates post 1992 and in particular after 1997 have become considerably less volatile compared to the rates observed prior to these reforms. We suggest that the interest rates observed during our sample 1997 to 2004 are sufficiently volatile for the EH to hold, but the rates are not too volatile as to invalidate the EH with a constant term premium. We offer this as a potential explanation for the stronger support we find for the EH in comparison to previous UK studies which use pre-1997 data, when interest rates were considerably more volatile. This change in the volatility of interest rates could in part be attributed to the Bank of England's credible anti-inflation policy, but appears primarily due to the stable economic climate during

this period described by the Governor of the Bank as the "nice decade".

Hence our findings for the UK add to those reported for the US, Germany and Denmark, that suggest favourable evidence for the EH is more likely under some monetary regimes than others, Mankiw and Miron (1986), Cuthbertson et al (2000b) and Christiansen et al (2003).

The objective of Chapter 4 is to examine the predictability of interest rates using an unrestricted VAR model that embeds the cointegration implied by the EH, and assess predictability using both statistical and economic value measures. From Chapter 2 the literature examining asset return predictability mainly focus their attention on stock predictability, e.g. Marquering and Verbeek (2004) and Guidolin and Timmermann (2005), with a few considering exchange rate predictability and the economic value of this predictability, Abhyankar et al (2005) and Garratt and Lee (2009). However, very little attention has been paid to decision-based forecast evaluation of interest rates and determining if there is economic value to interest rate predictability, which is where Chapter 4 contributes to the literature.

To my knowledge Della Corte et al (2008) is the only other to consider bond return predictability using an economic value criteria. They focus on testing the EH and seeing if there is economic value to departures from the EH. Whereas we do not seek to test the EH in this chapter, so use an unrestricted VAR model. We consider the importance of (1) predictability and parameter uncertainty in asset allocation, (2) generating density forecasts to capture the risk as well as the return about the asset and (3) the economic value to the investor of these interest rate forecasts. The innovation of this chapter is considering (1), (2) and (3) in the context of interest rates.

Chapter 5 extends earlier studies that examine stock return predictability by also allowing for the possibility of predictability in bond returns. These previous studies ignore bond return predictability and assume the T-bill rate is constant, with some considering if term structure variables have explanatory power, Pesaran and Timmer-

mann (1995) and Barberis (2000). The key innovation of this chapter is that we model the bond returns too, further we model the two returns jointly and evaluate their predictability using economic value.

Both Chapters 4 and 5 find that the investor's allocation is sensitive to the investment horizon, whether he assumes returns are predictable and to a lesser extent by parameter uncertainty over the investment horizon considered. Further, the findings of both chapters suggest that the conclusion of how well theory informed models perform, in terms of forecasting, is sensitive to the evaluation criteria used. In itself, as discussed in Chapters 2, 4 and 5 this is not a new empirical finding, but what is new is the asset return considered. We find support for the predictive power of theory informed models under a decision-based criterion using interest rates and stock returns. Whereas previous findings have focused on stock and exchange rate predictability.

In this thesis we test the EH, model and forecast the term structure using a theory informed model that embeds the cointegration implied by the EH, use these forecasts to inform asset allocation decisions and then assess asset return predictability using both statistical and decision-based methods. So here we draw upon, bring together and contribute to the areas of the term structure, density forecasting and forecast uncertainties, asset allocation and decision-based forecast evaluation.

6.3 Future Research

In Chapter 3 we find positive results for the EH using recent UK data. Future research may involve examining if this support is sensitive to the frequency and maturity of the data, and to alternative testing methods e.g. the newly developed Lagrange multiplier test as used by Sarno et al (2007) and Della Corte et al (2008).

We model the term structure using a range of time series models which can be summarised by the VAR framework and assess their forecasting ability using decision-

based methods. This set can be expanded to include ARCH/GARCH models, models that include macroeconomic variables and survey and market based forecasts. In which case, future work may examine if predictability and further the modelling approach chosen has economic value.

The fact that we model the term structure and stock returns using several alternative models in Chapters 3 to 5, highlights that there is model uncertainty since the true DGP is unknown. There is a growing body of literature that deals with forecast combination, that considers if forecasts should be combined and if so how. Our research deals with density forecasting and decision-based evaluation, there is potential to extend this research into the area of forecast combination. In this case, should the investor choose a single model to forecast and inform allocation decisions, or will some optimal combination of the forecasts result in a higher terminal wealth.

The results presented here lend support to the use of decision-based forecast evaluation. We consider here one application of decision-based methods, i.e. to assess forecasts of asset returns. But there is potentially a vast number of applications where these methods can be applied. For instance, for a private company that relies on survey and market based forecasts, this method may be used to determine which companies' forecasts generate the highest profits. Or when forecasting key macroeconomic variables that can be described by alternative models, decision-based methods may provide a means of discriminating between these alternatives.

6.4   Concluding Comments

This thesis examines the UK term structure at the short end, in relation to the questions addressed in this thesis we find:

1. Evidence to suggest that the EH holds for this recent sample of UK data.

2. More support using this dataset than previous UK studies, that test the EH over

a period when interest rates were comparatively more volatile.

3. Theory informed models to be best placed to describe the UK money market.

4. Under a statistical criterion the forecasting performance of the atheoretic random walk model is difficult to beat.

5. The investor's allocation is sensitive to the investment horizon, parameter uncertainty and the assumption of predictability.

6. Under decision-based methods the theory informed models do have predictive power and in many cases perform better than the random walk.

7. The forecast performance of the models is sensitive to the evaluation criteria used.

8. Predictability to have an economic value. Such that, by assuming returns are predictable and using a theory informed model, the investor gains in terms of wealth.

Our contributions to the literature are empirical. Chapter 3 presents new support for the EH using data post of the monetary policy regime changes in the UK. We suggest that these positive findings are due to interest rates being considerably less volatile post of these changes, than pre of these changes. This reduction in the volatility of interest rates may partly be explained by the anti-inflation stance taken by the Bank of England, but mostly by the stable conditions of the "nice decade". This finding implies that the volatility of the interest rates is important for the EH. Chapters 4 and 5 contribute to the asset return predictability and decision-based evaluation literature. Interest rate predictability has scarcely been examined using decision-based methods, this is addressed in Chapter 4. Further, Chapter 5 extends the stock predictability literature by considering interest rate predictability too. Chapters 4 and 5 not only forecast the term structure, but consider with how these forecasts are used in decision making. As stressed in Wright et al (1986) it is important to integrate forecasting and

decision making, and be aware of how the forecasts are to be used. As we show the most accurate forecast in terms of statistical criteria, may not be the best in terms of decision making.

This thesis presents new evidence of interest rate predictability. Previously, the exchange rate, interest rate and stock return forecasting literature finds atheoretic models, like a naive random walk, are difficult to beat under statistical measures implying that returns are not predictable. However, as discussed in Chapter 2, under decision-based evaluation methods evidence of stock return and exchange rate predictability, as captured by theory informed models, has been reported. We find support for theory informed models when we evaluate forecasts in terms of their economic value to the investor. That is, the investor gains in terms of a higher end-of-period wealth by assuming returns are predictable and using a theory informed model to forecast. Our findings, in support of the literature, illustrate that the conclusion drawn on the forecasting performance of theory informed models compared to atheoretic models is sensitive to the evaluation criterion used.

Statistical evaluation methods indicate that theory based models fail to satisfactorily capture reality and that rather than painstakingly trying to model reality, we are better off just using a no change model to forecast. However, our results taken with those presented in the literature, suggest that by using an alternative assessment criterion, one that considers the economic value of the forecast to the user, economic theory is in fact found to have predictive power and that this predictive power has a value.

The findings presented here highlight first, the importance of the volatility of interest rates when testing the EH. Second, the importance of density forecasting. In this thesis we apply density forecasting to investment decision making and illustrate (1) the appropriateness of these forecasts compared to point forecasts and (2) although analytically difficult to compute, density forecasts can be generated using simulation methods. Both (1) and (2) should serve to encourage greater use of density forecasts.

Third, the importance of and the need for evaluating forecasts using an appropriate criterion, i.e. an evaluation criterion that reflects the purpose for which the forecasts are ultimately intended. In our case this purpose is portfolio allocation, hence it is necessary to incorporate into the evaluation process the investor's feelings about risk and the economic value of the forecasts to the investor. This is done using decision-based methods. But statistical criteria do not take these factors into account, which arguably may deem them inadequate for judging forecasts.

Our results advocate the use of density forecasts that provide the entire distribution about the expected future value. And further, given the importance of forecast evaluation, the use of decision-based evaluation methods that assess forecasts from the user's perspective, i.e. in terms of loss and profits rather than forecast errors, since ultimately it is the user who will be using the forecast.

In short, we find economic theory to be important in explaining the term structure and evidence to suggest that the UK money market is consistent with the EH. Our findings demonstrate the importance of predictability and parameter uncertainty in asset allocation, generating density forecasts to capture the risk as well as the return of the asset, using a decision-based criterion to assess forecast accuracy and that there is economic value to assuming returns are predictable. To conclude we find economic theory to be significant for both modelling and forecasting the UK term structure at the short end.

Appendices

.1 Data Appendix

Here we provide details of the source, definitions and transformations of the data.

Chapters 3 and 4

Yield data: $r_t^{(n)}$

- Source: Bank of England (BoE)

  http://www.bankofengland.co.uk/statistics/yieldcurve/archive.htm

- Definition: Nominal government n-month spot interest rate obtained from 'UK Nominal Spot Curve'. The curve is estimated using gilt and gilt repo rates.

- Use Wednesday observations of the annualised and continuously compounded 1-, 3-, 6- and 12-month rates, where the n-month rate $r_t^{(n)} = \ln\left[1 + \left(R_t^{(n)}/100\right)\right]$.

- Notes on the BoE UK yield curves provides further details:

  http://www.bankofengland.co.uk/statistics/yieldcurve/

    notes%20on%20the%20bofe%20uk%20yield%20curvesV2.pdf

Chapter 5

Bond data: $r_t^{(n)}$

- Source: BoE, see above.

- Definition: As for Chapters 3 and 4, use the nominal government n-month spot interest rates.

- Transformation: Use the Wednesday observations of the annualised 1-, 3-, 6- and 12-month rates $R_t^{(n)}$, to construct the n-month rate expressed as a monthly rate $r_t^{(n)} = \ln\left[1 + \left(R_t^{(n)}/100\right)\right]^{1/12}$.

Stock data: r_t^s

- Source: Datastream, mneumonic=DSR1

- Definition: Data on FTSE All-Share Return Index RI_t, "..the return index presents the theoretical growth in value of a notional stock holding, the price of which is that of the selected price index. This holding is deemed to return a daily dividend, which is used to purchase new units of the stock at the current price. The gross dividend is used: $RI_t = RI_{t-1} * \frac{PI_t}{PI_{t-1}} * \left(1 + \frac{DY*f}{n}\right)$, where RI_t = return index on day t; RI_{t-1} = return index on day $t-1$; PI_t = price index on day t; PI_{t-1} = price index on day $t-1$; DY = dividend yield of the price index; f = grossing factor (normally 1) if the dividend yield is a net figure rather than a gross, f is used to gross up the yield; n = no of days in financial year (normally 260)*100. " datastream definition.

- Transformation: Use RI_t, Wednesday observations to compute the continuously compounded monthly return $r_t^s = \ln\left[RI_{t+4}/RI_t\right]$.

Dividend yield: dy_t

- Source: Datastream, mneumonic=DY

- Definition: Datastream definition "For sectors, dy is derived by calculating the total dividend amount for a sector and expressing it as a percentage of the total market value for the constituents of that sector. This provides an average of the individual yields of the constituents weighted by the market value, dividend yield (given by the total dividends over the value of the portfolio): $DY_t = \frac{\sum_1^n (D_t * N_t)}{\sum_1^n (P_t * N_t)} *$ 100, where DY_t = aggregate dividend yield on day t; D_t = dividend per share on day t; N_t = no of shares in issue on day t; P_t = unadjusted share price on day t; n = no. of constituents in index."

- Although this definition makes reference to DY on day t, a plot of the series shows that the dividend yield is between 2 and 4%. The magnitudes would suggest this is an annual measure.

- Use Wednesday observations of this series.

.2 Computing the Theoretical Spread

Here we illustrate how the unrestricted VAR can be used to compute the theoretical spread. Taking $n = 3$ and $m = 1$

$$
\begin{pmatrix}
s_t^{(3,1)} \\
\Delta r_t^{(1)} \\
s_{t-1}^{(3,1)} \\
\Delta r_{t-1}^{(1)} \\
\cdot \\
\cdot \\
s_{t-p+1}^{(3,1)} \\
\Delta r_{t-p+1}^{(1)}
\end{pmatrix}
=
\begin{pmatrix}
\mathbf{A}_1 & \mathbf{A}_2 & \cdot & \cdot & \cdot & \cdot & \mathbf{A}_{p-1} & \mathbf{A}_p \\
\mathbf{I}_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \cdot & 0 & & & & 0 & 0 \\
0 & 0 & \cdot & 0 & & & 0 & 0 \\
0 & & 0 & \cdot & 0 & & 0 & 0 \\
0 & & & 0 & \cdot & 0 & 0 & 0 \\
0 & & & & 0 & \cdot & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}_2 & 0
\end{pmatrix}
\begin{pmatrix}
s_{t-1}^{(3,1)} \\
\Delta r_{t-1}^{(1)} \\
s_{t-2}^{(3,1)} \\
\Delta r_{t-2}^{(1)} \\
\cdot \\
\cdot \\
s_{t-p}^{(3,1)} \\
\Delta r_{t-p}^{(1)}
\end{pmatrix}
+
\begin{pmatrix}
\epsilon_{1t} \\
\epsilon_{1t} \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix}
$$

From equations (3.1) and (3.2)

$$
\begin{aligned}
r_t^{(3)} &= \frac{1}{3}\left(r_t^{(1)} + E_t r_{t+1}^{(1)} + E_t r_{t+2}^{(1)}\right) \\
s_t^{(3,1)} &= \frac{2}{3}E_t \Delta r_{t+1}^{(1)} + \frac{1}{3}E_t \Delta r_{t+2}^{(1)}
\end{aligned}
$$

Using the chain rule of forecasting $\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t$ and $\mathbf{z}_{t+2} = \mathbf{A}^2\mathbf{z}_t$, so from equation (3.7) the expected changes in the short rates are

$$
\begin{aligned}
E_t\left(\Delta r_{t+1}^{(1)}\right) &= \mathbf{e2}' E_t\left(\mathbf{z}_{t+1}\right) = \mathbf{e2}'\mathbf{A}\mathbf{z}_t \\
E_t\left(\Delta r_{t+2}^{(1)}\right) &= \mathbf{e2}' E_t\left(\mathbf{z}_{t+2}\right) = \mathbf{e2}'\mathbf{A}^2\mathbf{z}_t
\end{aligned}
$$

From equation (3.8) the theoretical spread can be calculated as

$$s_t^{(3,1)*} = \left(\frac{2}{3} \mathbf{e2}' \mathbf{A} + \frac{1}{3} \mathbf{e2}' \mathbf{A}^2 \right) \mathbf{z}_t$$

Under the EH the actual and theoretical spreads are equal, such that

$$\mathbf{e1}' \mathbf{z}_t = \left(\frac{2}{3} \mathbf{e2}' \mathbf{A} + \frac{1}{3} \mathbf{e2}' \mathbf{A}^2 \right) \mathbf{z}_t$$

.3 VART to VAR in Levels

Here we describe how the VART model can be written as a VAR in levels. The VART is given by equation (2.25), which is composed of $\Delta r^{(m)}$ and $s^{(n,m)}$ that are both stationary, this can be written in levels as follows

$$
\begin{pmatrix} r_t^{(n)} - r_t^{(m)} \\ r_t^{(m)} - r_{t-1}^{(m)} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^{(n)} \\ r_t^{(m)} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} r_t^{(m)} \\ r_{t-1}^{(m)} \end{pmatrix}
$$

$$
= \beta + \mathbf{B}_1 \begin{pmatrix} r_{t-1}^{(n)} - r_{t-1}^{(m)} \\ r_{t-1}^{(m)} - r_{t-2}^{(m)} \end{pmatrix} + ... + \mathbf{B}_p \begin{pmatrix} r_{t-p}^{(n)} - r_{t-p}^{(m)} \\ r_{t-p}^{(m)} - r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix}
$$

$$
\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^{(n)} \\ r_t^{(m)} \end{pmatrix} = - \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} r_t^{(m)} \\ r_{t-1}^{(m)} \end{pmatrix} + \beta + \mathbf{B}_1 \begin{pmatrix} r_{t-1}^{(n)} - r_{t-1}^{(m)} \\ r_{t-1}^{(m)} - r_{t-2}^{(m)} \end{pmatrix} + ...
$$

$$
+ \mathbf{B}_p \begin{pmatrix} r_{t-p}^{(n)} - r_{t-p}^{(m)} \\ r_{t-p}^{(m)} - r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix}
$$

$$
= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^{(m)} \\ r_{t-1}^{(m)} \end{pmatrix} + \beta + \mathbf{B}_1 \begin{pmatrix} r_{t-1}^{(n)} - r_{t-1}^{(m)} \\ r_{t-1}^{(m)} - r_{t-2}^{(m)} \end{pmatrix} + ...
$$

$$
+ \mathbf{B}_p \begin{pmatrix} r_{t-p}^{(n)} - r_{t-p}^{(m)} \\ r_{t-p}^{(m)} - r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix}
$$

$$
\begin{pmatrix} r_t^{(n)} \\ r_t^{(m)} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^{-1} \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^{(m)} \\ r_{t-1}^{(m)} \end{pmatrix} + \beta + \mathbf{B}_1 \begin{pmatrix} r_{t-1}^{(n)} - r_{t-1}^{(m)} \\ r_{t-1}^{(m)} - r_{t-2}^{(m)} \end{pmatrix} + ... \right.
$$
$$
\left. + \mathbf{B}_p \begin{pmatrix} r_{t-p}^{(n)} - r_{t-p}^{(m)} \\ r_{t-p}^{(m)} - r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix} \right\}
$$
(1)

From equation (1) let

$$M_0 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

such that

$$\begin{pmatrix} r_t^{(n)} \\ r_t^{(m)} \end{pmatrix} = \mathbf{M}_0^{-1} \left\{ \begin{array}{l} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^{(m)} \\ r_{t-1}^{(m)} \end{pmatrix} + \beta + \mathbf{B}_1 \begin{pmatrix} r_{t-1}^{(n)} - r_{t-1}^{(m)} \\ r_{t-1}^{(m)} - r_{t-2}^{(m)} \end{pmatrix} + ... \\ \\ +\mathbf{B}_p \begin{pmatrix} r_{t-p}^{(n)} - r_{t-p}^{(m)} \\ r_{t-p}^{(m)} - r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix} \end{array} \right\}$$

$$= \mathbf{M}_0^{-1} \left\{ \begin{array}{l} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^{(m)} \\ r_{t-1}^{(m)} \end{pmatrix} + \beta \\ \\ +\mathbf{B}_1 \left[\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_{t-1}^{(n)} \\ r_{t-1}^{(m)} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} r_{t-1}^{(m)} \\ r_{t-2}^{(m)} \end{pmatrix} \right] + ... \\ \\ +\mathbf{B}_p \left[\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_{t-p}^{(n)} \\ r_{t-p}^{(m)} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} r_{t-p}^{(m)} \\ r_{t-p-1}^{(m)} \end{pmatrix} \right] \\ \\ + \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix} \end{array} \right\}$$

274

$$
= \mathbf{M}_0^{-1} \left\{
\begin{array}{l}
\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^{(m)} \\ r_{t-1}^{(m)} \end{pmatrix} + \beta \\[2ex]
+\mathbf{B}_1 \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_{t-1}^{(n)} \\ r_{t-1}^{(m)} \end{pmatrix} + \mathbf{B}_1 \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} r_{t-1}^{(m)} \\ r_{t-2}^{(m)} \end{pmatrix} + .. \\[2ex]
+\mathbf{B}_p \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_{t-p}^{(n)} \\ r_{t-p}^{(m)} \end{pmatrix} + \mathbf{B}_p \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} r_{t-p}^{(m)} \\ r_{t-p-1}^{(m)} \end{pmatrix} \\[2ex]
+ \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix}
\end{array}
\right\}
$$

$$
= \mathbf{M}_0^{-1} \left\{
\begin{array}{l}
\beta + \left[\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \mathbf{B}_1 \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} r_{t-1}^{(n)} \\ r_{t-1}^{(m)} \end{pmatrix} + \\[2ex]
\left[\mathbf{B}_1 \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} + \mathbf{B}_2 \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} r_{t-2}^{(n)} \\ r_{t-2}^{(m)} \end{pmatrix} + ... \\[2ex]
+ \left[\mathbf{B}_p \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \right] \begin{pmatrix} r_{t-p-1}^{(n)} \\ r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix}
\end{array}
\right\}
$$

thus in levels, as given by equation (2.26)

$$
\begin{pmatrix} r_t^{(n)} \\ r_t^{(m)} \end{pmatrix} = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \begin{pmatrix} r_{t-1}^{(n)} \\ r_{t-1}^{(m)} \end{pmatrix} + ... + \boldsymbol{\Phi}_{p+1} \begin{pmatrix} r_{t-p-1}^{(n)} \\ r_{t-p-1}^{(m)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}
$$

where

$$
\boldsymbol{\mu} = \mathbf{M}_0^{-1}\beta \quad ; \quad \boldsymbol{\Phi}_i = \mathbf{M}_0^{-1}\mathbf{M}_i \quad \text{for } i = 1 \text{ to } p+1
$$

275

$$\mathbf{M}_0^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad ; \quad \mathbf{M}_0^{-1} \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad ;$$

$$\mathbf{M}_1 = \mathbf{B}_0 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \mathbf{B}_1 \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad ; \quad \mathbf{M}_{p+1} = \mathbf{B}_p \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \quad ;$$

$$\mathbf{M}_i = \mathbf{B}_{i-1} \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} + \mathbf{B}_i \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad \text{for } i = 2 \text{ to } p$$

.4 Lag Structures

Below shows the lag structures under the VARD, VECM and VART models, for $p = 3$.

- VARD(p) Model

where $z_t = \left(r_t^{(12)}, r_t^{(6)}, r_t^{(3)}, r_t^{(1)} \right)'$

$$
\begin{aligned}
\Delta z_t &= a_o + \sum_{i=1}^{p} \theta_i \Delta z_{t-i} + \xi_t \\
\Delta z_t &= a_o + \theta_1 \Delta z_{t-1} + \theta_2 \Delta z_{t-2} + \theta_3 \Delta z_{t-3} + \xi_t \\
\Delta z_t &= a_o + \theta_1 (z_{t-1} - z_{t-2}) + \theta_2 (z_{t-2} - z_{t-3}) + \theta_3 (z_{t-3} - z_{t-4}) + \xi_t \qquad (2)
\end{aligned}
$$

- VECM(p)

Starting with an AR(p) where the variables in z_t are in levels

$$
\mathbf{z}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{z}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{z}_{t-2} + \ldots + \boldsymbol{\Phi}_{p+1} \mathbf{z}_{t-p-1} + \boldsymbol{\varepsilon}_t
$$

if $p = 3$

$$
\mathbf{z}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{z}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{z}_{t-2} + \boldsymbol{\Phi}_3 \mathbf{z}_{t-3} + \boldsymbol{\Phi}_4 \mathbf{z}_{t-4} + \boldsymbol{\varepsilon}_t
$$

to this $\pm \mathbf{z}_{t-1}; \pm \boldsymbol{\Phi}_2 \mathbf{z}_{t-1}; \pm \boldsymbol{\Phi}_3 \mathbf{z}_{t-1}; \pm \boldsymbol{\Phi}_4 \mathbf{z}_{t-1}; \pm \boldsymbol{\Phi}_3 \mathbf{z}_{t-2}; \pm \boldsymbol{\Phi}_4 \mathbf{z}_{t-2}; \pm \boldsymbol{\Phi}_4 \mathbf{z}_{t-3}; \pm \boldsymbol{\Phi}_4 \mathbf{z}_{t-4}$

$$
\mathbf{z}_t - \mathbf{z}_{t-1} = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{z}_{t-1} - \mathbf{z}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{z}_{t-1} - \boldsymbol{\Phi}_2 \mathbf{z}_{t-1} + \boldsymbol{\Phi}_3 \mathbf{z}_{t-3} - \ldots + \boldsymbol{\varepsilon}_t
$$

$$\Delta \mathbf{z}_t \;=\; \boldsymbol{\mu} - \left(I_k - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \boldsymbol{\Phi}_3 - \boldsymbol{\Phi}_4\right) \mathbf{z}_{t-1} - \boldsymbol{\Phi}_2 \left(\mathbf{z}_{t-1} - \mathbf{z}_{t-2}\right)$$

$$- \boldsymbol{\Phi}_3 \left(\mathbf{z}_{t-1} - \mathbf{z}_{t-2}\right) - \boldsymbol{\Phi}_4 \left(\mathbf{z}_{t-1} - \mathbf{z}_{t-2}\right) - \boldsymbol{\Phi}_3 \left(\mathbf{z}_{t-2} - \mathbf{z}_{t-3}\right)$$

$$- \boldsymbol{\Phi}_4 \left(\mathbf{z}_{t-2} - \mathbf{z}_{t-3}\right) - \boldsymbol{\Phi}_4 \left(\mathbf{z}_{t-3} - \mathbf{z}_{t-4}\right) + \boldsymbol{\varepsilon}_t$$

$$\Delta \mathbf{z}_t \;=\; \boldsymbol{\mu} - \Pi z_{t-1} - \boldsymbol{\Phi}_2 \Delta z_{t-1} - \boldsymbol{\Phi}_3 \Delta z_{t-1} - \boldsymbol{\Phi}_4 \Delta z_{t-1}$$

$$- \boldsymbol{\Phi}_3 \Delta z_{t-2} - \boldsymbol{\Phi}_4 \Delta z_{t-2} - \boldsymbol{\Phi}_4 \Delta z_{t-3} + \boldsymbol{\varepsilon}_t$$

this gives the VECM

$$\Delta \mathbf{z}_t = \boldsymbol{\mu} - \Pi z_{t-1} + \Gamma_1 \Delta z_{t-1} + \Gamma_2 \Delta z_{t-2} + \Gamma_3 \Delta z_{t-3} + \boldsymbol{\varepsilon}_t$$

where

$$\Gamma_1 = -\left(\boldsymbol{\Phi}_2 + \boldsymbol{\Phi}_3 + \boldsymbol{\Phi}_4\right); \;\; \Gamma_2 = -\left(\boldsymbol{\Phi}_3 + \boldsymbol{\Phi}_4\right); \;\; \Gamma_3 = -\boldsymbol{\Phi}_4$$

$$\Pi = \left(I_k - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \boldsymbol{\Phi}_3 - \boldsymbol{\Phi}_4\right)$$

$$\Delta z_t = \boldsymbol{\mu} + \Gamma_1(z_{t-1} - z_{t-2}) + \Gamma_2(z_{t-2} - z_{t-3}) + \Gamma_3(z_{t-3} - z_{t-4}) - \Pi z_{t-1} + \varepsilon_t \qquad (3)$$

- VART(p) (which is equivalent to the VECM)

$$
\begin{pmatrix} s_t^{(12,1)} \\ s_t^{(6,1)} \\ s_t^{(3,1)} \\ \Delta r_t^{(1)} \end{pmatrix} = \beta + \sum_{i=1}^{p} \mathbf{B}_i \begin{pmatrix} s_{t-i}^{(12,1)} \\ s_{t-i}^{(6,1)} \\ s_{t-i}^{(3,1)} \\ \Delta r_{t-i}^{(1)} \end{pmatrix} + \boldsymbol{\gamma}_t
$$

$$
\begin{pmatrix} s_t^{(12,1)} \\ s_t^{(6,1)} \\ s_t^{(3,1)} \\ \Delta r_t^{(1)} \end{pmatrix} = \beta + \mathbf{B}_1 \begin{pmatrix} s_{t-1}^{(12,1)} \\ s_{t-1}^{(6,1)} \\ s_{t-1}^{(3,1)} \\ \Delta r_{t-1}^{(1)} \end{pmatrix} + \mathbf{B}_2 \begin{pmatrix} s_{t-2}^{(12,1)} \\ s_{t-2}^{(6,1)} \\ s_{t-2}^{(3,1)} \\ \Delta r_{t-2}^{(1)} \end{pmatrix} + \mathbf{B}_3 \begin{pmatrix} s_{t-3}^{(12,1)} \\ s_{t-3}^{(6,1)} \\ s_{t-3}^{(3,1)} \\ \Delta r_{t-3}^{(1)} \end{pmatrix} + \boldsymbol{\gamma}_t
$$

$$
\begin{pmatrix} s_t^{(12,1)} \\ s_t^{(6,1)} \\ s_t^{(3,1)} \\ \Delta r_t^{(1)} \end{pmatrix} = \beta + \mathbf{B}_1 \begin{pmatrix} r_{t-1}^{(12)} - r_{t-1}^{(1)} \\ r_{t-1}^{(6)} - r_{t-1}^{(1)} \\ r_{t-1}^{(3)} - r_{t-1}^{(1)} \\ r_{t-1}^{(1)} - r_{t-2}^{(1)} \end{pmatrix} + \mathbf{B}_2 \begin{pmatrix} r_{t-2}^{(12)} - r_{t-2}^{(1)} \\ r_{t-2}^{(6)} - r_{t-2}^{(1)} \\ r_{t-2}^{(3)} - r_{t-2}^{(1)} \\ r_{t-2}^{(1)} - r_{t-3}^{(1)} \end{pmatrix}
$$

$$
+ \mathbf{B}_3 \begin{pmatrix} r_{t-3}^{(12)} - r_{t-3}^{(1)} \\ r_{t-3}^{(6)} - r_{t-3}^{(1)} \\ r_{t-3}^{(3)} - r_{t-3}^{(1)} \\ r_{t-3}^{(1)} - r_{t-4}^{(1)} \end{pmatrix} + \boldsymbol{\gamma}_t \tag{4}
$$

For all the estimated models to be comparable they must have the same lag structure, such that it is only the restrictions that are imposed on the various models that differentiates them. Hence it can be seen from equations (2), (3) and (4) that estimating all the models of order p will yield the same lag structure.

.5   End of Investment Period Cumulative Return

The cumulative return at the end of the investment horizon, $T + H$, from investing the proportion ω and $(1 - \omega)$ of initial wealth W_T in a sequence of rolling investments in m-period short bills and n-period long bills, respectively is

$$
\begin{aligned}
W_{T+H} &= \omega \prod_{i=1}^{s} \left(1 + R_{T+(i-1)m}^{(m)}\right)^{\frac{1}{s}} + (1 - \omega) \prod_{i=1}^{l} \left(1 + R_{T+(i-1)n}^{(n)}\right)^{\frac{1}{l}} \\
&= \omega \left\{ \left(1 + R_T^{(m)}\right)^{\frac{1}{s}} \ldots \left(1 + R_{T+(s-1)m}^{(m)}\right)^{\frac{1}{s}} \right\} \\
&\quad + (1 - \omega) \left\{ \left(1 + R_T^{(n)}\right)^{\frac{1}{l}} \ldots \left(1 + R_{T+(l-1)n}^{(n)}\right)^{\frac{1}{l}} \right\} \\
&= \omega \exp \left(\ln \left\{ \left(1 + R_T^{(m)}\right)^{\frac{1}{s}} \ldots \left(1 + R_{T+(s-1)m}^{(m)}\right)^{\frac{1}{s}} \right\} \right) \\
&\quad + (1 - \omega) \exp \left(\ln \left\{ \left(1 + R_T^{(n)}\right)^{\frac{1}{l}} \ldots \left(1 + R_{T+(l-1)n}^{(n)}\right)^{\frac{1}{l}} \right\} \right) \\
&= \omega \exp \left(\frac{1}{s} \ln \left(1 + R_T^{(m)}\right) + \ldots + \frac{1}{s} \ln \left(1 + R_{T+(s-1)m}^{(m)}\right) \right) \\
&\quad + (1 - \omega) \exp \left(\frac{1}{l} \ln \left(1 + R_T^{(n)}\right) + \ldots + \frac{1}{l} \ln \left(1 + R_{T+(l-1)n}^{(n)}\right) \right) \\
&= \omega \exp \left(\frac{1}{s} \left\{ \ln \left(1 + R_T^{(m)}\right) + \ldots + \ln \left(1 + R_{T+(s-1)m}^{(m)}\right) \right\} \right) \\
&\quad + (1 - \omega) \exp \left(\frac{1}{l} \left\{ \ln \left(1 + R_T^{(n)}\right) + \ldots + \ln \left(1 + R_{T+(l-1)n}^{(n)}\right) \right\} \right)
\end{aligned}
$$

using the approximation $\ln (1 + R) \approx r$

$$
\begin{aligned}
W_{T+H} &= \omega \exp \left(\frac{1}{s} \left(r_T^{(m)} + \ldots + r_{T+(s-1)m}^{(m)} \right) \right) + (1 - \omega) \exp \left(\frac{1}{l} \left(r_T^{(n)} + \ldots + r_{T+(l-1)n}^{(n)} \right) \right) \\
W_{T+H} &= \omega \exp \left(\frac{1}{s} \sum_{i=1}^{s} r_{T+(i-1)m}^{(m)} \right) + (1 - \omega) \exp \left(\frac{1}{l} \sum_{i=1}^{l} r_{T+(i-1)n}^{(n)} \right)
\end{aligned}
$$

.6 Multivariate Normal Distribution and Expected Utility

Here we show that the expected value of a non-linear variable is a function of both its mean and variance[1], and not solely of its mean.

If $\mathbf{X}$ is a $(q \times 1)$ vector of variables, such that $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, where $E(\mathbf{X}) = \boldsymbol{\mu}$ is a vector of means with the components $(E(\mathbf{X}_1), E(\mathbf{X}_2), ..., E(\mathbf{X}_q))' = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_q)'$, and $\boldsymbol{\Sigma}$ is the $(q \times q)$ covariance matrix with variances along the diagonal and covariances on the off diagonals. The random vector $\mathbf{X}$ is defined as having a multivariate normal distribution if and only if

$$\mathbf{c}'\mathbf{X} = c_1 \mathbf{X}_1 + ... + c_q \mathbf{X}_q$$

the above linear function is normal for all $\mathbf{c}$, given that $\mathbf{c}' = (c_1, c_2, ..., c_q)$. Further, it can be shown that from the moment generating function that the mean of the above non-linear transformation of $\mathbf{X}$ is a function of both the mean and the variance of $\mathbf{X}$

$$E(\exp(t'X)) = \exp\left(t'\mu + \frac{1}{2}t'\Sigma t\right)$$

In terms of asset returns, if $\mathbf{X}$ contains the variables that determine W_{T+H}, i.e. $r_{T+h}^{(n)}$ and $r_{T+h}^{(m)}$ for $h = 1, ..., H$, then the investor requires the variances and covariances of these forecast variables at each step ahead as well as their means. Thus he considers the uncertainty/risk about the return as well as the projected return itself. In essence, given the non-linear nature of the components of W_{T+H} decisions can not be made using the point forecasts of r_{T+h} only, but the entire joint probability distribution of the forecast values of r_{T+h} for $h = 1, ..., H$ are required to evaluate $E_T(\upsilon(W_{T+H}(\omega)) \mid \Omega_T)$.

---

[1] Reference made to notes written by Dr Gwenda Lewis, Statistics Department, University of New England, Armidale, New South Wales, Australia, http://turing.une.edu.au/~stat354/notes/node32.html.

.7 Using Stochastic Simulation to compute Density Forecasts based on the VAR Model

Here we describe how through stochastic simulation techniques an estimate of the probability density function of the forecasts can be obtained. The estimation procedure is discussed firstly by considering how the probability forecasts are calculated for given values of the parameters, and then by taking into account parameter uncertainty.

These methods are used in both Chapters 4 and 5. In Chapter 4 we consider the allocation between a n- and m-month T-bill, so require forecasts of $r_t^{(n)}$ and $r_t^{(m)}$. In Chapter 5 we consider the allocation between a 1-month bill and the Stock Index, so require forecasts of $r_t^{(1)}$ and r_t^s. In this appendix we will use the general notation r_t to denote the return the investor is interested in forecasting, which is $r_t^{(n)}$ and $r_t^{(m)}$ in Chapter 4, and $r_t^{(1)}$ and r_t^s in Chapter 5.

From equation (4.3), we can denote the maximum likelihood estimates of the model parameters $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\mu}}, \widehat{\mathbf{B}}_i, \widehat{\boldsymbol{\Sigma}}\right)$, for $i = 1$ to p. In the absence of parameter uncertainty, the investor assumes there is no uncertainty about the model parameters and they are fixed at the estimated values. Then the model is iterated forward to produce the point estimates of the h-step ahead forecasts, conditional on the observed data $\mathbf{X}_T$ and the estimated parameter values $\widehat{\boldsymbol{\theta}}$

$$\hat{\mathbf{x}}_{T+h} = \widehat{\boldsymbol{\mu}} + \sum_{i=1}^{p} \widehat{\mathbf{B}}_i \hat{\mathbf{x}}_{T+h-i} \tag{5}$$

for $h = 1, 2, ..., H, ...$ Using the initial values of the variables $\mathbf{x}_T, \mathbf{x}_{T-1}, ..., \mathbf{x}_{T-p+1}$, these forecasts are produced recursively.

First considering stochastic uncertainty only, ignoring parameter uncertainty, the forecast values of the variables $\mathbf{x}_{T+h}$ can be computed using stochastic simulations,

providing an estimate of $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$ from

$$\mathbf{x}_{T+h}^{(\widetilde{r})} = \widehat{\boldsymbol{\mu}} + \sum_{i=1}^{p} \widehat{\mathbf{B}}_i \widehat{\mathbf{x}}_{T+h-i} + \boldsymbol{\epsilon}_{T+h}^{(\widetilde{r})} \tag{6}$$

where $\mathbf{x}_{T+h}$ is the h-step ahead forecast. Given that H is the end of the investment period the investor is concerned with forecasts from $h = 1$ to H, $h > H$ step ahead forecasts can be generated but are not required here. Further, let $\widetilde{R}$ denote the total number of replications of the above simulation, $\widetilde{r} = 1$ to $\widetilde{R}$ and gives the $\widetilde{r}^{th}$ replication. For current and past values of $\mathbf{x}$, the actual values are used such that $\mathbf{x}_{T+h-i}^{(\widetilde{r})} = \mathbf{x}_{T+h-i}$, e.g. $\mathbf{x}_T^{(\widetilde{r})} = \mathbf{x}_T, \mathbf{x}_{T-1}^{(\widetilde{r})} = \mathbf{x}_{T-1}...$ for each replication.

To generate forecasts in the presence of parameter uncertainty the Monte Carlo procedure is used. First, the (in-sample) past values of $\mathbf{x}_t$ are simulated $\widetilde{H}$ times, i.e. simulate $\widetilde{H}$ 'histories' of $\mathbf{x}_t$, $t = 1, 2, ..., T$, denoted $\mathbf{x}_t^{(\widetilde{h})}$, $\widetilde{h} = 1, 2, ..., \widetilde{H}$.[2] Where

$$\mathbf{x}_t^{(\widetilde{h})} = \widehat{\boldsymbol{\mu}} + \sum_{i=1}^{p} \widehat{\mathbf{B}}_i \widehat{\mathbf{x}}_{t-i}^{(\widetilde{h})} + \boldsymbol{\epsilon}_t^{(\widetilde{h})} \tag{7}$$

the actual realised values of $\mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-p}$ are used for initial values, together with the estimated model parameters $\widehat{\boldsymbol{\theta}}$ obtained using the actual observed data.

With the $\widetilde{H}$ simulated histories for $\mathbf{x}_t$ i.e. $\mathbf{x}_1^{(\widetilde{h})}, \mathbf{x}_2^{(\widetilde{h})}, ..., \mathbf{x}_T^{(\widetilde{h})}$, such that for each past value of $\mathbf{x}$ there are $\widetilde{H}$ possible values, it is now possible to estimate the VAR(p) model given by equation (4.3) $\widetilde{H}$ times, yielding $\widetilde{H}$ sets of ML parameter estimates $\widehat{\boldsymbol{\beta}}^{(\widetilde{h})}, \widehat{\mathbf{B}}_i^{(\widetilde{h})}, \boldsymbol{\epsilon}_t^{(\widetilde{h})}$ and $\boldsymbol{\Sigma}^{(\widetilde{h})}$, one set of estimates for each Monte Carlo replication, where $i = 1, 2, ..., p$.

For each Monte Carlo replication, compute h-step ahead point forecasts of $\mathbf{x}_T$, where

---

[2] Note that $'\widetilde{r}'$ refers to the number of 'futures' generated in the simulation, whereas $'r'$ refers to the asset return. Equally $'\widetilde{h}'$ refers to the 'histories' generated and $'h'$ refers to the step ahead forecasts.

$\widetilde{R}$ replications of these forecasts are generated, i.e. for each of the $\widetilde{H}$ generated histories simulate $\widetilde{R}$ futures

$$\mathbf{x}_{T+h}^{(\widetilde{h},\widetilde{r})} = \widehat{\boldsymbol{\mu}}^{(\widetilde{h})} + \sum_{i=1}^{p} \widehat{\mathbf{B}}_i^{(\widetilde{h})} \widehat{\mathbf{x}}_{T+h-i}^{(\widetilde{h},\widetilde{r})} + \boldsymbol{\epsilon}_{T+h}^{(\widetilde{h},\widetilde{r})} \tag{8}$$

for $h = 1, 2, ..., H$; $\widetilde{r} = 1, 2, ..., \widetilde{R}$ and $\widetilde{h} = 1, 2, ..., \widetilde{H}$. Note that '$h$' refers to the horizon and '$\widetilde{h}$' to the number of histories generated, and $S = \widetilde{H} \times \widetilde{R} =$ total number of simulations.

The $\boldsymbol{\epsilon}_{T+h}^{(\widetilde{r})}, \boldsymbol{\epsilon}_t^{(\widetilde{h})}$ and $\boldsymbol{\epsilon}_{T+h}^{(\widetilde{h},\widetilde{r})}$'s can be drawn using either parametric or non-parametric methods, see GLPS (2006, pp. 166-168) for further details. Here parametric methods are utilised where the errors are assumed to be $i.i.d.N\left(0, \boldsymbol{\Sigma}\right)$ serially uncorrelated white noise errors, see the "Simulating Errors" section below.

These simulations provide an estimate of the predictive densities $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$ in the case where parameter uncertainty is ignored and $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$ when it is considered, so it is now possible to evaluate $E_T\left(\upsilon\left(W_{T+H}\right) \mid \Omega_T\right)$ for a range of portfolio weights ω. That is, in practice $\upsilon\left(W_{T+H}\left(\omega\right)\right)$ is computed $\widetilde{R}$ times for each value of ω, then the mean across these $\widetilde{R}$ replications is calculated, from which the investor chooses the weight ω that maximises the expected utility $E_T\upsilon\left(W_{T+H}\left(\omega\right)\right)$. Here ω takes values $0, 0.01, 0.02, ..., 0.99, 1$.

Computing the Predictive Densities

Here we describe how we compute the predictive densities using the method described in GL and GLPS, all of the steps detailed below are conducted for each of the models in Chapters 4 and 5. In the case where only stochastic uncertainty is considered the predictive density is $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T, \widehat{\boldsymbol{\theta}}\right)$ and where both stochastic and parameter uncertainty are accounted for the predictive density is $P\left(\mathbf{X}_{T+1,H} \mid \mathbf{X}_T\right)$.

Predictive Density under Stochastic Uncertainty only

1. Using the estimated model parameters $\widehat{\boldsymbol{\theta}}$, forecasts of the returns are generated $r_{T+h}^{(\widetilde{r})}$ for $h = 1, ..., H$ and $\widetilde{r} = 1, ..., \widetilde{R}$, where $\widetilde{R} = 50,000$.

2. From the above forecasts, values of $W_{T+H}^{(\widetilde{r},\omega)}$ can be calculated for each replication, with $\omega = 0, ..., 1$ increasing in steps of 0.01. So for each value of H, we have $\widetilde{R} \times 101$ values of $W_{T+H}^{(\widetilde{r},\omega)}$, where $H = 3, 6, 12, 18$ and 24 months.

3. These wealths are then used to calculate utility as given by the CRRA definition, $v_{T+H}^{(\widetilde{r},\omega,A)}$ where $A = 2, 5$ and 10. For the given values of ω, A and H the expected utility is given by averaging across the replications as follows

$$E_T v \left(W_{T+H}\right) = \frac{1}{\widetilde{R}} \sum_{\widetilde{r}=1}^{\widetilde{R}} v_{T+H}^{(\widetilde{r},\omega,A)}$$

4. Hence for a given investment horizon H and level of risk A, the investor selects that portfolio weight which maximises expected utility.

Predictive Density under Stochastic & Parameter Uncertainty

1. Using the estimated model parameters $\widehat{\boldsymbol{\theta}}$, in-sample values of $\mathbf{x}_T$ are simulated $\widetilde{H}$ times, for $t = 1, ..., T$ and $\widetilde{h} = 1, ..., \widetilde{H}..$, where $\widetilde{R} = 1000$ and $\widetilde{H} = 2000$.

2. Using each of these $\widetilde{H}$ 'histories' of $\mathbf{x}_T$, estimate the model. This yields $\widetilde{H}$ sets of parameter estimates $\widehat{\boldsymbol{\theta}}^{(\widetilde{h})}$, one for each history generated.

3. For each history compute $\widetilde{R}$ replications of the h-step ahead point forecasts of $\mathbf{x}_T$, where $\widetilde{r} = 1, ..., \widetilde{R}$.

4. Repeat steps 2-4 from the stochastic uncertainty only method above, for each history and its corresponding set of $\widetilde{R}$ simulated futures. Hence $\widetilde{H}$ sets of $E_T v \left(W_{T+H}\right)$ are calculated for given values of ω, A and H. The aim is to select the portfolio weight that maximises $E_T v \left(W_{T+H}\right)$ for a given investment horizon H and level of risk A, for each history $\widetilde{H}$.

Simulated Errors

As described by GLPS (2006, pp. 166) and Cuthbertson and Nitzsche (2004, pp. 648), the method used to generate these simulated errors is as follows:

1. Generate $q \times h$ draws from the assumed $i.i.d.N(0, \mathbf{I}_q)$ distribution for each $\widetilde{h}$ and $\widetilde{r}$. To give $e_{T+h}^{(\widetilde{h},\widetilde{r})}$ for $h = 1, ..., H; \widetilde{r} = 1, ..., \widetilde{R}$ and $\widetilde{h} = 1, ..., \widetilde{H}$. So for each $\widetilde{h}$ and $\widetilde{r}$ generate e_{T+i} as $i.i.d.N(0, \mathbf{I}_q)$.

2. Calculate $\boldsymbol{\epsilon}_{T+h}^{(\widetilde{h},\widetilde{r})}$ for $i = 1, ..., h$, where $\boldsymbol{\epsilon}_{T+h}^{(\widetilde{h},\widetilde{r})} = e_{T+h}^{(\widetilde{h},\widetilde{r})}.\widehat{\mathbf{P}}^{(\widetilde{h})}$ and $\widehat{\mathbf{P}}^{(\widetilde{h})}$ is the upper triangle choleski factor of $\widehat{\boldsymbol{\Sigma}}^{(\widetilde{h})}$. Further, $\widehat{\boldsymbol{\Sigma}}^{(\widetilde{h})} = \widehat{\mathbf{P}}^{(\widetilde{h})\prime}\widehat{\mathbf{P}}^{(\widetilde{h})}$ and $\widehat{\boldsymbol{\Sigma}}^{(\widetilde{h})}$ is an estimate of $\boldsymbol{\Sigma}$ computed in the $\widetilde{h}^{th}$ replication of the Monte Carlo. In the absence of parameter uncertainty $\boldsymbol{\epsilon}_{T+h}^{(\widetilde{r})} = e_{T+h}^{(\widetilde{r})}.\widehat{\mathbf{P}}'$, here $\widehat{\mathbf{P}}$ is the upper triangle choleski factor of $\widehat{\boldsymbol{\Sigma}}$.

.8 Properties of Stationary & Non-Stationary Processes

Here the properties of when the yields, $r_T^{(n)}$ are modelled as a non-stationary process, e.g. as under a Random Walk model, in comparison to when they are modelled as being stationary, e.g. as under a Autoregressive model of order 1, AR(1) model are described. We use r_T to denote $r_T^{(n)}$ for ease.

Random Walk with Drift Model (RW)

Here the bill returns are considered to be non-stationary and given by

$$r_T = \mu + r_{T-1} + e_t \tag{9}$$

where $e_t \sim i.i.d.\,(0, \sigma^2)$, through iteration

$$
\begin{aligned}
r_{T+1} &= \mu + r_T + e_{T+1} \\
r_{T+2} &= \mu + r_{T+1} + e_{T+2} = 2\mu + r_T + e_{T+1} + e_{T+2} \\
\therefore \quad r_{T+H} &= H\mu + r_T + \sum_{i=1}^{H} e_{T+i} \tag{10}
\end{aligned}
$$

The mean and variance of returns are

$$
\begin{aligned}
E_T\,(r_{T+H}) &= H\mu + r_T \tag{11} \\
Var\,(r_{T+H}) &= H\sigma^2 \tag{12}
\end{aligned}
$$

Note if it is assumed that the initial value of the series is zero, then $E_T\,(r_{T+H}) = H\mu$ as is often reported, see Cuthbertson and Nitzsche (2004, pp. 37). But here we assume r_T takes a value which is know at T. The cumulative returns over T to $T + H$, with

its corresponding mean and variance is given by

$$r_{T,T+H} = r_T+r_{T+1}+...+r_{T+H}$$

$$E_T\left(r_{T,T+H}\right) = r_T+E_T\left(r_{T+1}\right)+...+E_T\left(r_{T+H}\right)$$

$$= r_T+\mu+r_T+...+H\mu+r_T$$

$$= \left(1+H\right)r_T+\left(1+2+...+H\right)\mu \tag{13}$$

$$Var\left(r_{T,T+H}\right) = \sigma^2+2\sigma^2+...+H\sigma^2$$

$$= \left(1+2+...+H\right)\sigma^2 \tag{14}$$

AR(1) with a Constant Model

If the returns are assumed to be stationary and are modelled as being so by an AR(1) with a constant

$$r_T= \mu + \rho r_{T-1}+e_T \tag{15}$$

where $\mid\rho\mid<1$ and $e_t \sim i.i.d.\left(0,\sigma^2\right)$, in lag operator form

$$(1-\rho L)\,r_T = \mu+e_T$$

$$r_T = \left(\frac{1}{1-\rho L}\right)(\mu+e_T)$$

$$r_T = \left(1+\rho L+\rho^2 L^2+...\right)(\mu+e_T)$$

$$r_T = \left(1+\rho+\rho^2+...\right)\mu+\left(e_T+\rho e_{T-1}+\rho^2 e_{T-2}+...\right) \tag{16}$$

where $\left(\frac{1}{1-\rho L}\right)=(1+\rho L+\rho^2 L^2+...)$. Since μ is a constant the lag operator has no

effect. The unconditional mean is

$$E_T(r_T) = \left(1 + \rho + \rho^2 + ...\right)\mu \tag{17}$$

this is a geometric series that converges to $\frac{\mu}{1-\rho}$ in the long run. The conditional mean is found by taking expectations of equation (15)

$$E_T(r_T) = \mu + \rho r_{T-1} \tag{18}$$

the mean is conditional on past values and evolves with the series. From equation (16) the variance is

$$\begin{aligned}
Var(r_T) &= Var\left(e_T + \rho e_{T-1} + \rho^2 e_{T-2} + ...\right) \\
&= \sigma^2 + \rho^2\sigma^2 + \rho^4\sigma^2 + ... \\
&= \left(1 + \rho^2 + \rho^4 + ...\right)\sigma^2
\end{aligned} \tag{19}$$

which in the long run converges to $\frac{\sigma^2}{1-\rho^2}$. Using equation (15) and through continuous substitution of the lagged return on the RHS, an expression for r_{T+H} is

$$\begin{aligned}
r_{T+1} &= \mu + \rho r_T + e_{T+1} \\
r_{T+2} &= \mu + \rho r_{T+1} + e_{T+2} = \mu + \rho\mu + \rho^2 r_T + \rho e_{T+1} + e_{T+2} \\
\therefore \quad r_{T+H} &= \left(1 + \rho + \rho^2 + ... + \rho^{H-1}\right)\mu + \rho^H r_T + \sum_{i=1}^{H}\rho^{H-i}e_{T+i}
\end{aligned} \tag{20}$$

From equation (20)

$$E_T \left(r_{T+H} \right) = \left(1 + \rho + \rho^2 + ... + \rho^{H-1} \right) \mu + \rho^H r_T \qquad (21)$$

$$Var \left(r_{T+H} \right) = \left(1 + \rho^2 + \rho^4 + ... + \rho^{2H-2} \right) \sigma^2 \qquad (22)$$

as $H \longrightarrow \infty$ then the variance of the forecasts converges i.e. $Var \left(r_{T+H} \right) = \frac{\sigma^2}{1-\rho^2}$. The

mean and variance of the cumulative returns, using equation (20) is

$$
\begin{aligned}
r_{T,T+H} &= r_T + r_{T+1} + ... + r_{T+H} \\
E_T \left(r_{T,T+H} \right) &= r_T + E_T \left(r_{T+1} \right) + ... + E_T \left(r_{T+H} \right) \\
&= r_T + \left(\mu + \rho r_T \right) + ... + \left(1 + \rho + ... + \rho^{H-1} \right) \mu + \rho^H r_T \\
&= \left(1 + \rho + .. + \rho^H \right) r_T \\
&\quad + \left[1 + \left(1 + \rho \right) + .. + \left(1 + \rho + .. + \rho^{H-1} \right) \right] \mu \qquad (23) \\
Var \left(r_{T,T+H} \right) &= Var(r_T) + Var \left(r_{T+1} \right) + ... + Var \left(r_{T+H} \right) \\
&= 0 + \sigma^2 + \left(1 + \rho^2 \right) \sigma^2 + ... + \left(1 + \rho^2 + \rho^4 + ... + \rho^{2H-2} \right) \sigma^2 \\
&= \left[1 + \left(1 + \rho^2 \right) + ... + \left(1 + \rho^2 + \rho^4 + ... + \rho^{2H-2} \right) \right] \sigma^2 \qquad (24)
\end{aligned}
$$

The computed mean μ and variance σ^2 will be different for each model, so we denote

them $\mu_{RW,}$ σ^2_{RW} and $\mu_{AR,}$ σ^2_{AR} for the RW and AR models respectively. Comparing

the mean of the cumulative returns under both models, from above it can be seen that

the mean of cumulative returns under the AR model grows slower than that under the

RW. Using a simple numerical example, if $H = 4$ and $\rho = 0.9$, the mean under the

AR model is

$$\left(1 + \rho + ... + \rho^4\right) r_T + \left(1 + (1 + \rho) + ... + \left(1 + \rho + \rho^2 + \rho^3\right)\right) \mu_{AR} = 4.1 r_T + 9.05 \mu_{AR}$$

RW model is

$$\left(1 + H\right) r_T + (1 + 2 + ... + H) \mu = 5 r_T + 10 \mu_{RW}$$

Now comparing how the variance of the returns evolves over T to $T + H$ under each model. Under the RW the variance of the returns continues to grow linearly with H. However, the variance under the AR model is growing less than linearly. Such that, as $H \longrightarrow \infty$ the variance under the $RW = \infty \sigma_{RW}^2$, but the variance under the $AR = \frac{\sigma_{AR}^2}{1 - \rho^2}$ i.e. it converges. This will be translated through to the variance of the cumulative returns, with that under the AR model growing slower than the variance under the RW, as illustrated in the table below.

	$Var\left(r_{T+1}\right)$	$Var\left(r_{T+2}\right)$	$Var\left(r_{T+3}\right)$	...	$Var\left(r_{T+H}\right)$
RW	σ_{RW}^2	$2\sigma_{RW}^2$	$3\sigma_{RW}^2$	...	$H\sigma_{RW}^2$
AR	σ_{AR}^2	$\left(1 + \rho^2\right)\sigma_{AR}^2$	$\left(1 + \rho^2 + \rho^4\right)\sigma_{AR}^2$	...	$\left(1 + \rho^2 + ... + \rho^{2H-2}\right)\sigma_{AR}^2$

Note we are not making a direct comparison between the variances under each model (for a particular point in time or over a given investment horizon), since the computed variances from the RW model will differ from those under the AR model. We are merely highlighting how under the RW (where the returns are treated as being non-stationary) the variance of the returns grows linearly with H. However, when the returns are assumed to be stationary and modelled using an AR(1) the variance of the

returns not only grow less than linearly with H, but in the long run converges to some long-run value.

For stationary processes, the variance in the long run converges to some long-run value, so if the yields were modelled by a stationary VAR(p) then the variance of the returns will tend to a constant σ_∞^2 in the long run. We can then define

$$\alpha_i \sigma_\infty^2 = \sigma_{T+i}^2$$

where α_i is some unknown coefficient, that when multiplied by the long-run variance gives the variance at a particular point in time $T + i$. In the short run the variance about the returns may grow faster or slower than linearly depending upon the value of α_i. If $\alpha_i > 1$ then the variance will grow faster than linearly, if $\alpha_i < 1$ then the variance will grow slower than linearly. Only in the long run, if the VAR parameters are stationary then $\alpha_i = 1$, such that $\sigma_{T+i}^2 = \sigma_\infty^2$ i.e. the variance converges to its long-run value. Hence, in the long run the variance under the stationary VAR model will be less than that under the Random Walk model.

These derivations for the AR(1) model illustrate the properties of a stationary process, for comparison with a non-stationary process such as the RW. A stationary VAR would require more complex conditions for the parameters to be satisfied for stationarity. But in the long run the variance for a stationary VAR model would also converge to some long-run value. Consequently, in the long run the variance of the forecast distribution also grows less than linearly.

Bibliography

[1] Abhyankar, A., Sarno, L. & Valente, G. 2005, "Exchange rates and fundamentals: evidence on the economic value of predictability", Journal of International Economics, vol. 66, no. 2, pp. 325-348.

[2] Ang, A. & Bekaert, G. 2002, "Regime Switches in Interest Rates", Journal of Business & Economic Statistics, vol. 20, no. 2, pp. 163-182.

[3] Ang, A. & Bekaert, G. 2007, "Stock Return Predictability: Is it There?", Review of Financial Studies, vol. 20, no. 3, pp. 651-707.

[4] Ang, A. & Piazzesi, M. 2003, "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables", Journal of Monetary Economics, vol. 50, no. 4, pp. 745-787.

[5] Ang, A., Piazzesi, M. & Wei, M. 2006, "What does the yield curve tell us about GDP growth?", Journal of Econometrics, vol. 131, no. 1, pp. 359-403.

[6] Asteriou, D. & Hall, S. G. 2007, "Applied Econometrics", Palgrave MacMillan.

[7] Avramov, D. 2002, "Stock return predictability and model uncertainty", Journal of Financial Economics, vol. 64, no. 3, pp. 423-458.

[8] Bansal, R. & Zhou, H. 2002, "Term Structure of Interest Rates with Regime Shifts", The Journal of Finance, vol. 57, no. 5, pp. 1997-2043.

[9] Barberis, N. 2000, "Investing for the Long Run when Returns Are Predictable", The Journal of Finance, vol. 55, no. 1, pp. 225-264.

[10] Bean, C. 2003, "Inflation Targeting: The UK Experience", Bank of England Quarterly Bulletin, vol. 43, no. 4, pp. 479-494.

[11] Bean, C. 2004, "Inflation Targeting: The UK Experience", Perspektiven der Wirtschaftspolitik, vol. 5, no. 4, pp. 405-421.

[12] Bekaert, G. & Hodrick, R.J. 2001, "Expectations Hypotheses Tests", The Journal of Finance, vol. 56, no. 4, Papers and Proceedings of the Sixty-First Annual Meeting of the American Finance Association, New Orleans, Louisiana, pp. 1357-1394.

[13] Bekaert, G., Hodrick, R.J. & Marshall, D.A. 1997, "On biases in tests of the expectations hypothesis of the term structure of interest rates", Journal of Financial Economics, vol. 44, no. 3, pp. 309-348.

[14] Bidarkota, P.V. 1998, "The comparative forecast performance of univariate and multivariate models: an application to real interest rate forecasting", International Journal of Forecasting, vol. 14, no. 4, pp. 457-468.

[15] Boero, G. & Torricelli, C. 2002, "The information in the term structure of German interest rates", European Journal of Finance, vol. 8, no. 1, pp. 21-45.

[16] Boudry, W. & Gray, P. 2003, "Assessing the Economic Significance of Return Predictability: A Research Note", Journal of Business Finance & Accounting, vol. 30, no. 9, pp. 1305-1326.

[17] Britton, E., Fisher, P. & Whitley, J. 1998, "The Inflation Report projections: Understanding the fan chart", Bank of England Quarterly Bulletin, vol. 38, no. 1, pp. 30.

[18] Brooks, C. 2005, "Introductory Econometrics for Finance", Cambridge University Press.

[19] Brooks, C. & Persand, G. 2003, "Volatility Forecasting for Risk Management", Journal of Forecasting, vol. 22, no. 1, pp. 1-22.

[20] Brooks, C. & Rew, A.G. 2002, "Testing for non-stationarity and cointegration allowing for the possibility of a structural break: an application to EuroSterling interest rates", Economic Modelling, vol. 19, no. 1, pp. 65-90.

[21] Brooks, C. & Skinner, F. 2000, "What will be the risk-free rate and benchmark yield curve following European monetary union?", Applied Financial Economics, vol. 10, no. 1, pp. 59-69.

[22] Brooks, C. & Tsolacos, S. 2001, "Linkages between property asset returns and interest rates: evidence for the UK", Applied Economics, vol. 33, no. 6, pp. 711-719.

[23] Brown, C.R., Cyree, K.B., Griffiths, M.D. & Winters, D.B. 2008, "Further analysis of the expectations hypothesis using very short-term rates", Journal of Banking & Finance, vol. 32, no. 4, pp. 600-613.

[24] Campbell, J.Y. 1986, "A Defense of Traditional Hypotheses about the Term Structure of Interest Rates", The Journal of Finance, vol. 41, no. 1, pp. 183-193.

[25] Campbell, J.Y. 1987, "Stock returns and the term structure", Journal of Financial Economics, vol. 18, no. 2, pp. 373-399.

[26] Campbell, J. Y. 1991, "A Variance Decomposition for Stock Returns", Economic Journal, vol. 101, no. 405, pp. 157-179.

[27] Campbell, J.Y. 1995, "Some Lessons from the Yield Curve", The Journal of Economic Perspectives, vol. 9, no. 3, pp. 129-152.

[28] Campbell, J.Y. & Shiller, R.J. 1987, "Cointegration and Tests of Present Value Models", Journal of Political Economy, vol. 95, no. 5, pp. 1062-1088.

[29] Campbell, J.Y. & Shiller, R.J. 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View", The Review of Economic Studies, vol. 58, no. 3, Special Issue: The Econometrics of Financial Markets, pp. 495-514.

[30] Campbell, J. Y. & Viceira, L. M. 2003, "Strategic Asset Allocation Portfolio Choice for Long-Term Investors", Oxford University Press.

[31] Carriero, A., Favero, C.A. & Kaminska, I. 2006, "Financial factors, macroeconomic information and the Expectations Theory of the term structure of interest rates", Journal of Econometrics, vol. 131, no. 1-2, pp. 339-358.

[32] Chadha, J.S. & Nolan, C. 2001, "Inflation targeting, transparency and interest rate volatility: Ditching "monetary mystique" in the U.K.", Journal of Macroeconomics, vol. 23, no. 3, pp. 349-366.

[33] Chordia, T., Sarkar, A. & Subrahmanyam, A. 2005, "An Empirical Analysis of Stock and Bond Market Liquidity", Review of Financial Studies, vol. 18, no. 1, pp. 85-129.

[34] Christiansen, C., Engsted, T., Jakobsen, S. & Tanggaard, C. 2003, "An Empirical Study of the Term Structure of Interest Rates in Denmark, 1993-2002," Finance Working Papers 03-2, University of Aarhus, Aarhus School of Business, Department of Business Studies.

[35] Christoffersen, P.F. 1998, "Evaluating Interval Forecasts", International Economic Review, vol. 39, no. 4, pp. 841-862.

[36] Clarida, R.H., Sarno, L., Taylor, M.P. & Valente, G. 2003, "The out-of-sample success of term structure models as exchange rate predictors: a step beyond", Journal of International Economics, vol. 60, no. 1, pp. 61-83.

[37] Clarida, R.H., Sarno, L., Taylor, M.P. & Valente, G. 2006, "The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates", The Journal of Business, vol. 79, no. 3, pp. 1193-1224.

[38] Clemen, R.T. 1986, "Linear Constraints and the Efficiency of Combined Forecasts", Journal of Forecasting, vol. 5, no. 1, pp. 31-38.

[39] Clemen, R.T. 1989, "Combining forecasts: A review and annotated bibliography", International Journal of Forecasting, vol. 5, no. 4, pp. 559-583.

[40] Clements, M.P. 2004, "Evaluating the Bank of England Density Forecasts of Inflation", Economic Journal, vol. 114, no. 498, pp. 844-866.

[41] Clements, M. P. 2006, "Evaluating the survey of professional forecasters probability distributions of expected inflation based on derived event probability forecasts", Empirical Economics, vol. 31, no. 1, pp. 49-64.

[42] Clements, M. P. & Hendry, D. F. 1998, "Forecasting Economic Time Series", Cambridge University Press.

[43] Clements, M. P. & Hendry, D. F. 2004, "A Companion to Economic Forecasting", Blackwell Publishing.

[44] Clements, M.P. & Hendry, D.F. 2005a, "Evaluating a Model by Forecast Performance", Oxford Bulletin of Economics & Statistics, vol. 67, pp. 931-956.

[45] Clements, M.P. & Hendry, D.F. 2005b, "Guest Editors' Introduction: Information in Economic Forecasting", Oxford Bulletin of Economics & Statistics, vol. 67, pp. 713-753.

[46] Clements, M.P. & Smith, J. 2000, "Evaluating the Forecast Densities of Linear and Non-linear Models", Journal of Forecasting, vol. 19, no. 4, pp. 255.

[47] Clements, M.P. & Smith, J. 2002, "Evaluating multivariate forecast densities: a comparison of two approaches", International Journal of Forecasting, vol. 18, no. 3, pp. 397-407.

[48] Clements, M.P. & Taylor, N. 2003, "Evaluating Interval Forecasts of High-Frequency Financial Data", Journal of Applied Econometrics, vol. 18, no. 4, pp. 445-456.

[49] Cooray, A. 2003, "A test of the expectations hypothesis of the term structure of interest rates for Sri Lanka", Applied Economics, vol. 35, no. 17, pp. 1819-1827.

[50] Corradi, V. & Swason, N. R. 2006, "Predictive Density Evaluation", Chapter 5, in "Handbook of Forecasting", North-Holland, volume 1, edited by Elliot, G., Granger, C. W. J. & Timmermann, A..

[51] Cox, J.C., Ingersoll, J.E. & Ross, S.A. 1985, "A Theory of the Term Structure of Interest Rates", Econometrica, vol. 53, no. 2, pp. 385-407.

[52] Cuthbertson, K. 1996, "The Expectations Hypothesis of the Term Structure: The UK Interbank Market", The Economic Journal, vol. 106, no. 436, pp. 578-592.

[53] Cuthbertson, K. & Bredin, D. 2000a, "The Expectations Hypothesis of the Term Structure: The Case of Ireland", Economic and Social Review, vol. 31, no. 3, pp. 267-281.

[54] Cuthbertson, K., Hayes, S. & Nitzsche, D. 1998, "Interest rates in Germany and the UK: Cointegration and error correction models", Manchester School (14636786), vol. 66, no. 1, pp. 27.

[55] Cuthbertson, K., Hayes, S. & Nitzsche, D. 1996, "The Behaviour of Certificate of Deposit Rates in the UK", Oxford Economic Papers, vol. 48, no. 3, pp. 397-414.

[56] Cuthbertson, K., Hayes, S. & Nitzsche, D. 2000b, "Are German money market rates well behaved?", Journal of Economic Dynamics and Control, vol. 24, no. 3, pp. 347-360.

[57] Cuthbertson, K. & Nitzsche, D. 2003, "Long rates, risk premia and the overreaction hypothesis", Economic Modelling, vol. 20, no. 2, pp. 417.

[58] Cuthbertson, K. & Nitzsche, D. 2004, "Quantitative Financial Economics: Stocks, Bonds and Foreign Exchange ", John Wiley & Sons.

[59] Dai, Q. & Singleton, K.J. 2000, "Specification Analysis of Affine Term Structure Models", The Journal of Finance, vol. 55, no. 5, pp. 1943-1978.

[60] Dai, Q., Singleton, K.J. & Yang, W. 2007, "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields", Review of Financial Studies, vol. 20, no. 5, pp. 1669-1706.

[61] Deaves, R. 1996, "Forecasting Canadian Short-Term Interest Rates", The Canadian Journal of Economics / Revue canadienne d'Economique, vol. 29, no. 3, pp. 615-634.

[62] Della Corte, P., Sarno, L. & Thornton, D.L. 2008, "The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value", Journal of Financial Economics, vol. 89, no. 1, pp. 158-174.

[63] De Menezes, L.M., Bunn, D.W. & Taylor, J.W. 2000, "Review of guidelines for the use of combined forecasts", European Journal of Operational Research, vol. 120, no. 1, pp. 190-204.

[64] De Pooter, M., Ravazzolo, F. & van Dijk, D. 2007, "Predicting the Term Structure of Interest Rates, Incorporating parameter uncertainty, model uncertainty and macroeconomic information", Working Paper, Econometric Institute and Tinbergen Institute, Erasmus University Rotterdam.

[65] Dewachter, H. & Lyrio, M. 2006, "Macro Factors and the Term Structure of Interest Rates", Journal of Money, Credit and Banking, vol. 38, no. 1, pp. 119-140.

[66] Diebold, F. X. 2004, "Elements of Forecasting", 3rd Edition, South-Western.

[67] Diebold, F.X., Gunther, T.A. & Tay, A.S. 1998, "Evaluating Density Forecasts with Applications to Financial Risk Management", International Economic Review, vol. 39, no. 4, Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance, pp. 863-883.

[68] Diebold, F.X., Hahn, J. & Tay, A.S. 1999, "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange", Review of Economics & Statistics, vol. 81, no. 4, pp. 661.

[69] Diebold, F.X. & Li, C. 2006, "Forecasting the term structure of government bond yields", Journal of Econometrics, vol. 130, no. 2, pp. 337-364.

[70] Diebold, F.X. & Lopez, J.A. 1996, "Forecast evaluation and combination" in Handbook of Statistics, ed. G.S. Maddala and C.R. Rao, Elsevier, pp. 241-268.

[71] Diebold, F.X. & Lopez, J.A. 1996, Forecast Evaluation and Combination, National Bureau of Economic Research, Inc, NBER Technical Working Paper: 192.

[72] Diebold, F.X. & Mariano, R.S. 1995, "Comparing Predictive Accuracy", Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 253-263.

[73] Diebold, F.X., Rudebusch, G.D. & Boragan Aruoba, S. 2006, "The macroeconomy and the yield curve: a dynamic latent factor approach", Journal of Econometrics, vol. 131, no. 1, pp. 309-338.

[74] Dillen, H. 1997, "A model of the term structure of interest rates in an open economy with regime shifts", Journal of International Money and Finance, vol. 16, no. 5, pp. 795-819.

[75] Domínguez, E. & Novales, A. 2000, "Testing the expectations hypothesis in Eurodeposits", Journal of International Money and Finance, vol. 19, no. 5, pp. 713-736.

[76] Drakos, K. 2002, "A Daily View of the Term Structure Dynamics: some International Evidence", De Economist (0013-063X), vol. 150, no. 1, pp. 41-52.

[77] Dua, P., Raje, N. & Sahoo, S. 2008, "Forecasting Interest Rates in India", Margin: The Journal of Applied Economic Research, vol. 2, no. 1, pp. 1-41.

[78] Duarte, A., Venetis, I.A. & Paya, I. 2005, "Predicting real growth and the probability of recession in the Euro area using the yield spread", International Journal of Forecasting, vol. 21, no. 2, pp. 262-277.

[79] Duffee, G.R. 2002, "Term Premia and Interest Rate Forecasts in Affine Models", The Journal of Finance, vol. 57, no. 1, pp. 405-443.

[80] Elliot, G., Granger, C. W. J. & Timmermann, A. 2006, " Handbook of Forecasting", North-Holland, volume 1.

[81] Engsted, T. 1996, "The predictive power of the money market term structure", International Journal of Forecasting, vol. 12, no. 2, pp. 289-295.

[82] Engsted, T. & Tanggaard, C. 1994a, "A cointegration analysis of Danish zero-coupon bond yields", Applied Financial Economics, vol. 4, no. 4, pp. 265-278.

[83] Engsted, T. & Tanggaard, C. 1994b, "Cointegration and the US term structure", Journal of Banking & Finance, vol. 18, no. 1, pp. 167-181.

[84] Engsted, T. & Tanggaard, C. 1995, "The Predictive Power of Yield Spreads for Future Interest Rates: Evidence from the Danish Term Structure", Scandinavian Journal of Economics, vol. 97, no. 1, pp. 145.

[85] Ericsson, N. R. 2004, "Predictable Uncertainty in Economic Forecasting", Chapter 2, pp. 19-44, in "A Companion to Economic Forecasting" edited by Clements, M. P. & Hendry, D. F..

[86] Estrella, A. & Hardouvelis, G.A. 1991, "The Term Structure as a Predictor of Real Economic Activity", Journal of Finance, vol. 46, no. 2, pp. 555-576.

[87] Estrella, A. & Mishkin, F.S. 1997, "The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank", European Economic Review, vol. 41, no. 7, pp. 1375-1401.

[88] Evans, M.D.D. & Lewis, K.K. 1994, "Do stationary risk premia explain it all?: Evidence from the term structure", Journal of Monetary Economics, vol. 33, no. 2, pp. 285-318.

[89] Fama, E.F. 1984a, "The information in the term structure", Journal of Financial Economics, vol. 13, no. 4, pp. 509-528.

[90] Fama, E.F. 1984b, "Term premiums in bond returns", Journal of Financial Economics, vol. 13, no. 4, pp. 529-546.

[91] Fama, E.F. 1990, "Term-structure forecasts of interest rates, inflation and real returns", Journal of Monetary Economics, vol. 25, no. 1, pp. 59-76.

[92] Fama, E.F. & Bliss, R.R. 1987, "The Information in Long-Maturity Forward Rates", The American Economic Review, vol. 77, no. 4, pp. 680-692.

[93] Fama, E.F. & French, K.R. 1988a, "Dividend yields and expected stock returns", Journal of Financial Economics, vol. 22, no. 1, pp. 3-25.

[94] Fama, E. & French, K. 1988b, "Permanent and Temporary Components of Stock Prices", Journal of Political Economy, vol. 96, no. 2, pp. 246.

[95] Fama, E.F. & French, K.R. 1989, "Business conditions and expected returns on stocks and bonds", Journal of Financial Economics, vol. 25, no. 1, pp. 23-49.

[96] Fang, V. & Lee, V. C. S. 2003, "Testing the Expectations Hypothesis for Interest Rates Term Structure: Some Australian Evidence", V. Kumar et al (Eds): ICCSA 2003, LNCS 2669, pp. 189-198.

[97] Fauvel, Y., Paquet, A. & Zimmermann, C. 1999, "A Survey on Interest Rate Forecasting", CREFE Working Paper 87, University of Quebec.

[98] Fisher, I. 1930, "Theory of Interest", New York: Macmillan.

[99] Froot, K.A. 1989, "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates", The Journal of Finance, vol. 44, no. 2, pp. 283-305.

[100] Garratt, A. & Lee, K. 2009, "Investing under model uncertainty: Decision based evaluation of exchange rate forecasts in the US, UK and Japan", Journal of International Money and Finance, vol. In Press, Corrected Proof.

[101] Garratt, A., Lee, K., Hashem Pesaran, M. & Shin, Y. 2003, "A Long run structural macroeconometric model of the UK", Economic Journal, vol. 113, no. 487, pp. 412-455.

[102] Garratt, A., Lee, K., Pesaran, M.H. & Shin, Y. 2006, "Global and National Macroeconometric Modelling: A Long-Run Structural Approach", Oxford University Press, Oxford.

[103] Ghazali, N.A. & Soo-Wah Low. 2002, "The expectation hypothesis in emerging financial markets: the case of Malaysia", Applied Economics, vol. 34, no. 9, pp. 1147-1156.

[104] Gomez-Biscarri, J. 2009, "The predictive power of the term spread revisited: a change in the sign of the predictive relationship", Applied Financial Economics, vol. 19, no. 14, pp. 1131-1142.

[105] Gosell, T. F. & Kolb, R. W. 1997, "Accuracy of International Interest Rate Forecasts", The Financial Review, vol. 32, no. 3, pp. 431-448.

[106] Granger, C. W. J. 2003, "Time Series Analysis, Cointegration and Applications", Nobel Lecture.

[107] Granger, C. W. J. & Machina, M. J. 2006, "Forecasting and Decision Theory", Chapter 2, in " Handbook of Forecasting", North-Holland, vol. 1, edited by Elliot, G., Granger, C. W. J. & Timmermann, A..

[108] Granger, C.W.J. & Pesaran, H. 1996, A Decision Theoretic Approach to Forecast Evaluation, Department of Applied Economics, University of Cambridge, Cambridge Working Papers in Economics.

[109] Granger, C.W.J. & Pesaran, M. 2000, "Economic and Statistical Measures of Forecast Accuracy", Journal of Forecasting, vol. 19, no. 7, pp. 537.

[110] Granger, C.W.J. & Ramanathan, R. 1984, "Improved Methods of Combining Forecasts", Journal of Forecasting, vol. 3, no. 2, pp. 197-204.

[111] Guidolin, M. & Thornton, D.L. 2008, Predictions of short-term rates and the expectations hypothesis of the term structure of interest rates, European Central Bank, Working Paper Series: 977.

[112] Guidolin, M. & Timmermann, A. 2005, "Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns", Economic Journal, vol. 115, no. 500, pp. 111-143.

[113] Guidolin, M. & Timmermann, A. 2009, "Forecasts of US short-term interest rates: A flexible forecast combination approach", Journal of Econometrics, vol. 150, no. 2, pp. 297-311.

[114] Hafer, R.W., Hein, S.E. & MacDonald, S.S. 1992, "Market and survey forecasts of the three-month treasury-bill rate", Journal of Business, vol. 65, no. 1, pp. 123.

[115] Hall, A.D., Anderson, H.M. & Granger, C.W.J. 1992, "A Cointegration Analysis of Treasury Bill Yields", The review of economics and statistics, vol. 74, no. 1, pp. 116-126.

[116] Hall, S.G. & Henry, S.G.B. 1988, "Macroeconomic Modelling contributions to economic analysis", North-Holland Publishing Company.

[117] Hall, S.G. & Henry, S.G.B. 2000, "Independence of the Bank of England: an update", Economic Outlook, vol. 24, no. 2, pp. 5-11.

[118] Hall, S.G. & Mitchell, J. 2007, "Combining density forecasts", International Journal of Forecasting, vol. 23, no. 1, pp. 1-13.

[119] Hall, S.G. & Mitchell, J. 2009, "Recent developments in density forecasting", Chapter 5, Palgrave Handbook of Econometrics, Applied Econometrics, vol. 2, Edited by Mills, T. C. & Patterson, K.

[120] Hamilton, J.D. 1988, "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates", Journal of Economic Dynamics and Control, vol. 12, no. 2, pp. 385-423.

[121] Hamilton, J.D. & Kim, D.H. 2002, "A Reexamination of the Predictability of Economic Activity Using the Yield Spread", Journal of Money, Credit and Banking, vol. 34, no. 2, pp. 340-360.

[122] Hammond, G. 2009, "Inflation Targeting in the UK", Bank of England[3].

[123] Hardouvelis, G.A. 1994, "The term structure spread and future changes in long and short rates in the G7 countries: Is there a puzzle?", Journal of Monetary Economics, vol. 33, no. 2, pp. 255-283.

[124] Harris, R. & Sollis, R. 2003, "Applied Time Series Modelling and Forecasting", John Wiley & Sons.

[125] Harvey, C. R. 1989, "Forecasts of Economic Growth from the Bond and Stock Market", Financial Analysis Journal, vol. 45, no. 5, pp. 38-45.

[126] Hendry, D. F. & Ericsson, N. R. 2003, "Understanding Economic Forecasting", The MIT Press.

[127] Hicks, J. R. 1946, "Value and Capital", 2nd Edition, Oxford University Press.

[128] Hurn, A.S., Moody, T. & Muscatelli, V.A. 1995, "The Term Structure of Interest Rates in the London Interbank Market", Oxford Economic Papers, vol. 47, no. 3, pp. 418-436.

[129] Ioannides, M. 2003, "A comparison of yield curve estimation techniques using UK data", Journal of Banking & Finance, vol. 27, no. 1, pp. 1-26.

[130] Jardet, C. 2004, "Why did the term structure of interest rates lose its predictive power?", Economic Modelling, vol. 21, no. 3, pp. 509.

[131] Johansen, S.&. & Juselius, K. 1990, "Maximum Likelihood Estimation and Inference on Cointegration–with Applications to the Demand for Money", Oxford Bulletin of Economics & Statistics, vol. 52, no. 2, pp. 169-210.

---

[3]http://www.bcb.gov.br/Pec/Depep/Seminarios/2009_XISemAnualMetasInflBCB/
Arquivos/2009_XISemAnualMetasInflBCB_GillHammond.pdf

[132] Johansen, S. 1988, "Statistical analysis of cointegration vectors", Journal of Economic Dynamics and Control, vol. 12, no. 2-3, pp. 231-254.

[133] Johansen, S. 1991, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", Econometrica, vol. 59, no. 6, pp. 1551-1580.

[134] Jorion, P. & Mishkin, F. 1991, "A multicountry comparison of term-structure forecasts at long horizons", Journal of Financial Economics, vol. 29, no. 1, pp. 59-80.

[135] Kandel, S. & Stambaugh, R.F. 1996, "On the Predictability of Stock Returns: An Asset-Allocation Perspective", The Journal of Finance, vol. 51, no. 2, pp. 385-424.

[136] Kapetanios, G., Labhard, V. & Price, S. 2008, "Forecast combination and the Bank of England's suite of statistical forecasting models", Economic Modelling, vol. 25, no. 4, pp. 772-792.

[137] Kim, T., Mizen, P. & Chevapatrakul, T. 2008, "Forecasting changes in UK interest rates", Journal of Forecasting, vol. 27, no. 1, pp. 53-74.

[138] King, M. 2008, from "Recession danger is real, warns Mervyn King" by Robert Winnett, Telegraph Article 14th May 2008.

[139] Klein, R.W. & Bawa, V.S. 1976, "The effect of estimation risk on optimal portfolio choice", Journal of Financial Economics, vol. 3, no. 3, pp. 215-231.

[140] Koukouritakis, M. & Michelis, L. 2008, "The term structure of interest rates in the 12 newest EU countries", Applied Economics, vol. 40, no. 4, pp. 479-490.

[141] Kugler, P. 1988, "An Empirical Note on the Term Structure and Interest Rate Stabilization Policies", Quarterly Journal of Economics, vol. 103, no. 4, pp. 789-792.

[142] Kugler, P. 1996, "The term structure of interest rates and regime shifts: Some empirical results", Economics Letters, vol. 50, no. 1, pp. 121-126.

[143] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. & Shin, Y. 1992, "Testing the null hypothesis of stationarity against the alternative of a unit root", Journal of Econometrics, vol. 54, no. 1, pp. 159-178.

[144] Lange, R. 1999, The Expectations Hypothesis for the Longer End of the Term Structure: Some Evidence for Canada, Bank of Canada, Working Paper No. 99-20.

[145] Lanne, M. 1999, "Near Unit Roots and the Predictive Power of Yield Spreads for Changes in Long-Term Interest Rates", Review of Economics & Statistics, vol. 81, no. 3, pp. 393-398.

[146] Lee, K. & Shields, K. 2000, "Expectations Formation and Business Cycle Fluctuations: an Empirical Analysis of Actual and Expected Output in Uk Manufacturing, 1975-1996", Oxford Bulletin of Economics & Statistics, vol. 62, no. 4, pp. 463.

[147] Leitch, G. & Tanner, 1991, "Economic forecast evaluation: Profits versus the conventional error measures", American Economic Review, vol. 81, no. 3, pp. 580.

[148] Longstaff, F.A. 1990, "Time Varying Term Premia and Traditional Hypotheses about the Term Structure", The Journal of Finance, vol. 45, no. 4, pp. 1307-1314.

[149] Longstaff, F.A. 2000, "The term structure of very short-term rates: New evidence for the expectations hypothesis", Journal of Financial Economics, vol. 58, no. 3, pp. 397-415.

[150] Lopes, A. C. & Monteiro, O. S. 2007, "The Expectations Hypothesis of the Term Structure: Some Empirical Evidence for Portugal" MPRA Paper No. 3437, mpra.ub.uni-muenchen.de

[151] Lutz, F. A. 1940, "The Structure of Interest Rates", Quarterly Journal of Economics, vol. 55, pp. 33-65.

[152] MacDonald, R. & Speight, A.E.H. 1988, "The Term Structure of Interest Rates in the UK", Bulletin of Economic Research, vol. 40, no. 4, pp. 287-299.

[153] MacDonald, R. & Speight, A.E.H. 1991, "The term structure of interest rates under rational expectations: some international evidence", Applied Financial Economics, vol. 1, no. 4, pp. 211-221.

[154] Mankiw, N.G. 1986, "The Term Structure of Interest Rates Revisited", Brookings Papers on Economic Activity, vol. 1986, no. 1, pp. 61-110.

[155] Mankiw, N.G. & Miron, J.A. 1986, "The Changing Behavior of the Term Structure of Interest Rates", Quarterly Journal of Economics, vol. 101, no. 2, pp. 211-228.

[156] Mankiw, N.G. & Summers, L.H. 1984, "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates?", Brookings Papers on Economic Activity, vol. 1984, no. 1, pp. 223-242.

[157] Mariscal, I.B. & Howells, P. 2007, "Monetary Policy Transparency in the UK: The Impact of Independence and Inflation Targeting", International Review of Applied Economics, vol. 21, no. 5, pp. 603-617.

[158] Marquering, W. & Verbeek, M. 2004, "The economic value of predicting stock index returns and volatility", Journal of Financial and Quantitative Analysis, vol. 39, no. 2, pp. 407-429.

[159] McCracken, M. W. & West, K. D. 2004, "Inference About Predictive Ability", Chapter 14, pp. 299-321, in "A Companion to Economic Forecasting" edited by Clements, M. P. & Hendry, D. F..

[160] McMillan, D. G. 2002, "Interest rate spread and real activity: evidence for the UK", Applied Economic Letters, vol. 9, pp. 191-194.

[161] Meese, R. & Rogoff, 1983, "Empirical Change Rate Models of the Seventies", Journal of International Economics, vol. 14, no. 1, pp. 3-24.

[162] Merton, R.C. 1969, "Lifetime Portfolio Selection Under Uncertainty: the Continuous-Time Case", Review of Economics & Statistics, vol. 51, no. 3, pp. 247.

[163] Mills, T.C. 1991, "The term structure of UK interest rates: tests of the expectations hypothesis", Applied Economics, vol. 23, no. 4, pp. 599-606.

[164] Mishkin, F.S. 1988, "The Information in the Term Structure: Some Further Results", Journal of Applied Econometrics, vol. 3, no. 4, pp. 307-314.

[165] Mishkin, F.S. 1990, "The Information in the Longer Maturity Term Structure About Future Inflation", The Quarterly Journal of Economics, vol. 105, no. 3, pp. 815-828.

[166] Mitchell, J. & Hall, S.G. 2005, "Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR 'Fan' Charts of Inflation", Oxford Bulletin of Economics & Statistics, vol. 67, pp. 995-1033.

[167] Musti, S. & D'Ecclesia, R.L. 2008, "Term structure of interest rates and the expectation hypothesis: The euro area", European Journal of Operational Research, vol. 185, no. 3, pp. 1596-1606.

[168] Newbold, P. and Harvey, D. I. 2004, "Forecast Combination and Encompassing", Chapter 12, pp. 268-283, in "A Companion to Economic Forecasting" edited by Clements, M. P. & Hendry, D. F..

[169] Patterson, K. 2000, "An Introduction to Applied Econometrics: A Time Series Approach", Macmillan Press Ltd.

[170] Pesaran, M. H. & Skouras, S. 2004, "Decision-Based Methods for Forecast Evaluation", Chapter 11, pp. 241-267, in "A Companion to Economic Forecasting" edited by Clements, M. P. & Hendry, D. F..

[171] Pesaran, M.H. & Timmermann, A. 1995, "Predictability of Stock Returns: Robustness and Economic Significance", Journal of Finance, vol. 50, no. 4, pp. 1201-1228.

[172] Phillips, P.C.B. & Perron, P. 1987, "Testing for a Unit Root in Time Series Regression", Cowles Foundation, Yale University, Cowles Foundation Discussion Papers: 795R.

[173] Plosser, C.I. & Rouwenhorst, K.G. 1994, "International term structures and real economic growth", Journal of Monetary Economics, vol. 33, no. 1, pp. 133-155.

[174] Roberds, W., Runkle, D. & Whiteman, C.H. 1996, "A Daily View of Yield Spreads and Short-Term Interest Rate Movements", Journal of Money, Credit and Banking, vol. 28, no. 1, pp. 34-53.

[175] Robinson, E.S. 1998, "The term structure of Australian interest rates: Tests of the expectation hypothesis", Applied Economics Letters, vol. 5, no. 7, pp. 463.

[176] Rossi, M. 1996, "The Information Content of the Short End of the Term Structure of Interest Rates", Bank of England Working Paper No.55.

[177] Rudebusch, G.D. & Wu, T. 2008, "A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy", Economic Journal, vol. 118, no. 530, pp. 906-926.

[178] Samuelson, P.A. 1969, "Lifetime Portfolio Selection by Dynamic Stochastic Programming", Review of Economics & Statistics, vol. 51, no. 3, pp. 239.

[179] Sarno, L., Thornton, D. & Valente, G. 2005, "Federal Funds Rate Prediction", The Journal of Money, Credit and Banking, vol. 37, no. 3, pp. 449-471.

[180] Sarno, L., Thornton, D. & Valente, G. 2007, "The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields", Journal of Financial & Quantitative Analysis, vol. 42, no. 1, pp. 81-100.

[181] Sarno, L. & Valente, G. 2004, "Comparing the Accuracy of Density Forecasts from Competing Models", Journal of Forecasting, vol. 23, no. 8, pp. 541-557.

[182] Sarno, L. & Valente, G. 2005, "Empirical exchange rate models and currency risk: some evidence from density forecasts", Journal of International Money and Finance, vol. 24, no. 2, pp. 363-385.

[183] Shea, G.S. 1992, "Benchmarking the Expectations Hypothesis of the Interest-Rate Term Structure: An Analysis of Cointegration Vectors", Journal of Business & Economic Statistics, vol. 10, no. 3, pp. 347-366.

[184] Shiller, R.J. 1979, "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure", Journal of Political Economy, vol. 87, no. 6, pp. 1190.

[185] Soderlind, P., Soderstrom, U. & Vredin, A. 2005, "Dynamic Taylor rules and the predictability of interest rates", Macroeconomic Dynamics, vol. 9, no. 3, pp. 412-428.

[186] Sola, M. & Driffill, J. 1994, "Testing the term structure of interest rates using a stationary vector autoregression with regime switching", Journal of Economic Dynamics and Control, vol. 18, no. 3-4, pp. 601-628.

[187] Stekler, H.O. 1991, "Macroeconomic forecast evaluation techniques", International Journal of Forecasting, vol. 7, no. 3, pp. 375-384.

[188] Stock, J.H. & Watson, M.W. 2003, "Forecasting Output and Inflation: The Role of Asset Prices", Journal of Economic Literature, vol. 41, no. 3, pp. 788-829.

[189] Stock, J.H. & Watson, M.W. 2004, "Combination Forecasts of Output Growth in a Seven-Country Data Set", Journal of Forecasting, vol. 23, no. 6, pp. 405-430.

[190] Tay, A.S. & Wallis, K.F. 2000, "Density Forecasting", Journal of Forecasting, vol. 19, no. 4, pp. 235.

[191] Tay, A. S. & Wallis. K. F. 2004, "Density Forecasting: A Survey", Chapter 3, pp. 45-68, in "A Companion to Economic Forecasting" edited by Clements, M. P. & Hendry, D. F..

[192] Taylor, M.P. 1992, "Modelling the Yield Curve", The Economic Journal, vol. 102, no. 412, pp. 524-537.

[193] Thornton, D.L. 2004, "Testing the Expectations Hypothesis: Some New Evidence for Japan", Review (00149187), vol. 86, no. 5, pp. 21-39.

[194] Timmermann, A. 2006, "Forecast Combination", Chapter 4, in " Handbook of Forecasting", North-Holland, vol. 1, edited by Elliot, G., Granger, C. W. J. & Timmermann, A..

[195] Tootell, G.M.B. 2002, "The Bank of England's Monetary Policy", New England Economic Review, pp. 61-64.

[196] Tzavalis, E. 2004, "The term premium and the puzzles of the expectations hypothesis of the term structure", Economic Modelling, vol. 21, no. 1, pp. 73-93.

[197] Tzavalis, E. & Wickens, M.R. 1997, "Explaining the Failures of the Term Spread Models of the Rational Expectations Hypothesis of the Term Structure", Journal of Money, Credit and Banking, vol. 29, no. 3, pp. 364-380.

[198] Tzavalis, E. & Wickens, M. 1998, "A Re-examination of the Rational Expectations Hypothesis of the Term Structure: Reconciling the Evidence from Long-Run and Short-Run Tests", International Journal of Finance and Economics, vol. 3, no. 3, pp. 229-239.

[199] Verbeek, M. 2000, "A Guide to Modern Econometrics", John Wiley & Sons Ltd.

[200] Vereda, L., Lopes, H. & Fukuda, R. 2008, "Estimating VAR models for the term structure of interest rates", Insurance: Mathematics and Economics, vol. 42, no. 2, pp. 548-559.

[201] Vickers, J. 1998, "Inflation targeting in practice: The UK experience", Bank of England Quarterly Bulletin, vol. 38, no. 4, pp. 368.

[202] West, K. D. 2006, "Forecast Evaluation", Chapter 3, in " Handbook of Forecasting", North-Holland, vol. 1, edited by Elliot, G., Granger, C. W. J. & Timmermann, A..

[203] West, K., Edison, H. & Cho, D. 1993, "A utility-based comparison of some models of exchange rate volatility", Journal of International Economics, vol. 35, no. 1-2, pp. 23-45.

[204] Wright, D.J., Capon, G., Page, R., Quiroga, J., Taseen, A.A. & Tomasini, F. 1986, "Evaluation of Forecasting Methods for Decision Support", International Journal of Forecasting, vol. 2, no. 2, pp. 139-152.

[205] Yihong Xia 2001, "Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation", Journal of Finance, vol. 56, no. 1, pp. 205-246.