Financial Markets and the Macroeconomy: Cross-Sectional Returns, Time-Variation of Risk Premia, and Forecasting

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Wirtschaftswissenschaftlichen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt von

Andreas Schrimpf
aus Karlsruhe

2009

Dekan:	Prof. Dr. Kerstin Pull
Erstberichterstatter:	Prof. Dr. Joachim Grammig
Zweitberichterstatter:	Prof. Dr.-Ing. Rainer Schöbel
Tag der mündlichen Prüfung:	25. Juni 2009

ACKNOWLEDGEMENTS

This thesis is based on my research as a doctoral student at the Centre for European Economic Research (ZEW) in Mannheim whilst being an associate member of the DFG research training group at the University of Tübingen (Graduiertenkolleg "Unternehmensentwicklung, Marktprozesse und Regulierung in dynamischen Entscheidungsmodellen" sponsored by the German Research Foundation, DFG). Over this stimulating period of three and a half years, there were many who contributed to my academic progress and thus deserve special thanks.

First of all, I am grateful to my supervisor Prof. Dr. Joachim Grammig, who sparked my interest in empirical asset pricing and macro-finance ever since I enrolled in his Financial Econometrics class at the University of Tübingen. I benefited greatly from his guidance as a coauthor and from many insightful comments during the research on my thesis. I very much appreciate his continuous encouragement and the freedom I had for conducting my research. I would also like to thank Prof. Dr.-Ing. Rainer Schöbel and Prof. Dr. Werner Neus for everything they taught me about finance and for kindly agreeing to serve on my thesis committee.

I am also very indebted to the interaction, feedback and discussions with colleagues and coauthors in my close research environment at ZEW and elsewhere, in particular: Prof. Francois Laisney, Emanuel Mönch, Waldemar Rotfuss, Maik Schmeling, Peter Schmidt, Michael Schröder, Prof. Richard Stehle, Michael Schuppli and Qingwei Wang. A special thanks goes to Stefan Frey for sharing his GMM library for Gauss and to Prof. Jesper Rangvid for offering me the opportunity of a research visit at Copenhagen Business School where final work on this thesis was accomplished. Numerous other people have helped me with their comments and suggestions when the different chapters of this thesis (or earlier drafts) were presented at various international conferences, workshops and seminars.

Over the past years, I have also grown to highly appreciate the help of several interns and student assistants at ZEW, who often took workload off me, enabling me to focus on my research projects. In particular, I thank Zohal Hesami, Alexander Bank, Jörg Breddermann, Oliver Stahnke, Florian Mörth, Christoph Schinke, Dirk Rauscher and Frieder Mokinski for their excellent research assistance. Many thanks to Hela Hellerich for careful proof-reading of the different chapters of this thesis.

Last but not least, a supportive non-academic environment was invaluably helpful during the period of working on my Ph.D. thesis. I want to thank my family (especially my parents Hans and Gertrud) for their unconditional support, patience and advice. A special thank you goes to Carolin. Her cheerfulness and loving support helped me bear the ups and downs over the past years.

Andreas Schrimpf

Contents

List of Tables

LIST OF TABLES

List of Figures

PREFACE

General Motivation of the Thesis

This thesis contains three separate essays which empirically investigate various aspects of the relationship between financial markets and the real economy. The interface of finance and macroeconomics has fascinated me ever since being exposed to the famous book "Asset Pricing" by John Cochrane during the later stages of my studies in Tübingen. During my time as a doctoral student, I then had the opportunity to conduct my own research in this field. This thesis is the summary of this work.

According to the most basic models of (financial) economics, many phenomena in financial markets should bear a strong relationship with the macroeconomy. Probably one of the most classical examples is the consumption-based asset pricing model, which posits that assets should be priced according to the covariance of their returns with consumption growth. Another example, is the predictable variation of stock excess returns, which has generally been interpreted as evidence for time-varying risk-premia in financial markets. The conventional (risk-based) explanation for return predictability draws on time-variation in risk aversion over the business cycle, and thus also emphasizes the link between financial markets and the macroeconomy.

It is fair to say, however, that – despite longstanding research efforts – understanding the link between financial markets and the real economy still represents a major challenge for financial economists. In the past, prevalent theories of financial economics often had trouble accounting for the empirical facts. This is exemplified most forcefully by the empirical failure of the canonical consumption-based asset pricing model. It is well-known that the model has difficulties explaining the high level of US stock returns relative to the risk free rate at reasonable level of relative risk aversion, known as the "equity premium puzzle" (Mehra and Prescott, 1985), or to rationalize the historically high returns on value stocks relative to growth stocks (Lettau and Ludvigson, 2001b). Thus, given the empirical evidence, which often constitutes a challenge to well-established economic theories, a better understanding of the relationship between financial markets and the real economy is generally of major importance for both finance and economics. In the words of Cochrane (2007, p.6 and p.91)

> "The program of understanding the real, macroeconomic risks that drive asset prices (or the proof that they do not do so at all) is not some weird branch of finance; it is the trunk of the tree."

> "The challenge is straightforward: we need to understand what macroeconomic risks underlie the "factor risk premia", the average returns on special portfolios that finance research uses to characterize the cross-section of assets."

Over time, the macro-finance literature has accumulated substantial knowledge about several "stylized facts" which are often taken for granted by the academic community. The major goal of the different essays contained in this thesis is to critically reassess some of those major empirical findings in the literature. The "facts" which I put under scrutiny include: (i) the general belief that the conventional consumption-based model with power utility exhibits a poor performance in explaining asset prices (in particular for size and book-to-market sorted portfolios), (ii) the predictability of stock market excess returns – labeled as one of the "new facts" in finance by Cochrane (1999) – which has generally been interpreted as evidence for countercyclically evolving risk premia, and (iii) that asset prices (most prominently the slope of the yield curve) are useful predictors of real activity. In order to investigate these separate macro-finance issues, modern econometric tools are applied which strive to address some methodological issues and limitations of earlier empirical work. In this way, the broad goal of my thesis is to shed new light on some of the old questions in the macro-finance literature and to critically re-evaluate previous empirical findings.

Outline of the Thesis

Specifically, I investigate the following issues in this thesis: (i) The empirical performance of the consumption-based asset pricing model when the relevant risk for an investor is long-run consumption risk, (ii) stock return predictability in international stock markets in the presence of model uncertainty, and (iii) the reliability and the predictive power of the yield curve for real activity in the context of structural instability. In the following paragraphs, I will briefly describe how each of the three different chapters of this thesis is designed to contribute to the existing general knowledge in the macro-finance literature.

Long-Horizon Consumption Risk

Chapter 1 ("Long-Horizon Consumption Risk and the Cross-Section of Returns: New Tests and International Evidence") investigates whether measuring consumption risk over long horizons can improve the empirical performance of the Consumption CAPM for size and value premia in international stock markets (US, UK, and Germany).

It is a well-known "fact" in macro-finance that the standard consumption-based asset pricing model – relating contemporaneous consumption growth to asset returns – has serious problems in explaining the cross-sectional variation of returns (e.g., Cochrane, 1996; Lettau and Ludvigson, 2001b). As a reaction to the poor economic performance of the canonical model, a large amount of explanations and modifications have been put forth in the literature. These extensions, for instance, include new utility functions such as habit formation (e.g., Campbell and Cochrane, 1999) or recursive utility (Epstein and Zin, 1989) or a departure from standard assumptions such as complete markets (e.g., Constantinides and Duffie, 1996).

However, several recent studies exploring the basic insights of the power-utility consumption-based paradigm have reported encouraging steps forward (Cochrane, 2007, p.51). In particular, a recent contribution by Parker and Julliard (2005) suggests relating asset returns to consumption growth measured over longer horizons within a simple consumption-based framework with standard power-utility. The long-run empirical setup is robust against various arguments as to why consumption expenditure may be slow to adjust to changes in aggregate wealth. Besides, the model is closely related to the literature on long-run consumption risk, as it implies expressions for expected returns that are similar to the testable implications of long-run risk models with recursive utility such as Hansen, Heaton, and Li (2008). By explicitly accounting for consumption risk measured over longer horizons, Parker and Julliard's long-horizon LH-CCAPM successfully explains a large fraction of cross-sectional variation in expected returns across US size and book-to-market sorted portfolios.

This chapter revisits the ability of the LH-CCAPM to explain the cross-section of returns and contributes to the literature in several ways. First, by modifying Parker and Julliard's empirical approach in our econometric estimation of the asset pricing models, we take recent methodological concerns about the strong factor structure of value and size portfolios into account (Phalippou, 2007; Lewellen, Nagel, and Shanken, 2007). By this means, we provide a critical reassessment of the explanatory power of the long-run consumption based asset pricing framework for the famous “value puzzle”. Second, we provide new international evidence on the role of long-run consumption risks for asset pricing by investigating the model's explanatory power for the cross-section of equity returns in the United Kingdom and Germany. Overall, our results shed new light on the relative strengths and weaknesses of the long-run approach to asset pricing.

The main results of chapter 1 can be summarized as follows. Under our modified empirical approach, we find that long-horizon consumption risk falls short of providing a complete account of the cross-section of expected returns, especially the premium on value stocks. In this way, our findings suggest that the long-horizon consumption-based approach falls short of resolving the famous “value premium puzzle”, as claimed in the original paper by Parker and Julliard (2005). Nevertheless, measuring consumption risk over longer horizons achieves other important improvements, most notably a reduction of risk aversion estimates. Thus, our results suggest that more plausible parameter estimates – as opposed to lower pricing errors – can be regarded as the main achievement of the long-horizon Consumption CAPM.

Return Predictability under Model Uncertainty

Chapter 2 ("International Stock Return Predictability under Model Uncertainty") empirically investigates the question whether excess returns on aggregate stock market indices are predictable.[1] Whereas the focus of the first chapter was on the cross-sectional variation of expected returns across different stock portfolios, the second chapter deals with time-variation of expected returns, i.e., time-series aspects of the link between financial markets and the macroeconomy.

Return predictability has been labeled as one of the "new facts" in finance (Cochrane, 1999) and has generally been accepted as a typical feature of stock markets. The standard risk-based explanation is that there is time-variation in risk premia, such that stock market participants demand a premium for holding risky assets in "bad times" (e.g. during recessions). This issue is particularly important since it has far-reaching consequences for empirical as well as for theoretical modeling – e.g., conditional asset pricing models (Jagannathan and Wang, 1996; Lettau and Ludvigson, 2001b), or intertemporal asset pricing (Campbell and Vuolteenaho, 2004; Petkova, 2006) – and other issues of practical importance such as long-run asset allocation (e.g., Campbell and Viceira, 2002).

Empirical studies have found a plethora of variables to be informative about future excess returns in predictive regressions. In particular, valuation ratios (e.g., dividend yields) and interest rate related variables (e.g., short-term interest rates as well as default and term spreads) have featured prominently in predictive regressions, but also macroeconomic variables – e.g., the consumption-wealth ratio by Lettau and Ludvigson

[1] This chapter has originally been inspired by the recent debate on return predictability triggered by the influential paper by Goyal and Welch (2008). In the meantime, the article by Goyal and Welch (2008) and further papers by authors joining the debate have been published in a special issue of the Review of Financial Studies (Vol. 21, No.4, 2008).

(2001a) or more recently the output gap by Cooper and Priestley (2006) – have been used to predict returns.

Given the large number of variables proposed in the literature, a high amount of uncertainty exists regarding the right choice of state variables. Moreover, the fact that so many variables have found to be valuable predictors of returns naturally raises the concern that the apparent predictability documented in the extant literature may simply arise due to data-snooping rather than genuine variation of economic risk premia.

The aim of this chapter is therefore to explore the robustness of several predictive variables in international stock markets in the context of model uncertainty. We follow the path set by the seminal work by Cremers (2002) and Avramov (2002) and use Bayesian model averaging in order to account for model uncertainty. A novel feature of the model averaging approach used in this paper is to account for a potential finite-sample bias of the coefficients in the predictive regressions. This issue has not previously been addressed in work on return predictability using model averaging methods.

Drawing on an extensive international dataset covering major international stock markets, we find that interest-rate related variables are usually among the most prominent predictive variables, whereas valuation ratios generally perform rather poorly. There is also some evidence that risk premia vary with the output gap. Yet, predictability of market excess returns clearly weakens once model uncertainty is accounted for. We also document notable differences in the degree of in-sample and out-of-sample predictability across different international stock markets. This finding suggests that return predictability is not a uniform and a universal feature across international capital markets.

The Yield Curve as a Leading Indicator under Structural Instability

The relation between financial markets and the macroeconomy is not unidirectional. For instance, there is good reason to believe that asset prices – being forward-looking in nature – may contain predictive content for real activity. In *Chapter 3 ("A Reappraisal of the Leading Indicator Properties of the Yield Curve in the Presence of Structural Instability")*, the relation between the two research fields are examined from a reverse angle.

Among the financial predictive variables, the yield spread between the long- and short-term interest rate (slope of yield curve) has generally stood out as one of the most prominent variables. The extant literature has long since established its role as a leading indicator for future economic activity. However, in the recent literature concerns have been raised over the fact that the predictive performance of the term spread may be time-variant and that predictive regressions based on the yield spread may be subject to substantial model instability (Estrella, Rodrigues, and Schich, 2003). For instance, the predictive power may depend on underlying factors such as the form of the monetary policy reaction function or the relative importance of real and nominal shocks in the economy. Both factors potentially change over time, which raises the need to investigate the time-variation of the forecasting relationship in greater detail.

The main goal of this chapter is to investigate whether the yield spread still qualifies as a useful leading indicator in environments characterized by model instability. For this purpose we provide an extensive reexamination of the leading indicator properties of the yield curve. A main feature of our approach is to focus on the time-varying out-of-sample (OOS) forecasting properties of the yield curve. This is of particular relevance, since one may argue that the ultimate concern of market participants and policy makers is out-of-sample forecast accuracy as well as a good predictive performance towards the end of the sample period.

Our general finding in this chapter is that there is a substantial time-variation in the out-of-sample forecast performance of the yield curve for real activity. Moreover, we document a degradation in the predictive performance of the yield spread over the most recent sample period which has not been shown in the literature before. This finding holds true for all countries considered. We thus take a closer look at potential reasons for the time-variation of predictive power and forecast breakdowns by using several modern (in-sample) tests for parameter stability. Using these econometric tools, we provide strong evidence for structural instabilities that affect the predictive relationship.

Hence, we address the fundamental question whether the yield spread can still be regarded as a reliable leading indictor in the presence of structural instability. For the purpose of reexamining the leading indicator properties under structural change, we use optimal window selection techniques, which are designed for forecasting in unstable environments. We find that newly developed methods for forecasting in the presence of structural change generally help improve forecast accuracy. However, this does not change our conclusion that the yield curve has been losing its edge as a predictor of real activity in recent years.

Chapter 1

Long-Horizon Consumption Risk and the Cross-Section of Returns: New Tests and International Evidence*

ABSTRACT

This chapter investigates whether measuring consumption risk over long horizons can improve the empirical performance of the Consumption CAPM for size and value premia in international stock markets (US, UK, and Germany). We modify the estimation approach of Parker and Julliard (2005) taking commonalities in size and book-to-market sorted portfolios into account. Our results show that, contrary to the findings of Parker and Julliard, the model falls short of providing an accurate description of the cross-section of returns under our modified empirical approach. At the same time, however, measuring consumption risk over longer horizons typically yields lower risk-aversion estimates. Thus, our results suggest that more plausible parameter estimates – as opposed to lower pricing errors – can be regarded as the main achievement of the long-horizon Consumption CAPM.

*This chapter is based on a joint paper with Joachim Grammig (University of Tübingen) and Michael Schuppli (University of Münster). A revised version of the paper is accepted for publication at the *European Journal of Finance*.

1.1 Introduction

Understanding the behavior of asset prices and their relation to macroeconomic risks can be considered as one of the most fundamental issues in finance. As is well known, however, the traditional workhorse for studying the link between financial markets and the real economy – the consumption-based asset pricing model (CCAPM) – has failed to explain a number of stylized facts in finance such as the equity premium (Mehra and Prescott, 1985), asset return volatility (Grossman and Shiller, 1982) or value and size premia in the cross-section of expected returns (Cochrane, 1996; Lettau and Ludvigson, 2001b).[1] After a long series of poor empirical results starting with Hansen and Singleton (1982, 1983), more recent studies exploring the basic insights of the consumption-based asset pricing paradigm report encouraging advances (Cochrane, 2007, p.267).

In particular, a recent contribution by Parker and Julliard (2005) suggests to relate asset returns to consumption growth measured over longer horizons within a simple consumption-based framework with CRRA preferences. Such reasoning is in line with theoretical literature on long-run consumption risk. Seminal work by Bansal and Yaron (2004) suggests that equilibrium asset returns depend on investors' expectations about both short and long-run changes in consumption growth. Among other things, this result implies that the covariance of returns with contemporaneous consumption growth may understate the risk perceived by investors.[2] By explicitly accounting for consumption risk over longer horizons, Parker and Julliard's long-horizon (LH) CCAPM is able to explain a large fraction of cross-sectional variation in expected returns across US size and book-to-market sorted portfolios.[3]

In this paper, we provide new detailed evidence as to whether long-run consumption

[1]The consumption-based asset pricing model has its roots in the original articles by Rubinstein (1976), Lucas (1978), and Breeden (1979). We use the terms CCAPM and consumption-based model interchangeably in the remainder of the paper.

[2]Research on the long-run implications of the consumption-based asset pricing framework has constituted a rather prominent field in recent literature [e.g. Jagannathan and Wang (2007), Bansal, Dittmar, and Kiku (2007), Hansen, Heaton, and Li (2008) or Rangvid (2008)]. More detailed information on how our paper is related to the extant literature is provided in Section 1.2.2.

[3]We will abbreviate the long-horizon CCAPM to LH-CCAPM in the remainder of the text.

risk helps explain the cross-section of expected returns in international stock markets. In particular, we modify Parker and Julliard's empirical approach along two lines. First, we take into account recent criticism about the widespread use of size and book-to-market sorted portfolios in the empirical asset pricing literature (Phalippou, 2007; Lewellen, Nagel, and Shanken, 2007). In order to reduce the adverse effects of strong commonalities in size and book-to-market sorted portfolios, we follow the prescription of Lewellen, Nagel, and Shanken (2007) to include industry portfolios alongside with the conventionally used size and book-to-market portfolios. Second, we provide new international evidence by investigating the model's explanatory power for the cross-section of equity returns in the United Kingdom and Germany.

Our empirical findings shed new light on the relative merits of the long-horizon CCAPM when it comes to explaining the cross-section of returns in international stock markets. First, we find that under our modified empirical approach accounting for the strong common factor structure in size and book-to-market sorted portfolios, the model's ability to account for cross-sectional variation in returns is clearly limited. This result suggests that the good empirical performance on US test assets reported by Parker and Julliard (2005) may be somewhat overstated. Tests with size and book-to-market sorted portfolios from the UK and Germany further corroborate the US evidence. Second, we find that measuring consumption risk over longer horizons typically yields lower risk-aversion estimates. Thus, our results suggest that more plausible parameter estimates – as opposed to a higher cross-sectional R^2 – can be viewed as the main achievement of the long-horizon consumption-based approach.

Even though the long-run risk framework has important implications for the explanation of risk premia and asset price fluctuations, previous empirical studies surveyed by Bansal (2007) have almost exclusively focussed on the US stock market. By estimating the proposed consumption-based model on UK and German portfolio returns, our paper explores the universality of the LH-CCAPM approach and, more generally, the role of long-run consumption risk in these markets.

This issue is particularly interesting since the countries considered in our study differ in several institutional respects. While the US and the UK for instance are known to have a market-based financial system and high private stock-ownership, Germany has a bank-based system and the share of stocks in the net wealth position of German households is lower. Furthermore, some authors have argued that the well-known US "equity premium puzzle" (i.e. the inability of the consumption based approach to quantitatively explain the high level of aggregate stock market returns compared to the T-Bill rate) may to some extent be due to extraordinarily high historical stock returns in the US during the post-war period [See, e.g., the discussion in (Cochrane, 2007, p.266)].[4] By contrast, post WWII excess returns on the German stock market have been somewhat lower.

The remainder of the text is structured as follows. Section 1.2 reviews the basic long-horizon consumption risk approach and provides a discussion on the literature most closely related to our paper. Section 1.3 describes the empirical methods used for estimating and evaluating the different models. Section 1.4 presents our data and discusses empirical results based on GMM estimation. Finally, Section 1.5 concludes.

1.2 The Long-Horizon Consumption Risk Framework

1.2.1 Parker and Julliard's Basic Model

This section briefly reviews the long-horizon consumption-based asset pricing approach put forth by Parker and Julliard (2005). As a starting point, consider the traditional two-period consumption-based model. As is well known, the model implies Euler equations of the following form

$$\mathrm{E}_t\left[\delta \frac{u'(C_{t+1})}{u'(C_t)}\; R^e_{t+1}\right] = 0 \tag{1.1}$$

where $u(.)$ denotes current-period utility, δ the subjective time discount factor, and R^e_{t+1}

[4] Some financial economists also argue that expected excess returns are likely to be lower in the future, thus reducing the puzzle [See e.g. Fama and French (2000), Welch (2001)].

the excess return on a risky asset. Empirical tests of consumption-based models are typically based on moment conditions implied by variants of Equation (1.1). Parker and Julliard (2005) use the model's first order condition for the risk-free rate between points in time t+1 and t+1+S

$$u'(C_{t+1}) = \delta \mathbf{E}_{t+1}[R^f_{t+1,t+1+S}\ u'(C_{t+1+S})] \tag{1.2}$$

to substitute out period t+1 marginal utility in the above Euler equation. Assuming power utility and $\delta \approx 1$, Equation (1.1) can thus be rewritten as

$$\mathbf{E}_t \left[m^S_{t+1}\ R^e_{t+1}\right] = 0 \tag{1.3}$$

where $m^S_{t+1} = R^f_{t+1,t+1+S}\left(\frac{C_{t+1+S}}{C_t}\right)^{-\gamma}$ is the stochastic discount factor (SDF) and S denotes the horizon at which consumption growth is measured. As shown by Malloy, Moskowitz, and Vissing-Jørgensen (2006), a very similar stochastic discount factor can be derived in the Epstein and Zin (1989) recursive utility framework of Hansen, Heaton, and Li (2008). Using unconditional instead of conditional moments and rearranging yields an expression for the expected excess return

$$\mathbf{E}[R^e_{i,t+1}] = -\frac{\mathrm{Cov}\left[m^S_{t+1}, R^e_{i,t+1}\right]}{\mathbf{E}[m_{t+1}]}, \tag{1.4}$$

which is similar to the case of the standard model except that the excess return now depends on its covariance with marginal utility growth over a longer time-horizon. In other words, investors demand a higher risk premium on assets whose return is more positively correlated with consumption growth over a long horizon. Parker and Julliard (2005) refer to the covariance of an asset's excess return with the modified SDF as "ultimate consumption risk".

The model's asset pricing implications can be tested either by directly estimating the nonlinear specification given by Equation (1.3), or by using the representation given by (1.4). Alternatively, the model can be estimated in its linearized form: Applying a first-order log-linear approximation in the spirit of Lettau and Ludvigson (2001b) of the SDF yields

$$m_{t+1}^{S} = R_{t,t+1+S}^{f} - \gamma_S R_{t,t+1+S}^{f} \Delta c_{t+1+S}, \tag{1.5}$$

where $\Delta c_{t+1+S} = \ln\left(C_{t+1+S}/C_t\right)$ represents log consumption growth from t to t+1+S. Hence, the model using the linearized SDF in (1.5) can be interpreted as a linear two-factor model. Furthermore, assuming the risk-free rate to be constant between t and t+1+S, the linear approximation reduces to a single factor model where the pricing kernel is a function of log consumption growth over long horizons.

1.2.2 Related Literature and Further Motivation

An important aspect of the proposed long-horizon CCAPM is that, in addition to retaining the parsimony of the power utility specification, it does not impair the basic assumptions of the consumption-based asset pricing framework. Yet, at the same time, the approach is consistent with various arguments why the covariance of an asset's return with contemporaneous consumption growth may understate its risk due to slow consumption adjustment. First, a wide range of factors not considered in the basic model, such as different sources of income, housing and durable goods consumption, may enter the utility function. In this case, the utility function is non-separable in that marginal utility with respect to one argument will always depend on the value of the other arguments. In addition, some of the consumption goods entering the utility function may involve a commitment (Chetty and Szeidl, 2005). Obviously, the adjustment of durable goods and housing consumption requires households to incur considerable transaction costs. Moreover, many services such as telecommunications are typically subject to long-term contracts. These real-world features imply that aggregate consumption adjustment may be slow.

Second, due to market imperfections such as costs of gathering and processing information, agents' short-term behavior may deviate from utility-maximizing consumption smoothing. In the presence of such frictions, investors may not optimally adjust consumption or rebalance their portfolio if utility losses from non-optimal behavior are small in magnitude (Cochrane, 1989). Such "near-rational" behavior appears plausible especially in the short-run. Again, from an empirical point of view, the reaction of consumption to changes in aggregate wealth will probably not be reflected in quarterly observations so that long-horizon consumption growth provides a more exact measure of perceived consumption risk.

Furthermore, the CCAPM of Parker and Julliard (2005) is closely related to a growing body of literature suggesting that investors require a premium on long-run consumption risk in asset returns. Pioneering theoretical work by Bansal and Yaron (2004) models consumption and dividend growth as containing a small persistent predictable component. Therefore, current shocks to expected growth will affect expectations about consumption growth in both the short and long run. From a theoretical point of view, the proposed consumption and dividend process can be motivated by explicitly modeling a production economy as in Kaltenbrunner and Lochstoer (2007).[5] Bansal and Yaron (2004) show that in an economy with Epstein-Zin investor preferences, this additional source of risk helps to explain longstanding issues in finance such as the equity premium, low risk-free rates, high stock market volatility, and the predictive power of price-dividend ratios for long-horizon stock returns. In addition, the long-run risk framework has strong implications for the cross-section of expected asset returns. If representative agents are concerned about both short and long-run consumption risk, they will require higher risk premia on assets that are correlated with long-run consumption growth. Modeling dividend and consumption growth as a VAR, Bansal, Dittmar, and Lundblad (2005) determine the exposure of dividends to long-run consumption risk. They show that this exposure helps explain a large fraction of cross-sectional variation in returns across book-to-market, size and momentum portfolios. Other recent papers documenting the relevance of long-run consumption risk for determining equilibrium asset returns include Bansal, Dittmar, and Kiku (2007), Hansen, Heaton, and Li (2008),

[5] The existence of a persistent component in consumption and dividends is empirically confirmed by Bansal, Kiku, and Yaron (2007).

Malloy, Moskowitz, and Vissing-Jørgensen (2006), and Colacito and Croce (2007).

In sum, a large body of evidence for the US suggests that consumption growth measured over longer horizons may be an important risk factor explaining cross-sectional variation in returns. Indeed, Parker and Julliard (2005) show that the cross-sectional R^2 obtained when estimating the model on 25 US book-to-market and size portfolios increases with the horizon at which consumption growth is measured. In fact, the non-linear model explains up to 44% of the cross-sectional variation in average excess returns for a horizon of 11 quarters. In this respect, the model's performance is similar to the conditional CCAPM of Lettau and Ludvigson (2001b) and the Fama and French (1993) three factor model. This finding seems to suggest that long-run risk may help resolve the value premium puzzle.

Another prominent drawback of the canonical CCAPM with CRRA utility is that, given the observed risk premia, estimated coefficients of relative risk aversion are usually implausibly high (Hansen and Singleton, 1983). This aspect is at the center of recent work by Rangvid (2008), who tests an international LH-CCAPM using world-consumption growth as a risk factor on excess aggregate stock market returns from 16 developed capital markets. The author shows that risk aversion estimates for an internationally diversified investor decrease substantially to more plausible values if long-run consumption risk is taken into account. However, the beta-pricing version of the model has trouble explaining the cross-section of international stock index returns.

It is important to note that his empirical approach is based on the strong assumption of an international representative investor, integrated financial markets, and purchasing power parity. This paper, in contrast, analyzes the ability of the LH-CCAPM to explain the individual cross-section of stock returns in three major stock markets. Besides requiring weaker assumptions, looking at only three countries enables us to use detailed consumption data that distinguish expenditure on nondurable goods and services from durable goods (rather than having to rely on measures of total consumption). Moreover, it allows us to pin down pricing errors for individual stock portfolios formed on characteristics such as size and book-to-market equity ratios, which have been of particular interest in the empirical finance literature.

1.3 Empirical Methodology

In this section we outline our empirical approach for exploring the performance of the long-horizon consumption-based asset pricing framework. Moment restrictions necessary to estimate any model for the stochastic discount factor by the Generalized Method of Moments (GMM) can be derived from Euler equations similar to Equation (1.3). Nonetheless, we opt for the slightly different GMM estimation strategy employed by Parker and Julliard (2005), using moment conditions based on the expression for expected excess returns in Equation (1.4). There are three reasons for doing this: First, closely following Parker and Julliard's approach renders our empirical results comparable to theirs. Second, as we will illustrate below, their approach allows us to empirically disentangle a model's ability to explain the equity premium from its explanatory power for cross-sectional variation in stock returns. Third, this approach provides an intuitive interpretation of our GMM estimation results: Using the moment restrictions in Equation (1.4) implies that the difference between empirical and theoretical moments can be interpreted as errors in expected returns, which in turn are proportional to pricing errors. These pricing errors will be directly comparable across models. More specifically, consider the vector of unconditional moment restrictions

$$\mathrm{E}[h(\Theta_{t+1}, \mu_S, \gamma_S, \alpha_S)] = 0, \tag{1.6}$$

where Θ_{t+1} represents the data (the vector of N test asset excess returns and consumption growth), whereas the model parameters are given as μ_S (mean of the stochastic discount factor m_{t+1}^S) and γ_S (risk aversion parameter of the representative agent). For the nonlinear model introduced in Section 1.2.1, the (N+1) $\times$ 1 empirical moment function $h(.)$ is given by

$$h(\Theta_{t+1}, \mu_S, \gamma_S, \alpha_S) = \begin{bmatrix} \mathbf{R}^e_{t+1} - \alpha_S \iota_N + \frac{(m^S_{t+1} - \mu_S) R^e_{t+1}}{\mu_S} \\ m^S_{t+1} - \mu_S \end{bmatrix} \tag{1.7}$$

where $\mathbf{R}^e_{t+1}$ denotes the vector of N test asset excess returns and ι_N is an N-dimensional vector of ones.[6] Notice that the point estimate for α_S will be expressed in units of expected returns. Therefore, including the parameter in the moment function in Equation (1.7) allows us to directly determine the magnitude of a model's implied "equity premium puzzle", i.e., to investigate whether a candidate model is able to explain the overall level of test asset returns compared to the level of the risk-free rate.

We modify the estimation approach by Parker and Julliard (2005) in one important dimension. In a recent contribution, Lewellen, Nagel, and Shanken (2007) highlight the statistical problems associated with the common use of size and book-to-market sorted portfolios in the empirical asset pricing literature. In particular, given the strong factor structure of these portfolios, Lewellen, Nagel, and Shanken (2007) point out that any model incorporating factors that are strongly correlated with SMB and HML potentially produces a high cross-sectional R^2 when tested on these test assets. In order to avoid these problems, we expand the set of test assets to include industry portfolios along with the commonly used size and book-to-market sorted portfolios. This implies that our modified empirical approach provides a clearly tougher challenge for the candidate asset pricing models compared to Parker and Julliard (2005).

In addition to testing the nonlinear long-horizon consumption-based model, we also compare the empirical performance of the linearized LH-CCAPM in Equation (1.5) to traditional factor models such as the CAPM and the Fama and French (1993) model. The moment function for the three candidate factor models differs slightly from the nonlinear model, reflecting the linear approximation of the stochastic discount factor. Let $\mathbf{f}_{t+1}$ denote the vector of k factors, μ the vector of estimated factor means, and $\mathbf{b}$ the vector of coefficients measuring the marginal effect of the respective factors on the SDF. The (N+k) $\times$ 1 moment function can then be written as

[6]The last moment condition is intended to identify the mean of the SDF, i.e. there are N+1 moment conditions in total.

$$h(\Theta_{t+1}, \mu, \mathbf{b}, \alpha) = \begin{bmatrix} \mathbf{R}^e_{t+1} - \alpha_S \iota_N + \mathbf{R}^e_{t+1}(\mathbf{f}_{t+1} - \mu)'\mathbf{b} \\ \mathbf{f}_{t+1} - \mu \end{bmatrix}. \tag{1.8}$$

This moment function satisfies N+k unconditional moment restrictions given by

$$\mathrm{E}[h(\Theta_{t+1}, \mu, \mathbf{b}, \alpha)] = 0, \tag{1.9}$$

which can be used to estimate the parameters of the model by GMM. In this context, it is important to note that identification of the parameters of the linear model requires some normalization. Using demeaned factors in the moment function in Equation 1.8 achieves this, but it also implies that we have to correct standard errors for the fact that factor means are estimated along the way. Therefore we use the augmented moment function in Equation 1.8, which imposes additional restrictions on the deviation of factors from their estimated means.[7]

In general, the GMM framework allows for various choices of the matrix determining the weights of individual moments in the objective function. As discussed in detail in Cochrane (2005, Ch. 11), the particular choice of weighting matrix affects both statistical properties and economic interpretation of the estimates: Even though second or higher stage GMM estimates based on the optimal weighting matrix of Hansen (1982) are efficient, they are difficult to interpret economically as they imply pricing some random combination of reweighted portfolios. Instead, relying on first stage estimates with equal weights compromises efficiency while maintaining the economic interpretation of empirical tests. Therefore, our discussion of empirical results in Section 1.4 centers on first stage GMM estimates. In addition, we also report results from the "test of overidentifying restrictions" based on iterated GMM estimation as a test of overall model fit. An alternative advocated by Hansen and Jagannathan (1997) is to use the inverse of the second moment matrix of returns as a first stage weighting matrix. This approach allows us to compute the corresponding Hansen-Jagannathan distance, which serves as an additional metric for model comparison.

[7] For a detailed discussion of this issue see Cochrane (2005, Ch. 13) and Yogo (2006, Appendix C).

1.4 Empirical Analysis

1.4.1 Data

This section provides a detailed overview of the data used in this paper. Data on personal consumption expenditure are available from national institutions in the respective country: the US Bureau of Economic Analysis (BEA), the UK Office for National Statistics (ONS), and the Federal Statistical Office (Destatis) in Germany. As is customary in the literature on consumption-based asset pricing, we use a measure of household's consumption of non-durable goods and services obtained from the official statistics. We divide by quarterly population figures to express consumption in per capita terms. Finally, all consumption time-series are deflated by the respective consumer price index.

While data on different consumption categories (nondurables, durables and services) are readily available at the quarterly frequency for both the US and the UK, this is not the case for Germany. We therefore use detailed annual data on personal consumption expenditures for different items to construct the share of nondurables and services in total consumption per annum. In order to estimate quarterly per capita expenditure on nondurables and services, we assign the same share to all quarterly total expenditure observations within a given year.[8] Another important aspect is the effect of Germany's reunification on consumption data. We correct for the negative outlier in the one-period (per capita) consumption growth rate due to the reunification using interpolation as in Stock and Watson (2003). Longer-horizon growth rates are then based on the corrected series.

Our choice of test assets is mainly guided by two considerations. First, our aim is to analyze the ability of the long-horizon CCAPM to price the cross-section of stock returns in major financial markets outside the United States. Second, following the suggestions of Lewellen, Nagel, and Shanken (2007), we use a broad set of test assets

[8]We experimented with various other matching procedures including quadratic polynomials and cubic splines, but found only negligible differences.

including portfolios sorted on both book-to-market and size as well as industry. This choice is intended to avoid problems arising from strong commonalities in size and book-to-market sorted portfolios.

As is standard in the empirical literature, our set of test assets contains 25 US value and size portfolios introduced by Fama and French (1993). Similar portfolios capturing both size and value premia are constructed by Dimson, Nagel, and Quigley (2003) for the United Kingdom[9] and by Schrimpf, Schröder, and Stehle (2007) for Germany. The total number of listed stocks in the UK and Germany is much smaller than in the US. Therefore, in both cases, stocks are sorted into merely 16 portfolios in order to avoid potential biases in portfolio returns. For comparisons with traditional asset pricing models such as the CAPM and the Fama and French (1993) three factor model, we obtain data on market returns, the excess return of small over big market capitalization firms (SMB), and the excess return of high versus low book-to-market firms (HML) from the same sources.

Returns on ten US industry portfolios sorted according to SIC codes are available from Kenneth French's website.[10] In case of the UK, we use seven industry portfolios obtained from Datastream which are available for the longest possible sample period matching the one of the other UK test assets. Our industry portfolios for the German stock market are obtained from the German Finance Database (Deutsche Finanzdatenbank) maintained at the University of Karlsruhe.[11] We compute excess returns on all portfolios using a country-specific proxy for the risk-free rate: For the US and the UK, we use a 3-month T-bill rate and, in the case of Germany, a 3-month money market rate provided by Deutsche Bundesbank is used. Finally, we compute real returns using the respective national consumer price index (CPI).[12]

[9] Returns on the 16 portfolios as well as Market, HML and SMB factors can be downloaded from Stefan Nagel's webpage: `http://faculty-gsb.stanford.edu/nagel`

[10] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/`

[11] The sample periods for test asset returns cover 1947:Q2 - 2001:Q4 for the US, 1965:Q2 - 2001:Q1 for the UK, and 1974:Q2 - 2001:Q1 for Germany. The overall sample period, however, is longer due to the long-horizon consumption growth (up to S) aligned to the returns: US (2004:Q3), UK (2003:Q4), GER (2003:Q4).

[12] CPI data for the US, the UK and Germany are available from the BEA, the IMF International Financial Statistics and the OECD Economic Outlook, respectively.

1.4.2 Empirical Results: Non-Linear Model

As pointed out in Section 1.3, we estimate the nonlinear LH-CCAPM for each of the three markets separately using the Generalized Method of Moments (GMM). Our discussion of empirical results focuses mainly on three aspects: a candidate model's ability to explain the equity premium ($\hat{\alpha}$), the plausibility of the estimated risk-aversion parameter ($\hat{\gamma}$), and the cross-sectional explanatory power as reflected by the cross-sectional R^2 and pricing error plots.[13] In addition, we report results from J-tests based on iterated GMM estimates, the root mean squared error (RMSE) from first stage GMM estimation, and the HJ-distance metric proposed by Hansen and Jagannathan (1997).

Our results for the US, reported in Table 1.1, complement the evidence in Table 1 of Parker and Julliard (2005) and provide a reassessment of their findings under our modified empirical approach.[14] It is important to keep in mind that we use an expanded set of test assets by adding 10 industry portfolios to the usual 25 Fama-French portfolios. As evinced by Table 1.1, the risk-aversion estimate for the standard CCAPM (S=0) is rather large, mirroring previous results in the literature. It is worth noting, however, that the estimated RRA coefficient typically decreases to substantially lower values as we move from short to long-term consumption risk. Moreover, the precision of the estimates tends to increase with the horizon. As the significant $\hat{\alpha}$ estimates show, a major limitation of the LH-CCAPM is the failure to explain the "equity premium", i.e. the overall level of stock returns in relation to the risk-free rate. In contrast to results reported by Parker and Julliard (2005), its magnitude hardly declines as the consumption growth horizon increases. Thus, the model leaves unexplained a substantial fraction of the excess return of stocks over the risk-free rate.[15]

Most importantly, however, our results presented in Table 1.1 suggest that the singular

[13] Computation of the cross-sectional R^2 in the GMM estimation framework follows Jagannathan and Wang (1996) and Parker and Julliard (2005).

[14] In order to render our results comparable across countries, we limit the horizon at which long-run consumption risk is measured to 11 quarters.

[15] The J-test rejects all short and long-horizon specifications of the CCAPM, which is a common finding in the empirical asset pricing literature: Even the best performing models such as the Fama-French three factor model are often rejected by formal statistical tests [e.g. Lettau and Ludvigson (2001b)].

Table 1.1: Consumption Risk and US Stock Returns - Nonlinear LH-CCAPM

Horizon	$\hat{\alpha}$ (std. err.)	$\hat{\gamma}$ (std. err.)	R^2	RMSE	HJ-Dist. (p-value)	J (p-value)
0	0.022	45.410	0.07	0.529	0.587	112.417
	(0.005)	(59.882)			(0.000)	(0.000)
1	0.019	30.720	0.09	0.525	0.584	106.269
	(0.005)	(29.364)			(0.000)	(0.000)
3	0.018	22.575	0.09	0.523	0.588	112.372
	(0.006)	(22.189)			(0.001)	(0.000)
5	0.018	20.626	0.11	0.520	0.586	110.750
	(0.005)	(18.728)			(0.005)	(0.000)
7	0.018	20.719	0.14	0.508	0.584	109.739
	(0.005)	(15.657)			(0.009)	(0.000)
9	0.019	20.525	0.17	0.500	0.584	110.940
	(0.004)	(12.488)			(0.012)	(0.000)
11	0.019	20.391	0.20	0.493	0.579	107.299
	(0.004)	(11.031)			(0.028)	(0.000)

Note: The reported values for $\hat{\alpha}$, $\hat{\gamma}$, R^2, and the Root Mean Squared Error (RMSE) are computed using equal weights across portfolios (first stage GMM). The HJ-Distance is based on first stage GMM estimation using the weighting matrix proposed by Hansen and Jagannathan (1997), the J-statistic on iterated GMM estimation. The risk-free rate is assumed to be constant. The sample period is 1947:Q2 - 2001:Q4 for returns and 1947:Q2 - 2004:Q3 for quarterly consumption.

use of size and book-to-market portfolios [as in Parker and Julliard (2005)] may overstate the empirical performance of the long-horizon CCAPM: If we include industry portfolios in our set of test assets, as advocated by Lewellen, Nagel, and Shanken (2007), we only find moderate improvements of the consumption-based asset pricing approach as the horizon of long-horizon consumption risk increases. Accordingly, the estimated R^2 reaches a maximum of 20% at a horizon of eleven quarters, which is half the value reported by Parker and Julliard (2005) for the same horizon. Therefore, the main empirical success of the the LH-CCAPM seems to lie in more plausible estimates of the coefficient of relative risk-aversion, while the model's performance to explain the value premium still remains rather poor.

Next, we provide estimation results on the performance of the LH-CCAPM for the cross-section of returns in the UK and Germany, where previous literature on cross-

Table 1.2: Consumption Risk and UK Stock Returns - Nonlinear LH-CCAPM

Horizon	$\hat{\alpha}$ (std. err.)	$\hat{\gamma}$ (std. err.)	R^2	RMSE	HJ-Dist. (p-value)	J (p-value)
0	0.025	14.787	0.09	0.671	0.505	48.102
	(0.009)	(27.133)			(0.028)	(0.001)
1	0.024	3.685	0.01	0.700	0.501	45.177
	(0.009)	(22.583)			(0.034)	(0.002)
3	0.021	15.012	0.14	0.654	0.500	49.357
	(0.010)	(17.637)			(0.030)	(0.000)
5	0.023	5.651	0.05	0.686	0.498	47.964
	(0.008)	(14.625)			(0.035)	(0.001)
7	0.021	8.950	0.13	0.656	0.497	48.309
	(0.008)	(12.054)			(0.035)	(0.001)
9	0.023	4.517	0.07	0.680	0.499	47.405
	(0.007)	(11.782)			(0.029)	(0.001)
11	0.022	5.037	0.09	0.671	0.496	47.800
	(0.007)	(12.011)			(0.027)	(0.001)

Note: The reported values for $\hat{\alpha}$, $\hat{\gamma}$, R^2, and the Root Mean Squared Error (RMSE) are computed using equal weights across portfolios (first stage GMM). The HJ-Distance is based on first stage GMM estimation using the weighting matrix proposed by Hansen and Jagannathan (1997), the J-statistic on iterated GMM estimation. The risk-free rate is assumed to be constant. The sample period is 1965:Q2 - 2001:Q1 for returns and 1965:Q2 - 2003:Q4 for quarterly consumption.

sectional tests of consumption-based asset pricing models has been rather scarce.[16] Estimation results for the UK reported in Table 1.2 largely confirm our findings for the US. Even though the estimated coefficient of determination arrives at a peak at shorter consumption growth horizons of 3 and 7 quarters, the overall explanatory power of the LH-CCAPM remains comparably low. Moreover, the model cannot explain the overall level of UK stock returns. Nevertheless, the effect of long-horizon risk on risk-aversion estimates is again remarkable. If we measure consumption growth over a time period of at least 5 quarters following the return, the estimated risk-aversion coefficient declines to values around 5.

Table 1.3 summarizes the evidence on the empirical content of the long-horizon CCAPM framework for the German stock market. The results for the LH-CCAPM in Germany are rather in line with those for the US stock market discussed above. As evinced by the

[16]An exception is the work of Gao and Huang (2004), who use UK value and size portfolios, whereas other papers such as Hyde and Sherif (2005a,b) for the UK and Lund and Engsted (1996) for Germany estimate consumption-based models separately for each industry sector or market index.

Table, we find that the plausibility of parameter estimates varies with the consumption growth horizon. Most importantly, risk-aversion estimates tend to decline to more plausible levels as we increase the time period over which consumption growth is measured. However, this decrease is not monotonous. At the same time, the estimated cross-sectional R^2 also varies with the horizon and reaches a maximum of 22% for S=11.

Table 1.3: Consumption Risk and German Stock Returns - Nonlinear LH-CCAPM

Horizon	$\hat{\alpha}$ (std. err.)	$\hat{\gamma}$ (std. err.)	R^2	RMSE	HJ-Dist. (p-value)	J (p-value)
0	0.015	61.927	0.09	0.730	0.544	61.121
	(0.009)	(31.840)			(0.362)	(0.000)
1	0.013	59.990	0.16	0.701	0.545	43.436
	(0.008)	(36.956)			(0.317)	(0.017)
3	0.013	27.586	0.05	0.744	0.545	97.116
	(0.008)	(37.379)			(0.275)	(0.000)
5	0.013	11.850	0.05	0.745	0.552	44.760
	(0.008)	(27.171)			(0.216)	(0.013)
7	0.010	17.963	0.12	0.718	0.554	46.184
	(0.006)	(19.539)			(0.205)	(0.009)
9	0.012	11.482	0.09	0.726	0.551	45.088
	(0.006)	(16.736)			(0.203)	(0.012)
11	0.007	19.987	0.22	0.675	0.552	46.216
	(0.004)	(17.863)			(0.208)	(0.009)

Note: The reported values for $\hat{\alpha}$, $\hat{\gamma}$, R^2, and the Root Mean Squared Error (RMSE) are computed using equal weights across portfolios (first stage GMM). The HJ-Distance is based on first stage GMM estimation using the weighting matrix proposed by Hansen and Jagannathan (1997), the J-statistic on iterated GMM estimation. The risk-free rate is assumed to be constant. The sample period is 1974:Q2 - 2001:Q1 for returns and 1974:Q2 - 2003:Q4 for quarterly consumption.

Interestingly, even the canonical consumption-based model does not imply an "equity premium puzzle" for Germany. What is more, the relevant coefficient ($\hat{\alpha}$) is further reduced if long-horizon consumption risk is taken into account. Overall, the results for the UK and the German stock markets further corroborate our earlier conclusion that, even though the ability of the LH-CCAPM to account for size and value premia is rather limited, the modified model helps to obtain more sensible risk-aversion parameter estimates.

1.4.3 Empirical Results: Linearized Model

In order to facilitate comparison with traditional factor models for the stochastic discount factor, we also estimate the linearized version of the LH-CCAPM. Tables 1.4, 1.5, and 1.6 summarize estimation results assuming a constant risk-free rate, which implies a one-factor model where long-horizon consumption growth serves as the single risk factor. In general, estimates are in accordance with those obtained for the nonlinear model.

As discussed in the previous subsection, when required to price a broader cross-section of assets, the long-horizon risk CCAPM apparently has trouble explaining US excess returns (Table 1.4). Nevertheless, our results confirm those of Parker and Julliard (2005) in two other regards. First, the cross-sectional R^2 increases considerably for longer horizons. Second, GMM coefficient estimates suggest that the effect of consumption growth on the representative investor's stochastic discount factor is estimated more precisely if consumption risk is measured over longer time periods. Moreover, the estimate of the risk-aversion coefficient declines to more economically plausible values as the horizon S increases.

The explanatory power of the linearized LH-CCAPM for the cross-section of returns seems clearly weaker when tested on UK stock portfolios. Similar to estimation results for the nonlinear specification, the coefficient of determination is highest for horizons of 3 (12%) and 7 (9%) quarters. In addition, point estimates $\hat{b}$ suggest that the SDF is not systematically related to consumption risk, irrespective of the chosen horizon. Although implied risk-aversion estimates have high standard errors, they exhibit a considerable decline as we extend the horizon over which consumption risk is measured.

Table 1.4: Consumption Risk and US Stock Returns - Linearized LH-CCAPM: GMM Estimation

Horizon	$\hat{\alpha}$ (std. err.)	$\hat{b}$ (std. err.)	$\gamma_S^{implied}$ (std. err.)	R^2	RMSE	HJ-Dist. (p-value)	J (p-value)
0	0.020	72.090	52.321	0.11	0.517	0.474	119.988
	(0.005)	(62.263)	(32.565)			(0.014)	(0.000)
1	0.019	30.859	23.328	0.08	0.526	0.559	108.883
	(0.005)	(33.270)	(18.878)			(0.003)	(0.000)
3	0.019	21.694	14.902	0.09	0.524	0.586	113.831
	(0.005)	(21.972)	(10.241)			(0.002)	(0.000)
5	0.019	18.615	11.723	0.09	0.524	0.583	110.429
	(0.005)	(18.572)	(7.251)			(0.005)	(0.000)
7	0.018	19.783	10.785	0.12	0.514	0.582	110.177
	(0.005)	(17.042)	(4.946)			(0.009)	(0.000)
9	0.019	21.336	10.027	0.16	0.503	0.583	111.229
	(0.005)	(14.896)	(3.161)			(0.012)	(0.000)
11	0.018	22.635	9.269	0.20	0.491	0.578	109.098
	(0.005)	(13.076)	(2.079)			(0.027)	(0.000)

Note: The reported values for $\hat{\alpha}$, $\hat{b}$, $\gamma_S^{implied}$, R^2, and the Root Mean Squared Error (RMSE) are computed using equal weights across portfolios (first stage GMM). The HJ-Distance is based on first stage GMM estimation using the weighting matrix proposed by Hansen and Jagannathan (1997), the J-statistic on iterated GMM estimation. The sample period is 1947:Q2 - 2001:Q4 for returns and 1947:Q2 - 2004:Q3 for quarterly consumption.

Table 1.5: Consumption Risk and UK Stock Returns - Linearized LH-CCAPM: GMM Estimation

Horizon	$\hat{\alpha}$ (std. err.)	$\hat{b}$ (std. err.)	$\gamma_S^{implied}$ (std. err.)	R^2	RMSE	HJ-Dist. (p-value)	J (p-value)
0	0.025	17.255	15.771	0.09	0.671	0.505	48.135
	(0.009)	(34.993)	(29.159)			(0.028)	(0.001)
1	0.025	4.295	4.101	0.00	0.703	0.500	45.940
	(0.009)	(23.865)	(21.753)			(0.038)	(0.001)
3	0.021	15.448	11.531	0.12	0.661	0.504	49.325
	(0.010)	(19.431)	(10.654)			(0.027)	(0.000)
5	0.023	7.301	5.882	0.03	0.693	0.503	48.214
	(0.009)	(15.406)	(9.920)			(0.031)	(0.001)
7	0.021	7.960	5.890	0.09	0.672	0.504	48.869
	(0.009)	(12.344)	(6.660)			(0.032)	(0.001)
9	0.022	4.805	3.799	0.03	0.695	0.505	48.089
	(0.008)	(12.396)	(7.701)			(0.029)	(0.001)
11	0.021	5.436	4.001	0.03	0.694	0.506	48.291
	(0.008)	(11.685)	(6.275)			(0.026)	(0.001)

Note: The reported values for $\hat{\alpha}$, $\hat{b}$, $\gamma_S^{implied}$, R^2, and the Root Mean Squared Error (RMSE) are computed using equal weights across portfolios (first stage GMM). The HJ-Distance is based on first stage GMM estimation using the weighting matrix proposed by Hansen and Jagannathan (1997), the J-statistic on iterated GMM estimation. The sample period is 1965:Q2 - 2001:Q1 for returns and 1965:Q2 - 2003:Q4 for quarterly consumption.

Table 1.6: Consumption Risk and German Stock Returns - Linearized LH-CCAPM: GMM Estimation

Horizon	$\hat{\alpha}$ (std. err.)	$\hat{b}$ (std. err.)	$\gamma_S^{implied}$ (std. err.)	R^2	RMSE	HJ-Dist. (p-value)	J (p-value)
0	0.014	24.194	21.639	0.01	0.758	0.534	45.255
	(0.009)	(42.199)	(33.678)			(0.290)	(0.011)
1	0.015	76.771	44.036	0.16	0.701	0.527	46.526
	(0.010)	(50.094)	(16.159)			(0.382)	(0.008)
3	0.013	21.763	15.372	0.02	0.757	0.544	41.872
	(0.008)	(32.764)	(16.157)			(0.263)	(0.025)
5	0.013	15.934	10.996	0.03	0.754	0.551	42.907
	(0.008)	(29.381)	(13.839)			(0.221)	(0.020)
7	0.010	19.573	11.373	0.09	0.730	0.556	46.859
	(0.006)	(20.279)	(6.591)			(0.200)	(0.007)
9	0.012	13.894	8.527	0.05	0.743	0.555	45.818
	(0.006)	(17.544)	(6.404)			(0.193)	(0.010)
11	0.006	22.898	10.290	0.20	0.683	0.555	47.149
	(0.004)	(16.149)	(3.074)			(0.204)	(0.007)

Note: The reported values for $\hat{\alpha}$, $\hat{b}$, $\gamma_S^{implied}$, R^2, and the Root Mean Squared Error (RMSE) are computed using equal weights across portfolios (first stage GMM). The HJ-Distance is based on first stage GMM estimation using the weighting matrix proposed by Hansen and Jagannathan (1997), the J-statistic on iterated GMM estimation. The risk-free rate is assumed to be constant. The sample period is 1974:Q2 - 2001:Q1 for returns and 1974:Q2 - 2003:Q4 for quarterly consumption.

Results for the linearized version of the LH-CCAPM for the German stock market are provided in Table 1.6. As was the case for the nonlinear specification, the model has no problem explaining the the overall level of stock returns. Taking long-horizon risk into account improves the performance of the CCAPM in other respects. The empirical fit as measured by R^2 and RMSE is best for a consumption risk horizon of 11 quarters. Moreover, implied risk aversion appears to decrease with horizon (albeit in a non-monotonous fashion). If consumption risk is measured over 11 quarters, the coefficient of relative risk aversion is estimated at a rather low value of 10 which is half the point estimate obtained for the conventional CCAPM. Moreover, the significance of $\hat{b}$, the parameter measuring the effect of consumption growth on the SDF, is far higher for S=11 than for the canonical CCAPM.

All together, inference for the linearized LH-CCAPM suggests that long-horizon consumption risk helps improve the empirical performance of the consumption-based model to some extent. Even though detailed empirical results differ across countries, some common patterns emerge. Most notably, measuring consumption risk over several quarters following the return helps to obtain much more plausible estimates of the representative investor's risk-aversion coefficient. This result is in accordance with recent evidence presented by Rangvid (2008).

1.4.4 Comparison to Traditional Linear Factor Models

Empirical results for the linearized CCAPM can be directly compared to those for the Fama and French (1993) three-factor model and the traditional CAPM, which are summarized in Table 1.7.

Estimates for 35 US portfolios in Panel A are in line with previous evidence in the literature [e.g. Fama and French (1993) or Lettau and Ludvigson (2001b)]: While the Fama-French three factor model explains more than 50% of cross-sectional variation in returns, the standard CAPM performs extremely poorly. Accordingly, as shown in Figure 1.1, portfolio excess returns predicted by the CAPM appear to be almost unrelated to realized average excess returns. In contrast, fitted excess returns for the

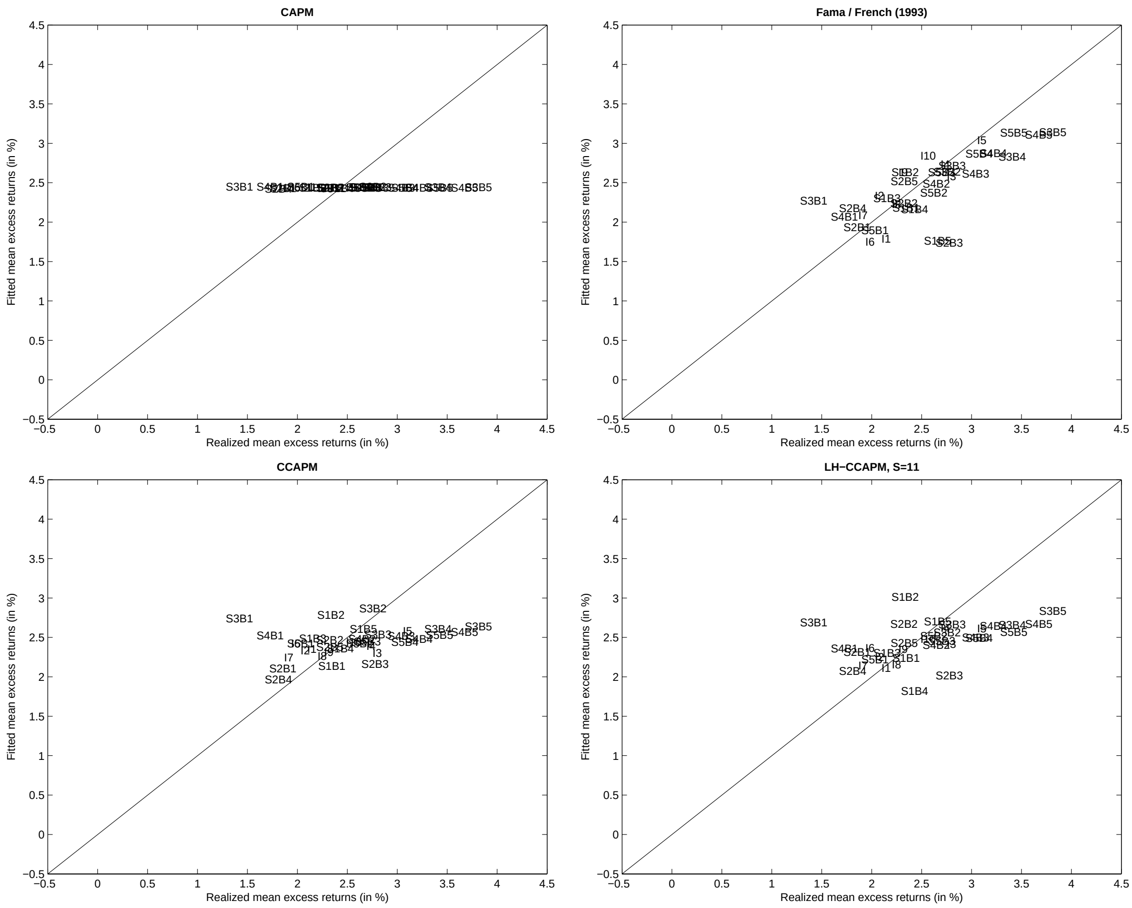
Fama-French model and, to a lesser extent, the LH-CCAPM line up more closely to the 45° line. At the same time, estimation results in Table 1.7 also indicate that, with the exception of HML, none of the proposed Fama-French factors seem to significantly affect the SDF of the representative US investor.

As illustrated in Figure 1.2, the high explanatory power of the Fama and French (1993) model typically found for the US is even higher for the cross-section of UK stock returns. First stage GMM estimates reveal that the model explains as much as 71% of cross-sectional variation in returns, compared to only 6% for the CAPM and 9% for the canonical CCAPM (S=0). However, as can be seen in Table 1.7, coefficients measuring the marginal impact of the respective financial risk factors on the SDF are not significant.

In the case of Germany (Panel C), the cross-sectional R^2 obtained for the long-run risk model - up to about 20% at 11 quarters - is clearly qualified by the high explanatory power of the three factor model (70%) and the CAPM (52%). Actually, the CAPM performs surprisingly well when tested on a cross-section of 28 industry, value and size portfolios, as reflected by significant $\hat{b}$ estimates. Nevertheless, the three factor model performs even better in that it provides an explanation for the overall level of returns relative to the risk-free rate and is not rejected by the test of overidentifying restrictions at the 5% significance level. Comparing all three models in terms of their explanatory power for German stock returns, the long-run consumption risk model does not provide any advantages over the two traditional linear models based on financial factors. Pricing error plots in Figure 1.3 confirm this conclusion as the magnitude of pricing errors is considerably lower for the three-factor model of Fama and French (1993).[17]

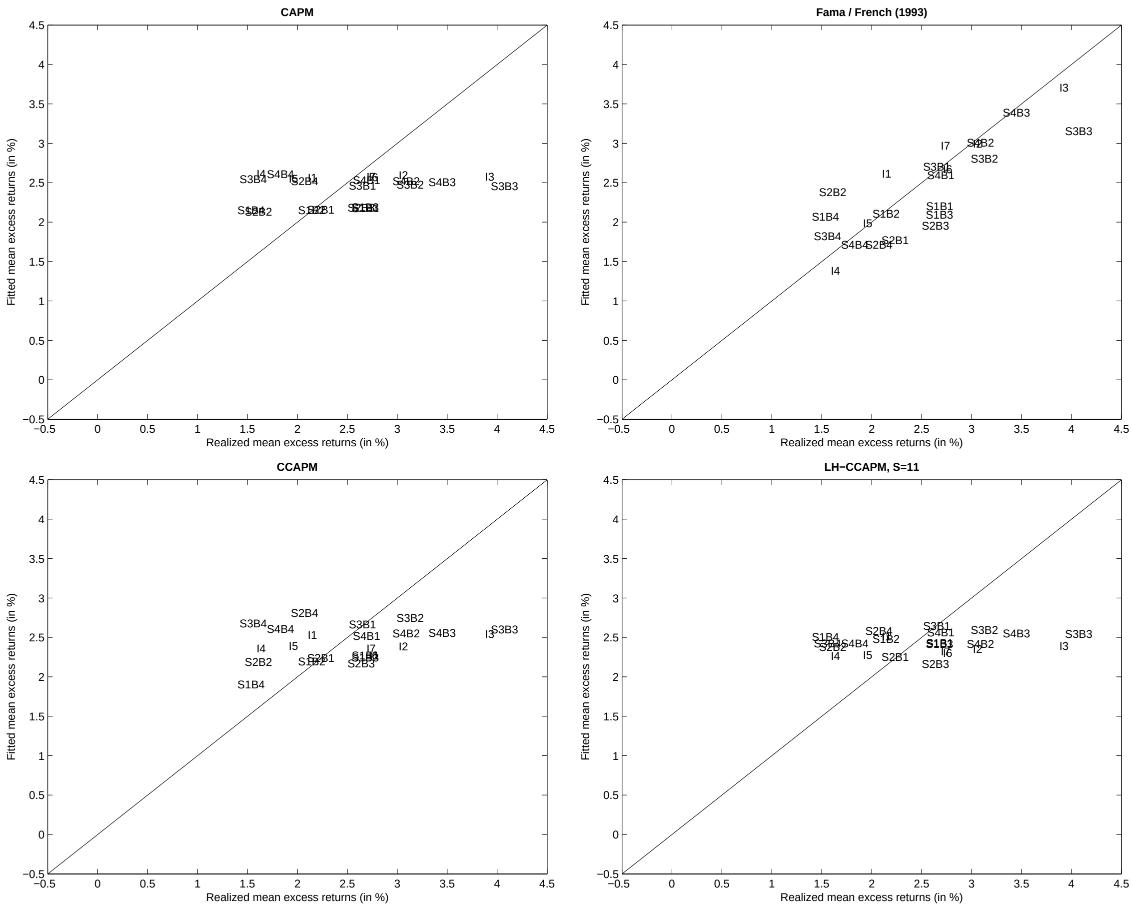
[17] However, models using macroeconomic factors will always be at a disadvantage to models using financial factors (Cochrane, 2007, p.7) due to a less precise measurement of macroeconomic variables. Moreover, these models allow for a more structural analysis of the economic determinants of risk premia, which typically cannot be delivered by models using merely financial factors.

Figure 1.1: Pricing Error Plots for US Stock Returns - Linearized LH-CCAPM and Traditional Linear Factor Models



Note: The figure compares realized excess returns on 25 value and size as well as 10 industry portfolios to those predicted by the CAPM, the Fama and French (1993) model, and the linearized LH-CCAPM (with constant risk-free rate) at various horizons. The portfolios are depicted in the following way: e.g. S1B1 refers to stocks in the smallest size and book-to-market Quintiles, while S5B5 refers to stocks in the largest size and book-to-market Quintiles; industry portfolios are depicted as I plus the corresponding industry number running from 1 to 10. Fitted excess returns are based on first stage GMM estimation with identity weighting matrix. The sample period is 1947:Q2 - 2001:Q4.

Figure 1.2: Pricing Error Plots for UK Stock Returns - Linearized LH-CCAPM and Traditional Linear Factor Models



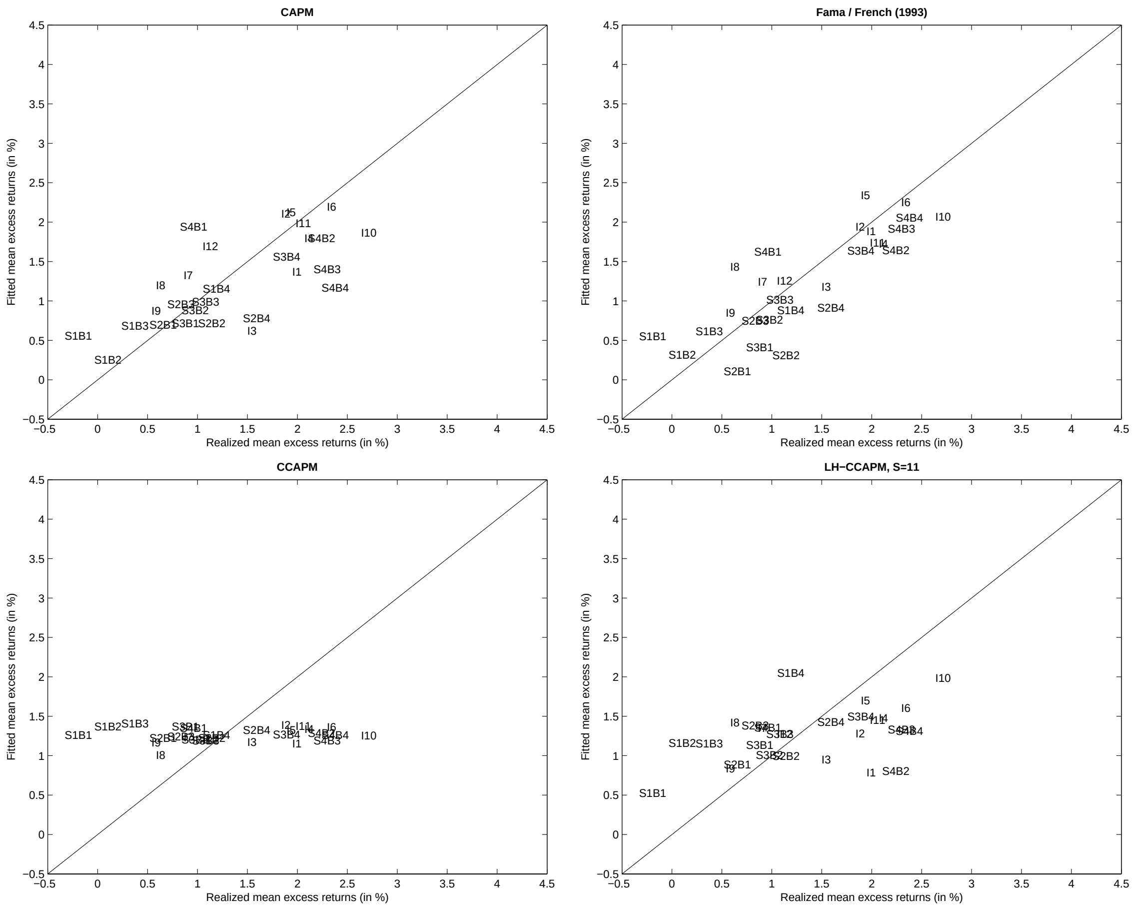
Note: The figure compares realized excess returns on 16 value and size as well as 10 industry portfolios to those predicted by the CAPM, the Fama and French (1993) model, and the linearized LH-CCAPM (with constant risk-free rate) at various horizons. The portfolios are depicted in the following way: e.g. S1B1 refers to stocks in the smallest size and book-to-market Quartiles, while S4B4 refers to stocks in the largest size and book-to-market Quartiles; industry portfolios are depicted as I plus the corresponding industry number running from 1 to 7. Fitted excess returns are based on first stage GMM estimation with identity weighting matrix. The sample period is 1965:Q2 - 2001:Q1.

Table 1.7: Traditional Linear Factor Models and German, UK and US Stock Returns - GMM Estimation

Model	$\hat{\alpha}$ (std. err.)	$\hat{b}_{m,e}$ (std. err.)	$\hat{b}_{SMB}$ (std. err.)	$\hat{b}_{HML}$ (std. err.)	R^2	RMSE	HJ-Dist. (p-value)	J (p-value)
				A. United States				
Fama-French	0.017	1.313	0.835	4.040	0.56	0.366	0.567	94.494
	(0.007)	(1.712)	(1.865)	(1.656)			(0.001)	(0.000)
CAPM	0.024	0.044			0.00	0.549	0.587	112.993
	(0.008)	(1.421)					(0.000)	(0.000)
				B. United Kingdom				
Fama-French	0.016	0.871	1.247	6.289	0.71	0.380	0.428	36.235
	(0.010)	(1.090)	(2.057)	(3.472)			(0.190)	(0.010)
CAPM	0.021	0.410			0.06	0.682	0.505	49.077
	(0.010)	(0.933)					(0.030)	(0.000)
				C. Germany				
Fama-French	−0.003	1.056	−4.731	3.342	0.70	0.419	0.515	35.499
	(0.007)	(2.132)	(3.810)	(2.547)			(0.322)	(0.061)
CAPM	−0.009	3.340			0.52	0.530	0.537	42.217
	(0.009)	(1.594)					(0.248)	(0.023)

Note: The reported values for $\hat{\alpha}$, $\hat{b}_{m,e}$, $\hat{b}_{SMB}$, $\hat{b}_{HML}$, R^2, and the Root Mean Squared Error (RMSE) are computed using equal weights across portfolios (first stage GMM). The HJ-Distance is based on first stage GMM estimation using the weighting matrix proposed by Hansen and Jagannathan (1997), the J-statistic on iterated GMM estimation. The sample period is 1974:Q2 - 2001:Q1 for Germany, 1965:Q2 - 2001:Q1 for the UK, and 1947:Q2 - 2001:Q4 for the US.

Figure 1.3: Pricing Error Plots for German Stock Returns - Linearized LH-CCAPM and Traditional Linear Factor Models



Note: The figure compares realized excess returns on 16 value and size as well as 10 industry portfolios to those predicted by the CAPM, the Fama and French (1993) model, and the linearized LH-CCAPM (with constant risk-free rate) at various horizons. The portfolios are depicted in the following way: e.g. S1B1 refers to stocks in the smallest size and book-to-market Quartiles, while S4B4 refers to stocks in the largest size and book-to-market Quartiles; industry portfolios are depicted as I plus the corresponding industry number running from 1 to 12. Fitted excess returns are based on first stage GMM estimation with identity weighting matrix. The sample period is 1974:Q2 - 2001:Q1.

Summing up, the empirical success of long-run consumption risk compared to the canonical CCAPM in terms of cross-sectional explanatory power is qualified by the astonishingly good performance of the factor model of Fama and French (1993).[18] At the same time, our results for the UK and the US confirm the bad performance of the CAPM typically found in empirical model comparisons. Surprisingly, we find that this model explains as much as 52% of cross-sectional variation in returns across German portfolios. In any case, measuring risk in stock returns as their covariance with long-run consumption growth leads to some – but generally limited – improvements over the canonical CCAPM in terms of overall empirical fit. Our results for international stock markets show that value and size premia still remain a major challenge for the LH-CCAPM.

1.5 Conclusion

Recent work by Parker and Julliard (2005) suggests that measuring consumption growth over several quarters following the return substantially improves the explanatory power of the consumption-based asset pricing paradigm. Their modified empirical setup is robust against various arguments as to why consumption expenditure may be slow to adjust to innovations in aggregate wealth. Besides, their model is closely related to the literature on long-run consumption risk, as it implies expressions for expected returns that are similar to the testable implications of long-run risk models with recursive utility such as Hansen, Heaton, and Li (2008).

Our work contributes to the literature on long-run consumption risks in three respects: First, by expanding the set of test assets to include industry portfolios, we take into account recent criticism regarding the widespread use of value and size portfolios as test assets (Phalippou, 2007; Lewellen, Nagel, and Shanken, 2007). Under our modified empirical approach, we find that long-horizon consumption risk falls short of providing a complete account of the cross-section of expected returns, especially the premium on value stocks. In this way, our findings suggest that the long-horizon consumption-based approach does not resolve the famous ”value premium puzzle”.

[18] A major disadvantage of Fama and French's three factor model is that there is still no full agreement in the literature about what the true risks underlying SMB and HML actually are. See, e.g., Petkova (2006) for a risk-based explanation in an empirical implementation of an ICAPM in the spirit of Merton (1973).

Second, evaluating the proposed CCAPM separately for three countries enables us to compare results across capital markets. In this sense, our findings provide additional out-of-sample evidence and address potential data-snooping concerns. Empirical results for Germany and the UK indicate that measuring consumption risk over longer horizons indeed helps increase the empirical performance of the CCAPM, albeit at modest levels. For both markets, estimated coefficients of determination remain below those obtained for the ad hoc factor model of Fama and French (1993).

Third, our analysis confirms the evidence of Parker and Julliard (2005), who find that point estimates of the investor's risk-aversion parameter vary with the time interval over which consumption growth is measured. In line with evidence reported by Rangvid (2008), we find that accounting for long-horizon consumption risk typically delivers more sensible estimates. This is true for all three equity markets considered in this study.

Summing up, accounting for long-horizon consumption risk within the CCAPM framework indeed seems to improve the model's cross-sectional explanatory power in certain ways. On the one hand, the model still falls short of providing an accurate description of size and value premia. On the other hand, the estimated risk aversion of an investor who is concerned about long-run consumption risk is much lower and therefore more plausible compared to the standard model. In this sense, long-horizon consumption risk appears to be a more accurate measure of macroeconomic risk factors in stock returns than contemporaneous consumption growth.

CHAPTER 2

INTERNATIONAL STOCK RETURN PREDICTABILITY UNDER MODEL UNCERTAINTY*

ABSTRACT

This chapter examines return predictability when the investor is uncertain about the right state variables. A novel feature of the model averaging approach used in this chapter is to account for finite-sample bias of the coefficients in the predictive regressions. Drawing on an extensive international dataset, we find that interest-rate related variables are usually among the most prominent predictive variables, whereas valuation ratios perform rather poorly. Yet, predictability of market excess returns weakens substantially, once model uncertainty is accounted for. We document notable differences in the degree of in-sample and out-of-sample predictability across different stock markets. Overall, these findings suggest that return predictability is not a uniform and a universal feature across international capital markets.

*This chapter is based on a joint paper with Qingwei Wang (University of Mannheim and Centre for European Economic Research).

2.1 Introduction

Empirical studies have asserted that a plethora of variables contain information about future excess returns in regressions of the form:

$$r_t = \alpha + \beta' x_{t-1} + u_t, \tag{2.1}$$

where r_t denotes the return of the aggregate stock market portfolio in excess of the risk-free rate, and x_{t-1} is a vector of predictive variables, such as the dividend yield, a term spread or certain macroeconomic variables.[1] Statistically significant β coefficients in Eq. (2.1) are interpreted as evidence for predictability and as evidence that risk premia are time-varying.[2]

Given the large number of variables proposed in the literature, a typical investor is confronted by a high degree of uncertainty on what the "right" state variables are. Moreover, the fact that so many variables have found to be valuable predictors of returns naturally raises the concern that the apparent predictability may well arise due to data-snooping rather than genuine variation of economic risk premia.[3] The aim of this paper is, therefore, to explore the robustness of several predictive variables in international stock markets in the context of model uncertainty. One of the major results of the paper is that few of the predictive variables put forth in the literature are truly robust predictors of returns. Second, substantial differences in the degree of in-sample and out-of-sample predictability can be observed across different stock markets.

[1]See e.g. Fama and French (1988), Fama and French (1989), Campbell and Shiller (1988a), Campbell and Shiller (1988b), Lettau and Ludvigson (2001a) etc.

[2]Based on the evidence for return predictability provided by the aforementioned articles, by the late 1990s the consensus among financial economists considered expected excess returns to be time-varying. In particular, predictability of market excess returns has been labeled as one of the "new facts in finance" (Cochrane, 1999).

[3]See e.g. Bossaerts and Hillion (1999), Ferson, Sarkissian, and Simin (2003) for critical views. Most notably, after a comprehensive out-of-sample forecast evaluation, Goyal and Welch (2008) come to the conclusion that knowledge of different state variables is of little use for a real-time investor. They interpret their findings as strong counterevidence against stock return predictability.

In this paper, we follow the spirit of the seminal work by Cremers (2002) and Avramov (2002) and use Bayesian model averaging in order to account for model uncertainty. Unlike the classical framework, the Bayesian approach does not assume the existence of a "true" model. By contrast, a-posteriori model probabilities can be derived for the different candidate models, which are then used to weight the coefficients accordingly in a composite model. In this way, model uncertainty can be accounted for in a coherent way.

A new feature of our approach is to account for finite-sample bias of the coefficients in the predictive regressions in a "frequentist" model averaging framework. A pure Bayesian model averaging framework as in Cremers (2002) and Avramov (2002) requires prior elicitation for the relevant parameters conditional on the different models. The specification of prior beliefs can be a problematic task when the set of models becomes very large.[4] Therefore, in order to reduce the impact of subjective prior information, we base our empirical study on Bayesian averaging of classical estimates (BACE) as in Sala-i-Martin, Doppelhofer, and Miller (2004). BACE can be seen as a limiting case of the Bayesian approach as the prior information becomes dominated by the data (See Leamer, 1978). Another less-attractive feature of the pure Bayesian model averaging approach as used by Cremers (2002) and Avramov (2002) is that it treats the predictive variables as exogenous, an assumption which is clearly invalid in the context of predictive regressions. How to conduct reliable inference in predictive regressions while taking the time-series properties of the predictive variables (such as the dividend yield) into account has been the subject of a great amount of recent research (See for instance Stambaugh, 1999; Campbell and Yogo, 2006; Lewellen, 2004; Amihud and Hurvich, 2004; Torous, Valkanov, and Yan, 2004; and Moon, Rubia, and Valkanov, 2006). In order to account for problems due to the persistence of the predictive variables, we estimate the models by classical OLS, where the coefficients are adjusted for finite-sample bias using the approach put forth in Amihud and Hurvich (2004). The bias-corrected coefficients in the particular models are then weighted by their posterior

[4] Avramov (2002) addresses this problem using an empirical Bayes approach which uses sample data for prior elicitation. In the Bayesian tradition, Cremers (2002) specifies subjective prior distributions based on different skeptical or optimistic beliefs about predictability.

model probabilities which are derived according to the BACE approach of Sala-i-Martin, Doppelhofer, and Miller (2004).

This paper also contributes to the existing literature by conducting a comprehensive analysis of stock return predictability in major international stock markets. It is fair to say that the profession's view on stock return predictability has been shaped for the most part by empirical studies on the US stock market. However, examining other important capital markets more closely may provide important additional insights, especially in a controversial field such as return predictability. Moreover, investigation of international markets also provides another way of guarding against data-snooping concerns. We thus examine the predictive performance of nine variables in a total of five international stock markets (France, Germany, Japan, United Kingdom, United States). Other important recent papers which provide evidence on international stock markets include Neeley and Weller (2000), Hjalmarsson (2004), Rapach, Wohar, and Rangvid (2005), Paye and Timmermann (2006), Giot and Petitjean (2006) or Ang and Bekaert (2007).[5] To the best of our knowledge, however, evidence on the effects of model uncertainty for return predictability in major international stock markets has been lacking so far.

There is a long list of variables which has been proposed in the literature on stock return predictability. In particular, valuation ratios such as the dividend yield or the earnings yield (e.g. Fama and French, 1988; Campbell and Shiller, 1988a; Lewellen, 2004), interest rate related variables such as short-term interest rates (e.g. Fama and Schwert, 1977; Hodrick, 1992; Ang and Bekaert, 2007) or default and term spreads (e.g. Campbell, 1987; Fama and French, 1989) have featured prominently in predictive regressions. Lamont (1998) has proposed the dividend-payout ratio as a predictive variable. The predictive power of stock market volatility has been studied by French, Schwert, and Stambaugh (1987). Pure macroeconomic variables used in predictive regressions include for instance the inflation rate (e.g. Fama, 1981), consumption-

[5] Hjalmarsson (2004) and Paye and Timmermann (2006) consider only four financial variables. Rapach, Wohar, and Rangvid (2005) focus merely on macroeconomic variables and do not consider financial valuation ratios. Giot and Petitjean (2006) consider finite-sample bias but do not address the issue of model uncertainty. Their set of predictive variables is limited to five financial variables.

wealth ratio (Lettau and Ludvigson, 2001a), price-GDP ratio (Rangvid, 2006), industrial production growth (e.g. Fama, 1990 or Avramov, 2002), and more recently the output gap (Cooper and Priestley, 2006). Variables motivated from a behavioral point of view (such as stock market sentiment as in Brown and Cliff, 2005) have also been shown to predict returns.

The brief review of the literature in the previous paragraph suggests that there is not much consensus on what the important variables are, or, put differently, that there is a tremendous model uncertainty in predictive regressions. In particular, some variables may appear significant in one specification and be insignificant in others, as researchers may only report their preferred specifications. As time elapses, more variables are sure to be added to the list of predictors.

While in-sample predictability is a debated topic, the question whether stock returns may be predictable out-of-sample (OOS) has been even more controversial. Empirical results on OOS predictability are mixed. Recently, several authors – most notably Goyal and Welch (2008) – argue against stock return predictability or time-varying risk premia based on the lacking evidence for out-of-sample predictability.[6] Campbell and Thompson (2007), however, find that once sensible restrictions are imposed on the predictive regression coefficients, the OOS forecast performance can be improved. It has also been argued that averaging forecasts of various models enhances out-of sample forecast performance substantially. Avramov (2002) finds that the out-of-sample performance of the weighted model is superior to the performance of models selected by information criteria and better than a naive benchmark. Another aim of the paper therefore is to look closer at the out-of-sample forecast performance of model averaging, in particular the time-variation of OOS performance in the spirit of Goyal and Welch (2008).

Our main results can be summarized as follows. Several notable differences with regard to return predictability are found across countries. We find that interest rate related

[6] Cochrane (2006) defends predictability based on the argument that even though predictability from the dividend-price ratio may be weak on statistical grounds, the fact that dividend growth is not predictable at all, may be interpreted as evidence that the variation of the dividend-price ratio is informative about future expected returns.

variables are usually among the most robust predictive variables in international stock markets, which corroborates recent results by Rapach, Wohar, and Rangvid (2005) and Ang and Bekaert (2007). Valuation ratios such as the dividend yield, however, perform rather poorly. There is also some evidence across countries that the output gap is related to expected returns and thus that risk premia vary with the state of the economy as pointed out recently by Cooper and Priestley (2006). The earnings yield often appears to be a more robust predictor than the dividend yield. Yet, predictability of market excess returns clearly weakens, once model uncertainty is accounted for. We only find some evidence for out-of-sample predictability by model averaging methods in the case of France but not for the remaining stock markets. Overall, our international analysis reveals that return predictability is not a uniform and a universal feature across international capital markets.

The remainder of the paper is structured as follows. Section II discusses the econometric framework of predictive regressions and how model uncertainty can be accounted for in a model averaging framework. Section III briefly discusses our data set. Empirical findings are discussed in Section IV. Section V concludes.

2.2 Methodology

In this paper we assess predictive ability in the conventional framework of predictive regressions. When there are multiple predictive variables (depending on the particular model $\mathcal{M}_j$), the predictive equation for future stock returns is given by

$$r_t = \alpha + \beta_j' x_{j;t-1} + u_{j,t}, \tag{2.2}$$

where r_t denotes the (log)-return on the market portfolio in excess of the (log) risk-free rate and $x_{j;t-1}$ is a k_j-dimensional vector of predictive variables, whose dimension and composition depends on the particular model $\mathcal{M}_j$. In total, we utilize κ different predictive variables which results in 2^κ different subsets, i.e. vectors of predictive variables $x_{j;t-1}$ $(j = 1, \cdots, 2^{\kappa})$. β_j is a k_j-dimensional vector of regression coefficients on the predictive variables. As is common in the extant literature, the vector of predictive variables is assumed to follow a first-order VAR:

$$x_{j;t} = \Theta_j + \Phi_j x_{j;t-1} + \nu_{j;t}. \tag{2.3}$$

Θ_j is a k_j-dimensional intercept and Φ_j is a $k_j \times k_j$ matrix with all eigenvalues smaller than one in absolute value to ensure stationarity of the process. The errors $(u_{j;t}, \nu'_{j;t})'$ are i.i.d. multivariate normal with mean zero.

2.2.1 Accounting for Model Uncertainty

We want to put ourselves in the position of an investor who is confronted by the voluminous literature on evidence for stock return predictability, yet is uncertain about which variables are actually of importance. In such a context, a Bayesian framework is attractive, since model uncertainty can be considered coherently. In a classical framework, however, the search for the "true model" usually implies running a series of model specification tests. Moreover, a classical approach is less appealing, because once a single model is determined, information in the remaining $2^{\kappa} - 1$ models is neglected. The approach taken in this paper is to combine the Bayesian feature of model averaging with coefficients estimated by classical OLS (BACE approach put forth by Sala-i-Martin, Doppelhofer, and Miller 2004).[7] The major advantage is that the BACE approach allows us correct for finite-sample bias of predictive slope coefficients, which is an issue previously neglected in the Bayesian model averaging literature as noted for instance by Stock and Watson 2004, p.34. Moreover, the approach largely avoids the drawback of the dependence on prior distributions (See Sala-i-Martin, Doppelhofer, and Miller 2004).

We explore the usefulness of $\kappa = 9$ candidate predictive variables in total, which implies that $2^{\kappa} = 512$ different model combinations are assessed. In a Bayesian framework,

[7]Bayesian and classical results are numerically identical when diffuse priors are specified.

posterior probabilities $p(\mathcal{M}_j|y)$ for each model $j = 1, \ldots, 2^\kappa$ can be derived. These posterior model probabilities are used in the Bayesian model averaging framework as weights of the composite model:

$$E[\beta|y] = \sum_{j=1}^{2^\kappa} p(\mathcal{M}_j|y)\beta_j|y, \tag{2.4}$$

where $\beta_j|y$ denotes the posterior mean of the predictive coefficients in the jth model. In the same way, the posterior standard deviation in the composite model is obtained from the corresponding diagonal element of the matrix

$$Var(\beta|y) = \sum_{j=1}^{2^\kappa} p(\mathcal{M}_j|y)[Var(\beta_j|y) + (\beta_j - E[\beta|y])(\beta_j - E[\beta|y])']. \tag{2.5}$$

Note that the posterior variance of the composite model in Eq. (2.5) contains essentially two components: the first term in the brackets accounts for estimation risk, whereas the second measures the variation of the predictive coefficients across the different models and thus accounts for model uncertainty.[8]

For determining the weights, the marginal likelihood for the different models $\mathcal{M}_j$ must be computed.[9] In the pure BMA framework, analytical solutions can be found only for certain prior distribution families.[10] In the "frequentist" model averaging framework of Sala-i-Martin, Doppelhofer, and Miller (2004), however, the marginal likelihood of a particular model is approximated using the Schwarz criterion as $exp(-0.5BIC_j)$. The posterior model probability for $\mathcal{M}_j$ can then be derived as

[8]Following Avramov (2002), we report posterior standard deviations with and without adjustment for model uncertainty in order to demonstrate the effects of accounting for model uncertainty in the inference.

[9]Mathematically, the marginal likelihoods can be obtained by integrating out the parameters from the combination of the likelihood and the prior conditional on the model.

[10]Avramov (2002), for instance, uses an "empirical Bayes" approach for prior elicitation, which uses data-information from the sample in order to determine the prior specification. Yet, such an approach can be criticized for using information of the dependent variable, which violates the rules of probability necessary for conditioning (Fernández, Ley, and Steel, 2001).

$$p(\mathcal{M}_j|y) = \frac{p(\mathcal{M}_j)exp(-0.5BIC_j)}{\sum_{i=1}^{2^\kappa} p(\mathcal{M}_i)exp(-0.5BIC_i)}, \tag{2.6}$$

where $p(\mathcal{M}_j)$ denotes the probability assigned to model j a-priori. As discussed in Sala-i-Martin, Doppelhofer, and Miller (2004), this formula can be derived in a standard g-prior framework taking the limit as the data-information increases relative to the prior information. Thus, using posterior model probabilities as in Eq. (2.6) essentially implies using a prior that becomes dominated by the data.

2.2.2 Finite-sample Bias in Predictive Regressions

In the following we outline our approach to correct for finite-sample bias in the BACE framework. In order to provide some intuition on the econometric problems arising from predictive variables which are not exogenous but rather predetermined, we first briefly review the single predictor case by Stambaugh (1999)

$$r_t = \alpha + \beta x_{t-1} + \epsilon_t, \tag{2.7}$$

where r_t denotes the (log)-return on the market portfolio in excess of the (log) risk-free rate and x_{t-1} is a predictive variable such as the dividend yield. The predictive variable itself is modeled as a first-order autoregressive process

$$x_t = \theta + \rho x_{t-1} + \xi_t. \tag{2.8}$$

The errors in Eq. (2.7) and Eq. (2.8) are assumed to be i.i.d. jointly normally distributed. Stambaugh (1999) then derives an analytical formula for the finite-sample bias of the predictive coefficient

$$E(\hat{\beta} - \beta) \approx \gamma E(\hat{\rho} - \rho), \tag{2.9}$$

where $\gamma = \frac{\sigma_{\epsilon\xi}}{\sigma_{\xi}^2}$ is the ratio of the covariance of the errors in both equations ($\sigma_{\epsilon\xi}$) and the variance (σ_{ξ}^2) of the error term ξ_t . As Eq. (2.9) shows, the bias of the predictive coefficients arises from the (downward) bias of the autoregressive parameter for the predictive variable $\hat{\rho}$ in combination with the correlation of the innovations in the predictive variable ξ_t and the error term ϵ_t in the predictive equation. The latter effect can be particularly severe in the case of valuation ratios (where the covariance between the shocks $\sigma_{\epsilon\xi}$ is typically strongly negative, which results in an upward bias of $\hat{\beta}$). A bias-corrected estimator $\hat{\beta}^s = \hat{\beta} + \hat{\gamma}(1 + 3\hat{\rho})/n$, where n denotes the sample size and $\hat{\gamma}$ is a sample estimate of γ, has been used e.g. by Giot and Petitjean (2006) in the single predictor case.

Since this paper is concerned about the issue of model uncertainty involving a multiplicity of variables, we work with the generalized case of multiple predictors as in Eq. (2.2) and Eq. (2.3). In order to obtain a bias-corrected estimator for the vector of predictive coefficients β_j in Eq. (2.2), we use the method recently put forth by Amihud and Hurvich (2004). Their approach amounts to running an augmented regression

$$r_t = \alpha + \beta_j' x_{j;t-1} + \phi_j' \nu_{j,t}^c + e_{j,t}, \tag{2.10}$$

which is equivalent to running the predictive regression in Eq. (2.2) augmented by a corrected $k_j \times 1$ residual series $\nu_{j,t}^c$. As shown by Amihud and Hurvich (2004), this procedure yields an unbiased estimator $\hat{\beta}_j^c$ for the vector of predictive coefficients. The residual series $\nu_{j,t}^c = x_{j;t} - (\hat{\Theta}_j^c + \hat{\Phi}_j^c x_{j;t-1})$ is based on a reduced-bias estimator for the autoregressive parameters $\hat{\Phi}_j$ in the multivariate autoregressive model in Eq. (2.3). Our estimate of $\hat{\Phi}_j^c$ follows the approach put forth by Amihud and Hurvich (2004) for the case when Φ_j is constrained to be diagonal.[11] Hence, the different series $x_{j,t}^i$ $(i = 1, \cdots, k_j)$ are considered separately. The individual error series are computed as $\nu_{j,t}^{c,i} = x_{j,t}^i - \hat{\theta}_i^c - \hat{\rho}_i^c x_{j,t-1}^i$. The autoregressive parameters are adjusted according to finite-sample bias by $\hat{\rho}_i^c = \hat{\rho}_i + (1 + 3\hat{\rho}_i)/n + 3(1 + 3\hat{\rho}_i)/n^2$. The reduced bias-estimator

[11] Allowing for a non-diagonal structure raises the need to estimate a multiplicity of parameters, in particular as k_j increases. This may result in a degradation of performance (See Amihud and Hurvich (2004)). We therefore impose a diagonal structure.

$\hat{\beta}_j^c$ is then obtained by regressing stock excess returns on the set of k_j lagged predictive and the corrected error proxies $\nu_{j,t}^i$ $(i = 1, \cdots, k_j)$. Standard errors for $\hat{\beta}_j^c$ are adjusted for the two-step procedure as proposed in Amihud and Hurvich (2004).

2.3 Empirical Results

2.3.1 Data

Our dataset comprises monthly and quarterly data for five international stock markets: France, Germany, Japan, United Kingdom and the United States. The dependent variables are (log) returns on broad stock indices in excess of the (log) short-term interest rate. Monthly summary statistics on the dependent variables and the predictive variables can be found in Table 2.1.

We assemble a data set of nine financial and macroeconomic predictive variables for the different international stock markets. The following variables comprise our set of predictors:

Interest rate variables: Difference between the yield on long-term government bonds and the three-month interest rate (term spread, TRM), short term interest rate relative to its 12-month backward-looking moving average (RTB), long-term government bond yield relative to its 12-month backward-looking moving average (RBR).

Valuation Ratios and other Financial Variables: Dividends paid over the past 12 months in relation to the current price (dividend yield, LDY) and earnings over the past 12 months in relation to the current price (earnings yield, LEY), both in logs. (Log) realized stock market volatility (LRV).

Macro Variables: Annual inflation rate (INF) based upon the Consumer Price Index, annual industrial production growth (IPG), estimate of the output gap obtained by the HP-filter (GAP).

The selection of variables is guided mainly by the previous US literature, as well as data availability. The main economic motivation for the different variables is that they are considered to be informative about future expected aggregate cash-flows in the economy or the discount rate applied to these cash-flows.[12] Hence, these variables have typically also featured prominently as state variables in empirical tests of intertemporal asset pricing models, e.g. Campbell (1996) or Campbell and Vuolteenaho (2004).

Table 2.1: Summary Statistics, Monthly

	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
France: 1973:02-2005:10										
MEAN	0.0044	1.0938	-0.0677	-0.0654	5.1294	0.9889	-6.1797	-3.3346	-2.5178	0.0456
STD	0.0621	1.2517	1.4400	0.8517	4.0892	4.4328	0.7752	0.3515	0.3275	2.8689
AC(1)	0.0798	0.9207	0.9171	0.9237	0.9966	0.8737	0.5835	0.9782	0.9673	0.8598
Germany: 1974:02-2004:12										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
MEAN	0.0031	1.3726	-0.1960	-0.0943	2.8285	1.2246	-6.5848	-3.7179	-2.7080	-0.1855
STD	0.0513	1.6839	1.1858	0.6146	1.8475	4.0470	0.9695	0.3530	0.2514	2.8891
AC(1)	0.0872	0.9723	0.9566	0.9054	0.9777	0.8178	0.7488	0.9824	0.9568	0.8354
Japan: 1973:02-2005:11										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
MEAN	0.0016	0.6874	-0.0750	-0.0850	3.0833	2.1889	-6.5518	-4.5379	-3.5609	-0.0424
STD	0.0522	1.1971	1.1642	0.6554	4.6170	6.2448	1.0202	0.5050	0.4687	4.1581
AC(1)	0.0838	0.9518	0.9611	0.9066	0.9890	0.9426	0.7242	0.9930	0.9905	0.9402
United Kingdom: 1973:01-2005:11										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
MEAN	0.0037	0.7989	-0.0258	-0.0796	6.6242	1.1718	-6.4382	-3.1629	-2.5142	0.0887
STD	0.0566	2.1353	1.5307	0.8933	5.2741	4.0886	0.8087	0.2748	0.3977	2.7035
AC(1)	0.1092	0.9774	0.9268	0.9089	0.9930	0.8562	0.6812	0.9747	0.9856	0.8691
United States: 1958:01-2005:12										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
MEAN	0.0044	1.6348	0.0000	0.0126	4.0387	3.0665	-6.6934	-3.5011	-2.7798	-0.1089
STD	0.0423	1.4360	1.1069	0.6257	2.7614	4.8664	0.8540	0.3962	0.3923	3.0940
AC(1)	0.0282	0.9493	0.9000	0.8764	0.9936	0.9609	0.8188	0.9920	0.9926	0.9637

Note: The table reports summary statistics of (log) market excess returns (EXRET) and predictive variables in five international stock markets. MEAN, STD, AC(1) denote the mean, standard deviation and first-order autocorrelation coefficient respectively. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Due to data availability, the different sample periods differ across markets. For most countries, the sample periods start in the early 1970s and end in mid 2000. The US

[12]Subsets of these variables are used for instance in Fama and Schwert (1977), Fama (1981), Fama and French (1988), Campbell and Shiller (1988a), Fama and French (1989)), Fama (1990), Hodrick (1992), Avramov (2002), Cremers (2002), Lewellen (2004), Rapach, Wohar, and Rangvid (2005), Cooper and Priestley (2006), Pastor and Stambaugh (2006), Ang and Bekaert (2007).

sample already starts in the late 1950s. Unfortunately, a default spread based on the yield difference of BAA and AAA rated corporate bonds (as used e.g. by Avramov 2002 or Cremers 2002) does not exist in the different international markets outside the US in a reasonable quality. For further detailed information on data sources and construction the reader is referred to Appendix A.

Table 2.1 provides monthly summary statistics on the mean, standard deviation and first-order autocorrelation of the particular state variables. The autocorrelation coefficients in the table show that some series, in particular valuation ratios and the inflation rate, exhibit a fairly strong degree of persistence. For this reason, taking the time series properties and potential finite-sample biases into account – as we do in this study – seems to be warranted.

2.3.2 In-sample Results: Return Predictability in International Stock Markets

First, we discuss the results of the in-sample analysis of return predictability in international stock markets. The only subjective element of the BACE approach is the choice of the a-priori expected model size $\bar{k}$, i.e. the researcher's belief of how many variables are a-priori likely to be included in the predictive model. We choose a rather moderate specification of this hyperparameter, consistent with the principle of parsimony prevailing in econometrics. We therefore set the a-priori expected model size to $\bar{k} = 2$ variables.[13] This implies a prior probability of inclusion of $\pi = 2/\kappa = 0.\bar{2}$ for each variable. The choice of the expected model size is linked to the a-priori model probability $p(\mathcal{M}_j)$ which is given as $p(\mathcal{M}_j) = \pi^{k_j}(1-\pi)^{\kappa - k_j}$.[14] It is important to note that a prior probability of inclusion smaller than 0.5 amounts to an a-priori down weighting of larger model specifications. This implies an additional penalty for highly parameterized models beside the penalty implied by the degree of freedom adjustment of the BIC.

[13]We discuss the sensitivity of the results to this choice of hyperparameter in section 3.3.

[14]In principle, one could also specify different prior probabilities of inclusion for the different variables based on economic considerations.

The tables for the different stock markets, which will be discussed in the following, are all organized in the same way. Panel A and C are based on monthly data while Panel B and D present results for quarterly data. Panel A and B report results for the composite model with bias-corrected slope coefficients. $\pi|y$ denotes the posterior probability of inclusion for each variable. The posterior probability of inclusion is defined as the total sum of the posterior probabilities of all models, in which the particular variable is included; it is computed as $\mathcal{C}'\mathcal{P}$, where $\mathcal{C}$ is a $2^\kappa \times \kappa$ matrix denoting inclusion (exclusion) of a particular variable in model j by 1 (0), and $\mathcal{P}$ is a $2^\kappa \times 1$ vector containing the posterior model probabilities $p(\mathcal{M}_j|y)$. Posterior means of the predictive coefficients in the weighted model based on Eq. (2.4) are reported in the second column of Panels A/B. The third and fourth column report posterior Bayesian t-ratios. Following Avramov (2002), we report both t-ratios based on posterior standard deviations which ignore model uncertainty and t-ratios adjusted for model uncertainty (see discussion in Section 2).

We also assess the robustness of the different predictive variables according to two other criteria. In Panels A/B we report the proportion of cases when the coefficient on a particular variable (every time it is included in one of the $j = 1, \cdots, 2^\kappa$ models) has the same sign as the posterior mean in the composite model (denoted as sgn prob. in the tables). Furthermore, we also report the fraction of cases across the different models when a classical t-statistic for the particular variable is greater than two in absolute value. This statistic serves as another indicator of the robustness or fragility of a particular predictive variable (Sala-i-Martin, Doppelhofer, and Miller, 2004). Panels C and D, presents the five top-performing model specifications which receive the highest posterior probability of all models. The models are defined by inclusion (1) or exclusion (0) of the specific variable. Moreover, the corresponding posterior model probabilities and the adjusted R^2 of the five top models are also reported.

France

Estimation results for the French stock market are provided in Table 2.2. As Panel A (monthly predictive regressions) shows, the only variable for which the posterior probability of inclusion $\pi|y$ rises, compared to the prior probability of inclusion, is the relative bond rate RBR. In the case of the other variables, inspection of the data leads us to retract our prior opinion about their usefulness. Panel C reports monthly results for the five best-performing model specifications. After having seen the data, the model which includes RBR as a single predictive variable receives a posterior model probability of more than 50%, which is greatly higher than the one of the next best model specifications. A negative relation of the realative bond rate and expected excess returns is reasonable from an economic point of view, given that higher yields on long-term bonds are typically reflected in a higher level of corporate loan rates and thus may have a negative impact on subsequent real activity. The relative bond rate together with the output gap is also significant according to a posterior t-ratio.

Robustness of a particular variable can also be assessed by the sign certainty probability which measures the fraction of cases where the coefficient on the particular variable (when included in one of the 2^{κ} Models) has the same sign as its coefficient in the weighted model. According to this criterion, the relative bond rate is again rather successful. The relative bond rate (RBR), the term spread (TRM), industrial production growth (IPG) and the output gap (GAP) all have sign certainty probabilities exceeding 90%, whereas several other popular predictors such as the dividend yield perform clearly worse. However, Table 2.2 also makes clear that none of the variables remains significant when the additional variability of estimates across models is accounted for.[15]

Panels B and D show that the evidence for predictability in the French stock market is somewhat weaker in the quarterly case. Again, only the relative bond rate receives a posterior probability of inclusion larger than $0.\bar{2}$. It is also worth noting that the

[15]This is a general result which holds for almost all predictive variables and almost all stock markets considered. In this way, we provide evidence consistent with Avramov (2002) that predictive regressions in finance are subject to a great deal of model uncertainty. Avramov also finds that many variables which appear to be significant, lose their significance once model uncertainty is considered.

earnings yield performs relatively well in-terms of sign certainty in the quarterly case.

Germany

Table 2.3 provides estimation results for the German stock market. As can be seen in Panel A and C of Table 2.3, predictability of monthly stock returns is fairly weak on statistical grounds. The case for predictability is clearly less pronounced than in the French stock market discussed in the previous subsection. The model receiving the highest posterior probability is the one without any lagged state variables (i.i.d. case). None of the variables in the monthly model receives a higher posterior inclusion probability compared to the prior inclusion probability of $\pi = 0.\bar{2}$. Among the variables considered only the relative bond rate (RBR) and the output gap (GAP) may be considered as significant according to a Bayesian t-ratio, but this does not hold true when the dispersion of coefficients across models is considered.

Similar to the French case, the relative bond rate is rather important in the quarterly regressions (Panel B of Table 2.3) where the probability of inclusion rises after having seen the data. Evidence for predictability with quarterly data is somewhat stronger than for monthly data. This can be seen from the result in Panel D that the most likely quarterly model is now the one which includes the relative bond rate. This model achieves an adjusted R^2 of about 5% in the quarterly regressions, which is quite high for the stock return predictability literature. Several variables appear quite robust with regard to sign certainty: The term spread (TRM), the relative bond rate (RBR), industrial production growth (IPG), and the two valuation ratios (LDY, LEY) have the same sign as the posterior mean in the composite model in more than 90% of all models in which they are included.

Table 2.2: Estimation Results, In-Sample: France

Panel A: Composite Model, Monthly							**Panel B: Composite Model, Quarterly**						
	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$		$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.033	0.010	1.205	0.557	0.969	0.027	TRM	0.036	0.028	1.008	0.511	0.949	0.004
RTB	0.042	-0.016	-1.535	-0.572	0.473	0.016	RTB	0.064	-0.070	-1.462	-0.601	0.625	0.023
RBR	0.743	-0.958	-3.494	-1.837	1.000	0.340	RBR	0.641	-2.400	-3.146	-1.461	1.000	0.414
INF	0.017	-0.001	-0.404	-0.272	0.180	0.117	INF	0.021	0.002	0.329	0.179	0.891	0.453
IPG	0.017	0.001	0.880	0.424	0.945	0.246	IPG	0.030	-0.002	-0.204	-0.127	0.039	0.238
LRV	0.042	0.022	1.311	0.581	0.441	0.000	LRV	0.031	0.024	0.598	0.376	0.250	0.000
LDY	0.003	0.001	0.245	0.197	0.434	0.000	LDY	0.015	0.034	0.772	0.438	0.500	0.000
LEY	0.003	0.001	0.190	0.159	0.742	0.000	LEY	0.018	0.051	0.890	0.475	0.910	0.000
GAP	0.116	-0.031	-2.347	-0.764	0.949	0.172	GAP	0.129	-0.104	-2.031	-0.732	0.816	0.027
Panel C: Top 5 Models, Monthly							**Panel D: Top 5 Models, Quarterly**						
TRM	0	0	0	0	0		TRM	0	0	0	0	0	
RTB	0	0	0	0	1		RTB	0	0	0	1	0	
RBR	1	0	0	1	0		RBR	1	0	0	0	1	
INF	0	0	0	0	0		INF	0	0	0	0	0	
IPG	0	0	0	0	0		IPG	0	0	0	0	0	
LRV	0	0	0	1	0		LRV	0	0	0	0	0	
LDY	0	0	0	0	0		LDY	0	0	0	0	0	
LEY	0	0	0	0	0		LEY	0	0	0	0	0	
GAP	0	0	1	0	0		GAP	0	0	1	0	1	
Prob	0.635	0.116	0.079	0.031	0.028		Prob	0.527	0.162	0.085	0.040	0.025	
$\bar{R}^2$	0.027	0.000	0.017	0.032	0.012		$\bar{R}^2$	0.065	0.000	0.038	0.027	0.068	

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 2.3: Estimation Results, In-Sample: Germany

Panel A: Composite Model, Monthly							**Panel B: Composite Model, Quarterly**						
	$\pi \| y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$		$\pi \| y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.029	0.006	1.198	0.554	0.934	0.469	TRM	0.038	0.021	0.980	0.504	0.922	0.418
RTB	0.025	-0.006	-1.064	-0.520	0.766	0.039	RTB	0.043	-0.032	-0.905	-0.473	0.777	0.066
RBR	0.188	-0.196	-2.418	-0.816	1.000	0.285	RBR	0.467	-1.896	-2.683	-1.103	1.000	0.527
INF	0.015	-0.001	-0.342	-0.254	0.488	0.000	INF	0.026	-0.004	-0.272	-0.212	0.504	0.000
IPG	0.017	-0.000	-0.241	-0.152	0.066	0.273	IPG	0.031	0.001	0.172	0.105	0.949	0.297
LRV	0.021	0.004	0.761	0.435	0.160	0.430	LRV	0.039	0.033	0.836	0.462	0.250	0.000
LDY	0.005	-0.003	-0.855	-0.462	1.000	0.004	LDY	0.007	-0.006	-0.292	-0.226	0.938	0.000
LEY	0.013	0.000	0.032	0.031	0.609	0.000	LEY	0.027	0.058	0.576	0.369	0.965	0.000
GAP	0.089	-0.016	-1.981	-0.707	0.816	0.133	GAP	0.144	-0.093	-1.995	-0.737	0.676	0.078
Panel C: Top 5 Models, Monthly							**Panel D: Top 5 Models, Quarterly**						
TRM	0	0	0	1	0		TRM	0	0	0	0	0	
RTB	0	0	0	0	1		RTB	0	0	0	0	1	
RBR	0	1	0	0	0		RBR	1	0	0	1	0	
INF	0	0	0	0	0		INF	0	0	0	0	0	
IPG	0	0	0	0	0		IPG	0	0	0	0	0	
LRV	0	0	0	0	0		LRV	0	0	0	0	0	
LDY	0	0	0	0	0		LDY	0	0	0	0	0	
LEY	0	0	0	0	0		LEY	0	0	0	0	0	
GAP	0	0	1	0	0		GAP	0	0	1	1	0	
Prob	0.635	0.163	0.075	0.022	0.020		Prob	0.364	0.322	0.093	0.026	0.024	
$\bar{R}^2$	0.000	0.013	0.008	0.002	0.001		$\bar{R}^2$	0.052	0.000	0.031	0.060	0.009	

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Japan

Results for the Japanese stock market are given in Table 2.4. As for Germany, there is no compelling evidence that monthly stock returns in Japan are predictable: The model with clearly the highest posterior probability in Panel C is the model with no explanatory variables (i.i.d.-model). The output gap (GAP) and the relative bond rate (RBR) are somewhat marginally important, but their explanatory power is fairly low. Note also that industrial production growth (IPG) and inflation (INF) are quite robust in terms of sign certainty probability.

With quarterly data, the evidence for predictability is even more modest. Again the model which does not include any predictors receives the highest probability a-posteriori. Only the output gap receives a higher posterior probability of inclusion than expected a-priori (Panel D of Table 2.4). However, model uncertainty again plays a substantial role as evinced by the adjusted Bayesian t-ratios. It is also worth noting that according to the sign certainty measure, the output gap must be considered as a rather fragile predictor.

United Kingdom

Table 2.5 reveals, that both for monthly and quarterly predictive regressions, the case for return predictability in the United Kingdom is quite weak. Panel C shows, that the largest posterior probability in the monthly regressions is assigned to the i.i.d.-model (as in the case of monthly regressions for Germany and Japan). Contrary to the countries discussed so far, interest rate variables do not show up among the most prominent predictors, which confirms the recent findings by Giot and Petitjean (2006) based on univariate return prediction models. By contrast, the dividend yield (LDY) has some predictive content for future stock returns in the UK. Yet, as before, accounting for model uncertainty greatly reduces the evidence for predictability and explanatory power of return prediction models in the UK is rather low.

Table 2.4: Estimation Results, In-Sample: Japan

Panel A: Composite Model, Monthly							**Panel B: Composite Model, Quarterly**						
	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$		$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.015	0.002	0.450	0.303	0.809	0.309	TRM	0.023	-0.003	-0.166	-0.137	0.434	0.066
RTB	0.055	-0.022	-1.694	-0.648	0.574	0.105	RTB	0.035	-0.021	-0.696	-0.377	0.582	0.102
RBR	0.192	-0.194	-2.479	-0.826	0.824	0.051	RBR	0.048	-0.083	-1.255	-0.571	0.664	0.000
INF	0.030	-0.002	-1.391	-0.592	1.000	0.371	INF	0.026	-0.003	-0.521	-0.345	0.883	0.031
IPG	0.023	0.001	0.944	0.424	0.910	0.336	IPG	0.041	0.007	0.941	0.405	0.895	0.348
LRV	0.022	-0.006	-1.017	-0.510	0.375	0.000	LRV	0.025	-0.013	-0.527	-0.348	0.703	0.027
LDY	0.003	-0.001	-0.584	-0.365	0.793	0.230	LDY	0.017	0.003	0.080	0.071	0.418	0.027
LEY	0.005	-0.001	-0.296	-0.185	0.238	0.480	LEY	0.020	0.012	0.307	0.235	0.840	0.500
GAP	0.256	-0.043	-2.603	-0.887	0.492	0.219	GAP	0.309	-0.182	-2.448	-0.909	0.645	0.195
Panel C: Top 5 Models, Monthly							**Panel D: Top 5 Models, Quarterly**						
TRM	0	0	0	0	0		TRM	0	0	0	0	0	
RTB	0	0	0	1	0		RTB	0	0	0	1	0	
RBR	0	0	1	0	0		RBR	0	0	1	0	0	
INF	0	0	0	0	1		INF	0	0	0	0	0	
IPG	0	0	0	0	0		IPG	0	0	0	0	1	
LRV	0	0	0	0	0		LRV	0	0	0	0	0	
LDY	0	0	0	0	0		LDY	0	0	0	0	0	
LEY	0	0	0	0	0		LEY	0	0	0	0	0	
GAP	0	1	0	0	0		GAP	0	1	0	0	1	
Prob	0.467	0.213	0.163	0.045	0.017		Prob	0.531	0.253	0.034	0.024	0.019	
$\bar{R}^2$	0.000	0.015	0.014	0.007	0.002		$\bar{R}^2$	0.000	0.037	0.007	0.001	0.047	

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

United States

As shown by Table 2.6, evidence for in-sample return predictability is clearly stronger in the US compared to other international stock markets such as Germany, Japan or the UK. Variables which appear important after having seen the data include the relative bond rate (RBR) and, most notably, the output gap (GAP). The output gap is the only variable which can be considered as a significant predictor once model uncertainty is accounted for. It receives a posterior probability of inclusion of more than 80%, which is a substantial upward revision of the prior probability of inclusion.[16] The output gap also appears to be a less fragile predictor in the US compared to the other countries. It is also worth noting that the earnings yield (LEY) provides more explanatory power than the dividend yield (LDY). Several other variables – such as the relative bond rate (RBR), inflation (INF), and industrial production growth (IPG) – are important when model uncertainty is ignored, but lose their significance once model uncertainty is considered.

When we consider predictive models at a quarterly horizon, the output gap (GAP) again appears as an important variable a-posteriori and also survives the model uncertainty adjustment. Also note that the relative bond rate is less important in the quarterly regressions. Panels A and B further show that the earnings yield appears to be very robust with regard to sign certainty, which holds both in the monthly and the quarterly models.

[16]Thus, our results corroborate the results of the recent paper by Cooper and Priestley (2006) who find that risk-premia are varying with the output-gap. Good economic conditions as measured by the output gap are associated with low risk premia.

Table 2.5: Estimation Results, In-Sample: United Kingdom

Panel A: Composite Model, Monthly	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$	**Panel B: Composite Model, Quarterly**	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.017	0.001	0.660	0.399	0.902	0.156	TRM	0.029	0.004	0.558	0.302	0.762	0.105
RTB	0.014	-0.000	-0.138	-0.089	0.113	0.152	RTB	0.018	-0.001	-0.095	-0.049	0.098	0.281
RBR	0.026	-0.010	-1.345	-0.580	0.781	0.148	RBR	0.026	-0.011	-0.708	-0.405	0.785	0.063
INF	0.016	0.000	0.509	0.317	0.594	0.008	INF	0.023	0.002	0.817	0.426	0.805	0.066
IPG	0.021	0.002	1.219	0.556	0.926	0.398	IPG	0.023	0.002	0.771	0.387	0.875	0.328
LRV	0.021	0.006	0.883	0.451	0.250	0.012	LRV	0.033	-0.019	-1.056	-0.520	0.922	0.414
LDY	0.203	0.363	1.725	0.724	0.633	0.020	LDY	0.559	3.974	2.169	1.101	0.992	0.289
LEY	0.043	0.030	0.993	0.509	1.000	0.480	LEY	0.083	0.269	1.429	0.619	1.000	0.441
GAP	0.073	-0.016	-2.038	-0.701	0.316	0.313	GAP	0.041	-0.023	-1.747	-0.621	0.320	0.273

Panel C: Top 5 Models, Monthly						**Panel D: Top 5 Models, Quarterly**					
TRM	0	0	0	0	0	TRM	0	0	0	0	0
RTB	0	0	0	0	0	RTB	0	0	0	0	0
RBR	0	0	0	0	1	RBR	0	0	0	0	0
INF	0	0	0	0	0	INF	0	0	0	0	0
IPG	0	0	0	0	0	IPG	0	0	0	0	0
LRV	0	0	0	0	0	LRV	0	0	0	0	1
LDY	0	1	0	0	0	LDY	1	0	0	0	1
LEY	0	0	0	1	0	LEY	0	0	1	0	0
GAP	0	0	1	0	0	GAP	0	0	0	1	0
Prob	0.604	0.185	0.060	0.039	0.019	Prob	0.481	0.268	0.070	0.032	0.019
$\bar{R}^2$	0.000	0.015	0.007	0.008	0.002	$\bar{R}^2$	0.062	0.000	0.036	0.016	0.063

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 2.6: Estimation Results, In-Sample: United States

Panel A: Composite Model, Monthly							**Panel B: Composite Model, Quarterly**						
	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$		$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.014	0.001	0.387	0.254	0.789	0.059	TRM	0.022	-0.000	-0.035	-0.026	0.246	0.172
RTB	0.025	-0.007	-1.450	-0.578	0.898	0.168	RTB	0.025	0.000	0.025	0.013	0.910	0.258
RBR	0.302	-0.278	-3.127	-0.982	1.000	0.809	RBR	0.060	-0.089	-1.622	-0.642	0.336	0.035
INF	0.086	-0.031	-3.787	-0.846	0.918	0.457	INF	0.011	-0.002	-0.912	-0.470	0.816	0.059
IPG	0.059	-0.006	-2.443	-0.701	0.852	0.254	IPG	0.028	-0.001	-0.142	-0.067	0.289	0.148
LRV	0.010	0.000	0.017	0.014	0.117	0.000	LRV	0.067	0.058	1.235	0.574	1.000	0.078
LDY	0.030	0.011	0.695	0.286	0.527	0.172	LDY	0.001	0.001	0.512	0.330	0.508	0.000
LEY	0.117	0.228	3.211	0.838	0.996	0.594	LEY	0.006	0.014	1.445	0.593	1.000	0.090
GAP	0.805	-0.181	-3.904	-2.179	1.000	0.621	GAP	0.938	-0.659	-3.829	-2.824	0.996	0.457
Panel C: Top 5 Models, Monthly							**Panel D: Top 5 Models, Quarterly**						
TRM	0	0	0	0	0		TRM	0	0	0	0	1	
RTB	0	0	0	0	0		RTB	0	0	0	0	0	
RBR	0	1	1	0	0		RBR	0	0	1	0	0	
INF	0	0	0	1	1		INF	0	0	0	0	0	
IPG	0	0	0	1	0		IPG	0	0	0	0	0	
LRV	0	0	0	0	0		LRV	0	1	0	0	0	
LDY	0	0	0	0	0		LDY	0	0	0	0	0	
LEY	0	0	0	1	1		LEY	0	0	0	0	0	
GAP	1	1	0	0	1		GAP	1	1	1	0	1	
Prob	0.544	0.136	0.103	0.034	0.025		Prob	0.777	0.055	0.037	0.024	0.016	
$\bar{R}^2$	0.029	0.037	0.023	0.046	0.046		$\bar{R}^2$	0.069	0.081	0.072	0.000	0.065	

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

2.3.3 Sensitivity to the Choice of Hyperparameter

The previous discussion of in-sample predictability and differences in the relevance of particular predictors across countries was based on a fairly moderate expected model size of two variables. In this sub-section, we analyze the robustness of our main findings to the specific choice of this hyperparameter $\bar{k}$ which is linked to the prior probability of inclusion π. For this purpose, we check whether our earlier conclusions on the relevance of a particular variable – as measured by a posterior probability of inclusion $\pi|y$ exceeding the prior probability of inclusion π – are affected by the choice of the expected model size. Table 2.7 reports posterior probabilities of inclusion of the predictor variables for different prior probabilities of inclusion π corresponding to model sizes with $\bar{k} = 2,\ 4,\ 6$ and 8 variables.

As shown in Table 2.7, our main conclusions on the relevance of a specific predictor are largely unaffected by the choice of the expected model size. Panel A for France, for instance shows that the relative bond rate can be considered an important predictor for almost all choices of prior probabilities of inclusion. There is not a single case where another predictive variable becomes relevant for a different choice of π , i.e. that there is an upward revision of the probability of inclusion after having seen the data. The same result holds true for the German (Panel B) and the British (Panel D) stock market. Results on the usefulness of the output gap in Japan are only slightly dependent on the choice of π but no other variable shows an upward revision of the probability of inclusion for different choices of hyperparameters. Results for quarterly predictive regressions for the US stock market (Panel E) are also largely unaffected. In the monthly case, however, the earnings yield and the inflation rate play a more prominent role in larger models, while the relative bond rate only serves as a significant predictor in the case of small expected model sizes.

Table 2.7: Sensitivity, Hyperparameter Expected Model Size

	Monthly				Quarterly			
π	0.222	0.444	0.667	0.889	0.222	0.444	0.667	0.889
Panel A:	France							
TRM	0.033	0.080	0.167	0.376	0.036	0.085	0.167	0.362
RTB	0.042	0.066	0.112	0.282	0.064	0.103	0.170	0.359
RBR	0.743	0.819	0.857	0.908	0.641	0.745	0.805	0.878
INF	0.017	0.043	0.098	0.299	0.021	0.054	0.122	0.348
IPG	0.017	0.049	0.125	0.422	0.030	0.069	0.145	0.383
LRV	0.042	0.113	0.250	0.597	0.031	0.082	0.182	0.475
LDY	0.003	0.004	0.005	0.006	0.015	0.020	0.027	0.055
LEY	0.003	0.004	0.006	0.008	0.018	0.033	0.056	0.107
GAP	0.116	0.174	0.280	0.592	0.129	0.194	0.291	0.538
Panel B:	Germany							
TRM	0.029	0.070	0.139	0.324	0.038	0.090	0.175	0.362
RTB	0.025	0.057	0.109	0.283	0.043	0.087	0.164	0.380
RBR	0.188	0.358	0.529	0.763	0.467	0.649	0.770	0.902
INF	0.015	0.039	0.088	0.241	0.026	0.069	0.155	0.419
IPG	0.017	0.047	0.110	0.344	0.031	0.085	0.204	0.552
LRV	0.021	0.054	0.115	0.271	0.039	0.098	0.198	0.418
LDY	0.005	0.010	0.017	0.033	0.007	0.011	0.014	0.024
LEY	0.013	0.033	0.069	0.171	0.027	0.063	0.124	0.275
GAP	0.089	0.182	0.300	0.553	0.144	0.255	0.413	0.731
Panel C:	Japan							
TRM	0.015	0.039	0.091	0.287	0.023	0.062	0.142	0.395
RTB	0.055	0.093	0.142	0.308	0.035	0.080	0.157	0.390
RBR	0.192	0.307	0.410	0.616	0.048	0.102	0.189	0.452
INF	0.030	0.071	0.147	0.391	0.026	0.067	0.145	0.372
IPG	0.023	0.072	0.174	0.473	0.041	0.125	0.287	0.651
LRV	0.022	0.062	0.145	0.399	0.025	0.065	0.134	0.295
LDY	0.003	0.004	0.006	0.008	0.017	0.029	0.039	0.071
LEY	0.005	0.008	0.011	0.009	0.020	0.032	0.036	0.044
GAP	0.256	0.419	0.572	0.804	0.309	0.535	0.718	0.890
Panel D:	UK							
TRM	0.017	0.043	0.094	0.266	0.029	0.071	0.155	0.460
RTB	0.014	0.035	0.085	0.298	0.018	0.043	0.102	0.402
RBR	0.026	0.061	0.126	0.346	0.026	0.066	0.152	0.476
INF	0.016	0.041	0.094	0.300	0.023	0.052	0.112	0.356
IPG	0.021	0.071	0.184	0.506	0.023	0.061	0.139	0.401
LRV	0.021	0.052	0.111	0.310	0.033	0.091	0.199	0.470
LDY	0.203	0.347	0.438	0.453	0.559	0.704	0.782	0.886
LEY	0.043	0.073	0.090	0.098	0.083	0.109	0.131	0.222
GAP	0.073	0.149	0.255	0.488	0.041	0.059	0.086	0.236
Panel E:	US							
TRM	0.014	0.036	0.081	0.238	0.022	0.058	0.132	0.336
RTB	0.025	0.048	0.098	0.260	0.025	0.059	0.129	0.330
RBR	0.302	0.399	0.438	0.567	0.060	0.133	0.264	0.518
INF	0.086	0.349	0.684	0.928	0.011	0.037	0.123	0.429
IPG	0.059	0.208	0.389	0.570	0.028	0.061	0.121	0.277
LRV	0.010	0.028	0.059	0.113	0.067	0.171	0.364	0.750
LDY	0.030	0.088	0.207	0.501	0.001	0.001	0.002	0.011
LEY	0.117	0.415	0.754	0.957	0.006	0.019	0.069	0.230
GAP	0.805	0.734	0.663	0.722	0.938	0.956	0.964	0.982

Note: The table contains detailed results on the sensitivity of estimation results with respect to the choice of the expected model size. For different prior probabilities of inclusion π corresponding to model sizes with 2, 4, 6 and 8 variables the posterior probabilities of inclusion are reported. The predictors include the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

2.3.4 Out-of-Sample Analysis of Return Predictability

The question whether predictability of stock returns exists out-of-sample (OOS) has been a much debated topic and results in the literature are mixed.[17] There are several theoretical reasons why OOS performance of stock return prediction models may be poor. Cochrane (2006), for instance shows by simulations that even in a world where risk premia are truely time-varying, the results of Goyal and Welch (2008) will occur frequently. Inoue and Kilian (2004) argue that in-sample predictability tests are more powerful than out-of sample tests and are therefore more trustworthy when assessing the existence of a predictive relationship. Another reason for poor OOS predictability may be temporal instability of the return prediction models.[18] We address the latter issue by studying the time-variation of OOS forecast errors in international stock markets using Net-SSE plots in the spirit of Goyal and Welch (2008).

It is not the purpose of this paper to discuss the entire debate in the literature or to take a particular side. Rather, we are interested in a thorough investigation of the performance of model averaging in the context of OOS predictability of excess returns. Avramov (2002), for instance, argues that averaging the forecasts of the different competing models in a Bayesian model averaging framework can substantially improve the out-of-sample forecast performance. Therefore, the main motivation of our analysis in this subsection is to reassess the findings by Avramov (2002) in the context of major international stock markets.

For the purpose of evaluating OOS forecast performance, we estimate the 2^{κ} models using a recursive scheme. The first ten years are used as initialization period. Afterwards, the models are estimated recursively. We compare the performance of several (conditional) models to the results of an unconditional (or naive) benchmark model which takes the prevailing historical mean as the forecast of the future excess return. The model-based forecasts include Bayesian averaging of OLS coefficients adjusted

[17] The recent predictability debate has been spurred by the question whether the documented (limited) in-sample predictability is of any use for an investor in real-time. See the different conclusions obtained by e.g. Goyal and Welch (2008) and Campbell and Thompson (2007).

[18] See also the recent papers by Paye and Timmermann (2006), Dangl, Halling, and Randl (2006) and Ravazzolo, Paap, van Dijk, and Franses (2006).

for finite-sample bias (BACE-adj), a conventional Bayesian model averaging approach (BMA) with g-prior specification[19], the individual model which receives the highest posterior model probability according to BMA (denoted as TOP), and an all-inclusive specification (ALL). Following Bossaerts and Hillion (1999), we also assess the performance of individual models selected by the conventional model selection criteria: Akaike criterion (AIC), Schwarz criterion (BIC), as well as the adjusted R^2. The corresponding (pseudo-) OOS forecasts are then evaluated according to several criteria for assessing forecast accuracy.

Table 2.8 reports the results of the evaluation of OOS performance for our international set of stock markets. The evaluation of forecast accuracy uses standard criteria. ME denotes the mean prediction error. Testing the significance of the ME amounts to testing the unbiasedness of the forecasts. Theil's U (TU) is the ratio of the mean square prediction error (MSPE) of the particular model-based forecast to the one of the naive benchmark model.[20] In order to provide an evaluation of directional accuracy of forecasts obtained by model averaging, we also report the fraction of times the direction of the dependent variable is correctly predicted by the model (denoted as Hit in the table). PT denotes the test-statistic for directional accuracy proposed by Pesaran and Timmermann (1992). Net-SSE plots are depicted in Figure 2.1. These graphs display the cumulated sum of the squared forecast errors of the benchmark model minus the squared forecast errors of the model of interest. One can use these plots to infer how the OOS performance of the predictive model evolves over time and where major forecast breakdowns occur. Periods where the line in the graph is upward sloping represent times when the conditional model outperforms the naive model in terms of squared forecast errors.

As the evaluation of the monthly forecasts in Table 2.8 shows, out-of-sample predictabil-

[19]The approach is similar to Cremers (2002). However, rather than motivating the g hyperparameter from economic reasoning, we follow recommended practice and set this parameter to $g = \max\{n, \kappa^2\}^{-1}$, where n denotes the sample size (See Fernández, Ley, and Steel 2001 or Koop 2003).

[20]Note that TU is merely a descriptive criterion. In the case of nested models, the mean square prediction error MSPE of the smaller nested model is expected to be smaller than the MSPE under the null of equal predictive power, a point raised by Clark and West (2007). This is due to the fact that the larger model needs to estimate parameters which are zero in population, which introduces noise in the forecasts.

Table 2.8: Estimation Results: Out-of-sample, Monthly

Panel A:	France						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	0.0010	0.0014	0.0015	-0.0050	-0.0012	0.0015	-0.0035
t-stat	0.2845	0.4114	0.4190	-1.4247	-0.3421	0.4314	-1.0058
TU	0.9947	0.9959	0.9993	1.0025	1.0019	0.9986	1.0035
Hit	0.5978	0.5941	0.5646	0.5830	0.5535	0.5720	0.5720
PT	1.0524	0.9375	0.3858	0.9105	0.3264	0.6230	0.6579
Panel B:	Germany						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	-0.0003	-0.0003	-0.0003	-0.0040	-0.0017	-0.0003	-0.0039
t-stat	-0.0693	-0.0837	-0.0883	-1.0591	-0.4659	-0.0924	-1.0324
TU	1.0034	1.0038	1.0087	1.0278	1.0221	1.0085	1.0309
Hit	0.5221	0.5181	0.5100	0.5542	0.4940	0.5100	0.5221
PT	-0.6551	-0.7653	-0.6806	0.8108	-1.0818	-0.6806	-0.2041
Panel C:	Japan						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	-0.0010	-0.0013	-0.0021	-0.0011	-0.0029	-0.0021	-0.0011
t-stat	-0.2827	-0.3769	-0.5894	-0.3138	-0.8317	-0.6030	-0.3255
TU	1.0034	1.0047	1.0034	1.0095	1.0095	1.0038	1.0054
Hit	0.5257	0.5257	0.5037	0.4853	0.5037	0.5037	0.5257
PT	-0.1541	-0.1541	-0.5811	-0.9397	-0.4146	-0.5811	0.4359
Panel D:	UK						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	0.0023	0.0047	0.0083	0.0112	0.0104	0.0083	0.0114
t-stat	0.8009	1.6394	2.8486	3.8104	3.5598	2.8549	3.8779
TU	1.0032	1.0093	1.0287	1.0517	1.0390	1.0289	1.0495
Hit	0.5678	0.4396	0.4322	0.4542	0.4359	0.4322	0.4322
PT	0.0810	-2.0820	-0.9730	0.0165	-0.5822	-0.9730	-0.5855
Panel E:	US						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	-0.0005	-0.0005	0.0007	-0.0009	0.0006	0.0007	-0.0003
t-stat	-0.2439	-0.2472	0.3159	-0.4444	0.2806	0.3149	-0.1522
TU	1.0010	1.0009	1.0129	1.0118	1.0065	1.0115	1.0117
Hit	0.5507	0.5485	0.5088	0.5220	0.4934	0.5066	0.5132
PT	0.6817	0.6526	-0.2493	0.0433	-0.8345	-0.3297	-0.1500

Note: The table reports evaluation results of out-of-sample performance of different predictive models (monthly data). After 10 years of initialization, the models are estimated recursively. BACE-adj uses the forecasts of the weighted model whose coefficients are adjusted for finite-sample bias. BMA is based on a pure Bayesian model averaging framework with a g-prior specification. TOP denotes the forecast by the model specification which receives the highest posterior model probability according to BMA. ALL is the all-inclusive specification. AIC, BIC, $\bar{R}^2$ are based on the best models selected by the Akaike, Schwarz criterion or adjusted R^2, respectively. ME denotes the mean prediction error (t-statistic reported below). TU is the ratio of the root mean square error of the particular model-based forecast to the one of the naive benchmark model. Hit denotes the fraction of times the direction of the dependent variable is correctly predicted by the model. PT denotes the test-statistic for directional accuracy by Pesaran and Timmermann (1992).

ity of monthly stock returns is generally very limited. Moreover, notable differences of OOS return predictability can be detected across countries. Table 2.8 also shows that the BACE approach with bias adjustment generally compares rather favorably in terms of forecast accuracy compared to conventional Bayesian model averaging for most stock markets.

The results for the French stock market, presented in Panel A of Table 2.8, show some evidence for out-of-sample predictability. This is consistent with the in-sample results for the composite model, where also the evidence was stronger compared to other capital markets (such as the UK or Germany). Panel A also shows that model averaging approaches (BACE-adj, BMA) typically outperform the naive model and model selection criteria in terms of MSPE, i.e. have a Theil's U (TU) smaller than one. All model-based forecasts generally appear to be unbiased for the French case. The Net-SSE plot (a) in Figure 2.1 shows the relative OOS performance of the forecasts produced by the BACE-adj model over time.[21] As shown by the graph, the model has produced lower squared forecast errors relative to the benchmark up to about 2000. In the aftermath of the climax of the internet boom no outperformance relative to the naive benchmark can be detected anymore.

In the case of Germany (Panel B of Table 2.8), BACE-adj and BMA generally do a better job compared to other model specifications, but are not able to outperform the i.i.d. model in terms of MSPE. This is consistent with the modest results for in-sample predictability in Table 2.3, where little evidence for return predictability was detected at a monthly horizon. The Net-SSE plot (b) in Figure 2.1 shows that OOS predictability has been clearly stronger in the 1990s, where lagged state variables contributed to lower squared prediction errors relative to the benchmark. Also note that, similar to the French case, return prediction models did not provide better forecast accuracy than the benchmark since the height of the new economy boom until the end of the sample.

For the Japanese stock market the case for OOS predictability is also fairly weak, as Panel C of Table 2.8 reveals: forecasts of the naive model generally produce a lower

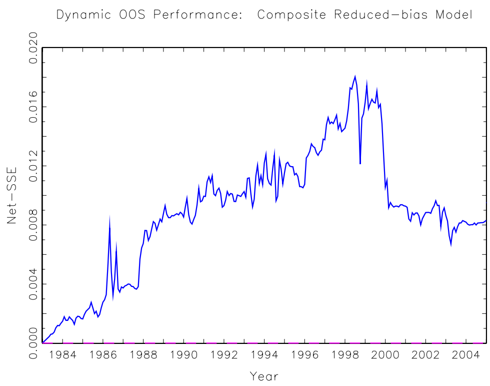
[21]Net-SSE plots based on the BMA approach are generally quite similar.

MSPE than models conditioning on predictive variables. This is confirmed by the Net-SSE plot (c) of Figure 2.1. The plot shows a decline of OOS forecast performance of the weighted model forecast from the early 1990s onwards. Analogously to Germany and Japan, OOS predictability in the United Kingdom (Panel D of Table 2.8) is very poor. Moreover, the United Kingdom is the only stock market where conditional models produce forecasts with a substantial bias (however less pronounced when model averaging techniques are used). Also note that the model averaging methods (BACE-adj and BMA) again outperform the other selection criteria but fail to outperform the naive model in terms of mean-square prediction error.

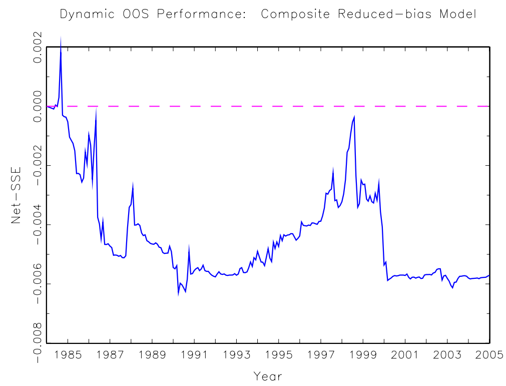
Evaluation results for the US stock market are given in Panel E of Table 2.8. Contrary to the in-sample regressions, out-of-sample predictability of US excess returns is rather poor. Hence, our OOS results are more in line with Goyal and Welch (2008) than Avramov (2002). The Net-SSE plot for the United States in (e) of Figure 2.1 illustrates the time-variation in the degree of OOS predictability. In particular, a steady decline of predictability since the late 1980s can be recognized. This is consistent with other studies for the US documenting poor return predictability over the 1990s (e.g. Paye and Timmermann, 2006; Ang and Bekaert, 2007).

Results for quarterly market excess returns are quite similar to the monthly case and are therefore provided in the Appendix B. We do not find much evidence that OOS predictability increases with the horizon of the forecast. Quite to the contrary, OOS predictability is somewhat weaker than the OOS predictability in the monthly case (e.g. for the US). Again, France is the only stock market where out-of-sample return predictability by model averaging methods can be observed (Panel A of Table 2.9). Results for the German stock market (Panel B of Table 2.9) are quite similar to the monthly case. However, modest evidence of market timing possibility can be found for quarterly models. This happens in particular for highly parameterized models (i.e. ALL, AIC, $\bar{R}^2$), with significant PT-statistics at the 10% level. Quarterly results for Japan (Panel C) and UK (Panel D) are very similar to the monthly case. For the US stock market (Panel E), evidence for OOS predictability with quarterly data is weaker compared to the monthly case. According to the Net-SSE plot for the US in Figure 2.2,

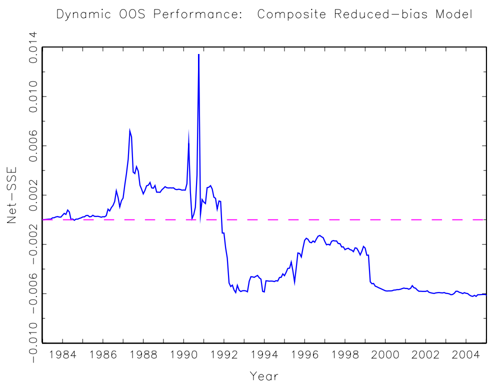
Figure 2.1: Time-variation of Out-of-sample Performance, Net-SSE Plots, Monthly



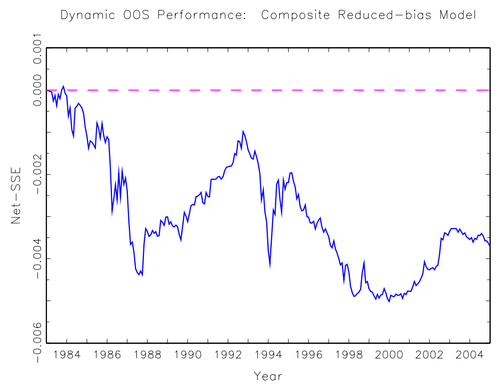
(a) France



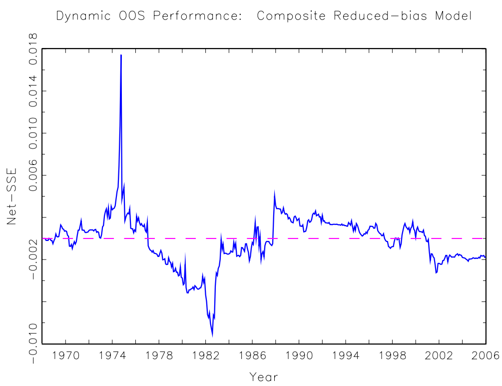
(b) Germany



(c) Japan



(d) United Kingdom



(e) United States

Note: The figure shows Net-SSE plots for the aggregate stock market following Goyal and Welch (2003). Net-SSE is the cumulated difference of squared forecast errors of the unconditional benchmark model (i.i.d. model) and the conditional model (BACE-adj): Net-SSE$(\tau) = \sum_{t=1}^{\tau}(e^2_{uc,t} - e^2_{c,t})$, where $e_{uc,t}$ is the forecast error of the unconditional benchmark, and $e_{c,t}$ is the error of the conditional model. A decrease of the slope represents a better forecast performance of the unconditional model at the particular point in time.

a substantial forecast breakdown can be detected in the aftermath of the first oil price shock (around 1974). As evinced by Figure 2.2, OOS performance of return prediction models in the US has been poor over most of the 1990s consistent with previous studies mentioned before.

2.4 Conclusion

This paper explores stock return predictability in international stock markets in the context of model uncertainty. A Bayesian averaging of classical estimates (BACE) approach is used to account for the tremendous uncertainty of a typical investor in order to find out what the important predictive variables are. This approach is combined with a finite-sample bias correction which accounts for the persistence of the usually employed state variables. Using a comprehensive dataset for international stock markets allows us to gain fresh insights into the empirical evidence for return predictability, which has so far been mainly based on results for the US stock market.

We find substantial differences across countries in terms of return predictability. Evidence for in-sample predictability is stronger for France and the United States compared to the other countries. In the French case also a (modest amount) of out-of-sample predictability can be detected. Out-of-sample predictability by model averaging methods appears to be more accurate for monthly than for quarterly data. Consistent with Avramov (2002), we find that model averaging often produces better OOS forecasts than individual models based on selection criteria. Nevertheless, we also document a substantial amount of time-variation of OOS forecast performance by averaged forecasts.

Two variables appear to be quite robust predictors across countries: the relative bond rate and the output gap. The latter is the only variable which also remains a significant predictor of market excess returns in the US, once model uncertainty is accounted for. The earnings yield often appears to be a more robust predictive variable than the dividend yield. In general, however, our results show that evidence for in-sample predictability for the excess returns in international equity markets is substantially weakened once model uncertainty is accounted for.

The model averaging approach accounting for finite-sample bias employed in this paper may be useful beyond the context of return predictability. In the field of macroeconomic forecasting (e.g. inflation or real activity), for instance, also a large amount of model uncertainty exists and the typical predictors often exhibit a fairly strong degree of persistence (cf. Stock and Watson, 2004). Moreover, another promising subject for future research would be to link the evidence for time-variation in expected returns with the cross-sectional variation of expected returns. An international analysis under model uncertainty with size and book-to-market sorted portfolios may provide additional insights into the particular risks which are relevant to investors.

Appendix A. Data Description

This section of the appendix provides a more detailed description of the stock returns as well as the predictive variables used in our analysis. The original data are monthly but we also report estimation results using quarterly data. Information on the sample periods for the international stock markets can be found in Table 2.1.

Excess returns: The dependent variables for the international stock markets are taken from various sources. In the case of Germany, the return on the DAFOX is used, which is a broad stock index published for research purposes by Karlsruher Kapitalmarktdatenbank. It comprises all German stocks traded in the top segment (Amtlicher Handel) of the Frankfurt stock exchange. For the US, the value-weighted return on the CRSP market portfolio is employed.[22] For the other stock markets, we use broad stock market indexes by Datastream. Excess returns are constructed by subtracting a risk-free rate proxy. When available, a 3-month T-Bill is used as the risk-free rate proxy. Otherwise, a three-month money market rate is used. Interest rates are taken from the Reuters-Ecowin database. In the case of Germany, the money market rate for three-month deposits obtained from the time series database of Deutsche Bundesbank is used as our proxy for the risk-free rate.

Interest rate related variables: The term spread (TRM) is defined as the difference of the yield on long-term government bonds and the short-term interest rate (3-month). The necessary yield curve and interest rate data were obtained from the time series databases of Deutsche Bundesbank (Germany), St. Louis Fed (USA), Econstats (France, United Kingdom and Japan). Following much of the extant literature, the relative short-term interest rate (RTB) is calculated as the short-term interest rate minus its 12-month backward looking moving average. The relative long-term bond rate (RBR) is calculated as the long-term government bond yield minus its 12-month backward looking moving average.

[22]We would like to thank Amit Goyal and Ivo Welch for providing these data on their webpages.

Valuation ratios and other financial variables: The time series of dividend yields (LDY) and earnings yield (LEY) are defined as dividends (earnings) over the past 12 months in relation to the current price. Both series are used in logs, which improves their time-series properties as noted by Lewellen (2004). The US data are taken from Amit Goyal's webpage, while the rest of the valuation ratios refer to the broad stock market indexes provided by Datastream. Realized stock market volatility (LRV) is computed as the sum of the squared daily stock returns and is also used in logs.

Macroeconomic variables: The annual inflation rate (INF) is calculated from the seasonally -adjusted Consumer Price Index (CPI). Another macroeconomic variable is the annual growth rate of industrial production (IPG). The time series of the CPI as well as industrial production for the calculation of industrial production growth (IPG) and the output gap (OPG) measure are taken from the IMF/IFS database and were obtained from Reuters-Ecowin. Following Cooper and Priestley (2006), we construct the output gap measure by applying the filter by HP-filter to the logarithmic series of industrial production. As in Cooper and Priestley (2006), the smoothing parameter is set to 128800 for the monthly data and 1600 for the quarterly data. The cyclical component of the series is taken as the output gap.

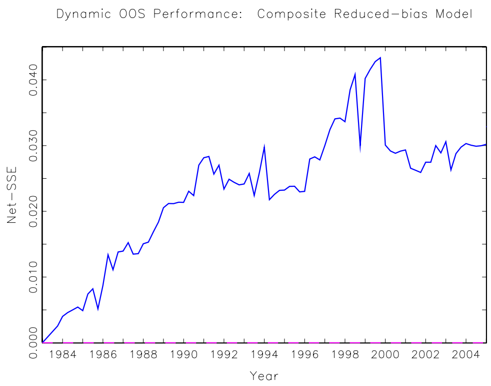
Appendix B. Out-of-Sample Results at the Quarterly Horizon

Table 2.9: Estimation Results: Out-of-sample, Quarterly

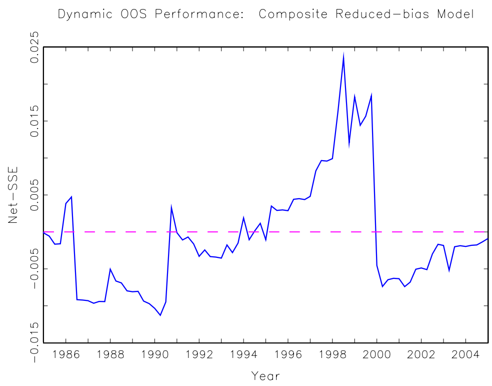
Panel A:	France						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	-0.0007	0.0045	0.0055	-0.0198	-0.0083	0.0029	-0.0131
t-stat	-0.0608	0.3652	0.4345	-1.6140	-0.6705	0.2270	-1.0596
TU	0.9864	0.9914	1.0158	1.0020	1.0054	1.0173	1.0060
Hit	0.7416	0.6966	0.6517	0.7079	0.7303	0.6629	0.6966
PT	0.6404	0.1086	0.1834	0.6767	1.1980	0.2841	0.7432
Panel B:	Germany						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	-0.0021	-0.0016	0.0026	-0.0090	-0.0016	0.0036	-0.0055
t-stat	-0.1628	-0.1235	0.1933	-0.6634	-0.1179	0.2764	-0.4104
TU	1.0004	1.0012	1.0031	1.0343	1.0032	0.9985	1.0192
Hit	0.5926	0.5926	0.5679	0.6420	0.6296	0.5432	0.6173
PT	0.1036	0.1036	0.7742	1.6455	1.8985	0.5957	1.4920
Panel C:	Japan						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	-0.0077	-0.0078	-0.0071	-0.0104	-0.0087	-0.0068	-0.0071
t-stat	-0.6123	-0.6179	-0.5535	-0.8190	-0.6875	-0.5282	-0.5553
TU	1.0040	1.0053	1.0210	1.0197	1.0106	1.0270	1.0188
Hit	0.5955	0.6067	0.5955	0.4944	0.6180	0.5955	0.5618
PT	-0.2404	-0.1190	-0.2404	-1.1983	0.7746	-0.2404	0.0810
Panel D:	UK						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	0.0081	0.0171	0.0300	0.0273	0.0271	0.0310	0.0261
t-stat	0.8604	1.8151	3.1203	2.6937	2.7411	3.1558	2.6423
TU	1.0092	1.0260	1.0839	1.1266	1.1010	1.1075	1.0970
Hit	0.7191	0.5955	0.5056	0.5281	0.5056	0.4944	0.5169
PT	0.0000	-0.1719	0.8151	1.0346	0.6477	0.7056	0.7572
Panel E:	US						
	BACE-adj	BMA	TOP	All	AIC	BIC	$\bar{R}^2$
ME	-0.0003	0.0013	0.0025	-0.0026	0.0035	0.0036	0.0017
t-stat	-0.0430	0.1880	0.3643	-0.3642	0.5132	0.5137	0.2491
TU	1.0252	1.0233	1.0307	1.0453	1.0155	1.0285	1.0287
Hit	0.6333	0.5933	0.5867	0.5600	0.5933	0.5933	0.5867
PT	0.2928	-0.4239	0.2397	-0.5101	1.0035	0.5237	0.4235

Note: The table reports evaluation results of out-of-sample performance of different predictive models (quarterly data). After 10 years of initialization, the models are estimated recursively. BACE-adj uses the forecasts of the weighted model whose coefficients are adjusted for finite-sample bias. BMA is based on a pure Bayesian model averaging framework with a g-prior specification. TOP denotes the forecast by the model specification which receives the highest posterior model probability according to BMA. ALL is the all-inclusive specification. AIC, BIC, $\bar{R}^2$ are based on the best models selected by the Akaike, Schwarz criterion or adjusted R^2, respectively. ME denotes the mean prediction error (t-statistic reported below). TU is the ratio of the root mean square error of the particular model-based forecast to the one of the naive benchmark model. Hit denotes the fraction of times the direction of the dependent variable is correctly predicted by the model. PT denotes the test-statistic for directional accuracy by Pesaran and Timmermann (1992).

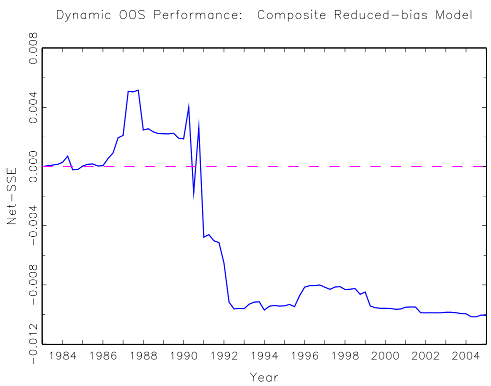
Figure 2.2: Net-SSE Plots, Quarterly



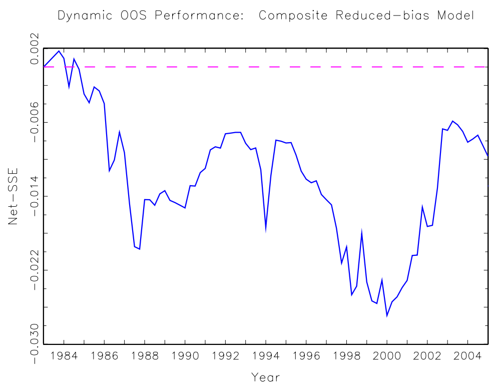
(a) France



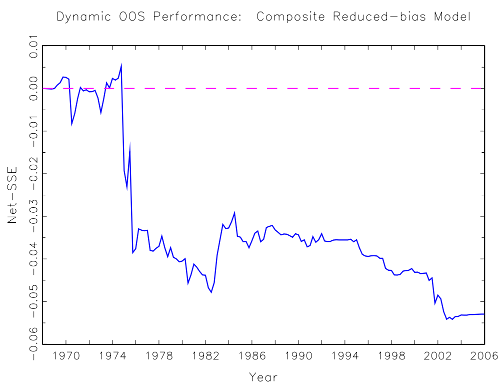
(b) Germany



(c) Japan



(d) United Kingdom



(e) United States

Note: The figure shows Net-SSE plots for the aggregate stock market following Goyal and Welch (2008). Net-SSE is the cumulated difference of squared forecast errors of the unconditional benchmark model (i.i.d. model) and the conditional model (BACE-adj): Net-SSE$(\tau) = \sum_{t=1}^{\tau}(e_{uc,t}^2 - e_{c,t}^2)$, where $e_{uc,t}$ is the forecast error of the unconditional benchmark, and $e_{c,t}$ is the error of the conditional model. A decrease of the slope represents a better forecast performance of the unconditional model at the particular point in time.

Chapter 3

A Reappraisal of the Leading Indicator Properties of the Yield Curve in the Presence of Structural Instability*

ABSTRACT

This chapter provides an extensive reexamination of the leading indicator properties of the yield curve. We study whether the yield spread still qualifies as a useful predictor of real activity in the presence of model instability and forecast breakdowns. Multiple break tests provide strong evidence for structural change and allow us to pin down the exact dates associated with these breaks. We find that window selection methods newly developed for forecasting in the presence of structural change offer some improvements in terms of forecast accuracy. Overall, our results strongly suggest, however, that the yield curve has been losing its edge as a predictor of output growth in recent years.

*This chapter is based on a joint paper with Qingwei Wang (University of Mannheim and Centre for European Economic Research). A revised version of the paper is accepted for publication in the *International Journal of Forecasting*.

3.1 Introduction

The slope of the yield curve is one of the most widely followed economic variables. The alertness of professional economists, market watchers and central bankers can largely be ascribed to the bulk of empirical literature which has documented the term spread's usefulness for predicting future GDP growth.[1] However, recently concerns have been raised over the fact that the predictive performance of the term spread may be time-variant and that predictive regressions based on the yield spread may suffer from parameter instability (e.g. Estrella, Rodrigues, and Schich, 2003; Stock and Watson, 2003; Giacomini and Rossi, 2006).

The main goal of this paper is therefore to investigate whether the yield spread still qualifies as a useful leading indicator in environments characterized by model instability. We mainly focus on two issues in the paper: i) the out-of-sample (OOS) forecast performance of the yield spread (slope of the yield curve) for real activity and, of particular importance, how this OOS predictive performance evolves over time. ii) we investigate whether newly developed window selection techniques for environments characterized by structural breaks (put forth in a recent article by Pesaran and Timmermann, 2007) may help enhance the empirical performance of the yield curve for forecasting. Given the major focus of the previous literature on the empirical relationship between the yield curve and subsequent output growth in the US, we consider international data from Canada, Germany, and the UK as additional "hold-out samples" to examine the usefulness of the yield curve as a leading indicator.[2]

While the in-sample predictive performance of the yield curve is well studied and established, the time-varying nature of the relationship is comparatively unexplored and has only received attention in recent years. A major motivation of this paper

[1]The general finding in the literature is that an inverted yield curve precedes periods of slow economic growth (See e.g. the contributions by Harvey, 1989; Stock and Watson, 1989; Estrella and Hardouvelis, 1991; Hamilton and Kim, 2002, etc.).

[2]Several papers have shown that the yield spread also serves as a significant predictor for real activity in countries outside the US (See e.g. Jorion and Mishkin, 1991; Plosser and Rouwenhorst, 1994; Bernard and Gerlach, 1998; Stock and Watson, 2003; Benati and Goodhart, 2007). In the same vein, our paper also provides insights through an international perspective. This can help mitigate potential data-snooping concerns due to repeated visits of the US dataset.

is therefore to take a closer look at the time-varying forecasting performance of the yield curve for real output growth. The main economic rationale for the yield spread's predictive power is that it serves as an indicator for the effectiveness of monetary policy or the stance of the monetary policy (See e.g. Estrella, Rodrigues, and Schich, 2003). If the central bank raises short-term interest rates and market participants expect this policy to be effective in curbing inflation in the long run, long-term rates (averages of future expected short rates according to the expectations hypothesis) should rise in a smaller proportion. Thus, a restrictive monetary policy tends to flatten the yield curve and at the same time slows down the economy.[3] However, there are strong theoretical reasons to believe that the relationship may vary over time. As noted by Estrella, Rodrigues, and Schich (2003), for instance, the predictive power may depend on underlying factors such as the form of the monetary policy reaction function or the relative importance of real and nominal shocks in the economy. Both factors may be subject to variation over time, which raises the need to investigate the time-variation of the forecasting relationship in greater detail.

As yet, most of the papers addressing the issue of model instability focus on an in-sample analysis of time-varying predictive ability, using mainly sub-sample analysis (e.g. Stock and Watson, 2003), parameter stability tests (e.g. Estrella, Rodrigues, and Schich, 2003) or time-varying parameter models (e.g. Benati and Goodhart, 2007). However, one may argue that the ultimate concern for market participants and policy makers is out-of-sample forecast accuracy as well as a good predictive performance towards the end of the sample period. Hence, our paper distinguishes itself from the remaining literature with its explicit focus on the time-varying out-of-sample (OOS) forecasting properties of the yield curve. We first illustrate the dynamics of forecasting ability via diagnostic plots displaying the evolution of squared forecast errors over time compared to those of a benchmark model. This approach has recently been put forth by Goyal and Welch (2008) in the field of stock return predictability. Using these tools, we document a substantial amount of time-variation in the OOS predictive accuracy of the yield spread which has not previously been shown in the literature. Our findings also

[3]See Estrella (2005) for a formal rational expectations model providing a theoretical account of the relationship.

suggest that the relative OOS forecast accuracy of models based on the yield spread has diminished substantially towards the end of the sample period, which holds true in almost all countries considered.

We thus take a deeper look at potential reasons for this degradation of predictive power and forecast breakdowns by running several modern (in-sample) tests for parameter stability in order to back up the OOS evidence by further formal tests. In particular, we apply the structural break test by Elliott and Müller (2006) – testing the null of parameter stability against the alternative of an unknown number of breaks – as well as structural break tests allowing for multiple breaks developed by Bai and Perron (1998, 2003). These tests largely corroborate our out-of-sample results. We find that the relation of the yield curve and output growth is subject to substantial instabilities in all countries considered.

Hence, it seems natural to investigate whether methods of optimal forecast window selection – which have recently been put forth by Pesaran and Timmermann (2007) for situations where structural breaks are present – yield a better forecast accuracy when the predictive regressions are plagued by parameter instabilities. According to our findings, these optimal window selection methods typically do a good job in reducing the bias of forecast errors. There is also some (though not uniform) evidence on improvements regarding forecast error variance. However, our main finding that the OOS forecast capacity of the yield curve has diminished towards the end of the sample period at an international level still holds under these modified forecasting schemes. Hence, accounting for the existence of structural breaks via optimal window selection methods does not suffice to prevent the poor performance of the yield spread as a leading indicator over most of the 1990s.

The remainder of this paper is structured as follows. Section 3.2 contains a brief overview of our data and provides a reexamination of the leading indicator properties of the yield spread. The main focus is the assessment of time-varying out-of-sample forecast power. In Section 3.3 we discuss the results of structural break tests and the forecast performance of window selection methods designed for environments characterized by model instability. This allows us to judge whether the yield curve still qualifies as a useful leading indicator in environments characterized by structural change. Section 3.5 concludes.

3.2 The Predictive Power of the Yield Spread: A Reexamination

This section reexamines the predictive power of the yield spread for real activity in Canada, Germany, the UK and the US.[4] First, we corroborate the typical result in literature that the yield spread is a significant and strong (in-sample) predictor of real output growth over horizons $k = 4, \ldots, 8$ quarters. This result is confirmed by out-of-sample evaluation statistics. Most importantly, however, we document a strong degree of time-variation in the OOS predictive performance and present evidence of a degradation in the relative OOS forecast performance of the yield spread in all countries considered.

3.2.1 Data Overview

Our dataset comprises time series of real GDP, three-month interest rates i^{short}, long-term government bond yields i^{long} for Canada, Germany, the United Kingdom, and the United States. Data were obtained mainly from the following sources: national central banks, Datastream (national sources as well as the IMF-IFS database), and Reuters-Ecowin. The sample period ranges from 1962:Q1 to 2006:Q2. During this sample period the necessary data are available for all four countries, which facilitates a cross-country comparison of the leading indicator properties of the yield curve. Further detailed information on the data, their sources and data transformation is provided in Appendix 3.5.

[4]In this section we also lay out several of the econometric techniques used in the paper. They include bootstrap-based inference in (in-sample) predictive regressions as well as the OOS forecast evaluation methods applied in the paper.

3.2.2 In-sample Predictive Regressions

Following the vast majority of the extant literature (e.g Estrella and Hardouvelis, 1991; Stock and Watson, 2003), we use predictive regressions in order to investigate the informational content of the yield curve for future real GDP growth. The predictive regression is based on the direct multi-step forecasting approach and takes the following form:

$$y_{t+k} = \beta_0 + \beta_1' z_t + \beta_2' x_t + \epsilon_{t+k}. \tag{3.1}$$

y_{t+k} denotes the (log) growth rate of real GDP from t to $t+k$, $y_{t+k} = (400/k)\ln(Y_{t+k}/Y_t)$, where Y_t is the level of real GDP as of period t. We refer to y_{t+k} as cumulative real GDP growth hereafter. z_t contains the specific yield curve variable we are interested in. Our main focus is on the term spread, which is defined as the difference of the long term government bond yield and the short term interest rate: $i^{long} - i^{short}$. x_t represents (a vector of) additional control variables. In particular, we use the lagged quarterly growth rate of real GDP as an additional predictor as in Stock and Watson (2003) in order to judge the predictive content of the yield curve beyond the information contained in the past history of the dependent variable.

Despite the apparent simplicity of the predictive linear regression in Equation (3.1), the approach is plagued by econometric problems due to overlapping observations of the dependent variable. A common remedy for this problem is to use kernel-based HAC standard errors, e.g. according to Hansen and Hodrick (1980) or Newey and West (1987), which are robust against heteroskedasticity and serial correlation. Although these commonly applied estimators of the long-run covariance matrix deliver consistent estimates, recent evidence suggests that they do not perform well in small sample sizes typically encountered in predictive regressions (See e.g. Goncalves and White, 2005; Ang and Bekaert, 2007). We therefore use a moving block bootstrap (MBB) methodology which is particularly suitable in a finite-sample setting with dependent data.[5]

Table 3.1 presents estimation results on the predictive power of the term spread for Canada, Germany, the UK, and the US. The sample period is 1962:Q1-2006:Q2. All estimation results are based on models including a constant and lagged output growth (for the sake of brevity only the estimated coefficient on the term spread $\hat{\beta}_1^{TS}$ is reported). The term spread is defined as the difference between the long term interest rate and the three-month interest rate.

Panel A of Table 3.1 displays results when the dependent variable is defined as cumulative real GDP growth, and when the forecasting horizons (denoted by k) are 4, 6 and 8 quarters. Besides the conventional statistics, we also provide a "bootstrap p-value" for the R^2 of the predictive regression (denoted as $\%[\bar{R}_b^2 > \bar{R}^2]$ in the table) which is calculated as the fraction of times that the adjusted R^2 in bootstrap samples (generated under the null of no predictability) exceeds the adjusted R^2 of the regression.[6] Overall, we obtain the well-known picture from previous studies. The term spread has a significant (in-sample) predictive power for real activity. The coefficient on the term spread $\hat{\beta}_1^{TS}$ is positive and significant, which holds across all countries and (almost) all considered forecasting horizons. Similarly, the adjusted R^2 shows the model's significant predictive ability. The predictive power appears to be particularly strong in the case of Canada and Germany, where the term spread's coefficient is highly significant even for horizons up to 8 quarters. Note, however, that the predictive power of the yield spread is relatively weak in the UK and refers only to a horizon below 8 quarters.

Although cumulative GDP growth is commonly used as the dependent variable, it is also interesting to consider marginal GDP growth as the dependent variable, since one

[5] In a simulation study, Goncalves and White (2005) show that inference based on the MBB may be considerably more accurate in small samples compared to standard kernel-based HAC standard errors. Contrary to a parametric bootstrap (as used e.g. by Kilian (1999) for inference in predictive regressions), the MBB is a non-parametric bootstrap which draws blocks of re-sampled observations randomly with replacement from the time series of original observations. As recommended by Goncalves and White (2005), we use a data-driven block length, following the procedure by Andrews (1991).

[6] In this context, we use a parametric bootstrapping scheme based on an assumed DGP for the predictors as individual AR(1)-processes (see e.g. Kilian, 1999; Mark, 1995; Rapach, Wohar, and Rangvid, 2005, for similar approaches).

Table 3.1: Predictive Regressions for Real GDP Growth using the Yield Spread: Empirical Results

Panel A: Cumulative Real GDP Growth					Panel B: Marginal Real GDP Growth				
Dependent Variable: $y_{t,t+k} = 400/k[ln(Y_{t+k}/Y_t)]$					Dependent Variable: $y_{t+k-h,t+k} = 400/h[ln(Y_{t+k}/Y_{t+k-h})]$				
Horizon: k=4	CAN	GER	UK	US	**Horizon k=6**	CAN	GER	UK	US
$\hat{\beta}_1^{TS}$	0.711	0.741	0.321	0.677	$\hat{\beta}_1^{TS}$	0.660	0.664	0.279	0.594
t-stat. (BS)	[3.59]	[4.17]	[1.93]	[2.92]	t-stat. (BS)	[2.73]	[3.53]	[1.99]	[2.54]
p-val. (BS)	(0.00)	(0.00)	(0.02)	(0.00)	p-val. (BS)	(0.01)	(0.00)	(0.02)	(0.00)
$\bar{R}^2$	0.306	0.244	0.092	0.210	$\bar{R}^2$	0.207	0.191	0.081	0.105
$\%[\bar{R}_b^2] > \bar{R}^2$	(0.00)	(0.00)	(0.04)	(0.00)	$\%[\bar{R}_b^2] > \bar{R}^2$	(0.01)	(0.01)	(0.11)	(0.06)
Horizon: k=6	CAN	GER	UK	US	**Horizon k=8**	CAN	GER	UK	US
$\hat{\beta}_1^{TS}$	0.665	0.712	0.284	0.620	$\hat{\beta}_1^{TS}$	0.481	0.528	0.139	0.370
t-stat. (BS)	[3.00]	[4.29]	[1.88]	[2.61]	t-stat. (BS)	[2.22]	[3.02]	[0.77]	[2.15]
p-val. (BS)	(0.01)	(0.00)	(0.04)	(0.00)	p-val. (BS)	(0.01)	(0.00)	(0.33)	(0.00)
$\bar{R}^2$	0.309	0.265	0.107	0.187	$\bar{R}^2$	0.098	0.123	0.008	0.053
$\%[\bar{R}_b^2] > \bar{R}^2$	(0.00)	(0.00)	(0.07)	(0.01)	$\%[\bar{R}_b^2] > \bar{R}^2$	(0.11)	(0.06)	(0.71)	(0.23)
Horizon: k=8	CAN	GER	UK	US	**Horizon k=10**	CAN	GER	UK	US
$\hat{\beta}_1^{TS}$	0.602	0.635	0.231	0.527	$\hat{\beta}_1^{TS}$	0.337	0.258	-0.050	0.066
t-stat. (BS)	[3.01]	[4.45]	[1.49]	[2.48]	t-stat. (BS)	[1.57]	[1.11]	[-0.23]	[0.33]
p-val. (BS)	(0.01)	(0.00)	(0.11)	(0.01)	p-val. (BS)	(0.07)	(0.13)	(0.75)	(0.61)
$\bar{R}^2$	0.295	0.272	0.074	0.153	$\bar{R}^2$	0.046	0.022	0.005	0.002
$\%[\bar{R}_b^2] > \bar{R}^2$	(0.00)	(0.00)	(0.18)	(0.04)	$\%[\bar{R}_b^2] > \bar{R}^2$	(0.29)	(0.43)	(0.82)	(0.93)

Note: The table displays estimation results for predictive regressions for real activity in Canada (CAN), Germany (GER), the United Kingdom (UK), and the United States (US). $\hat{\beta}_1^{TS}$ is the estimate of the coefficient on the term spread. All estimation results are based on models including a constant and lagged output growth (estimates not reported for brevity). The dependent variable is defined as (annualized) cumulative real GDP growth (Panel A) and (annualized) marginal real GDP growth (Panel B, h=4). The forecasting horizon is denoted by k. t-stat. (BS) is the t-statistic based on MBB standard errors with 99,999 replications, and p-val. (BS) represents the bootstrap p-value (based on studentization). $\bar{R}^2$ denotes the adjusted R^2, and $\%[\bar{R}_b^2] > \bar{R}^2$ denotes the fraction of times where the bootstrap $\bar{R}_b^2$ exceeds $\bar{R}^2$ (based on a parametric bootstrap with 9,999 replications). Sample period: 1962:Q1-2006:Q2.

can assess how far into the future the predictive power of the yield curve can reach (e.g. Estrella and Hardouvelis, 1991; Dotsey, 1998; Hamilton and Kim, 2002). Marginal GDP growth is defined as $y_{t+k-h,t+k} = 400/h[ln(Y_{t+k}/Y_{t+k-h})]$. Our results in Table 1 (Panel B) focus on marginal GDP growth over the past four quarters ($h = 4$) and forecasting horizons of $k = 6$, 8 and 10 quarters.[7]

As depicted in Table 3.1 (Panel B), the marginal predictive power of the term spread declines substantially when the predictive horizon increases. The adjusted R^2 and the significance of the coefficients indicate that the predictive power refers mostly to a horizon up to 8 quarters and vanishes at the 10 quarter horizon. Again, the weakest results are observed for the UK, where the information in the yield spread refers only to a shorter horizon up to 6 quarters.

3.2.3 Out-of-sample Performance

We now investigate the capacity of the yield curve to predict real activity out-of-sample (OOS). The first 10 years (1962:Q1-1972:Q1) are used as an initialization period for the models, afterwards forecasts are generated using a recursive scheme (i.e., an expanding forecasting window). This provides us with $n = T - m - k - 1$ OOS forecasts of real GDP growth, where m represents the length of the initialization period and T denotes the overall sample size.

In Table 3.2 and Table 3.5 we provide several forecast evaluation statistics. First, we report the mean forecast error and the corresponding bootstrapped standard error (also based on the MBB). A significant mean forecast error can be interpreted as evidence against the hypothesis of forecast unbiasedness. We also report results from traditional Mincer-Zarnowitz (1969) regressions, where the realizations of real GDP growth are regressed on a constant and the corresponding forecasts. According to these statistics, the better the forecasting model, the closer the intercept $\hat{a}$ should be to zero and the

[7] By definition, the results for cumulative GDP growth and marginal GDP growth are the same if both k and h are set to four, so we omit a forecast horizon of $k = 4$ in Panel B of Table 1.

slope $\hat{b}$ should be to one.[8] Another simple descriptive measure of forecast evaluation is Theil's U, which is the ratio of the root-mean squared error (RMSE) of the prediction model to the RMSE of the benchmark model. As is common in the literature nowadays (e.g. Stock and Watson, 2003 or Ang, Piazzesi, and Wei, 2006) we use an AR(1) as the benchmark. If the forecast of the model is superior to the benchmark (given a quadratic loss), Theil's U should be less than one.

We mainly base our inference regarding superior OOS predictability on the test recently proposed by Clark and West (2007). This test is designed for comparing a parsimonious null model to a larger model which nests the null model, as is the case in our context. The central idea of the Clark-West test is to adjust the mean squared forecast error of the larger unrestricted model.[9] In our context, we test whether the difference of the mean squared forecast error (MSFE) of the AR(1)- benchmark model (Model 0) $\hat{\sigma}_0^2$ and the adjusted mean squared forecast error $\hat{\sigma}_1^2\text{−adj}$ of the model of interest (Model 1) is equal to zero against the alternative of superior forecast accuracy of the prediction model (one-sided test). Clark and West (2007) suggest to adjust the MSFE of the larger model as follows

$$\hat{\sigma}_1^2\text{−adj} \quad = \quad n^{-1}\sum_{t=m+1}^{T-k}(y_{t+k}-\hat{f}^{(1)}_{t,t+k})^2 - n^{-1}\sum_{t=m+1}^{T-k}(\hat{f}^{(0)}_{t,t+k}-\hat{f}^{(1)}_{t,t+k})^2, \tag{3.2}$$

where the GDP growth forecast (k-quarter ahead) based on the information set at time t is denoted as $\hat{f}^{(1)}_{t,t+k}$ for the case of the (unrestricted) model of interest and $\hat{f}^{(0)}_{t,t+k}$ for the case of the benchmark model. n is the number of OOS predictions: $n = T-m-k-1$. Note that the first term in Equation (3.2) corresponds to the usual mean squared forecast

[8]However, it is well-known that the condition $\hat{a} = 0, \hat{b} = 1$ only represents a necessary but not sufficient condition for unbiasedness (Clements and Hendry, 1998, p.57). Hence, we do not report results of the joint test but merely report Mincer-Zarnowitz regression results along with the direct test whether the mean forecast error is equal to zero.

[9]The reason for the adjustment put forth by Clark and West (2007) is that – under the null hypothesis that the additional regressors in the larger model are not necessary for forecasting – there is the need to estimate parameters of the unrestricted model that are zero in population, which introduces noise in the forecast.

error of the (unrestricted) model of interest, and the second term is the adjustment term discussed above. In order to test whether Clark-West's MSFE-adj (defined as $\hat{\sigma}_0^2 - \hat{\sigma}_1^2-\mathrm{adj}$) is equal to zero, we again use the MBB for inference to take account of serial correlation.[10]

Table 3.2: Out-of-Sample Performance of the Yield Spread: Forecast Evaluation Statistics

Horizon: k=4	CAN		GER		UK		US	
Mean Forecast Error	-1.12	(0.27)	-0.63	(0.31)	-0.07	(0.43)	-1.23	(0.25)
Theil's U	0.87		0.85		1.05		0.97	
MSFE-adj	3.59**		2.94**		0.22		3.77***	
Mincer-Zarnowitz: $\hat{a}$	-0.81	(0.81)	0.38	(0.66)	2.27	(0.87)	-0.12	(0.64)
Mincer-Zarnowitz: $\hat{b}$	0.93	(0.16)	0.61	(0.20)	0.01	(0.33)	0.74	(0.12)
Mincer-Zarnowitz: R^2	0.38		0.15		0.00		0.38	
Horizon: k=6	CAN		GER		UK		US	
Mean Forecast Error	-1.24	(0.28)	-0.73	(0.37)	-0.17	(0.58)	-1.22	(0.30)
Theil's U	0.88		0.84		1.07		0.97	
MSFE-adj	2.92**		2.77***		0.01		2.91***	
Mincer-Zarnowitz: $\hat{a}$	-0.80	(1.02)	0.60	(0.67)	2.74	(0.87)	-0.24	(0.73)
Mincer-Zarnowitz: $\hat{b}$	0.90	(0.21)	0.49	(0.18)	-0.23	(0.29)	0.76	(0.14)
Mincer-Zarnowitz: R^2	0.38		0.12		0.01		0.38	
Horizon: k=8	CAN		GER		UK		US	
Mean Forecast Error	-1.27	(0.25)	-0.76	(0.38)	-0.18	(0.54)	-1.02	(0.22)
Theil's U	0.90		0.83		1.09		0.93	
MSFE-adj	2.17**		2.22**		-0.27		2.11***	
Mincer-Zarnowitz: $\hat{a}$	-0.84	(1.28)	0.55	(0.65)	3.31	(0.85)	-0.28	(0.81)
Mincer-Zarnowitz: $\hat{b}$	0.90	(0.27)	0.50	(0.18)	-0.45	(0.30)	0.82	(0.17)
Mincer-Zarnowitz: R^2	0.34		0.12		0.05		0.35	

Note: This table presents various evaluation statistics of out-of-sample forecast performance of the yield spread for real activity. A recursive forecasting scheme is used. The first 10 years (1962:Q1-1972:Q1) are used as initialization period. Theil's U is the ratio of the RMSE of the models based on the term spread and the RMSE of the AR(1)-benchmark model. MSFE-adj is the difference of the MSFE of the benchmark and the adjusted mean squared forecast error according to Clark and West (2007) (*, **, *** denotes significance of Clark-West's test statistic for testing equal predictive performance at the 10%, 5%, and the 1% level). Coefficients and R^2 of Mincer-Zarnowitz regressions are also reported. Bootstrapped standard errors (MBB with 99,999 replications) are given in parentheses.

Table 3.2 summarizes the results of OOS forecast evaluation for the model with cumulative real GDP growth (over forecast horizons of $k = 4, \ldots, 8$) as the dependent variable and a constant, the term spread and lagged output growth as regressors. Inspection of Table 3.2 reveals that forecasts based on the yield spread are usually upward biased. In all countries (except the UK) a significant overprediction of real output growth can be detected. However, Theil's U and the Clark/West test indicate a superior out-

[10] Critical values from the standard normal distribution can be used to test the significance of MSFE-adj. Simulations in Clark and West (2007) show that their test using MSFE-adj with standard normal critical values is as accurate as other competing tests, while the power is as good or better.

of-sample performance of the model including the spread over the benchmark model for Canada, Germany and the United States. The poor out-of-sample performance in the United Kingdom does not come as a surprise, given its comparatively weak in-sample performance in Table 1. For the other three countries, the success of the yield spread for out-of-sample forecasting is evident even for forecast horizons of 8 quarters. The findings in Table 3.2 broadly corroborate results of OOS forecasting experiments conducted elsewhere in the literature (e.g. Stock and Watson, 2003; Duarte, Venetis, and Paya, 2005; Giacomini and Rossi, 2006) which have typically concluded that there is a good OOS forecast performance of models using the yield spread relative to the benchmark model.

As discussed before, however, there are several reasons to conjecture that the forecasting relationship may be time-varying. Thus, in the following we shed some light on time-variation of the relative OOS performance of the yield spread as a predictor of real activity. This allows us to reexamine the yield curve's usefulness as a leading indicator in particular towards the end of the sample period, which is of ultimate concern for market participants. We investigate the time-variation of OOS performance using diagnostic plots, which are motivated by the recent work of Goyal and Welch (2008) in the context of stock return predictability.[11] To our knowledge such an analysis – making the relative forecast performance over time transparent – has been lacking in the literature so far.

Following Goyal and Welch (2008), we plot the cumulative sum of squared forecast errors from a benchmark model minus the squared errors from the prediction model

$$\mathrm{Net} - \mathrm{SSE}(\tau_0, \tau_1) = \sum_{t=\tau_0}^{\tau_1} [(y_{t+k} - \hat{f}^{(0)}_{t,t+k})^2 - (y_{t+k} - \hat{f}^{(1)}_{t,t+k})^2], \tag{3.3}$$

where τ_0 is the starting date and τ_1 is the end date at which the Net-SSE is evaluated. $\hat{f}^{(0)}_{t,t+k}$ ($\hat{f}^{(1)}_{t,t+k}$) are forecasts generated by the benchmark model (term spread model).

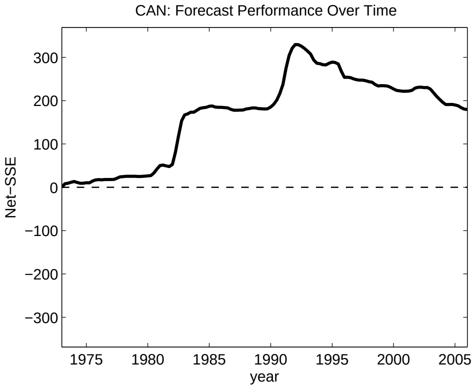
[11]In an extensive analysis for the US stock market, Goyal and Welch (2008) question the existence of stock return predictability based on their finding of poor OOS performance over time.

When Net-SSE is above the zero horizontal line, it indicates that the model of interest outperforms the benchmark model (i.e. by producing lower squared forecast errors) up to the period τ_1. This graph is a rather informative diagnostic for comparing the relative performance of competing models over time.

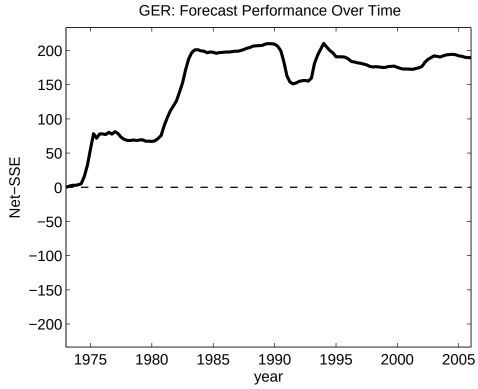
Figure 3.1 (based on a forecast horizon of 4 quarters) depicts how the OOS performance of prediction models using the term spread model evolves over time relative to the AR(1) benchmark. All four panels in Figure 3.1 indicate a strong time-variation of the forecast performance. More concretely, Canada, US, and Germany all experience a rather good forecast performance of the term spread in the early sub-sample period (1970s and 1980s). In these periods, models including the yield spread typically outperformed the AR(1)-benchmark in terms of forecast accuracy. However, as the Net-SSE plots forcefully demonstrate, for Canada (a) and the US (d) the OOS forecast performance has deteriorated thereafter. Clearly, this calls into question the practical usefulness of the yield spread as a predictor of real activity in those countries in the most recent period. Similarly, in the case of Germany (b) no clear improvements of including the yield spread in prediction models can be observed over the 1990s. As depicted by the Net-SSE plot for the United Kingdom, the term spread has generally proved to be a rather poor predictor out-of-sample throughout almost the whole sample period. However, there are some periods (early 1980s and early 1990s) in which including the term spread actually lowered squared forecast errors.

These results extend previous findings of a degradation of predictive performance of the yield curve in the United States (already noted by Dotsey, 1998 or Stock and Watson, 2003) by adding an international perspective and by making forecast breakdowns more transparent through an explicit focus on OOS forecasting. Having illustrated the time-variation of the OOS performance and forecast breakdowns, it thus seems natural to investigate the role of structural breaks for periods of breakdowns of forecast performance in greater detail. This is the purpose of the next section.

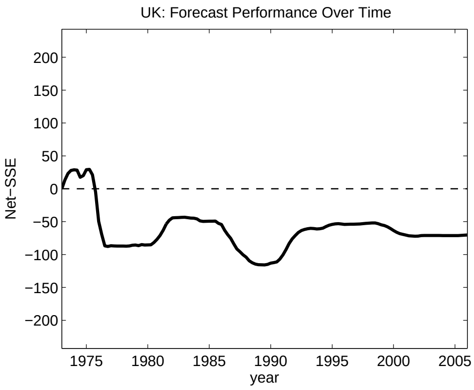
Figure 3.1: Time-varying Forecast Performance, Net-SSE, k=4



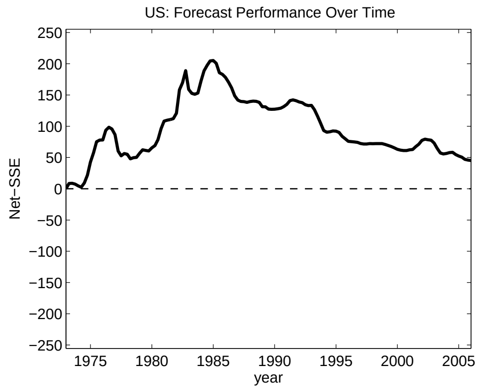
(a) Canada



(b) Germany



(c) UK



(d) US

Note: The figure shows Net-SSE plots following Goyal and Welch (2008). Net-SSE is the cumulated difference of squared forecast errors of AR(1) benchmark model and the prediction model including the yield spread and lagged GDP growth: Net-SSE$(\tau_0, \tau_1) = \sum_{t=\tau_0}^{\tau_1}(e_{b;t}^2 - e_{m;t}^2)$, where $e_{b;t}$ is the forecast error of the benchmark, and $e_{m;t}$ is the error of the prediction model. A decrease of the slope represents a better forecast performance of the benchmark model at the particular point in time.

3.3 Empirical Analysis of Model Instability and Forecast Breakdowns

We now investigate the stability of the empirical relationship between the yield curve and real activity. First, we briefly outline empirical methods, i.e., structural break tests allowing for multiple breaks and forecast window selection methods in the presence of breaks. Our empirical results reported in this section provide strong evidence that, indeed, the relationship between the yield spread and output growth is subject to substantial structural change in all countries of this study. Newly proposed window selection methods offer improvements by reducing the bias of forecast errors and (in some cases) forecast error variance. Nevertheless, using these methods is not enough to prevent the deterioration of predictive power of the yield spread in the recent period.

3.3.1 Econometric Methods: Structural Break Tests and Window Selection for Forecasting

Predictive regressions for output growth using the yield spread as in Equation (3.1) may be subject to potential structural instability. In particular, different monetary policy regimes (e.g., whether the central bank is more concerned by the output gap or deviations of inflation from the target) could be the reason for such a structural change affecting the predictive relation. When structural change is strong enough, standard inference becomes misleading. Moreover, the question of how to select the estimation window in the presence of structural breaks arises, which is of ultimate importance from a forecaster's perspective.

Contrary to previous papers (Estrella, Rodrigues, and Schich, 2003; Giacomini and Rossi, 2006), we consider recently developed structural break tests allowing for multiple structural breaks at an unknown date under the alternative. These tests have been developed by Bai and Perron in a series of articles (Bai and Perron, 1998, Bai and Perron, 2003, and Bai and Perron, 2006) and allow us to pin down the dates associated with the identified (multiple) breaks.[12] More concretely, by allowing the parameters in Equation (3.1) to vary across $r+1$ regimes, we consider a predictive regression of the following form:

$$y_{t+k} = \beta'_{0,j} + \beta'_{1,j} z_t + \beta'_{2,j} x_t + \epsilon_{t+k}. \qquad (t = T_{j-1}+1, ..., T_j) \tag{3.4}$$

where $j = 1, .., r+1$, and r is the number of breaks in the linear regression. Note that Equation (3.4) implies a splitting of the sample into r partitions. For each of the r partitions within the set of admissible partitions, the least squares estimates of $\beta_{i,j}, (i = 0, 1, 2)$ and the corresponding sum of squared residuals are obtained. Then, the break date estimates $\hat{T}_1, ..., \hat{T}_r$ are selected as the ones globally minimizing the sum of squared residuals. Bai and Perron (1998) also consider a test of the null hypothesis of l breaks against the alternative $l+1$ breaks by proposing a $SupF(l+1|l)$ test statistic. If the reduction of the sum of squared residuals is significant, the null hypothesis of l breaks is rejected in favor of the alternative of $l+1$ breaks.

Sometimes interest lies on the question whether there is general instability of the relationship and not on the exact number of breaks. To test the null hypothesis of no break against an alternative hypothesis of an unknown number of breaks up to a given upper bound R, Bai and Perron (1998) propose two double maximum statistics. The double maximum statistics have weights a_r reflecting priors on how likely various numbers of breaks r might occur

$$Dmax = max_{1 \leq r \leq R} \;\; a_r SupF_T(r). \tag{3.5}$$

There are no precise theoretical guidelines about the choice of a_r. A simple and obvious candidate is to use a uniform weight, which leads to the so-called "UDmax" statistic. Alternatively, weights can be chosen such that the marginal p-values are equal across

[12] Neither tests used by Estrella, Rodrigues, and Schich (2003) [supLM-Test by Andrews (1993) and PR-test by Ghysels, Guay, and Hall (1997)] nor the ones conducted by Giacomini and Rossi (2006) allow for multiple breaks in the predictive relationship.

values of r (See Bai and Perron, 1998, p.59). This version of double maximum statistic is labeled as “WDmax".

Based on an extensive simulation study, Bai and Perron (2006) recommend a preferred strategy for structural break testing in the presence of multiples breaks. First, the UDmax and WDmax statistics are used to detect whether at least one break is present. If this is the case, then the number of breaks l is identified by an examination of the $SupF(l+1|l)$ tests, where l is associated with the break dates that minimize the global sum of squared residuals. We closely adhere to this strategy in our empirical application.

A simulation study conducted by Paye and Timmermann (2006) finds that the UDmax as well as the SupF statistic can have size distortions under some circumstances.[13] They find, instead, that the structural break test recently proposed by Elliott and Müller (2006) performs better in those cases. Drawing on the similarities between the concepts of “structural breaks" and “random coefficients", Elliott and Müller (2006) propose to test the null hypothesis that $\beta_t = 0$ for any t, where $\beta = (\bar{\beta} + \beta_t)$ against the alternative hypothesis $\beta_t \neq 0$ for some $t > 1$. This test statistic is easy to compute and is labeled as $\widehat{qLL}$. For the purpose of completeness, we also provide the $\widehat{qLL}$ statistic in addition to the Bai-Perron tests.[14]

When forecasting time series by predictive regressions that are subject to structural breaks, care has to be taken since breaks can severely affect the model's out-of-sample performance. This difficulty can be addressed by a careful selection of the estimation window. Intuitively, one should estimate the model only with the data available after the most recent break. However, as pointed out in a recent article by Pesaran and Timmermann (2007), this conventional wisdom is not necessarily optimal since there can be a tradeoff between forecast error bias and forecast error variance. Theoretical

[13]More concretely, they consider a predictive regression, where the regressors follow an AR(1) process. When the predictors are persistent and the innovations in the predictive regression and those of the AR(1) regression are strongly correlated, size distortions of the tests can be substantial.

[14]A detailed description of the steps for computing $\widehat{qLL}$ can be found in Elliott and Müller (2006, p.914). We use the GAUSS code provided by David E. Rapach for running the structural break tests. We thank David E. Rapach for providing the code via his web page: http://pages.slu.edu/faculty/rapachde/Research.htm

and simulation results by Pesaran and Timmermann (2007) suggest that the forecasting performance can typically be improved if (some) pre-break information is included.

However, there is typically a substantial estimation uncertainty regarding the exact timing and the size of breaks in real-time, particularly when the breaks occur close to the boundary of the data. For this reason, Pesaran and Timmermann (2007) propose several forecast schemes which are based on a combination of forecasts from different estimation windows, instead of a single estimation window.[15] These approaches require a minimum of $\underline{\omega}$ observations for estimating the parameters of the forecasting models. The last $\tilde{\omega}$ observations of the estimation period are reserved for a ("pseudo") OOS evaluation of the different forecasts based on different sizes of the estimation window. For each potential starting point w of the estimation window, a set of forecasts is generated which are evaluated according to their MSFE within the evaluation window $\tilde{\omega}$. Then one can combine forecasts from different estimation windows $\hat{f}_{t+k,w}$, where the weights are proportional to the inverse of the associated ("pseudo") MSFE in the evaluation window

$$\hat{f}_{t+k}^{weighted}(\underline{\omega},\tilde{\omega}) = \frac{\sum_{w=1}^{t-\underline{\omega}-\tilde{\omega}}(\hat{f}_{t+k,w}MSFE(w|t,\tilde{\omega})^{-1})}{\sum_{w=1}^{t-\underline{\omega}-\tilde{\omega}}MSFE(w|t,\tilde{\omega})^{-1}}. \tag{3.6}$$

A more parsimonious approach is to put equal weight on all forecasts regardless of the corresponding MSFE, which means that no evaluation of the forecasts within the evaluation window $\tilde{\omega}$ is needed. We denote the equally weighted forecast as "pooled forecast". As noted by Pesaran and Timmermann (2007), the MSFE-weighted forecast and the pooled forecast may work better if the breaks are small. Alternatively, one can use a weight of one for the forecast based on the estimation window w which produces the lowest MSFE within the evaluation window, and a weight of zero for all other forecasts. This (so-called) "cross-validation" approach is more likely to work well if there is a single break which is well defined and large.

[15] Pesaran and Timmermann (2007) provide simulation results which show that their combination approaches often work better than methods ignoring the presence of breaks. This is in line with the typical result in the forecasting literature that forecast combinations often improve upon a single forecast (See e.g., Timmermann, 2006).

3.3.2 Empirical Results

Table 3.3 provides estimation results for different structural break tests. The predictive regression of real GDP growth includes a constant, the term spread, and lagged GDP growth as regressors, and the sample period covers 1962:Q1-2006:Q2 for all countries.

In Panel A, the $\widehat{qLL}$ statistic proposed by Elliott and Müller (2006) and the UDMax and WDmax statistics proposed by Bai and Perron (1998) test the null hypothesis of no break against the alternative hypothesis of at least one break. All three test statistics are significant at the 1% level for all countries. This provides strong evidence that the predictive relationship between the yield spread and GDP growth has been affected by structural change during our sample period.

Given the strong evidence for structural breaks, we follow the recommendation by Bai and Perron (2006) and conduct $SupF(l+1|l)$ tests to identify the number and timing of structural breaks.[16] Panel B reports the results of these tests. According to the $SupF(l+1|l)$ tests, three breaks are detected for Canada and the UK, and four breaks are found for Germany and the US.

Table 3.4 reports estimated break dates and the associated confidence intervals.[17] When taking a closer look at estimated break dates, several interesting patterns emerge. Some of the estimated break dates can be linked to particular phases of the business cycle, well-known unanticipated events (such as the German reunification) or changes in the monetary regime. For example, two out of the three break dates (1980:Q4 and 1990:Q4) for Canada are very close to the particular peaks of the business cycle, as reported in Demers and MacDonald (2007). For Germany, the break dates identified in 1989:Q4 and 1993:Q3 may be linked to the German unification, which is a typical example of a real shock, and the turmoil in the European Monetary System after Germany's reunification boom. The break in 1999:Q2 in Germany can be ascribed to the European

[16] We impose the maximum number of breaks R to be five, and chose a trimming parameter of 0.15 for the construction and critical values of these tests, as recommended by Bai and Perron (2006).

[17] Results of structural break tests for other horizons also indicate substantial instability. The results are not reported for the sake of brevity but can be provided by the authors upon request.

Table 3.3: Structural Break Tests: Predictive Regressions for Real GDP Growth, k=4

	Panel A: SB-tests			**Panel B: Bai/Perron SupF Test**				
Country	$\widehat{qLL}$	UDmax	WDmax	SupF(1|0)	SupF(2|1)	SupF(3|2)	SupF(4|3)	SupF(5|4)
Canada	-42.54***	38.19***	64.80***	32.22***	28.77***	22.64***	4.78	–
Germany	-53.83***	201.18***	234.57***	38.87***	48.17***	45.57***	45.01***	0.02
UK	-66.43***	45.52***	69.89***	45.49***	25.35***	22.12***	9.88	–
US	-73.36***	54.88***	102.85***	43.37***	40.38***	17.20**	54.82***	8.04

Note: The table reports results from several structural break tests. The tests are based on models including a constant, the term spread and lagged output growth as regressors. The forecast horizon is k=4 quarters. The $\widehat{qLL}$ statistic by Elliott and Müller (2006) tests the null hypothesis of no structural break against the alternative hypothesis of an unknown number of breaks. Inference on other test statistics is based on critical values in Bai and Perron (1998). An upper bound of $R = 5$ for the number of breaks is imposed. Accordingly, we use a trimming parameter of 0.15 for the construction and critical values of these tests, as recommended by Bai and Perron (2006). UDmax and WDmax test the null of no break against an unknown number of breaks up to $R = 5$. $SupF(l + 1|l)$ tests the null hypothesis of l breaks against the alternative $l + 1$ breaks. *, **,*** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 3.4: Identification of Break Dates (Bai/Perron Test Procedure)

Country	Lower Bound	Break Date	Upper Bound
Canada	1979:3	1980:4	1981:3
Canada	1990:2	1990:4	1992:1
Canada	1996:3	1997:3	1998:2
Germany	1985:2	1986:1	1986:2
Germany	1989:3	1989:4	1990:1
Germany	1993:1	1993:3	1993:4
Germany	1998:3	1999:2	2000:1
UK	1967:3	1969:1	1970:2
UK	1986:4	1987:3	1989:3
UK	1995:2	1997:3	1998:1
US	1968:2	1969:1	1969:4
US	1983:1	1983:4	1986:4
US	1991:2	1991:4	1992:3
US	1998:1	1999:1	1999:2

Note: This table reports estimates of break dates and corresponding confidence intervals. The break dates and the number of breaks are obtained as global minimizers of the sum of squared residuals [See Bai and Perron (1998) for further details].

Monetary Union. In the case of the UK, the break identified in 1997:Q3 could be related to the regain of the independence of the Bank of England. Structural breaks found in the United States seem to be mostly related to business cycle turning points rather than monetary regimes. All four break dates identified by the tests are close to either a peak or a trough dated by NBER's Business Cycle Dating Committee.[18]

Overall, our findings on the timing of breaks for Germany and the US are somewhat different from the results of Estrella, Rodrigues, and Schich (2003) who found no evidence for breaks in Germany and only weak evidence in the case of the US. Their results are obtained by applying the supLM-Test of Andrews (1993) and the PR-test of Ghysels, Guay, and Hall (1997) to a sample from January 1967 and December 1997. They also impose a rather large trimming parameter (25%), which implies that breaks in the more recent period could not be detected. By contrast, our results are based on more powerful recently developed tests which allow for multiple structural breaks.[19] It is noteworthy, however, that we detect a break in 1983:Q4 for the US which is very close to the (single) break identified in Estrella, Rodrigues, and Schich (2003). We find additional breaks in 1991:Q4 and 1999:Q1, which were not possible for Estrella, Rodrigues, and Schich (2003) to detect given their sample period, trimming parameter and methodology. Similarly, the two breaks (1993:Q3 and 1999:Q2) which we find in the case of Germany could not be detected by Estrella, Rodrigues, and Schich (2003) for the same reason.

Given the strong evidence for structural breaks affecting the in-sample predictive regression, a natural question appears: how is the out-of-sample forecasting performance affected by these breaks? We address this question by using forecast combination methods with different window lengths put forth by Pesaran and Timmermann (2007) for forecasting in the presence of structural change.

Table 3.5 presents an evaluation of the out-of-sample performance using various forecast schemes: a standard recursive scheme (no combination, expanding window size),

---

[18]See http://www.nber.org/cycles/cyclesmain.html

[19]Moreover, we also have a longer sample period available and impose a smaller trimming parameter.

MSFE-weighted forecast combination (weighted forecasts from different estimation windows with weights determined by the inverse of the MSFE in the evaluation window), cross-validation (forecast from the single window with the lowest MSFE in the evaluation period) and pooled forecast (average of forecasts based on different estimation windows).[20]

The results for forecast window selection methods in the presence of breaks are rather similar for Canada, Germany and US. All these combination schemes typically produce forecasts with a substantially reduced bias. This is what should be expected given the arguments in Pesaran and Timmermann (2007, p.138-139). However, only rather modest improvements can be found in terms of forecast error variance as evinced by Theil's U or other evaluation statistics. Among the combination schemes, the pooled forecast tends to generate a rather small forecast error variance, although it often has a larger bias. The cross-validation approach – only based on a forecast using a single estimation window – is typically the most fragile.[21] Despite reducing the bias in forecast errors, our results suggest that accounting for structural breaks by using optimal window selection methods is not sufficient to prevent the deterioration of OOS forecast accuracy of the yield spread documented in the previous section. Indeed, also when these more sophisticated forecasting schemes are used, the degradation of OOS forecast performance of the yield spread still continues to hold, as evinced by Figure 3.2.

[20]The minimum window length was set to 12 observations (3 years), the evaluation period was set to 16 observations (4 years).

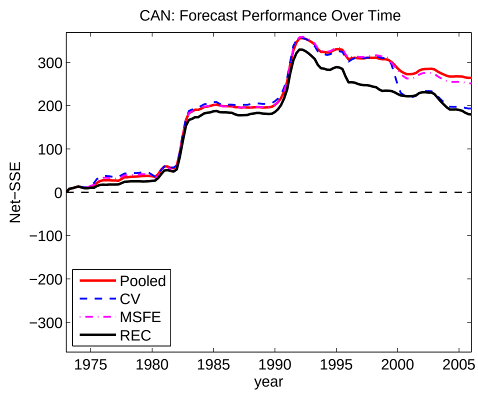
[21]This finding links to the general result in the forecasting literature that simple combination approaches often produce better forecasts compared to single forecasts or very sophisticated combination approaches (See e.g. Timmermann, 2006).

Table 3.5: Window Selection under Model Instability: Forecasting Evaluation Statistics (OOS)

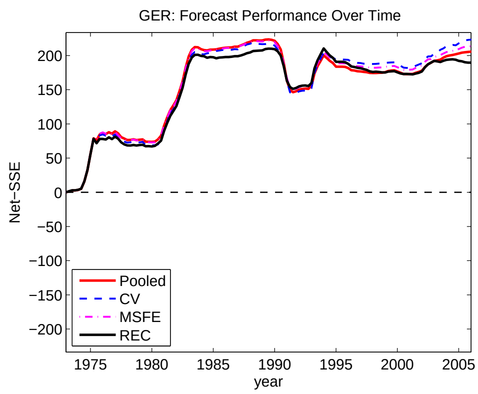
Panel A: CAN	**Recursive**		**MSFE-weighted**		**Cross-Validation**		**Pooled**	
Mean Forecast Error	-1.12	(0.25)	-0.55	(0.30)	-0.48	(0.35)	-0.66	(0.24)
Theil's U	0.87		0.81		0.86		0.80	
MSFE-adj	3.59***		4.37**		4.27***		4.25***	
Mincer-Zarnowitz: $\hat{a}$	-0.81	(0.81)	0.20	(0.81)	0.71	(0.87)	-0.17	(0.75)
Mincer-Zarnowitz: $\hat{b}$	0.93	(0.16)	0.79	(0.18)	0.66	(0.20)	0.87	(0.17)
Mincer-Zarnowitz: R^2	0.38		0.31		0.25		0.35	
Panel B: GER	**Recursive**		**MSFE-weighted**		**Cross-Validation**		**Pooled**	
Mean Forecast Error	-0.63	(0.33)	-0.41	(0.31)	-0.36	(0.34)	-0.46	(0.32)
Theil's U	0.85		0.83		0.82		0.84	
MSFE-adj	2.94**		3.29***		3.50***		3.09**	
Mincer-Zarnowitz: $\hat{a}$	0.38	(0.66)	0.54	(0.61)	0.56	(0.56)	0.52	(0.70)
Mincer-Zarnowitz: $\hat{b}$	0.61	(0.20)	0.59	(0.20)	0.60	(0.19)	0.59	(0.23)
Mincer-Zarnowitz: R^2	0.15		0.14		0.15		0.13	
Panel C: UK	**Recursive**		**MSFE-weighted**		**Cross-Validation**		**Pooled**	
Mean Forecast Error	-0.07	(0.59)	0.01	(0.65)	-0.14	(0.51)	0.16	(0.52)
Theil's U	1.05		1.11		1.16		1.06	
MSFE-adj	0.22		0.13		0.00		0.67	
Mincer-Zarnowitz: $\hat{a}$	2.27	(0.87)	2.41	(0.71)	2.42	(0.65)	2.04	(0.80)
Mincer-Zarnowitz: $\hat{b}$	0.01	(0.33)	-0.04	(0.24)	-0.05	(0.20)	0.12	(0.31)
Mincer-Zarnowitz: R^2	0.00		0.00		0.00		0.01	
Panel D: US	**Recursive**		**MSFE-weighted**		**Cross-Validation**		**Pooled**	
Mean Forecast Error	-1.23	(0.27)	-0.64	(0.28)	-0.64	(0.27)	-0.74	(0.28)
Theil's U	0.97		0.86		0.91		0.89	
MSFE-adj	3.77***		5.13***		5.15***		4.74***	
Mincer-Zarnowitz: $\hat{a}$	-0.12	(0.64)	0.57	(0.47)	0.77	(0.45)	0.51	(0.52)
Mincer-Zarnowitz: $\hat{b}$	0.74	(0.12)	0.67	(0.09)	0.61	(0.09)	0.67	(0.10)
Mincer-Zarnowitz: R^2	0.38		0.38		0.35		0.36	

Note: This table compares evaluation statistics for OOS forecasts based on different window selection methods designed for environments characterized by model instability. For the ease of comparison, OOS forecasts based on the conventional expanding window are also repeated (Recursive). The forecast horizon is k=4 quarters. The first 10 years (1962:Q1-1972:Q1) are used as an initialization period for estimating the parameters of the different models. The window selection schemes include: i) weighted forecasts according to squared OOS-forecast errors in the evaluation window (MSFE-weighted), ii) single-best window with the lowest MSFE in the evaluation period (Cross-validation), and iii) average of forecasts based on different estimation windows (Pooled). Theil's U is the ratio of the RMSE of the models based on the term spread and the RMSE of the AR(1)-benchmark model. MSFE-adj is the difference of the MSFE of the benchmark and the adjusted mean squared forecast error according to Clark and West (2007) (*, **, *** denotes significance of Clark-West's test statistic for testing equal predictive performance at the 10%, 5%, and the 1% level). Coefficients and R^2 of Mincer-Zarnowitz regressions are also reported. Bootstrapped standard errors (MBB with 99,999 replications) are given in parentheses.

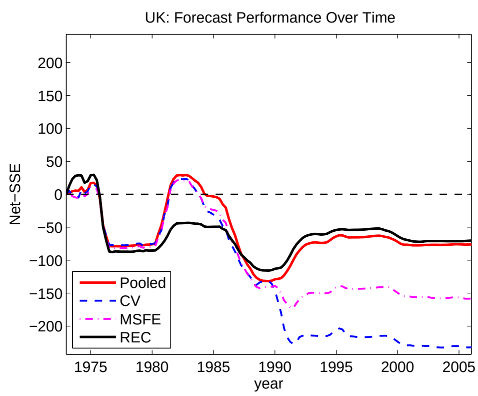
Figure 3.2: Time-varying Forecast Performance (Net-SSE), Window Selection Methods



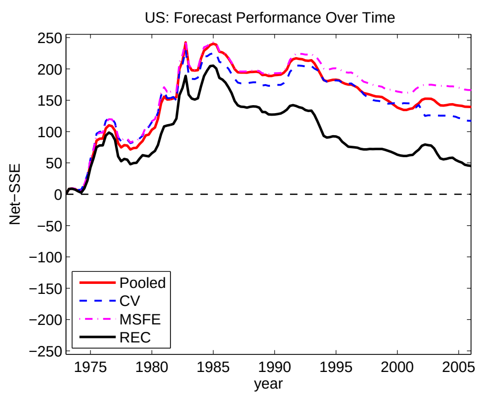
(a) Canada



(b) Germany



(c) UK



(d) US

Note: The figure shows Net-SSE plots for forecasts based on different window selection methods. The forecast horizon is 4 quarters. Net-SSE is the cumulated difference of squared forecast errors of AR(1) benchmark model and the prediction model including the yield spread and lagged GDP growth: Net-SSE$(\tau_0, \tau_1) = \sum_{t=\tau_0}^{\tau_1}(e_{b;t}^2 - e_{m;t}^2)$, where $e_{b;t}$ is the forecast error of the benchmark, and $e_{m;t}$ is the error of the prediction model. A decrease of the slope represents a better forecast performance of the benchmark model at the particular point in time.

3.4 The Role of Other Yield Curve Information

The findings of the previous section suggest that the term spread has been losing its edge as a leading indicator in recent years. Hence, the question emerges whether this finding applies to the yield curve as a whole. There are several reasons why other components of the yield curve may contain information for future real activity beyond the yield spread. First, it can be shown that the term spread can be expressed as the sum of a (risk-neutral) expectations hypothesis component and a term premium component (See e.g. Hamilton and Kim, 2002). Hence, simply using the slope of the yield curve for forecasting implies that potentially useful information contained by the yield curve could be neglected.[22] Second, the level of the short rate could be considered as an alternative measure of the stance of the monetary policy, which may also qualify as a useful predictor as emphasized by Ang, Piazzesi, and Wei (2006).

Therefore, in order to study the role of additional yield curve information for forecasting real activity, we investigate the role of the short rate as well as a measure of time-varying bond return risk premia. This allows us to analyze whether using these variables in predictive regressions with or without the yield spread can be beneficial from a forecasting perspective. Following Wright (2006), we use the bond return forecasting factor by Cochrane and Piazzesi (2005) (denoted as CP-factor) as our proxy for time-varying bond risk premia. Hence, our risk premia proxy is a measure of bond *return* risk premia instead of the *yield* risk premia, which would be needed to decompose the yield spread. We choose to use bond *return* risk premia instead of theoretically more desirable yield risk premia in the face of substantial estimation uncertainties associated with (long-end) yield curve decompositions (See Cochrane and Piazzesi, 2007 for further details). Moreover, as noted by Wright (2006), the Cochrane-Piazzesi factor is correlated with term premia estimates obtained by other alternative methods (based on (arbitrage-free) affine term structure models).

Table 3.6 displays estimation results of alternative model specifications. The sample period covers 1972:Q4-2006:Q2 (Canada, Germany, US) and 1979:Q1-2006:Q2 (UK).[23]

[22]Hamilton and Kim (2002), for instance, decompose the predictive power of the term spread into an expectations hypothesis and a term premium component using instrumental variables to identify the expected path of future short rates. However, their approach using leads of short-term interest rates cannot be used for real-time forecasting, which is the focus of this paper.

[23]The sample periods are restricted by the availability of zero bond data covering a whole range of maturities. Note that data on zero bond yields for Canada (necessary to compute bond risk premia) are only available from official sources for a rather short period. For this reason we omit models including the CP-factor from the table in the case of Canada.

Table 3.6: Predictive Content of the Term Spread and other Yield Curve Variables

Panel A: CAN						Panel B: GER					
Model	(1)	(2)	(3)	(4)	(5)	Model	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}_1^{TS}$	0.711			0.385		$\hat{\beta}_1^{TS}$	0.546			0.425	0.666
t-stat. (BS)	[3.58]			[1.41]		t-stat. (BS)	[3.45]			[1.69]	[2.61]
p-val. (BS)	(0.00)			(0.06)		p-val. (BS)	(0.00)			(0.05)	(0.00)
$\hat{\beta}_1^{SR}$		-0.322		-0.198		$\hat{\beta}_1^{SR}$		-0.257		-0.094	
t-stat. (BS)		[-3.02]		[-1.44]		t-stat. (BS)		[-1.89]		[-0.51]	
p-val. (BS)		(0.01)		(0.05)		p-val. (BS)		(0.04)		(0.43)	
$\hat{\beta}_1^{CP}$						$\hat{\beta}_1^{CP}$			0.103		-0.059
t-stat. (BS)						t-stat. (BS)			[1.35]		[-0.60]
p-val. (BS)						p-val. (BS)			(0.11)		(0.36)
$\hat{\beta}_2$	0.120	0.090		0.092		$\hat{\beta}_2$	0.016	0.038	0.050	0.018	0.012
t-stat. (BS)	[2.21]	[1.61]		[1.73]		t-stat. (BS)	[0.32]	[0.75]	[1.08]	[0.37]	[0.23]
p-val. (BS)	(0.01)	(0.07)		(0.03)		p-val. (BS)	(0.62)	(0.34)	(0.17)	(0.59)	(0.70)
$\bar{R}^2$	0.306	0.318		0.346		$\bar{R}^2$	0.168	0.125	0.050	0.170	0.170
$\%[\bar{R}_b^2] > \bar{R}^2$	(0.00)	(0.00)		(0.00)		$\%[\bar{R}_b^2] > \bar{R}^2$	(0.01)	(0.04)	(0.15)	(0.04)	(0.03)
Panel C: UK						Panel D: US					
Model	(1)	(2)	(3)	(4)	(5)	Model	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}_1^{TS}$	0.475			0.268	0.670	$\hat{\beta}_1^{TS}$	0.876			0.836	0.859
t-stat. (BS)	[2.70]			[1.74]	[3.00]	t-stat. (BS)	[4.23]			[2.87]	[2.30]
p-val. (BS)	(0.01)			(0.03)	(0.01)	p-val. (BS)	(0.00)			(0.00)	(0.04)
$\hat{\beta}_1^{SR}$		-0.271		-0.206		$\hat{\beta}_1^{SR}$		-0.225		-0.033	
t-stat. (BS)		[-2.17]		[-1.56]		t-stat. (BS)		[-1.75]		[-0.26]	
p-val. (BS)		(0.07)		(0.11)		p-val. (BS)		(0.08)		(0.69)	
$\hat{\beta}_1^{CP}$			-0.050		-0.241	$\hat{\beta}_1^{CP}$			0.183		0.007
t-stat. (BS)			[-0.38]		[-1.82]	t-stat. (BS)			[2.29]		[0.04]
p-val. (BS)			(0.61)		(0.07)	p-val. (BS)			(0.02)		(0.92)
$\hat{\beta}_2$	0.121	0.085	0.176	0.076	0.091	$\hat{\beta}_2$	0.075	0.115	0.061	0.074	0.073
t-stat. (BS)	[1.12]	[0.79]	[1.46]	[0.70]	[0.83]	t-stat. (BS)	[1.35]	[1.64]	[1.00]	[1.25]	[1.39]
p-val. (BS)	(0.28)	(0.44)	(0.20)	(0.43)	(0.36)	p-val. (BS)	(0.07)	(0.10)	(0.19)	(0.11)	(0.06)
$\bar{R}^2$	0.233	0.292	0.070	0.325	0.321	$\bar{R}^2$	0.322	0.134	0.178	0.318	0.317
$\%[\bar{R}_b^2] > \bar{R}^2$	(0.01)	(0.00)	(0.13)	(0.00)	(0.00)	$\%[\bar{R}_b^2] > \bar{R}^2$	(0.00)	(0.04)	(0.01)	(0.00)	(0.00)

Note: The table displays estimation results of predictive regressions using model specifications with different yield curve variables. All estimation results are based on models including a constant and lagged GDP growth. The dependent variable is defined as (annualized) cumulative real GDP growth. The forecasting horizon is 4 quarters. t-stat. (BS) is based on MBB standard errors with 99,999 replications, and p-val. (BS) represents the bootstrap p-value (based on studentization). $\bar{R}^2$ denotes the adjusted R^2, and $\%[\bar{R}_b^2] > \bar{R}^2$ denotes the fraction of times where the bootstrap $\bar{R}_b^2$ exceeds $\bar{R}^2$ (based on a parametric bootstrap with 9,999 replications). Coefficient estimates refer to the term spread ($\hat{\beta}_1^{TS}$), the short rate ($\hat{\beta}_1^{SR}$), bond risk premia ($\hat{\beta}_1^{CP}$) and lagged GDP growth ($\hat{\beta}_2$). Sample periods: 1962:Q1-2006:Q2 (Canada), 1972Q4-2006:Q2 (Germany, US), 1979:Q1-2006:Q2 (UK).

As shown by the table, the short rate appears as a significant predictor of real activity for every country considered. The negative coefficient is consistent with the reasoning that an increase of short rate imposes higher costs of investment and is associated with a subsequent slowdown of economic growth. In case of the UK it is noteworthy that the short rate outperforms the term spread in terms of the predictive $\bar{R}^2$. When combining the term spread with the short rate, however, we find that the short rate typically tends to lose much its predictive ability while the spread in most cases maintains its predictive power, consistent with Plosser and Rouwenhorst (1994). Similarly, bond risk premia (as proxied by the Cochrane/Piazzesi factor) generally have a rather limited predictive content. Only in the US we find a significant effect of bond return risk premia which disappears however when the term spread is controlled for. These (in-sample) findings suggest that the major informational content of the yield curve for real activity refers to the slope.

In order to judge the usefulness of alternative yield curve variables for OOS forecasting, we provide evaluation statistics in Table 3.7 for the different model specifications. The table shows that the short rate (Model 2) produces forecasts outperforming the naive model (Theil's U smaller than one and significant Clark-West statistics) similar to the yield spread. The table shows further, however, that forecasts using the yield spread (Model 1) tend to be more accurate. A notable exception, is the UK where there is evidence that the short rate is the better yield curve variable for forecasting. Including both the spread and the short rate generally leads to a degradation in forecast performance. Similarly, the forecast performance of models including return risk premia (Models 3 and 5) is not encouraging.

Based on both in-sample and out-of-sample results, we conclude that the short rate and bond risk premia generally have a rather limited predictive ability and that the term spread typically plays a dominant role.[24] This implies that accounting for additional yield curve information is unlikely to prevent the deterioration of the predictive content

[24]Regarding our conclusions on the role of the short rate, our results differ from those of Ang, Piazzesi, and Wei (2006), who found an increased role of the short rate as a predictor of US output growth in recent years. Our results are more in line with Plosser and Rouwenhorst (1994), which suggests that the short rate plays a different role in models imposing no-arbitrage restrictions as in Ang, Piazzesi, and Wei (2006).

Table 3.7: Out-of-Sample Forecast Evaluation: Yield Spread and other Yield Curve Variables

Model (1)	CAN		GER		UK		US	
Mean Forecast Error	-1.12	(0.25)	-0.63	(0.30)	-0.07	(-0.50)	-1.23	(0.26)
Theil's U	0.87		0.85		1.05		0.97	
MSFE-adj	3.59**		2.94**		0.22		3.77***	
Mincer-Zarnowitz: $\hat{a}$	-0.81	(0.81)	0.38	(0.66)	2.27	(0.87)	-0.12	(0.64)
Mincer-Zarnowitz: $\hat{b}$	0.93	(0.16)	0.61	(0.20)	0.01	(0.33)	0.74	(0.12)
Mincer-Zarnowitz: R^2	0.38		0.15		0.00		0.38	
Model (2)	**CAN**		**GER**		**UK**		**US**	
Mean Forecast Error	-0.52	(0.59)	-1.06	(0.34)	0.36	(0.43)	0.20	(0.65)
Theil's U	0.90		0.93		0.90		1.05	
MSFE-adj	5.13**		3.71**		5.18**		5.66**	
Mincer-Zarnowitz: $\hat{a}$	0.92	(0.92)	0.27	(0.56)	1.16	(0.53)	1.81	(0.46)
Mincer-Zarnowitz: $\hat{b}$	0.59	(0.21)	0.56	(0.15)	0.59	(0.17)	0.43	(0.10)
Mincer-Zarnowitz: R^2	0.20		0.21		0.26		0.21	
Model (3)	**CAN**		**GER**		**UK**		**US**	
Mean Forecast Error			-1.14	(0.32)	-0.25	(0.41)	-0.45	(0.38)
Theil's U			1.00		1.00		1.01	
MSFE-adj			-0.02		0.09*		-0.01	
Mincer-Zarnowitz: $\hat{a}$			1.73	(1.95)	3.95	(1.22)	2.09	(1.08)
Mincer-Zarnowitz: $\hat{b}$			0.07	(0.69)	-0.64	(0.50)	0.26	(0.33)
Mincer-Zarnowitz: R^2			0.00		0.02		0.01	
Model (4)	**CAN**		**GER**		**UK**		**US**	
Mean Forecast Error	-0.56	(0.48)	-0.90	(0.39)	0.33	(0.40)	-0.45	(0.58)
Theil's U	0.90		0.95		0.91		1.02	
MSFE-adj	5.46**		4.45**		5.14**		6.67**	
Mincer-Zarnowitz: $\hat{a}$	0.90	(0.87)	0.61	(0.51)	1.19	(0.54)	1.36	(0.43)
Mincer-Zarnowitz: $\hat{b}$	0.59	(0.20)	0.47	(0.12)	0.56	(0.17)	0.47	(0.09)
Mincer-Zarnowitz: R^2	0.22		0.18		0.25		0.32	
Model (5)	**CAN**		**GER**		**UK**		**US**	
Mean Forecast Error			-0.61	(0.37)	-0.06	(0.49)	-1.21	(0.23)
Theil's U			0.86		1.05		0.97	
MSFE-adj			3.09**		0.20		3.72***	
Mincer-Zarnowitz: $\hat{a}$			0.52	(0.63)	2.31	(0.87)	-0.08	(0.62)
Mincer-Zarnowitz: $\hat{b}$			0.56	(0.19)	0.00	(0.30)	0.73	(0.12)
Mincer-Zarnowitz: R^2			0.14		0.00		0.37	

Note: This table presents various statistics of forecast evaluation (forecast horizon k=4 quarters). Different model specifications based on different yield curve variables (term spread, short rate, return risk premia) are estimated. The model specifications are given as

(1) Const, Term Spread, Lagged Output Growth
(2) Const, Short Rate, Lagged Output Growth
(3) Const, Bond Return Risk Premia (Cochrane-Piazzesi Factor), Lagged Output Growth
(4) Const, Term Spread, Short rate, Lagged Output Growth
(5) Const, Term Spread, Bond Return Risk Premia (Cochrane-Piazzesi Factor), Lagged Output Growth

of the yield curve for real activity in the recent period.

3.5 Conclusion

In this paper we study whether the yield curve can still be regarded as a useful leading indicator in forecasting environments characterized by structural change. Studying the out-of-sample forecast accuracy of models using the yield spread over time relative to a naive benchmark model, we are able to identify periods of particularly good and bad performance. Our general finding is that there is a substantial degradation in the out-of-sample forecast performance of the yield curve for real activity. This result holds for all countries considered in the study (Canada, Germany, UK, and the US).

Another contribution of our paper is to investigate how parameter stability affects the forecasting relationship. Using structural break tests allowing for multiple breaks under the alternative, we find clear evidence for instabilities and are able to pin down the dates associated with structural change. Moreover, we consider how to optimally choose the forecasting estimation window in the presence of such breaks. For this purpose, we use newly developed forecast combination methods by Pesaran and Timmermann (2007) which also use pre-break information for forecasting. While these methods help reduce the bias of forecast errors, they only produce minor improvements in terms of a reduced forecast error variance. Hence, our overall results suggest that the relationship of the yield curve and real activity has become clearly weaker in recent years at the international level.

Our work can still be extended along the following lines. In particular, it would be interesting to investigate further whether the model instabilities and time-variation of out-of-sample forecast performance identified in this paper can be explained by monetary regime shifts or by rather different aspects such as declining output volatility. Another promising area would be to disentangle yield risk premia from the expectations-hypothesis component of the yield spread [building upon the earlier work by Hamilton and Kim (2002)]. The existing literature still falls short of an analysis whether separating the effects is helpful for out-of-sample forecast accuracy. For this purpose, yield risk premia are needed, which can be reliably estimated in real-time without much estimation error. Given the substantial estimation uncertainties noted by Cochrane and Piazzesi (2007), obtaining such decompositions still poses a great challenge. We leave these interesting issues for future research.

Appendix A: Data Description

Table 3.8: Details on Data Construction

Variable	Data Source	Details on Data Construction
Panel A: Canada		
Real GDP	Datastream	Seasonally adjusted time series of real GDP from Statistics Canada.
Long-term interest rate	Datastream/IMF-IFS	Long-term government bond yield (10 years to maturity) from Statistics Canada
Short-term interest rate	Datastream/IMF-IFS	Three-month T-bill rate.
Panel B: Germany		
Real GDP	Reuters-Ecowin	Seasonally adjusted time series of real GDP (Stat. Bundesamt). The outlier in the growth rate of real GDP due to the reunification (1991:Q1) is adjusted by interpolation as in Stock and Watson (2003): the corresponding observation is replaced by the median of the three previous and the three following observations. Long-horizon growth rates are calculated using the one-step growth rates.
Long-term interest rate	Datastream/IMF-IFS	Long term government bond yield (9-10 years to maturity)
Short-term interest rate	Datastream/IMF-IFS	Three-month Money Market Rate calculated from Bundesbank data.
Panel C: UK		
Real GDP	Datastream	Seasonally adjusted time series of real GDP growth (ONS).
Long-term interest rate	Datastream/IMF-IFS	Long term government bond yield (20 years to maturity).
Short-term interest rate	Datastream/IMF-IFS	Treasury-bill rate calculated from Bank of England data.
Panel D: USA		
Real GDP	Datastream	Seasonally adjusted time series of real GDP growth (BEA).
Long-term interest rate	Federal Reserve	Market yield on U.S. Treasury securities with 10-year constant maturity
Short-term interest rate	Federal Reserve	Three-month Treasury-bill rate. Monthly data are transformed into quarterly data.

Note: The sample period is usually 1962:Q1-2006:Q2 unless otherwise indicated.

Appendix B: Estimating Return Risk Premia

This section provides a brief description on the estimation of our measure of time-varying bond risk premia, which is the bond return forecast factor by Cochrane and Piazzesi (2005) (so called CP-factor). First, it is useful to define (one-year) holding period returns (i.e. from t to t+4 quarters) on longer term bonds with n years to maturity as $hpr_{t+4}^{(n)} = p_{t+4}^{(n-1)} - p_t^{(n)}$, where $p_t^{(n)}$ denotes the log price of a bond maturing in n years. By subtracting the one-year interest rate, excess returns $rx_{t+4}^{(n)} = hpr_{t+4}^{(n)} - y_t^{(1)}$ are obtained.

Under the expectations hypothesis, bond excess returns should not be predictable. As shown by Cochrane and Piazzesi (2005), building on previous results by Fama and Bliss (1987), a single combination of forward rates $f_t^{(0,1)}, \dots, f_t^{(m-1,m)}$ is a significant predictor of (one-year) bond excess returns of bonds of all maturities ($n = 2, \dots, m$):[25]

$$rx_{t+4}^{(n)} = \beta_0^{(n)} + \beta_1^{(n)} f_t^{(0,1)} + \dots + \beta_m^{(n)} f_t^{(m-1,m)} + \epsilon_{t+4}^{(n)}, \tag{3.7}$$

where $f_t^{(n-1,n)}$ are forward rates implied by the yield curve: $f_t^{(n-1,n)} = p_t^{(n-1)} - p_t^{(n)}$. The CP-factor as of period t is obtained as the fitted values of a regression of the average of $rx_{t+4}^{(n)}$ over all maturities ($n = 2, \dots, m$) on the term structure of forward rates.[26] Thus, the CP-factor can be regarded as a measure of (one-year) bond return risk premia. In order to avoid look-ahead bias and to make sure that only information truly available to the forecaster as of period t is used, we use a recursively fitted CP-factor as the measure of return risk premia.

[25] Drawing on the Fama/Bliss yield curve data, Cochrane and Piazzesi (2005) consider maturities ranging from 2 to 5 years. Tang and Xia (2005) and Cochrane and Piazzesi (2007) also show that the main results extend to longer maturities and other datasets.

[26] In our implementation, we follow Tang and Xia (2005) in our selection of forward rates ($f_t^{(0,1)}$, $f_t^{(2,3)}$, $f_t^{(9,10)}$) due to multicollinearity problems when neighboring forwards rates are used. See also Cochrane and Piazzesi (2007) for more details on this issue.

General Conclusions

In this thesis various aspects of the relationship between financial markets and the macroeconomy were explored from an empirical perspective. The goal was to shed new light on three different macro-finance issues.

The first chapter investigated the performance of the consumption-based asset pricing model when the relevant consumption risk is measured over long-horizons. There are several reasons suggesting that long-run consumption risk is relevant for asset pricing. For instance, measurement issues, non-separabilities or infrequent consumption adjustment imply that long-horizon consumption growth – as advocated by Parker and Julliard (2005) – may serve as a better measure of consumption risk than contemporaneous consumption growth. Contrary to Parker and Julliard (2005), our empirical asset pricing tests take into account the recent critique of asset pricing tests raised by Lewellen, Nagel, and Shanken (2007) as well as their suggested remedies. Our results generally suggest that more plausible parameter estimates rather than a better fit for the cross-section can be regarded as the major success of the long-horizon consumption-based model.

There are several fields of research which are promising to explore in the future based on the findings in chapter one. First, from a theoretical perspective it may be desirable to base the empirical analysis on moment conditions derived from a preference-based approach with Epstein/Zin utility and a process for consumption growth exhibiting a small predictable component as in Hansen, Heaton, and Li (2008). In light of the fact that the long-run risk framework has been very popular in recent years and several authors have argued that it helps solving long-standing asset pricing puzzles, it may be very promising to bring this model to an international asset pricing context. It may also be very useful to gain a better understanding of the long-run risk framework by studying predictive regressions for long-run consumption growth by common factors of portfolio returns. I am currently exploring these novel issues in another piece of work.

In the second chapter, I investigated the predictability of excess returns in international stock markets. The vast literature on return predictability has identified a large amount of variables which have been found to predict returns. However, often it is assumed that the investor selects a-priori the particular variables or a particular combination of these variables for predicting returns. This is a rather restrictive assumption, since there is no guarantee that the combination of variables that the investor chooses is the right one. Thus, in this chapter, I use a Bayesian model averaging framework in order to account for the uncertainty about the relevance of a particular variable or combination of variables for predicting returns. In addition, my empirical approach accounts for potential biases arising from the strong persistence of the typical predictor variables. Based on an extensive international dataset, I document notable differences in the degree of return predictability across different stock markets. Overall, the findings of this chapter suggest that return predictability is not a uniform and a universal feature across international capital markets.

There are several potentially interesting directions for further research based on the findings reported in the second chapter of the thesis. Thus far, my analysis was based on a rather conventional set of predictor variables, mainly due to restrictions of data-availability for international equity markets. In recent years, however, a plethora of macro variables has been motivated as predictors for returns.[27] It would be interesting to investigate the robustness of predictive relations and model uncertainty using these newly proposed predictors. Another interesting field would be to study the performance of alternative bias correction methods in a comprehensive Monte Carlo study. Many alternative approaches have been put forth in the literature and the empirical setups typically differ (single vs. multiple predictors, one-step vs. multi-step forecasting etc.).

[27]These variables include e.g. the consumption-wealth ratio by Lettau and Ludvigson (2001a), the housing collateral as in Lustig and van Nieuwerburgh (2005), the price-output ratio by Rangvid (2006), the labor income to consumption ratio of Santos and Veronesi (2006) or the expenditure share on non-housing consumption by Piazzesi, Schneider, and Tuzel (2007).

Hence, more knowledge on the accuracy of the different methods in the different setups would be of very useful for future empirical research.

Chapter 3 investigated the predictive properties of the yield curve for real GDP growth in the context of structural instability. This issue is of particular relevance since the so-called term spread (difference between long-term and short-term interest rates) is commonly perceived as one of the most prominent predictors of real activity. In particular, the purpose of this chapter was to study whether the yield spread still qualifies as a useful predictor of real activity in environments characterized by model instability and forecast breakdowns. Using multiple break tests, this chapter provided strong evidence that the predictive relation has been subject to substantial structural change. Moreover, the findings suggest that window selection methods newly developed for forecasting in the presence of structural change offer some improvements in terms of forecast accuracy. Overall, the results reported in this chapter strongly suggest, however, that the yield curve has been losing its edge as a predictor of output growth in recent years.

The work in chapter 3 could still be extended along the following lines. One interesting avenue would be to investigate in greater detail whether the model instabilities and time-variation of out-of-sample forecast performance identified in chapter 3 can be explained by changes in monetary regimes or by rather different aspects such as declining output volatility. Another promising area would be to disentangle yield risk premia from the expectations-hypothesis component of the yield spread in order to investigate whether separating their contributions is helpful for better out-of-sample prediction. For this purpose, yield risk premia are needed, which can be reliably estimated in real-time without much estimation error.[28]

Overall, the thesis provides a critical reassessment of existing empirical findings and facts in the macro-finance literature. A general theme of my results is that some major findings or "facts" which are often taken for granted in the literature are substantially attenuated once we put the finding under scrutiny using an appropriate

[28] One possibility to obtain model-free estimates of yield risk premia could be to use a VAR-approach as in Ludvigson and Ng (2007).

econometric methodology. As outlined in the paragraphs above – though there may have been progress in some regards – many challenges and interesting issues still remain unresolved in the macro-finance literature. Seeking answers to these research questions constitutes an interesting agenda for future research.

BIBLIOGRAPHY

AMIHUD, Y., AND C. HURVICH (2004): "Predictive Regressions: A Reduced-Bias Estimation Method," *Journal of Financial and Quantitative Analysis*, 39, 813–842.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–58.

ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

ANG, A., AND G. BEKAERT (2007): "Stock Return Predictability: Is It There?," *Review of Financial Studies*, 20(3), 651–707.

ANG, A., M. PIAZZESI, AND M. WEI (2006): "What Does the Yield Curve Tell Us about GDP Growth?," *Journal of Econometrics*, 131, 359–403.

AVRAMOV, D. (2002): "Stock Return Predictability and Model Uncertainty," *Journal of Financial Economics*, 64(3), 423–458.

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66(1), 47–78.

——— (2003): "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18, 1–22.

——— (2006): *Econometric Theory and Practice: Frontiers of Analysis and Applied Research,* chap. Multiple Structural Change Models: A Simulation Analysis, pp. 212–237. Cambridge University Press.

BANSAL, R. (2007): "Long-Run Risk and Financial Markets," NBER Working Paper 13196, National Bureau of Economic Research.

BANSAL, R., R. DITTMAR, AND D. KIKU (2007): "Cointegration and Consumption Risks in Asset Returns," Working paper, University of Pennsylvania.

BANSAL, R., R. F. DITTMAR, AND C. T. LUNDBLAD (2005): "Consumption, Dividends, and the Cross Section of Equity Returns," *Journal of Finance*, 60(4), 1639–1672.

BANSAL, R., D. KIKU, AND A. YARON (2007): "Risks For the Long Run: Estimation and Inference," Working paper, University of Pennsylvania.

BANSAL, R., AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

BENATI, L., AND C. GOODHART (2007): "Investigating Time-Variation in the Marginal Predictive Power of the Yield Spread," ECB Working Paper 802, Journal of Economic Dynamics and Control (forthcoming).

BERNARD, H., AND S. GERLACH (1998): "Does the Term Structure Predict Recessions? The International Evidence," *International Journal of Finance and Economics*, 3, 195–215.

BOSSAERTS, P., AND P. HILLION (1999): "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *Review of Financial Studies*, 12(2), 405–428.

BREEDEN, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–96.

BROWN, G. W., AND M. T. CLIFF (2005): "Investor Sentiment and Asset Valuation," *Journal of Business*, 78, 405–440.

CAMPBELL, J. Y. (1987): "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

——— (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 298–345.

CAMPBELL, J. Y., AND J. H. COCHRANE (1999): "By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.

CAMPBELL, J. Y., AND R. J. SHILLER (1988a): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1(3), 195–228.

——— (1988b): "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 43(3), 661–676.

CAMPBELL, J. Y., AND S. THOMPSON (2007): "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?," NBER, Working Paper No. 11468, Review of Financial Studies (forthcoming).

CAMPBELL, J. Y., AND L. M. VICEIRA (2002): *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press, Oxford UK.

CAMPBELL, J. Y., AND T. VUOLTEENAHO (2004): "Bad Beta, Good Beta," *American Economic Review*, 94, 1249–1275.

CAMPBELL, J. Y., AND M. YOGO (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81(1), 27–60.

CHETTY, R., AND A. SZEIDL (2005): "Consumption Commitments: Neoclassical Foundations for Habit Formation," Working paper, UC Berkeley.

CLARK, T. E., AND K. D. WEST (2007): "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138, 291–311.

CLEMENTS, M., AND D. HENDRY (1998): *Forecasting Economic Time Series*. Cambridge Unviversity Press.

COCHRANE, J. H. (1989): "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives," *American Economic Review*, 79(3), 319 – 337.

——— (1996): "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104(3), 572–621.

——— (1999): "New Facts in Finance," *Economic Perspectives*, 23, 36 – 58.

——— (2005): *Asset Pricing*. Princeton University Press, Princeton, NJ.

COCHRANE, J. H. (2006): "The Dog That Did Not Bark: A Defense of Return Predictability," Working Paper, University of Chicago, Review of Financial Studies (forthcoming).

COCHRANE, J. H. (2007): "Financial Markets and the Real Economy," in *Handbook of the Equity Premium*, ed. by R. Mehra, pp. 237–325. Elsevier.

COCHRANE, J. H., AND M. PIAZZESI (2005): "Bond Risk Premia," *American Economic Review*, 95, 138–160.

——— (2007): "Decomposing the Yield Curve," Working paper, University of Chicago.

COLACITO, R., AND M. M. CROCE (2007): "Risks For The Long Run And The Real Exchange Rate," Discussion paper, Stern School of Management.

CONSTANTINIDES, G. M., AND D. DUFFIE (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104(2), 219–240.

COOPER, I., AND R. PRIESTLEY (2006): "Time-varying Risk Premia and the Output Gap," *Review of Financial Studies*, forthcoming.

CREMERS, K. J. M. (2002): "Stock Return Predictability: A Bayesian Model Selection Perspective," *Review of Financial Studies*, 15(4), 1223–1249.

DANGL, T., M. HALLING, AND O. RANDL (2006): "Equity Return Prediction: Are Coefficients Time Varying?," Working Paper, University of Vienna.

DEMERS, F., AND R. MACDONALD (2007): "The Canadian Business Cycle: A Comparison of Models," Working Paper 07-38, Bank of Canada.

DIMSON, E., S. NAGEL, AND G. QUIGLEY (2003): "Capturing the Value Premium in the United Kingdom," *Financial Analysts Journal*, 59(6), 35 – 45.

DOTSEY, M. (1998): "The Predictive Content of the Interest Rate Term Spread for Future Economic Growth," *Economic Quarterly*, 84, 31–51.

DUARTE, A., I. A. VENETIS, AND I. PAYA (2005): "Predicting Real Growth and the Probability of Recession in the Euro Area Using the Yield Spread," *International Journal of Forecasting*, 21, 261–271.

ELLIOTT, G., AND U. K. MÜLLER (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73(4), 907–940.

EPSTEIN, L., AND S. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption Growth and Asset Returns I: A Theoretical Framework," *Econometrica*, 57, 937–969.

ESTRELLA, A. (2005): "Why Does the Yield Curve Predict Output and Inflation?," *The Economic Journal*, 115, 722–744.

ESTRELLA, A., AND G. HARDOUVELIS (1991): "The Term Spread as a Predictor of Real Economic Activity," *Journal of Finance*, 46, 555–576.

ESTRELLA, A., A. P. RODRIGUES, AND S. SCHICH (2003): "How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States," *Review of Economics and Statistics*, 85, 629–644.

FAMA, E. F. (1981): "Stock Returns, Real Activity, Inflation, and Money," *American Economic Review*, 71, 545–565.

——— (1990): "Stock Returns, Expected Returns, and Real Activity," *Journal of Finance*, 45(4), 1089–1108.

FAMA, E. F., AND R. R. BLISS (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

FAMA, E. F., AND K. R. FRENCH (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22(1), 3–25.

——— (1989): "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, 25(1), 23–49.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

——— (2000): "The Equity Premium," University of Chicago.

FAMA, E. F., AND G. W. SCHWERT (1977): "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115–146.

FERNÁNDEZ, C., E. LEY, AND M. F. J. STEEL (2001): "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 100, 381–427.

FERSON, W. E., S. SARKISSIAN, AND T. T. SIMIN (2003): "Spurious Regressions in Financial Economics," *Journal of Finance*, 58, 1393–1413.

FRENCH, K., G. W. SCHWERT, AND R. F. STAMBAUGH (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

GAO, P., AND K. X. HUANG (2004): "Aggregate Consumption-Wealth Ratio and the Cross-Section of Stock Returns: Some International Evidence," Working Paper RWP 04-07, Federal Reserve Bank of Kansas City, Kansas City, MO.

GHYSELS, E., A. GUAY, AND A. HALL (1997): "Predictive Tests for Structural Change with Unknown Breakpoint," *Journal of Econometrics*, 82, 209–233.

GIACOMINI, R., AND B. ROSSI (2006): "How Stable is the Predictive Performance of the Yield Curve for Output Growth ?," *Oxford Bulletin of Economics and Statistics*, 68, 783–795.

GIOT, P., AND M. PETITJEAN (2006): "International Stock Return Predictability: Statistical Evidence and Economic Significance," CORE Discussion paper 2006/88.

GONCALVES, S., AND H. WHITE (2005): "Bootstrap Standard Error Estimates for Linear Regressions," *Journal of the American Statistical Association*, 100, 970–979.

GOYAL, A., AND I. WELCH (2003): "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49, 639–654.

——— (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1456–1508.

GROSSMAN, S. J., AND R. J. SHILLER (1982): "The Determinants of the Variability of Stock Market Prices," *American Economic Review*, 71, 222–227.

HAMILTON, J. D., AND D. H. KIM (2002): "A Reexamination of the Predictability of Economic Activity Using the Yield Spread," *Journal of Money, Credit and Banking*, 34, 340–360.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

HANSEN, L. P., J. C. HEATON, AND N. LI (2008): "Consumption Strikes Back?: Measuring Long-Run Risk," Working paper, University of Chicago, Journal of Political Economy (forthcoming).

HANSEN, L. P., AND R. J. HODRICK (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88(5), 829–853.

HANSEN, L. P., AND R. JAGANNATHAN (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52(2), 557–590.

HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(2), 1296–1286.

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–265.

HARVEY, C. R. (1989): "Forecasts of Economic Growth from the Bond and Stock Markets," *Financial Analysts Journal*, 49, 38–45.

HJALMARSSON, E. (2004): "On the Predictability of Global Stock Returns," Working paper, Journal of Financial and Quantitative Analysis (forthcoming).

HODRICK, R. J. (1992): "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5(3), 357–386.

HYDE, S., AND M. SHERIF (2005a): "Consumption Asset Pricing Models: Evidence from the UK," *The Manchester School*, 73(3), 343 – 363.

——— (2005b): "Don't Break the Habit: Structual Stability Tests of Consumption Asset Pricing Models in the UK," *Applied Economics Letters*, 12, 289–296.

INOUE, A., AND L. KILIAN (2004): "In-Sample or Out of Sample Tests of Predictability: Which One Should We Use?," *Econometric Reviews*, 23(4), 371–402.

JAGANNATHAN, R., AND Y. WANG (2007): "Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns," *Journal of Finance*, 62(4), 1623–1661.

JAGANNATHAN, R., AND Z. WANG (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51(1), 3–53.

JORION, P., AND F. MISHKIN (1991): "A Multicountry Comparison of Term-Structure Forecasts at Long Horizons," *Journal of Financial Economics*, 29, 59–80.

KALTENBRUNNER, G., AND L. A. LOCHSTOER (2007): "Long-Run Risk Through Consumption Smoothing," Working paper, London Business School.

KILIAN, L. (1999): "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?," *Journal of Applied Econometrics*, 14(5), 491–510.

KOOP, G. (2003): *Bayesian Econometrics*. Wiley, Chichester.

LAMONT, O. (1998): "Earnings and Expected Returns," *Journal of Finance*, 53(5), 1563–1587.

LEAMER, E. E. (1978): *Specification Searches: Ad Hoc Inference with Non-Experimental Data*. Wiley.

LETTAU, M., AND S. LUDVIGSON (2001a): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

LETTAU, M., AND S. LUDVIGSON (2001b): "Ressurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1287.

LEWELLEN, J. (2004): "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, 74(2), 209–235.

LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2007): "A Skeptical Appraisal of Asset-Pricing Tests," Working paper, Stanford University, Journal of Financial Economics (forthcoming).

LUCAS, ROBERT E., J. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429–1445.

LUDVIGSON, S., AND S. NG (2007): "Macro Factors in Bond Risk Premia," Working paper, New York University, Review of Financial Studies (forthcoming).

LUND, J., AND T. ENGSTED (1996): "GMM and Present Value Tests of the C-CAPM: Evidence from the Danish, German, Swedish and UK Stock Markets," *Journal of International Money and Finance*, 15(4), 497 – 521.

LUSTIG, H. N., AND S. G. VAN NIEUWERBURGH (2005): "Housing Collateral, Consumption Insurance, and Risk Premia: an Empirical Perspective," *Journal of Finance*, 60(3), 1167–1219.

MALLOY, C. J., T. J. MOSKOWITZ, AND A. VISSING-JØRGENSEN (2006): "Long-Run Stockholder Consumption Risk and Asset Returns," Working paper, Northwestern University.

MARK, N. C. (1995): "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review*, 85(1), 201–218.

MEHRA, R., AND E. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

MERTON, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41(5), 867–887.

MINCER, J., AND V. ZARNOWITZ (1969): *Economic Forecasts and Expectations,* chap. The Evaluation of Economic Forecasts. National Bureau of Economic Research.

MOON, R., A. RUBIA, AND R. VALKANOV (2006): "Long-Horizon Regressions when the Predictor is Slowly Varying," Working Paper.

NEELEY, C. J., AND P. WELLER (2000): "Predictability in International Asset Returns: A Reexamination," *Journal of Financial and Quantitative Analysis*, 35(4), 601–620.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

PARKER, J. A., AND C. JULLIARD (2005): "Consumption Risk and the Cross Section of Expected Returns," *Journal of Political Economy*, 113(1), 185 – 222.

PASTOR, L., AND R. STAMBAUGH (2006): "Predictive Systems: Living with Imperfect Ppredictors," Working Paper, GSB Chicago.

PAYE, B. S., AND A. TIMMERMANN (2006): "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13(3), 274–315.

PESARAN, M., AND A. TIMMERMANN (1992): "A Simple Non-Parametric Test of Predictive Performance," *Journal of Business and Economic Statistics*, 10, 461–465.

PESARAN, M. H., AND A. TIMMERMANN (2007): "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics*, 127(1), 134–161.

PETKOVA, R. G. (2006): "Do the Fama-French Factors Proxy for Innovations in Predictive Variables?," *Journal of Finance*, 61(2), 581–612.

PHALIPPOU, L. (2007): "Can Risk-Based Theories Explain the Value Premium?," *Review of Finance*, 11, 143 – 166.

PIAZZESI, M., M. SCHNEIDER, AND S. TUZEL (2007): "Housing, Consumption and Asset Prices," *Journal of Financial Economics*, 83, 531–569.

PLOSSER, C. I., AND K. G. ROUWENHORST (1994): "International Term Structures and Real Economic Growth," *Journal of Monetary Economics*, 33, 133–156.

RANGVID, J. (2006): "Output and Expected Returns," *Journal of Financial Economics*, 81, 595–624.

——— (2008): "The Long-Run World Consumption Risk of International Stock Markets," Working paper, Copenhagen Business School.

RAPACH, D. E., M. E. WOHAR, AND J. RANGVID (2005): "Macro Variables and International Stock Return Predictability," *International Journal of Forecasting*, 21, 137–166.

RAVAZZOLO, F., R. PAAP, D. VAN DIJK, AND P. H. FRANSES (2006): "Bayesian Model Averaging in the Presence of Structural Breaks," Working Paper, Erasmus University Rotterdam.

RUBINSTEIN, M. (1976): "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics*, 7, 407–425.

SALA-I-MARTIN, X., G. DOPPELHOFER, AND R. I. MILLER (2004): "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94, 813–835.

SANTOS, T., AND P. VERONESI (2006): "Labor Income and Predictable Stock Returns," *Review of Financial Studies*, 19(1), 1–44.

SCHRIMPF, A., M. SCHRÖDER, AND R. STEHLE (2007): "Cross-Sectional Tests of Conditional Asset Pricing Models: Evidence from the German Stock Market," *European Financial Management*, 13(5), 880–907.

STAMBAUGH, R. F. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

STOCK, J. H., AND M. W. WATSON (1989): *NBER Macroeconomics Annual 1989,* chap. New Indexes of Coincedent and Leading Economic Indicators, pp. 352–394. NBER.

——— (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 151, 788–829.

——— (2004): "An Empirical Comparison of Methods for Forecasting Using many Predictors," Working paper, Harvard University.

TANG, H., AND Y. XIA (2005): “An International Examination of Affine Term Structure Models and the Expectations Hypothesis,” Working paper, Wharton School.

TIMMERMANN, A. (2006): “Forecast Combinations,” in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: North-Holland.

TOROUS, W., R. VALKANOV, AND S. YAN (2004): “On Predicting Returns with nearly Integrated Explanatory Variables,” *Journal of Business*, 77, 937–966.

WELCH, I. (2001): “The Equity Premium Consensus Forecast Revisited,” Working Paper, Brown University.

WRIGHT, J. H. (2006): “The Yield Curve and Predicting Recessions,” Working paper, Federal Reserve Board.

YOGO, M. (2006): “A Consumption-Based Explanation of Expected Stock Returns,” *Journal of Finance*, 61(2), 539–580.