Qingwei Wang

Investor Beliefs and Forecast Evaluation

Inaugural-Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

Vorgelegt im Sommersemester 2010

Dekan: Prof. Dr. Hans H. Bauer

Referent: Prof. Dr. Ernst Maug

Korreferent: Prof. Dr. Stefan Ruenzi

Tag der mündlichen Prüfung: 03. März 2010

ACKNOWLEDGEMENTS

Writing this dissertation is an exceptional journey with both sweets and joys. Yet it won't be possible to get finished without support of numerous individuals.

At the outset, I would like to thank my supervisor, Prof. Ernst Maug, for his support and patience in the past four years. I am fortunate to have the opportunity to write the dissertation under his supervision. He has constantly encouraged me to remain focused on achieving my goal, inspiring and guiding me to pursue an academic career in finance. I appreciate his thorough and enlightening comments on my papers.

I am also grateful to Prof. Stefan Ruenzi, a member of my dissertation committee. He was very diligent at commenting my article, suggesting new approaches and guiding my research direction. I benefit tremendously from his encouragement and illuminating advice.

My thanks go also to Dr. Michael Schröder. During the course of my stay as a researcher at the Centre for European Economic Research, I have benefited greatly from interactions with him, both from joint research projects and from the joint research paper. I am indebted to him for providing me the freedom for conducting my research.

Many local "peers" at the Centre for European Economic Research and University of Mannheim have contributed towards shaping this thesis. I am indebted to Prof. Francois Laisney, Waldemar Rotfuss, Sandra Schmidt for their critical comments and helpful suggestions. I also thank my coauthors, Chritian Dick, Prof. Bernd Fitzenberger, Karsten Kohn, Pei Kuang, Christoph Schneider, Andreas Schrimpf and Michael Schröder, whose insights and interactions keep the joint research a particularly enjoyable experience.

Many thanks to Hela Hellerich, Richard Hinz, Nataliya Matosova, Frieder Mokinski and Jonas Vogt for their excellent research assistance. Hela Hellerich has provided careful proof-reading of the different chapters of this thesis.

Finally, I wish to extend my special thanks to my parents, Fangying Li and Dezhi Wang, and my sister, Yun Wang, for their love and encouragement in the past years. I also thank my parents in-laws for their invaluable support. Of course, I am immensely grateful to my wife, Ru Xie, who provided a constant source of love, comfort, and stability through the ups and downs during the course of this thesis.

Qingwei Wang

Contents

List of Tables

LIST OF TABLES

LIST OF FIGURES

PREFACE

Introduction

Investor beliefs and forecast evaluation have emerged gradually to become major interest of my research during my Ph.D. study. The forward-looking nature of economic agents in decision-making, which drives a wedge between economics and natural science, rendering agent's belief or forecast a central and fantastic object of research. In modern finance theory expectation plays a key role. Asset prices, for example, clearly depend on expected future price. Therefore, how investor expectations are formed, and what are the effects they may have on decisions and financial market outcomes have become issues of central importance.

Forming a "right" belief is usually a challenging task for investors. In the real world a common problem faced by agents is that the underlying true model is unknown. Economic theory often provides no clear-cut guidelines on the exact way of forming one's "correct" belief. Sometimes agents need to formulate expectations on the basis of available data with desired forecasting models. Yet to find the "right" model to form beliefs means a model search if there are multiple forecasts to choose from. Forecast evaluation becomes essential for decision making in these situations. A failure to make correct evaluation creates biased beliefs, which can in turn affect the economic outcomes.

Additional complication comes from the fact that relationships among economic variables are time varying, although a stable relationship is repeatedly assumed in economics and finance literature. How to form a better belief in such an environment deserves a close scrutiny.

This dissertation examines these issues under several specific but important situations. The first chapter studies the impact of investors' biased belief and its cross-investor dispersion on skewness of future aggregate market return. The measures of biased belief and its dispersion is constructed from one class of commonly used forecasting models in financial market: technical analysis. To assess whether such forecasts are indeed biased, the second essay examines the extent of bias due to the search for superior trading strategies, a common practice engaged by investors and researchers. The third chapter focuses on forecast evaluation. It examines the forecasting performance of yield curve variables for economic growth, especially in an environment characterized by model 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 literature.

Belief Induced Market Skewness

Chapter 1 ("Sentiment, Convergence of Opinion, and Market Crash") investigates whether investor beliefs help forecast the return asymmetry. More concretely, it examines whether two features of beliefs among trend chasing investors, investor sentiment and differences of opinion, predict the conditional skewness of aggregate market returns.

Asymmetry is prevalent in stock returns. In particular, market returns "go up by the stairs and down by the elevator". This phenomena mirrors the empirical evidence of negative skewness, which is often taken as a measure of crash risk for the aggregate stock market. More interestingly, market sometimes crashes without substantial news to justify it. A famous example of this kind is the "Black Monday" in October 1987.

Several theories have been proposed to explain the economic mechanism underlying the observed skewness, such as stochastic bubbles theory [Blanchard and Watson (1982)], leverage effect, volatility feedback effect. Most of the empirical literature, however, are unable to establish these economic links for the aggregate market return. For example, Charoenrook and Daouk (2008) find evidence inconsistent with leverage effect and volatility feedback effect.

Recent developments in behavior finance augment the standard theory with biased expectations such as investor sentiment [e.g., De Long, Shleifer, Summers, and Waldmann (1990)], or with heterogeneity in expectations [e.g., Hong and Stein (2003)], to facilitate an alternative explanation. One prominent example of investor sentiment in theoretical papers is the technical trading, where investors form their beliefs based on mechanical trading rules without consideration of fundamentals. Technical analysis generates also differences of opinion when different investors apply different trading strategies. This heterogeneity stems solely from the application of different models and only on public information, which distinguishes itself from studies of information asymmetry. It is therefore interesting to examine the implications of investor sentiment and differences of opinion on asset price dynamics due to the application of technical analysis.

The main goal of this chapter is to examine whether investor sentiment and disagreement among trend-chasing investors forecast skewness in daily aggregate stock market returns. I make two contributions to the literature. First, I construct a new measure of investor sentiment and differences of opinion among trend-chasing investors from forecasts of a spectrum of trend-chasing strategies. The new sentiment indicator correlates with classical investor sentiment indicators. Its advantage lies in its clear interpretation as an example of demand shift without a fundamental rationalization. Furthermore, it can be constructed for a much longer time period for which other sentiment indicators are not available. The novel way of constructing investor sentiment and differences of opinion can be applied to various asset markets, as long as asset price is available. Second, I empirically show that both investor sentiment and differences of opinion forecast skewness in daily aggregate stock market returns. In particular, the role of differences of opinion depends on the states of average investor sentiment: it positively forecasts market skewness in a optimistic state, but negatively forecasts it in a pessimistic state. I provide an explanation for the role of difference of opinion by augmenting the theory of Abreu and Brunnermeier (Econometrica 2003) with heterogeneous beliefs among trend-chasing investors. I argue that convergence of opinions among trend-chasers in a bullish state indicates that the price run-up is unlikely to be sustained since fewer investors can remain net buyers in the future. Therefore rational arbitragers coordinate their attack on the bubble, leading to a market crash. Vice versa, the convergence of opinion in a bearish state promotes coordinated purchases among rational arbitragers, leading to a strong recovery.

Illusory Profitability of Technical Analysis?

Investor sentiment as biased belief may stem from psychological bias such as overconfidence, confirmation bias, etc. It can also come from incapability of conducting proper statistical inference. One inference problem is that agents tend to ignore the extensive search conducted before inference, leading to data mining bias. *Chapter 2 ("How Illusory is the Profitability of Technical Analysis?")* empirically investigates whether the often-held belief among investors about the usefulness of technical analysis as a useful market timing tool is a biased belief, where the bias stems mainly from data mining bias.

The profitability of technical trading rules is a conundrum. On the one hand, technical analysis only needs public available information like asset prices. If the market is (weakly) efficient, as widely believed among financial economists, rational investors should quickly arbitrage away the profits, and therefore leave no room for technical analysis. Yet at least 90% of experienced traders place some weight on it in costly trading activity [Taylor and Allen (1992)].

Empirical evidence is more supportive for the technical analysis, which seems to resolve this conundrum to a high extent. For example, Park and Irwin (2007) surveys 95 modern studies and finds that 56 studies report profitability, 20 studies report negative results, while the rest indicate mixed results. These results, when taken at face value, tend to reject the market efficiency hypothesis and lend support for the traders' practice. Economists at large maintain that the profitability of technical analysis is illusory and the empirical evidence is mostly due to data mining bias. For example, Jensen (1967) highlights this problem:

> ... if we begin to test various mechanical trading rules on the data we can be virtually certain that if we try enough rules with enough variants we will eventually find one or more which would have yielded profits (even adjusted for any risk differentials) superior to a buy-and-hold policy. But, and this is the crucial question, does this mean the same trading rule will yield superior profits when actually put into practice?

In this chapter we highlight the danger of data mining bias by quantifying its effect using several recently developed statistical tests. These tests enable us to identify not only whether the best trading strategy is profitable, but also how many seemingly profitable strategies have genuine predictive ability or pure luck. Studying 25988 trading strategies for emerging foreign exchange markets, we find thousands of profitable rules if we ignore the data snooping bias. The profitability disappears, however, when data snooping bias is taken into account. Our results indicate that the profitability of technical analysis documented in the literature is vastly overstated due to data snooping bias, and the belief of many investors is biased due to the incapability of correct inference.

We also show that out-of-sample tests in such a context have substantial data snooping bias. Such tests are often proposed as an effective way to alleviate the bias due to data mining. However, when investors search for profitable trading strategies both in-sample and out-of-sample, out-of-sample tests become vulnerable to data snooping too. We highlight this danger in our empirical study.

The Yield Curve as a Leading Indicator under Structural Instability

In *Chapter 3 ("A Reappraisal of the Leading Indicator Properties of the Yield Curve in the Presence of Structural Instability")*, we provide an extensive re-examination of the leading indicator properties of the yield curve in four major developed countries.

PREFACE

Term spread (difference between the long- and short-term interest rate) is among the most prominent financial variables for predicting real activity. An inverted yield curve is commonly believed to signal recessions [See e.g. Estrella and Hardouvelis (1991) and more recent work by Hamilton and Kim (2002) or Stock and Watson (2003)]. However, recent literature has raised the concern 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). The economic rationale for the potential instability stems from, for instance, the form of the monetary policy reaction function or the relative importance of real and nominal shocks in the economy. As these factors potentially change substantially over time (e.g., a change in monetary policy regime), it raises the need to investigate the time-variation of the forecasting relationship in greater detail.

We provide an extensive reexamination of the leading indicator properties of the yield curve in environments characterized by model instability. Unlike the well studied and established in-sample predictive performance of the yield curve, our approach focuses explicitly on its time-varying out-of-sample (OOS) forecasting properties. 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.

We find that there is a substantial time-variation in the out-of-sample forecast performance of the yield curve for real activity. Moreover, we document that the information contained in the yield curve has substantially declined in past years, which has not been shown in the literature before. This finding holds true for all countries considered. To shed some light on potential reasons for the time-variation of predictive power and forecast breakdowns, we use several modern (in-sample) tests for parameter stability and link identified break dates to real and nominal shocks. Using these econometric tools, we provide strong evidence for structural instabilities that affect the predictive relationship.

Our paper raises 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 employ 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

Sentiment, Convergence of Opinion, and Market Crash

ABSTRACT

I introduce a novel proxy of investor sentiment and differences of opinion among trend-chasing investors to forecast skewness in daily aggregate stock market returns. The new proxy is an easy-to-construct, real-time measure available at different frequencies for more than a century. Empirically I find that negative skewness is most pronounced when investors have experienced high sentiment. The role of differences of opinion depends on the states of average investor sentiment: it positively forecasts market skewness in an optimistic state, but negatively forecasts it in a pessimistic state. Conceptually, I provide an explanation for the role of differences of opinion by augmenting the theory of Abreu and Brunnermeier (2003) with heterogeneous beliefs among trend-chasing investors. I argue that convergence of opinion in an optimistic state indicates that the price run-up is unlikely to be sustained since fewer investors can remain net buyers in the future. Therefore rational arbitrageurs coordinate their attack on the bubble, leading to a market crash. Vice versa, the convergence of opinion in a pessimistic state promotes coordinated purchases among rational arbitrageurs, leading to a strong recovery.

"When everyone thinks alike, everyone is likely to be wrong"

Neill (1997)

1.1 Introduction

Asymmetry is prevalent in stock returns. In particular, market returns "go up by the stairs and down by the elevator". This phenomena mirrors the empirical evidence of negative skewness, which is often taken as a measure of crash risk for the aggregate stock market [e.g., Chen, Hong, and Stein (2001), Hong and Stein (2003)]. Several theories have been proposed to explain the economic mechanism underlying the observed skewness. Most of the empirical literature, however, is unable to establish these economic links for the aggregate market return.

In this paper I study the role of investor sentiment and differences of opinion in forecasting skewness in the aggregate daily stock market returns. Motivated by experimental, empirical, and survey evidence that investors chase the trend,[1] I derive both measures of investor sentiment and differences of opinion from forecasts of a spectrum of commonly used trend following trading strategies. I find that negative (positive) skewness is most pronounced when investors have experienced high (low) sentiment. The role of differences of opinion depends on the states of average investor sentiment: when trend-chasing investors are on average pessimistic, differences of opinion negatively forecast market skewness; when they are on average optimistic, differences of opinion positively forecast market skewness. These results hold regardless of whether or not investors are allowed to learn and select desired trading strategies based on various measures of past performance.

Why investor sentiment? Our history has witnessed investor manias during famous bubbles, such as the Tulip mania and the South Sea Company bubble, which were

[1]See Andreassen and Kraus (1988), Hommes, Sonnemans, Tuinstra, and Van de Velden (2005), and Haruvy, Lahav, and Noussair (2007) for examples of experimental evidence, and Frankel and Froot (1988), Taylor and Allen (1992) and Gehrig and Menkhoff (2004) for survey evidence. Griffin, Harris, and Topaloglu (2006) show that the actual trades of day traders follow the trend.

followed by a crash. The notion of "Irrational Exuberance" in Shiller (2005) suggests that investors are inspired by past performances of stock market and become more optimistic, bidding up prices further. As the price cannot deviate too much from the fundamental, this eventually leads to a crash. Therefore, one would expect that high investor sentiment predicts a subsequent crash, which is confirmed by my findings.

The role of differences of opinion is much more controversial. Theoretical models of differences of opinion deliver different predictions. Hong and Stein (2003) predict that disagreement intensity decreases skewness, while Xu (2007) predicts that it increases skewness. Empirical studies fail to find that detrended turnover, a proxy for differences of opinion, forecasts market skewness. Unlike those studies, I focus on differences of opinion among trend-chasing investors and examine its implication for skewness. To this end, I construct a new measure of differences of opinion from the forecasts of trend-following trading strategies. Furthermore, I conjecture that the role of differences of opinion can change in different regimes of average investor sentiment. Intuitively, a situation in which almost everyone believes the market will go down has conceivably different implications for the market movement than a situation in which almost everyone believes the market will go up, although the differences of opinion are very low in both cases. My empirical evidence supports this conjecture.

What explains the distinct role of disagreement across sentiment regimes? I argue that differences of opinion resemble the sustainability of bubbles/market downturn, facilitating the coordination among rational arbitragers. Two key ingredients underlie this argument: limited wealth or margin restriction of investors and the synchronization problem of rational arbitragers. Investors usually have limited wealth or margin restrictions and cannot buy/sell without limitations. They hold one of the following beliefs: bearish, bullish, or neutral, which can be changed when observing new price information. Arbitragers can attack a bubble, but they have difficulty in temporally coordinating their strategies since they do not know when others will attack. Bubbles persist due to this synchronization problem [Abreu and Brunnermeier (2003)]. I augment this theory with heterogeneous beliefs among trend-chasing investors. I argue that convergence of opinions among trend-chasing investors can help rational arbitrageurs to coordinate their actions. The intuition is the following: when almost everyone is optimistic, the average belief is positive and the dispersion of beliefs is very small. Since the investors who became optimistic at an early stage have fully invested during a bubble, fewer investors can continue to buy. A bubble is therefore unlikely to be sustained. Understanding the absence of further momentum, and understanding that other rational arbitragers are likely to perceive it in a similar way, rational arbitragers can coordinate their attack on the bubble. If they do, a market crash is expected. On the other hand, when the average belief is positive, a high dispersion of beliefs implies that many trend-following investors have neutral or bearish beliefs, which can become bullish to fuel the rally when the price continues to increase. Therefore, the bubble can be sustained, and rational arbitragers are likely to choose to ride the bubble. Similarly, the convergence of opinion in the low sentiment period indicates that the market downturn is unlikely to be sustained, so the rational arbitragers coordinate to buy. As a result, a strong rebound is expected.

Constructing a sentiment indicator from forecasts of trend-following strategies deserves a more in-depth explanation. First, forecasts of trend-following trading strategies are widely held as deviations from fundamentals. Theoretical papers often take technical analysis as an example of investor sentiment [e.g., Shleifer and Summers (1990)]. Empirically, although earlier studies often report significant profitability by applying technical analysis, recent studies show that little profitability remains once the data snooping bias is controlled for [e.g., Sullivan, Timmermann, and White (1999)]. It is therefore appealing that forecasts of trend-following trading strategies represent investor sentiment rather than changes in fundamentals. I empirically validate this conjecture by showing significant correlations between trend-chasing sentiment and other common proxies of investor sentiment. Second, forecasts of trend-following trading strategies capture the sentiment of a wide range of people, such as technical analysts, investors who pick their stocks according to the recommendation of newsletters[2], or investors who entrust their money to other trend-followers.

[2]In his interview with Forbes, the editor of Investor's Intelligence stated that: "Most [newsletters] are trend followers...". See Forbes.com, March 19, 2002.

The new proxy of investor sentiment and disagreement is an easy-to-construct, real-time measure, which is available at different frequencies for more than a century. It enables me to study interesting episodes, such as the Roaring Twenties and the Great Depression, whereas other commonly-used proxies, such as closed-end fund discounts are available only since the 1960s. It can be potentially useful for testing models of investor sentiment and differences of opinion for various asset markets as well as other countries, especially when other proxies are sparse.

The rest of this paper is structured as follows. Section 1.2 contains a brief overview of literature. Section 1.3 presents the data and the universe of the trading strategies. Section 1.4 discusses empirical results on the forecasting ability of sentiment and differences of opinion. Section 1.5 concludes.

1.2 Literature

A number of theories have been proposed within a representative agent framework to explain negative skewness in aggregate stock market returns. Early models include the leverage effect, the volatility feedback effect and the stochastic bubble. The leverage effect [Black (1976) and Christie (1982)] suggests that the financial and operating leverage of the firm rises when its stock price drops, followed by an increase in subsequent stock return volatility. When the stock price increases, however, the return volatility is reduced due to a decline in the leverage. The asymmetric response of volatility to change in return renders negative skewness. Alternative theories by Pindyck (1984), French, Schwert, and Stambaugh (1987), and Campbell and Hentschel (1992) propose a volatility feedback effect by relating volatility and risk premium to the arrival of either good or bad news. Both good news and bad news increase volatility, and hence the risk premium. However, the risk premium offsets good news while amplifying bad news, resulting in a deeper drop at the arrival of bad news than of good news. The stochastic bubble theory by Blanchard and Watson (1982) postulates that negative skewness arises from the popping of bubbles, a rare event that leads to very large negative returns.

Hong and Stein (2003) deviate from the previous models by incorporating heterogeneity among agents. They explain skewness through the revelation of bad news hidden by short-sales constraints when investors hold different opinions. Since the higher the dispersion of beliefs, the higher trading volume will be, this model implies that trading volume negatively forecasts skewness. Xu (2007) adopts a similar framework, but investors disagree on the precision of a publicly observed signal. He shows that the equilibrium asset price is a convex function of the signal, due to the different information sensitivity of high and low precision investors. The model predicts that disagreement intensity increases contemporaneous skewness.

The empirical literature, however, is unable to establish these links for the aggregate market return. For example, Chen, Hong, and Stein (2001), Hueng and McDonald (2005), and Charoenrook and Daouk (2008) do not find that detrended turnover forecasts market skewness.[3] Chen, Hong, and Stein (2001) and Xu (2007) find a negative relationship between skewness and past returns, which is consistent with the stochastic bubble theory of Blanchard and Watson (1982) and the price convexity theory of Xu (2007). However, Charoenrook and Daouk (2008) find returns are more negatively skewed following an increase in stock prices and are more positively skewed following a decrease in stock prices. They argue that these findings cannot be coherently explained by leverage effect and volatility feedback effect theories, which predict more negatively skewed returns following a stock price decline; this is only partly consistent with the stochastic bubble theory, which predicts more negatively skewed returns following a period of stock price increase.

This paper relates to the literature on the asset pricing implications of investor sentiment. Empirical studies have examined the role of investor sentiment for near-term and long-term returns, volatility and trading volume. One challenge of these studies arises from the difficulty of measuring sentiment. As a belief unjustified by fundamentals, sentiment is usually not directly observable. Existing literature has relied on proxies such as closed-end fund discounts [Lee, Shleifer, and Thaler (1991)], consumer

[3] Chen, Hong, and Stein (2001) and Charoenrook and Daouk (2008) find that detrended turnover forecasts conditional skewness of individual stocks.

confidence index [Lemmon and Portniaguina (2006) and Qiu and Welch (2006)]. Other papers call for the use of investor survey as a direct measure [see Brown and Cliff (2005), Qiu and Welch (2006)]. Unlike those papers, this paper constructs a novel sentiment indicator from forecasts of popular technical trading rules. The advantage of using trend-following forecasts lies in its clear interpretation as an example of demand shift without a fundamental rationalization. Furthermore, it can be simulated as long as the past asset price information is available, and thus has the potential to expand investor sentiment data to a longer history and more countries.

This paper can also be viewed as an empirical test for the economic impact of technical trading. Technical analysis attempts to use past prices and perhaps other related summary statistics to forecast price movements in order to make investment decisions. It has enjoyed wide popularity among traders for a long time. Existing literature on technical analysis focuses almost exclusively on the profitability of various trading rules and the implications for market efficiency. In contrast, I focus on how technical trading can be related to skewness of aggregate stock return, which has not yet been studied empirically in the literature.[4]

1.3 Data

I consider three samples in this paper. The baseline sample is the Dow Jones Industry Average (DJIA) between January 1952 and December 2008. The DJIA is available for a longer sample period. However, to construct a refined sentiment index (which is explained below), the daily federal fund rate is needed to obtain the excess return. As it has only been available since 1952, it puts constraint on the length of the baseline sample. To examine the robustness of the emprical findings, DJIA (1900-1951) is used as a "hold-out" sample, although the refined sentiment indicator is not considered. For an additional out-of-sample test, I conduct the same analysis on the S&P 500 index between 1964 and 2008. Both the DJIA (1952-2008) and the S&P

[4] A notable exception is Brunnermeier, Nagel, and Pedersen (2008), who argue that negative skewness in the currency market is due to a sudden unwinding of carry trades. Nagel (2004) also studies the impact of trading strategies, though his focus is on trading volume.

500 index are obtained from Datastream. The DJIA (1900-1951) is downloaded from http://www.analyzeindices.com/dow-jones-history.shtml. The daily federal funds rate is from the Federal Reserve Bank.

Dependent Variables

I use the "coefficient of skewness" as my baseline measure of skewness. It is calculated using the daily return from $t+1$ to $t+h$ as follows:

$$\mathit{SKEWNESS} = \frac{\frac{1}{h}\sum_{i=t+1}^{t+h}(r_i - \overline{r})^3}{\left(\frac{1}{h}\sum_{i=t+1}^{t+h}(r_i - \overline{r})^2\right)^{3/2}} \tag{1.1}$$

where h is the number of observations during this period, r_i is the daily return at time i, and $\overline{r}$ is the average daily return over this period. Note that a large negative value of SKEWNESS corresponds to a left skewed distribution, indicating that market return during this period is more "crash prone".

Following Chen, Hong, and Stein (2001), I use an alternative measure of asymmetry of stock returns, denoted as DUVOL for "down-to-up volatility". To calculate DUVOL from time $t+1$ to $t+h$, I first obtain the average return for this period. Then I separate the days during this period into a group whose daily return is above the average return (up days) and a group whose daily return is below the average returns (down days). DUVOL is then computed as the log of the ratio of the standard deviation on the down days to the one on the up days:

$$\mathit{DUVOL} = log\left\{\frac{(n_u - 1)\sum_{DOWN}(r_i - \overline{r})^2}{(n_d - 1)\sum_{UP}(r_i - \overline{r})^2}\right\} \tag{1.2}$$

where n_u and n_d are the number of days in up days group and down days group. The use of DUVOL is motivated by the concern that, for a relatively small sample, the calculation of SKEWNESS above is prone to estimation errors from calculating third moments. DUVOL, on the other hand, involves only the estimation of second moments, which therefore mitigates this concern.

Throughout this paper, I consider predicting skewness or DUVOL over 30 days horizon unless otherwise stated. Admittedly, the choice of this particular forecast horizon is somewhat arbitrary. It rather reflects an attempt to balance two considerations: on the one hand, shorter horizon is in principle deemed more interesting [Chen, Hong, and Stein (2001)]; on the other hand, estimation of skewness variables over such a horizon invites estimation errors. Table 1.1 shows that the average skewness (DUVOL) over 30 days of daily returns is slightly positive (negative). Their standard deviations are usually high relative the mean, indicating high variations over time. Also, SKEWNESS and DUVOL are notably highly correlated with a negative coefficient of -95%, as shown in Table 1.2. This result is similar to the findings in Chen, Hong, and Stein (2001). Table 1.2 also shows that Skewness is negatively correlated with the return over past 30 days. Taken at face value, it suggests that a smaller skewness follows a higher past return, which is consistent with the stochastic bubble theory of Blanchard and Watson (1982) and the price convexity theory of Xu (2007).

Table 1.1: Summary Statistics

This table presents the summary statistics. "Return" refers to the daily gross return of DJIA (in percentage). "SD" is the standard deviation of returns from time t to $t+30$. "SKEWNESS" and "DUVOL" are the measures of return asymmetry calculated using the daily return from time t to $t+30$ according to Equation 1.1 and 1.2. "Sentiment" and "Disagreement" are the (equally weighted) average and the standard deviation of the forecasts from the trading strategies for time t. "Sentiment_w" and "Disagreement_w" are the average and the standard deviation of the forecasts weighted according to the past two years of excess returns. Note that performance-weighted forecasts are re-scaled by 1000 times for the ease of reporting coefficients in the regression studies.

	Return_30	SD_30	Skewness	DUVOL	Sentiment	Disagreement	Sentiment_w	Disagreement_w
Mean	0.006	0.008	0.009	-0.019	0.191	0.850	0.216	1.441
Std. Dev.	0.052	0.005	0.557	0.331	0.445	0.123	0.816	0.998
Min	-0.387	0.002	-4.372	-1.661	-0.845	0.449	-4.330	0.297
Max	0.188	0.057	3.598	2.111	0.886	0.998	3.125	12.870
Skewness	-0.917	4.475	-0.786	0.117	-0.423	-0.782	-0.475	2.939
Kurtosis	6.721	37.387	6.713	3.510	1.975	2.701	3.611	19.810

Sentiment and Disagreement

I construct new measures of investor sentiment and differences of opinion from a spectrum of trading strategies. Admittedly, the choice of the universe of trading

Table 1.2: Correlation Coefficient Matrix

This table presents the correlation coefficient matrix. "Return_30" is the gross return over the period of $t-30$ to t. "SD_30" is the standard deviation of daily return from time $t-30$ to t. "Skewness" and "DUVOL" are the measures of return asymmetry calculated using the daily return from time t to $t+30$ according to Equation 1.1 and 1.2. "Sentiment" and "Disagreement" is the (equally weighted) average and the standard deviation of forecast from the trading strategies for time t. "Sentiment_W" and "Disagreement_w" is the average and the standard deviation of the forecasts weighted according to the past two years excess returns. Note that performance-weighted forecasts are re-scaled by 1000 times for the ease of reporting coefficients in the regression studies.

Variables	Return_30	SD_30	Skewness	DUVOL	Sentiment	Disagreement	Sentiment_w	Disagreement_w
Return_30	1.000							
SD_30	-0.354	1.000						
Skewness	-0.137	0.061	1.000					
DUVOL	0.153	-0.078	-0.951	1.000				
Sentiment	0.707	-0.359	-0.237	0.255	1.000			
Disagreement	-0.258	0.132	0.121	-0.127	-0.429	1.000		
Sentiment_w	0.516	-0.262	-0.206	0.215	0.727	-0.276	1.000	
Disagreement_w	-0.030	-0.038	0.075	-0.072	-0.069	0.277	-0.143	1.000

strategies is to some extent at the disposal of the researchers. To avoid the concern that the empirical results are driven by choosing a desired trading universe, I simply use the same universe of trading rules as in Qi and Wu (2006), which nests nearly all the trading rules studied in the top three finance journals. These trading rules have been in use for a long time, and are used in current financial web sites such as Yahoo Finance. They also enjoy wide popularity among the finance media such as the Wall Street Journal and the Financial Times.

More specifically, the trading rule universe includes Filter Rules, Moving Average, Trading Range Break (or Support and Resistance) Rules, and Channel Breakout Rules. As is common in the technical analysis literature, all these trading strategies generate one of the three trading recommendations, which assumes a value of 1 (buy signal), 0 (no position recommended), or -1 (sell signal). These strategies have been used for a long time, and have frequently been studied in the literature[5]. The total number of strategies I consider is 2127. The definitions of these trading strategies follow Sullivan, Timmermann, and White (1999) and Qi and Wu (2006). For detailed definitions, please refer to the appendix in Chapter 2. The parameters of these trading rules are provided in the appendix at the end of this chapter.

[5]More discussions of these trading strategies can be found in Sullivan, Timmermann, and White (1999)

I consider two ways of constructing the measures of investor sentiment and differences of opinion. The first one, denoted as "SENTIMENT" throughout the paper, is a simple average of forecasts from all the trading strategies at each time t. This amounts to assigning an equal weight to each trading strategy, without taking its past performance into account. Similarly, "DISAGREEMENT" is obtained as the standard deviation (a common measure of differences of opinion in the literature) of the forecasts from all the trading strategies. Alternatively, to capture the idea that better performing strategies are more likely to be used, I construct a refined indicator of sentiment and differences of opinion, which weights the trading signals according to the past performance of the corresponding strategy. More concretely, for each time t, the weight for each trading strategy with a positive mean excess return (over the daily federal fund rate) in the evaluation period equals the proportion of the mean excess return relative to the sum of mean excess returns from all these profitable strategies. The unprofitable strategies during the two-year evaluation period are weighted with zero. I use "SENTIMENT_W" hereafter to denote the sentiment weighted by two-year mean excess returns. "DISAGREEMENT_W" is analogously defined.

An inspection of Table 1.1 shows that the mean of SENTIMENT and the SENTIMENT_W are both positive[6], indicating that the forecasts from trend-following strategies are on average optimistic. The correlation between SENTIMENT and SENTIMENT_W is as high as 73%. Both SENTIMENT and SENTIMENT_W are highly correlated with the skewness variables, with an absolute value of correlation coefficient larger than 25%. DISAGREEMENT and DISAGREEMENT_W, on the other hand, have much lower correlations with the skewness variables. Note that trend chasing strategies are likely to have positive signals if past returns increase. Therefore one would expect SENTIMENT and SENTIMENT_W to be positively correlated with the past 30 days of return. I find a correlation coefficient of 71% and 52%, respectively. This strongly suggests a need to control for the past returns if one wants to isolate the effects of sentiment and disagreement beyond the past returns.

[6]Note that performance-weighted forecasts are rescaled by 1000 times to facilitate reporting coefficients in the regression studies.

Validation of Sentiment Indicator

To validate the new investor sentiment indicator, I provide its pairwise correlation with other commonly used proxies for investor sentiment. These sentiment proxies include lagged value-weighted dividend premium, IPO volume, lagged first-day return on IPOs, lagged NYSE turnover from NYSE Factbook, closed-end fund discount, and new issued debt and equity. Baker and Wurgler (2007) provide detailed descriptions and discussions of these variables.[7] These sentiment indicators are at monthly frequency and available since the 1950s or the 1960s till 12/2005. I also consider the top-down sentiment index of Baker and Wurgler (2007) based on first principal component of above six (standardized) sentiment proxies over 1962-2005 data. In addition, I include the monthly Consumer Confidence Index between Jan. 1978 and Dec. 2008, which is obtained from the Michigan Consumer Research Center, and a weekly Bull-Bear spread from Lowerrisk.com. Consumer confidence index has been considered as a sentiment indicator by Lemmon and Portniaguina (2006) and Qiu and Welch (2006), among others. The Bull-Bear spread is calculated as the difference between the proportion of investors holding bullish opinion and the proportion of investors holding bearish opinions. It is based on the weekly investor sentiment data from an online survey conducted by Lowerrisk.com from May, 1997 to July, 2006. To calculate the correlation with other monthly sentiment indicators, I consider the third version of the new sentiment index, "SENTIMENT_AVG", which is the monthly average of daily SENTIMENT. As other sentiment variables, except the Bull-Bear spread, are in monthly frequency, both SENTIMENT and SENTIMENT_W are taken at the end of the month.[8] In order to obtain the correlation with the Bull-Bear spread, both SENTIMENT and SENTIMENT_W are taken on the same day as the survey day.

Table 1.3 reports the correlation coefficient and the p-value from testing the null hypothesis that two sentiment indicators are independent. It shows that all three versions of new sentiment indicator significantly correlate with other investor sentiment

[7] I thank Jeffrey Wurgler for providing these data at his web page.

[8] Similar correlations are obtained when SENTIMENT and SENTIMENT_W are taken at the beginning of the month.

Table 1.3: Correlation with Other Sentiment Indicators

This table presents the pairwise correlation of the sentiment index from trend-following trading strategies with other commonly used sentiment index. “pdnd_lag” is the lagged value-weighted dividend premium, “nipo” is IPO volume, “ripo_lag” is the lagged first-day return on IPOs, “turn_lag” is the lagged NYSE turnover from NYSE Factbook, “cefd” is the closed-end fund discount, and “sd”(“se”) the new issued debt (equity). These sentiment indicators are at monthly frequency and available since the 1950s or the 1960s until 12/2005. “sent_bw” is the Sentiment index in Baker and Wurgler (2007) based on first principal component of six (standardized) sentiment proxies over the 1962-2005 data. “bbspread” is the Bull-Bear spread calculated using the weekly investor sentiment data from Lowerrisk.com (05/1997-07/2006). SENTIMENT is the (equally weighted) average of forecast from the trading strategies. “cci” is the Consumer Confidence Index obtained from the Michigan Consumer Research Center, which has been available at monthly frequency since 1978. SENTIMENT_W is the average of forecast weighted according to past two years excess return. Both SENTIMENT and SENTIMENT_W are taken at the end of month. SENTIMENT_AVG is the monthly average of the daily sentiment index SENTIMENT. Spearman correlation coefficient is reported with P-value in parenthesis.

	pdnd_lag	nipo	ripo_lag	turn_lag	cefd	se	sd	sent_bw	bb_spread	cci
SENTIMENT	-0.119	0.310	0.137	0.148	-0.082	0.175	0.133	0.113	0.433	0.119
	(0.006)	(0.000)	(0.002)	(0.000)	(0.070)	(0.000)	(0.001)	0.013	(0.000)	(0.011)
SENTIMENT_W	-0.130	0.367	0.120	0.185	-0.233	0.216	0.198	0.116	0.364	0.152
	(0.002)	(0.000)	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)	0.011	(0.000)	(0.001)
SENTIMENT_AVG	-0.126	0.344	0.197	0.175	-0.087	0.204	0.133	0.123	–	0.152
	(0.003)	(0.000)	(0.000)	(0.000)	(0.054)	(0.000)	(0.001)	0.007	–	(0.001)

indicators, with the corresponding p-value almost always smaller than 1%. The signs of the coefficients are expected since higher values of lagged value-weighted dividend premium and closed-end fund discount proxy for lower investor sentiment, whereas higher values of other indicators proxy for higher investor sentiment. There is also a significant and positive correlation between the new sentiment index and the top-down sentiment index of Baker and Wurgler (2007), which is perceived to be less prone to the noise in individual sentiment indicators. Remarkably, the correlation of the Bull-Bear spread with the new sentiment indicator is about 40%. Since the Bull-Bear spread is from an online survey, it provides a more convincing validation of the new sentiment through direct opinions of investors (similar arguments can be found in Qiu and Welch (2006)).

1.4 Empirical Results

In this section I examine the role of the investor sentiment and differences of opinion in forecasting skewness of market return. I start with a sorting-based approach, which examines the skewness in different quintiles of sentiment and disagreement. Then I proceed with a regression-based analysis to account for potential effects from control variables. The empirical results are better explained when results from both approaches are taken together.

1.4.1 Skewness in Different Quintiles of Sentiment and Disagreement

I use $Skew_t^{t+h}$ to denote SKEWNESS or DUVOL of DJIA calculated from daily returns from t to $t+h$ (see Equation 1.1 and 1.2). I sort $Skew_t^{t+h}$ by SENTIMENT at time t into quintiles of equal number of observations, with quintile 1 as the lowest quintile and quintile 5 as the highest quintile. For each quintile, the mean of $Skew_t^{t+h}$ within the quintile is reported. I also report a "t" statistic obtained from testing whether mean $Skew_t^{t+h}$ equals zero. A similar sorting procedure is applied to DISAGREEMENT, and the corresponding mean of $Skew_t^{t+h}$ and t-statistics are calculated.

Table 1.4: Breakdown by Sorts of Sentiment and Disagreement

This table reports the break-down results for dependent variables in each quintile of SENTIMENT or DISAGREEMENT. "SKEWNESS_30" ("SKEWNESS_60") is the skewness (Equation 1.1) of daily return obtained from time t to $t+30$ ($t+60$). "DUVOL_30" ("DUVOL_60") is the DUVOL (Equation 1.2) of daily return obtained from time t to $t+30$ ($t+60$). Panel A of this table reports the average SKEWNESS or DUVOL when SENTIMENT is sorted into quintiles of equal number of observations. "t-stat" is the t-value obtained from testing whether the mean of SKEWNESS/DUVOL equals zero. Panel B reports similar results when the sorting procedure is applied to the DISAGREEMENT.

Panel A: Sort by SENTIMENT

	SKEWNESS_30		DUVOL_30		SKEWNESS_60		DUVOL_60	
Quintile	mean	t-stat	mean	t-stat	mean	t-stat	mean	t-stat
1	0.20	21.82	-0.14	-23.29	0.20	21.36	-0.13	-24.79
2	0.08	7.99	-0.06	-10.10	0.04	4.45	-0.03	-7.18
3	-0.03	-2.93	0.00	0.26	-0.08	-6.62	0.02	3.63
4	-0.11	-10.25	0.06	9.77	-0.15	-13.86	0.07	14.71
5	-0.17	-15.53	0.09	14.75	-0.30	-21.00	0.13	23.85

Panel B: Sort by DISAGREEMENT

	SKEWNESS_30		DUVOL_30		SKEWNESS_60		DUVOL_60	
Quintile	mean	t-stat	mean	t-stat	mean	t-stat	mean	t-stat
1	-0.12	-10.99	0.06	9.44	-0.25	-17.28	0.10	17.41
2	-0.03	-2.74	0.00	0.44	-0.07	-6.12	0.02	3.30
3	0.01	0.91	-0.02	-3.32	-0.02	-2.00	-0.00	-0.79
4	0.04	4.22	-0.04	-6.29	0.02	1.61	-0.02	-5.03
5	0.06	5.90	-0.05	-8.30	0.03	2.83	-0.03	-6.62

Table 1.4 reports the results for $Skew_t^{t+h}$ of 30 days and 60 days of the baseline sample (DJIA, 1952-2008).[9] Panel A indicates that as the SENTIMENT becomes larger, the SKEWNESS declines uniformly, suggesting that the higher the sentiment, the more likely the crash will be. When the investor sentiment is in the highest quintile (quintile 5), the average 30 days skewness is -0.17, about twenty times higher in absolute value than the unconditional skewness (0.009 in Table 1.1). An increase in DISAGREEMENT in Panel B corresponds to a decline of SKEWNESS. Notably, when DISAGREEMENT is in the lowest quintile, the average SKEWNESS is much lower than the other four quintiles, showing higher crash risk. Similar results are obtained for DUVOL with an

[9] Similar results can be found for the SKEWNESS/DUVOL for other length of trading days, and for the other two samples: DJIA (1900-1951) and S&P 500 (1964-2008).

expected reverse pattern.

The monotonic relationship between skewness and disagreement obscures the possible role of the investor sentiment. Differences of opinion are very low both in the very high sentiment period and very low sentiment period. Intuitively, however, a situation with almost everyone believes the market will go up has considerable different implication than a situation that almost everyone believes the market will go down. Motivated by this intuition, I report the mean skewness and the t-value when the SENTIMENT is above zero (optimistic state) or below zero (pessimistic state).

Table 1.5: Breakdown by Sorts of Disagreement in Different States of Sentiment

This table reports the break-down results for dependent variables conditional on SENTIMENT is below zero (pessimistic state) or above zero (optimistic state). "SKEWNESS_30" ("SKEWNESS_60") is the skewness (Equation 1.1) of daily return obtained from time t to $t + 30$ ($t + 60$). "DUVOL_30" ("DUVOL_60") are the DUVOL (Equation 1.2) of daily return obtained from time t to $t + 30$ ($t + 60$). Panel A of this table reports the average SKEWNESS or DUVOL when DISAGREEMENT is sorted into quintiles of equal number of observations and when the average sentiment is pessimistic. "t-stat" is the t-value obtained from testing whether the mean of SKEWNESS/DUVOL equals zero. Panel B reports similar results when DISAGREEMENT_W is sorted into quintiles of equal number of observations and when the average sentiment is optimistic.

Panel B: Sort by DISAGREEMENT in Pessimistic State

	SKEWNESS_30		DUVOL_30		SKEWNESS_60		DUVOL_60	
Quintile	mean	t-stat	mean	t-stat	mean	t-stat	mean	t-stat
1	0.38	19.01	-0.28	-20.33	0.37	19.99	-0.24	-19.60
2	0.25	16.07	-0.18	-17.25	0.26	16.44	-0.17	-19.21
3	0.15	9.78	-0.10	-10.61	0.15	9.90	-0.10	-11.59
4	0.10	6.78	-0.06	-6.71	0.08	5.21	-0.05	-6.33
5	0.09	5.82	-0.06	-7.11	0.06	3.82	-0.05	-6.30

Panel A: Sort by DISAGREEMENT in Optimistic State

	SKEWNESS_30		DUVOL_30		SKEWNESS_60		DUVOL_60	
Quintile	mean	t-stat	mean	t-stat	mean	t-stat	mean	t-stat
1	-0.19	-16.61	0.11	16.58	-0.33	-21.58	0.15	25.75
2	-0.14	-11.30	0.07	10.64	-0.20	-14.55	0.09	15.72
3	-0.07	-4.89	0.03	3.21	-0.12	-8.47	0.05	7.91
4	0.01	0.37	-0.02	-2.59	-0.02	-1.78	-0.01	-1.10
5	0.04	2.80	-0.04	-4.81	0.01	0.36	-0.02	-3.19

Table 1.5 reveals an interesting pattern. When trend-following investors are on average

pessimistic (Panel A), an increase in DISAGREEMENT is associated with a monotonic decline in SKEWNESS, indicating that the lower the disagreement, the more likely a large rebound will occur. The positive SKEWNESS is particularly pronounced in the lowest quintile, with a value of 0.38, i.e., more than forty times higher than the unconditional skewness of 0.009 in Table 1.1. When trend-chasing investors are on average optimistic (Panel B), DISAGREEMENT is in general positively(negatively) associated with the SKEWNESS (DUVOL), which is just the opposite from the results in Panel A. Relative to the unconditional skewness, the large negative average skewness of -0.19 at the lowest quintile indicates that the crash risk is very high when the disagreement is extremely low. These results suggest that a convergence of opinion is likely to be associated with a crash in optimistic states. Nevertheless, these results are still descriptive and ignore other potential predictors of skewness. Therefore, I now turn to regression analysis for a more rigorous investigation.

1.4.2 Regression Analysis

Similar to Chen, Hong, and Stein (2001), and Charoenrook and Daouk (2008), I use standard predictive regressions to investigate the role of the sentiment and differences of opinion for forecasting skewness of subsequent market returns. It takes the following form:

$$\textit{Skew}_t^{t+h} = \beta_0 + \beta_1' X_t + \beta_2' Z_t + \epsilon_t^{t+h}, \tag{1.3}$$

where z_t is either a measure of investor sentiment or differences of opinion or both. X_t contains a vector of control variables similar to Chen, Hong, and Stein (2001), such as past returns and volatility of past returns.

Due to overlapping observations of the dependent variable, this approach is plagued by econometric problems of serial correlation in the residuals. I adjust the p-value using Newey-West standard errors [Newey and West (1987a)], which are robust against heteroscedasticity and serial correlation. I also consider a moving block bootstrap methodology, which is particularly suitable in a setting with highly dependent data. The consistency of the MBB standard error estimator has recently been proven by Goncalves and White (2005). MBB is a (non-parametric) bootstrap which draws blocks of re-sampled observations randomly with replacement from the time series of original observations, where the block length can be fixed or data-driven.[10] The results are similar when using different standard errors, so I only report Newey-West standard errors for the sake of brevity.

Another potential concern is whether the dependent variables, the sentiment, and disagreement variables contain a unit root. I provide the unit root test results for a forecasting horizon of 30 days[11] in table 1.6 for the baseline sample. Three unit root tests have been conducted: the Augmented Dickey-Fuller test, the Phillips-Perron test, and the DF-GLS test, which performs a modified Dickey-Fuller t test for a unit root in which the series has been transformed by a generalized least-squares regression. All tests reject the existence of a unit root at the 1% level. The results for the other two samples are qualitatively similar.

Table 1.6: Unit Root Test

This table presents the results from unit root tests. ADF is the augmented Dickey-Fuller unit-root test, PPerron is the Phillips-Perron unit-root test that a variable has a unit root, and DF-GLS performs a modified Dickey-Fuller t test for a unit root in which the series has been transformed by a generalized least-squares regression. $***, **, *$ indicate significance at the 1%, 5%, and 10% levels.

Tests	SKEWNESS	DUVOL	SENTIMENT	DISAGREEMENT	SENTIMENT_W	DISAGREEMENT_W
ADF	-13.81***	-13.07***	-5.08***	-10.17***	-17.03***	-14.87***
PPerron	-12.50***	-11.99***	-10.82***	-16.19***	-15.64***	-11.40***
DF-GLS	-12.14***	-10.57***	-4.93***	-2.58***	-17.04***	-15.06***

[10] As recommended by Goncalves and White (2005), I use a data-driven block length, following the procedure by Andrews (1991).

[11] Unit root tests become less significant as the forecasting horizon grows. Still, no unit root can be detected for the horizons we considered (up to 360 days).

Baseline Estimation Results

In the baseline estimation, I forecast *Skew* over thirty days' with daily regressions for the baseline sample (DJIA 1952-2008). In the regressions reported by Table 1.7, I consider three sentiment states. "Both" indicates that the whole sample is used for regression, while "Pessimistic"/"Optimistic" indicates that only the observations with "SENTIMENT" below/above zero are considered. The first and fourth columns of Panel A indicate that the investor sentiment has a strong forecasting ability for the skewness for both specifications, with a negative sign and significant at the 1% level. The higher the trend-following investor sentiment, the lower skewness will be, showing a higher crash risk. Disagreement, however, does not have a significant incremental predictive power for future skewness beyond sentiment when both sentiment states are considered together. This stands in contrast to the theoretical predictions of Hong and Stein (2003), but is consistent with the empirical findings of Chen, Hong, and Stein (2001), Hueng and McDonald (2005), and Charoenrook and Daouk (2008), who find no significant relationship between conditional skewness of aggregate market returns and detrended turnover, a proxy for the extent of the differences of opinion.

Splitting the sample into optimistic state and pessimistic state reveals a differential predictive role of DISAGREEMENT across different sentiment states. When investors are on average pessimistic (second column in Panel A), DISAGREEMENT negatively forecasts the subsequent skewness. This result, when taken at face value, seems to support the model of Hong and Stein (2003), which predicts that higher differences of opinion are related to a higher crash risk. Recall that, however, even in the highest quintile of differences of opinion during a pessimistic state, the average skewness is positive (Table 1.4). Therefore a higher disagreement does not seem to predict a crash, which is by definition a large negative skewness. Instead, it is more evident that convergence of opinion is associated with a higher conditional skewness (large rebound in returns). Hence, I interpret the negative coefficient of DISAGREEMENT as evidence that convergence of opinion in a pessimistic state forecasts a large rebound.

When trend-chasing investors are on average optimistic (third column of Panel A), the coefficient of DISAGREEMENT_W becomes positive. Since in the optimistic state, the SKEWNESS is mostly negative in different quintiles (Table 1.4), the positive coefficient is best interpreted as lower disagreement forecasting larger negative skewness. That is, when trend-chasing investors' opinions converge in an optimistic state, it forecasts a subsequent higher crash risk. Taking the results in the optimistic state and pessimistic state together, neither Hong and Stein (2003) nor Xu (2007) explains the findings in the pessimistic and optimistic periods at the same time.

What explains the differential role of differences of opinion then? One explanation is that differences of opinion reflect the sustainability of an ongoing bubble or market downturn, which can be used by rational arbitragers to coordinate their attack at the bubble/market downturn. Consider a market with rational arbitragers similar to Abreu and Brunnermeier (2003), each of whom is small and unable to move the market alone. Arbitragers have a synchronization problem in temporarily coordinating with other rational arbitragers. For example, during a bubble period in which the average belief is usually positive, the arbitrager has the option to either ride the bubble or attack the bubble. Attacking a bubble can result in a loss if other arbitragers continue to buy. Therefore, an arbitrager's decision hinges upon her belief about what other arbitragers will do. Note also that trend-chasing investors require a different extent of price change to change their beliefs. For example, a 2% increase from a recent low in the stock price may invite a long position from one trend follower, but may not be sufficient for another investor to become bullish if she relies on at least a 3% increase to ensure her confidence in an up-trend. The later investor can become a buyer if the price continues to increase. A high dispersion of belief during an optimistic state among trend chasers implies that many trend-following investors are of more pessimistic/neutral type, but can subsequently become optimistic and then drive the price up further. The change of belief is likely to happen because a recent positive return leads to a net buy by trend followers (in line with the average optimistic sentiment), which can in turn push the price higher, causing the yet non-optimistic types to become optimistic. Therefore, a bubble is likely to be sustained when disagreement is high and trend chasers are on average optimistic. In this case, the rational arbitragers are likely to choose to ride

Table 1.7: Forecasting the Aggregate Stock Market Crash (DJIA 1952-2008)

The sample period runs from January 1952 to December 2008 and is based on returns on DJIA. "Both" indicates that the whole sample is used for regression, while "Pessimistic"/"Optimistic" indicates that only the observations with "SENTIMENT" below/above zero (pessimistic/optimistic state) are considered. The dependent variable is the SKEWNESS (Equation 1.1) or DUVOL (Equation 1.2) of daily returns calculated from t to $t+30$. "SENTIMENT" and "DISAGREEMENT" are the sentiment and disagreement defined as the average and the standard deviation of equally weighted forecasts without taking past performance into account. "skew_past_30" is the lagged dependent variable. "Return_30" is the gross return over the period of $t-30$ to t, and "Return_60" is the gross return over the period of $t-60$ to $t-30$. Other lagged returns are similarly defined. "SD_past_30" is the lagged realized standard deviation, calculated from the daily returns between $t-30$ and t. * significant at 10%; ** significant at 5%; *** significant at 1%.

	Panel A: SKEWNESS				Panel B: DUVOL			
Sentiment State	Both	Pessimistic	Optimistic	Both	Both	Pessimistic	Optimistic	Both
SENTIMENT	-0.364***			-0.369***	0.229***			0.232***
	(0.071)			(0.072)	(0.041)			(0.043)
DISAGREEMENT		-0.956***	0.531***	-0.027		0.715***	-0.307***	0.022
		(0.272)	(0.183)	(0.137)		(0.172)	(0.106)	(0.081)
SKEW_past_30	0.124***	0.106**	0.125***	0.124***				
	(0.033)	(0.041)	(0.033)	(0.034)				
DUVOL_past_30					0.149***	0.125***	0.152***	0.149***
					(0.032)	(0.039)	(0.032)	(0.032)
Return_30	0.562	0.095	0.223	0.575	-0.340	-0.189	0.130	-0.351
	(0.510)	(0.667)	(0.649)	(0.513)	(0.313)	(0.403)	(0.381)	(0.316)
Return_60	0.707	0.908	-0.069	0.709	-0.443*	-0.454	0.025	-0.445*
	(0.477)	(0.650)	(0.466)	(0.475)	(0.259)	(0.350)	(0.265)	(0.259)
Return_90	-0.105	-0.357	-0.222	-0.103	-0.012	0.318	-0.015	-0.013
	(0.343)	(0.417)	(0.375)	(0.342)	(0.199)	(0.245)	(0.218)	(0.199)
SD_past_30	-0.594	-5.533	12.000*	-0.607	0.179	3.874	-10.612***	0.185
	(3.654)	(3.977)	(6.896)	(3.661)	(2.264)	(2.419)	(3.879)	(2.262)
Constant	0.075**	1.081***	-0.596***	0.099	-0.056**	-0.795***	0.358***	-0.076
	(0.035)	(0.250)	(0.166)	(0.126)	(0.022)	(0.157)	(0.096)	(0.074)
Adjusted R^2	0.07	0.07	0.03	0.07	0.09	0.08	0.05	0.09
N	13685	4598	9084	13685	13685	4598	9084	13685

the bubble. By contrast, there is less of a chance for the price run-up to be sustained when the dispersion of belief converges in an optimistic state. This is due to the fact that fewer trend-following investors who still hold non-optimistic beliefs can join the party, while previous optimistic investors have margin restrictions and used up their buying capacity during the bubble. Understanding this, and understanding that other rational arbitragers are likely to perceive the same way, the convergence of opinion helps arbitragers to attack the bubble jointly. As a result, the market crashes. Vice versa, the convergence of opinion in the low sentiment period indicates that the market downturn is unlikely to be sustained and that the rational arbitragers coordinate to buy, hence, a strong recovery is expected.

For the control variables, past returns in most cases do not have a significant forecasting power, which seems inconsistent with the findings in Chen, Hong, and Stein (2001) that negative skewness is more pronounced following positive returns. However, in untabulated regressions, once SENTIMENT and DISAGREEMENT are omitted from the regression, past returns become significant and negative predictors of future skewness. Results in Table 1.7 suggests that much of the effect from past returns has been captured by the SENTIMENT and/or DISAGREEMENT when they are included jointly in the regressions. Lagged SKEWNESS is positive and significant, indicating persistency in the skewness. Volatility of past returns is included as in Chen, Hong, and Stein (2001) to address the concern that SENTIMENT or DISAGREEMENT forecasts volatility, which in turn is reflected in skewness such that we are probably forecasting volatility instead of skewness. As argued by Chen, Hong, and Stein (2001), past realized volatility is probably the best univariate predictor for future volatility, hence controlling for it helps to alleviate this concern. Nevertheless, the coefficient of the volatility of past returns is seldom significant.

Panel B of Table 1.7 reports results with DUVOL as a dependent variable following Chen, Hong, and Stein (2001), which is calculated using daily returns over thirty trading days. They corroborate well the findings with SKEWNESS as the dependent variable. The signs of the coefficients on SENTIMENT and DISAGREEMENT are opposite to those in Panel A, which is expected since SKEWNESS and DUVOL are highly negatively correlated. The opposite sign of the coefficients on DISAGREEMENT in the second and third columns of Panel B confirms the differential role of differences of opinion across pessimistic and optimistic states.

So far I have used the whole baseline sample to investigate the role of investor sentiment and disagreement. To address the robustness of the results in different sub-samples, I have conducted a sub-sample analysis by splitting the sample into two sub-periods: before 1980 and after 1980, which divides the sample into roughly equal-sized sub-samples. I find that the results in each sub-sample are qualitatively similar to the whole sample analysis. Detailed results are not included for the sake of space, but are available from the author upon request.

Another concern that may emerge is the extreme price movement on the "Black Monday", October 19, 1987, which might dominate the findings. In an untabulated regression which excludes October 1987, I find that the results virtually do not change.

Out-of-sample Test

As argued in the introduction, this paper has introduced a new way of constructing investor sentiment and differences of opinion through commonly used technical trading rules. Since the forecasts of trading rules need only past prices as the input, it enables me to examine the roles of sentiment and disagreement in forecasting market skewness as long as market index information is available. This greatly expands the potential data for such an analysis since commonly used sentiment indicators go back only to the 1960s.

In the following, I examine the robustness of the results by conducting the same analysis on the sample of DJIA between 1900 and 1951. This sample period includes interesting episodes such as the Roaring Twenties and the Great Depression.

Table 1.8 shows that the results are remarkably similar to those in Table 1.7. Investor sentiment negatively (positively) predicts future SKEWNESS (DUVOL). The role of differences of opinion depends on the average investor sentiment. It predicts nega-

Table 1.8: Forecasting the Aggregate Stock Market Crash (DJIA 1900-1951)

The sample period runs from January 1900 to December 1951 and is based on return on DJIA. "Both" indicates that the whole sample is used for regression, while "Pessimistic"/"Optimistic" indicates that only the observations with "SENTIMENT" below/above zero (pessimistic/optimistic state) are considered. The dependent variable is the SKEWNESS (Equation 1.1) or DUVOL (Equation 1.2) of daily returns calculated from t to $t+30$. "SENTIMENT" and "DISAGREEMENT" are the sentiment and disagreement defined as the average and the standard deviation of equally weighted forecasts without taking past performance into account. "skew_past_30" is the lagged dependent variable. "Return_30" is the gross return over the period of $t-30$ to t, and "Return_60" is the gross return over the period of $t-60$ to $t-30$. Other lagged returns are similarly defined. "SD_past_30" is the lagged realized standard deviation, calculated from the daily returns between $t-30$ and t. * significant at 10%; ** significant at 5%; *** significant at 1%.

	SKEWNESS				DUVOL			
Sentiment State	Both	Pessimistic	Optimistic	Both	Both	Pessimistic	Optimistic	Both
SENTIMENT	-0.343***			-0.350***	0.227***			0.231***
	(0.074)			(0.075)	(0.045)			(0.046)
DISAGREEMENT		-1.495***	0.867***	-0.116		0.908***	-0.510***	0.061
		(0.287)	(0.259)	(0.202)		(0.172)	(0.156)	(0.120)
SKEW_past_30	0.101***	0.115***	0.080***	0.100***				
	(0.026)	(0.030)	(0.030)	(0.026)				
DUVOL_past_30					0.110***	0.136***	0.075**	0.109***
					(0.026)	(0.031)	(0.030)	(0.026)
Return_30	0.138	0.742*	-0.243	0.179	-0.054	-0.418	0.282	-0.076
	(0.338)	(0.411)	(0.487)	(0.353)	(0.207)	(0.254)	(0.281)	(0.215)
Return_60	0.443*	0.972***	-0.326	0.456*	-0.312*	-0.622***	0.188	-0.318**
	(0.261)	(0.299)	(0.349)	(0.265)	(0.160)	(0.177)	(0.213)	(0.162)
Return_90	-0.110	-0.068	-0.240	-0.095	0.018	0.008	0.108	0.010
	(0.223)	(0.240)	(0.316)	(0.228)	(0.131)	(0.143)	(0.188)	(0.134)
SD_past_30	10.220***	12.138***	14.186***	10.181***	-6.036***	-6.974***	-8.959***	-6.016***
	(2.471)	(3.009)	(3.749)	(2.456)	(1.525)	(1.797)	(2.356)	(1.518)
Constant	-0.249***	1.115***	-1.085***	-0.152	0.150***	-0.690***	0.654***	0.099
	(0.036)	(0.245)	(0.219)	(0.167)	(0.022)	(0.150)	(0.131)	(0.100)
Adjusted R^2	0.07	0.06	0.04	0.07	0.08	0.07	0.04	0.08
N	15413	6753	8652	15413	15413	6753	8652	15413

tively (positively) SKEWNESS (DUVOL) in a pessimistic state, but predicts positively (negatively) in an optimistic state.

Past returns occasionally have positive significant coefficients when predicting future market skewness. When both investor sentiment and differences of opinion are omitted from the regression, past returns have again negative coefficients as in Chen, Hong, and Stein (2001). Past volatility has a significant predictive power in all model specifications compared to the weak predictive ability in the sample of DJIA during 1952-2008. It indicates a higher crash risk when past return volatility is high. Similar findings can be found in Chen, Hong, and Stein (2001). Finally, the model fit reflected in Adjusted R^2 is similar to the one in Table 1.7, and comparable to the findings in Chen, Hong, and Stein (2001).

Additional out-of-sample tests have been conducted for the S&P 500 index during 1964 to 2008. DJIA is computed from the stock prices of 30 of the largest and most widely held public companies in the United States, which are presumably the most liquid stocks. It is interesting to examine whether the above results hold for S&P 500, which includes smaller firms and is hence less liquid than the DJIA.

Table 1.9 shows that the investor sentiment continues to be a strong predictor of future return asymmetry, whose coefficient is significant at the 1% level when forecasting either SKEWNESS or DUVOL. The predictive ability of DISAGREEMENT, on the other hand, becomes quite weak. Nevertheless, the direction of predictive ability of DISAGREEMENT remains the same as before, despite less statistical significance in its coefficients. That is, it negatively (positively) predicts SKEWNESS (DUVOL) in a pessimistic state, and positively (negatively) predicts SKEWNESS (DUVOL) in an optimistic state. Note that the forecasting horizon is 30 days. When the forecasting horizon is extended to 60 days, 90 days or 120 days, DISAGREEMENT regains its significance, . Given the fact theoretical models do not provide guidance for the exact choice of the forecasting horizon, and the observation that the signs of the coefficients are the same for 30 days and longer forecasting horizons, I conclude that the results based on S&P 500 are consistent with the findings based on DJIA. Nevertheless I report only results for a forecasting horizon of 30 days in the paper for the consistency of exposition. Results for other forecasting horizons are available from the author upon request.

Monthly regressions

The above regressions are conducted at a daily frequency. This has the advantage of improving the statistical significance due to the large number of observations. Furthermore, investors are unlikely to care about crash risk only once a month, rather, they get alert once they find it is likely to occur during their daily trading. Still, one may argue that daily changes have more noise, and major episodes develop over months or even years. I therefore run monthly regressions. Table 1.10 reports the results. The SENTIMENT or DISAGREEMENT is taken at the end of month, while the SKEWNESS to be forecasted is from daily returns of the next 30 days. Although the statistical significance becomes weaker, the monthly regression yields consistent results as in the daily regression, indicating that the noise in daily returns or other variables cannot be the reasons for driving the previous results.

Regressions with Refined Measures of Sentiment and Disagreement

Investors are likely to engage in model selection. They can choose the models with more success in the past. Furthermore, investors who used unsuccessful trading rules can be driven out of the market or intentionally stay out of the market. This implies that better performing strategies are more likely to be used among trend-chasing investors. To capture this idea, I construct a refined indicator of sentiment and differences of opinion, which weights the trading signals according to the past performance of corresponding strategies. I examine whether the predictive role of investor sentiment and differences of opinion continue to hold for the refined measures.

Table 1.11 and Table 1.12 report the regression results for DJIA (1952-2008) and S&P 500 (1964-2008) respectively. The results for the DJIA (1900-1951) are not reported

Table 1.9: Forecasting the Aggregate Stock Market Crash (S&P 500 Index 1964-2008)

The sample period runs from January 1964 to December 2008 and is based on returns on the S&P 500 Index. "Both" indicates that the whole sample is used for regression, while "Pessimistic"/"Optimistic" indicates that only the observations with "SENTIMENT" below/above zero (pessimistic/optimistic state) are considered. The dependent variable is the SKEWNESS (Equation 1.1) or DUVOL (Equation 1.2) of daily returns calculated from t to $t + 30$. "SENTIMENT" and "DISAGREEMENT" are the sentiment and disagreement defined as the average and the standard deviation of equally weighted forecasts without taking past performance into account. "skew_past_30" is the lagged dependent variable. "Return_30" is the gross return over the period of $t - 30$ to t, and "Return_60" is the gross return over the period of $t - 60$ to $t - 30$. Other lagged returns are similarly defined. "SD_past_30" is the lagged realized standard deviation, calculated from the daily returns between $t - 30$ and t. * significant at 10%; ** significant at 5%; *** significant at 1%.

	SKEWNESS				DUVOL			
Sentiment State	Both	Pessimistic	Optimistic	Both	Both	Pessimistic	Optimistic	Both
SENTIMENT	-0.408***			-0.446***	0.225***			0.247***
	(0.084)			(0.087)	(0.051)			(0.052)
DISAGREEMENT		-0.528	0.253	-0.245		0.378*	-0.092	0.147
		(0.370)	(0.268)	(0.182)		(0.217)	(0.158)	(0.112)
SKEW_past_30	0.117***	0.188***	0.087**	0.110***				
	(0.037)	(0.045)	(0.040)	(0.038)				
DUVOL_past_30					0.140***	0.211***	0.109***	0.133***
					(0.037)	(0.046)	(0.040)	(0.038)
Return_30	0.172	-0.577	-0.312	0.236	0.080	0.446	0.508	0.044
	(0.523)	(0.684)	(0.784)	(0.524)	(0.323)	(0.405)	(0.440)	(0.322)
Return_60	0.533	0.465	0.101	0.532	-0.305	-0.205	-0.039	-0.303
	(0.535)	(0.825)	(0.513)	(0.532)	(0.286)	(0.411)	(0.295)	(0.284)
Return_90	-0.393	-0.486	-0.575	-0.392	0.217	0.304	0.330	0.216
	(0.369)	(0.577)	(0.363)	(0.368)	(0.221)	(0.329)	(0.219)	(0.220)
SD_past_30	-5.253	-7.953*	1.382	-5.539	2.976	5.667**	-3.190	3.107
	(3.720)	(4.468)	(7.577)	(3.747)	(2.391)	(2.708)	(3.929)	(2.394)
Constant	0.126***	0.703**	-0.280	0.332**	-0.085***	-0.488***	0.115	-0.208**
	(0.041)	(0.316)	(0.215)	(0.156)	(0.026)	(0.184)	(0.130)	(0.096)
Adjusted R^2	0.08	0.08	0.01	0.08	0.08	0.08	0.02	0.08
N	11212	3424	7782	11212	11212	3424	7782	11212

Table 1.10: Forecasting the Aggregate Stock Market Crash at Monthly Frequency (DJIA 1952-2008)

The sample period runs from January 1952 to December 2008 and is based on returns on DJIA. “Both” indicates that the whole sample is used for regression, while “Pessimistic”/“Optimistic” indicates that only the observations with “SENTIMENT” below/above zero (pessimistic/optimistic state) are considered. The dependent variable is the SKEWNESS (Equation 1.1) or DUVOL (Equation 1.2) of daily returns calculated from t to $t+30$. “SENTIMENT” and “DISAGREEMENT” are the sentiment and disagreement defined as the average and the standard deviation of equally weighted forecasts without taking past performance into account. “skew_past_30” is the lagged dependent variable. “Return_30” is the gross return over the period of $t-30$ to t, and “Return_60” is the gross return over the period of $t-60$ to $t-30$. Other lagged returns are similarly defined. “SD_past_30” is the lagged realized standard deviation, calculated from the daily returns between $t-30$ and t. * significant at 10%; ** significant at 5%; *** significant at 1%.

	SKEWNESS				DUVOL			
Sentiment State	Both	Pessimistic	Optimistic	Both	Both	Pessimistic	Optimistic	Both
SENTIMENT	-0.349***			-0.340***	0.192***			0.192***
	(0.089)			(0.095)	(0.053)			(0.056)
DISAGREEMENT		-0.771	0.716**	0.060		0.593*	-0.371**	0.005
		(0.472)	(0.319)	(0.200)		(0.303)	(0.183)	(0.119)
SKEW_past_30	0.107**	0.073	0.116**	0.108***				
	(0.041)	(0.066)	(0.054)	(0.042)				
DUVOL_past_30					0.135***	0.106	0.141***	0.135***
					(0.041)	(0.070)	(0.052)	(0.042)
Return_30	0.414	-0.772	0.432	0.388	-0.093	0.614	0.017	-0.095
	(0.746)	(1.126)	(1.217)	(0.751)	(0.440)	(0.719)	(0.694)	(0.444)
Return_60	0.783	0.125	0.553	0.781	-0.364	0.251	-0.295	-0.364
	(0.542)	(0.809)	(0.783)	(0.542)	(0.321)	(0.517)	(0.450)	(0.322)
Return_90	-0.152	-0.500	-0.239	-0.153	0.137	0.419	0.203	0.137
	(0.490)	(0.673)	(0.701)	(0.491)	(0.291)	(0.431)	(0.403)	(0.291)
SD_past_30	1.491	-7.251	13.321	1.542	-2.014	4.955	-12.361*	-2.011
	(5.443)	(6.935)	(11.475)	(5.450)	(3.224)	(4.429)	(6.569)	(3.228)
Constant	0.041	0.870**	-0.786***	-0.013	-0.025	-0.644**	0.435***	-0.029
	(0.054)	(0.438)	(0.276)	(0.186)	(0.032)	(0.281)	(0.158)	(0.110)
Adjusted R^2	0.06	0.03	0.03	0.06	0.07	0.05	0.04	0.07
N	630	211	419	630	630	211	419	630

since the refined measure needs to calculate the excess return over daily federal fund rate, which has only been available since 1952. Still, when ignoring the federal fund rate in the calculation of excess returns, I find that the results for DJIA 1900-1951 are qualitatively similar to those for DJIA 1952-2008.

An inspection of Table 1.11 indicates that, for the sample of DJIA 1952-2008, the role of investor sentiment is unaffected by applying the refined investor sentiment indicator (SENTIMENT_W). Its coefficient is significant at 1% level for all the specifications for whenever it is included in the regression. Refined measure of differences of opinion (DISAGREEMENT_W) predicts in the same direction as in Table 1.7, although in the pessimistic state the coefficient is not significant. Note that in untabulated regressions at 90 days forecasting horizon, DISAGREEMENT_W regains its statistical significance at the 1% level in the pessimistic state. For the sample of the S&P 500 between 1964 and 2008, the refined sentiment indicator has again a strong predictive power. The refined measure of differences of opinion has significant predictive power for both pessimistic and optimistic states. Taking these results together, I conclude that the regression results with refined measures of sentiment and disagreement support the major findings of this paper.

1.5 Conclusion

This paper provides empirical evidence that both investor sentiment and differences of opinion have a robust forecasting power for aggregate market skewness. High sentiment forecasts market crash. The role of differences of opinion depends on the status of investor sentiment. When trend-chasing investors are on average optimistic, differences of opinion negatively forecast the market skewness; when they are on average pessimistic, differences of opinion positively forecast the market skewness.

I provide an explanation for the role of differences of opinion by augmenting the theory of Abreu and Brunnermeier (2003) with heterogeneous beliefs among trend-chasing investors. I argue that convergence of opinion in an optimistic state indicates that the

Table 1.11: Forecasting the Aggregate Stock Market Crash with Learning (DJIA 1952-2008)

The sample period runs from January 1952 to December 2008 and is based on the returns on DJIA. "Both" indicates that the whole sample is used for regression, while "Pessimistic"/"Optimistic" indicates that only the observations with "SENTIMENT" below/above zero (pessimistic/optimistic state) are considered. The observations with "SENTIMENT_W" above zero (optimistic state) and below zero (pessimistic state) are considered separately. The dependent variable is the SKEWNESS (Equation 1.1) of daily returns calculated from t to $t+30$. "SENTIMENT_W" and "DISAGREEMENT_W" are the performance-based sentiment and disagreement, defined as the average and the standard deviation of the forecasts, weighted according to the past two years excess returns. "skew_past_30" is the lagged dependent variable. "Return_30" is the gross return over the period of $t-30$ to t, and "Return_60" is the gross return over the period of $t-60$ to $t-30$. Other lagged returns are similarly defined. "SD_past_30" is the lagged realized standard deviation, calculated from the daily returns between $t-30$ and t. * significant at 10%; ** significant at 5%; *** significant at 1%.

	SKEWNESS				DUVOL			
Sentiment State	Both	Pessimistic	Optimistic	Both	Both	Pessimistic	Optimistic	Both
SENTIMENT_W	-0.103***			-0.098***	0.062***			0.059***
	(0.027)			(0.027)	(0.015)			(0.015)
DISAGREEMENT_W		-0.030	0.064***	0.023		0.017	-0.038***	-0.012
		(0.020)	(0.020)	(0.019)		(0.012)	(0.011)	(0.011)
SKEW_past_30	0.135***	0.122***	0.111***	0.134***				
	(0.034)	(0.043)	(0.032)	(0.033)				
DUVOL_past_30					0.160***	0.147***	0.136***	0.159***
					(0.032)	(0.041)	(0.031)	(0.032)
Return_30	-0.846**	-1.315***	-0.303	-0.855**	0.567**	0.922***	0.381	0.572**
	(0.352)	(0.401)	(0.466)	(0.352)	(0.223)	(0.257)	(0.264)	(0.224)
Return_60	-0.178	0.268	-0.232	-0.160	0.125	0.080	0.075	0.115
	(0.416)	(0.519)	(0.404)	(0.415)	(0.230)	(0.298)	(0.232)	(0.230)
Return_90	-0.630**	-0.406	-0.416	-0.612*	0.325*	0.402*	0.099	0.315*
	(0.316)	(0.365)	(0.345)	(0.317)	(0.186)	(0.230)	(0.199)	(0.187)
SD_past_30	-0.919	-8.482**	11.080*	-0.554	0.439	7.098***	-9.728***	0.237
	(3.953)	(3.482)	(6.014)	(3.921)	(2.525)	(2.163)	(3.249)	(2.518)
Constant	0.049	0.253***	-0.221***	0.012	-0.040	-0.173***	0.142***	-0.019
	(0.038)	(0.052)	(0.053)	(0.045)	(0.024)	(0.033)	(0.032)	(0.028)
Adjusted R^2	0.06	0.04	0.03	0.07	0.08	0.05	0.04	0.08
N	13685	4707	8978	13685	13685	4707	8978	13685

Table 1.12: Forecasting the Aggregate Stock Market Crash with Learning (S&P 500 Index 1964-2008)

The sample period runs from January 1964 to December 2008 and is based on the returns on S&P 500 Index. "Both" indicates that the whole sample is used for regression, while "Pessimistic"/"Optimistic" indicates that only the observations with "SENTIMENT" below/above zero (pessimistic/optimistic state) are considered. The observations with "SENTIMENT_W" above zero (optimistic state) and below zero (pessimistic state) are considered separately. The dependent variable is the SKEWNESS (Equation 1.1) of daily returns calculated from t to $t+30$. "SENTIMENT_W" and "DISAGREEMENT_W" are the performance-based sentiment and disagreement, defined as the average and the standard deviation of the forecasts, weighted according to the past two years excess returns. "skew_past_30" is the lagged dependent variable. "Return_30" is the gross return over the period of $t-30$ to t, and "Return_60" is the gross return over the period of $t-60$ to $t-30$. Other lagged returns are similarly defined. "SD_past_30" is the lagged realized standard deviation, calculated from the daily returns between $t-30$ and t. * significant at 10%; ** significant at 5%; *** significant at 1%.

	SKEWNESS				DUVOL			
Sentiment State	Both	Pessimistic	Optimistic	Both	Both	Pessimistic	Optimistic	Both
SENTIMENT_W	-0.072***			-0.072***	0.034**			0.035**
	(0.027)			(0.027)	(0.016)			(0.016)
DISAGREEMENT_W		-0.048*	0.106***	0.057***		0.032*	-0.060***	-0.031**
		(0.027)	(0.021)	(0.021)		(0.017)	(0.013)	(0.013)
SKEW_past_30	0.133***	0.137***	0.081**	0.127***				
	(0.037)	(0.047)	(0.036)	(0.036)				
DUVOL_past_30					0.152***	0.162***	0.102***	0.147***
					(0.037)	(0.047)	(0.036)	(0.036)
Return_30	-1.393***	-1.442***	-0.378	-1.227***	0.981***	1.016***	0.413	0.891***
	(0.378)	(0.455)	(0.494)	(0.382)	(0.242)	(0.302)	(0.285)	(0.242)
Return_60	-0.307	-0.085	0.213	-0.164	0.184	0.129	-0.147	0.106
	(0.498)	(0.653)	(0.469)	(0.497)	(0.267)	(0.363)	(0.262)	(0.267)
Return_90	-0.916***	-0.734	-0.481	-0.830**	0.523**	0.464	0.242	0.476**
	(0.353)	(0.474)	(0.351)	(0.355)	(0.210)	(0.292)	(0.212)	(0.211)
SD_past_30	-5.234	-12.710***	4.590	-2.950	3.018	8.747***	-4.830	1.743
	(4.254)	(3.755)	(6.406)	(4.312)	(2.765)	(2.403)	(3.431)	(2.827)
Constant	0.085*	0.336***	-0.244***	-0.020	-0.061**	-0.223***	0.139***	-0.005
	(0.044)	(0.067)	(0.062)	(0.057)	(0.028)	(0.042)	(0.037)	(0.036)
Adjusted R^2	0.06	0.06	0.04	0.07	0.07	0.07	0.05	0.07
N	11212	3415	7797	11212	11212	3415	7797	11212

price run-up is unlikely to be sustained since fewer investors can remain net buyers in the future. Therefore rational arbitrageurs coordinate their attack on the bubble, leading to a market crash. Vice versa, the convergence of opinion in a pessimistic state promotes coordinated purchases among rational arbitrageurs, leading to a strong recovery. Admittedly, the explanation is tentative and informal. Therefore, it calls for a rigorous model to incorporate the states of investor sentiment into the differences of opinion framework.

The novel way of constructing trend-chasing investor sentiment and differences of opinion can be applied to various asset markets, as long as trend-chasing behavior is prevalent. It has the potential to greatly expand the availability of sentiment indicators and measures of differences of opinion to a much longer history and to countries where data on other sentiment indicators are limited. Thus, an immediate extension of this paper would be to examine whether our results hold for other asset markets, such as stock markets of other countries or foreign exchange markets. Another interesting extension would be to study how the trend-chasing investor sentiment and disagreement help explain the cross-sectional variation in individual stocks. I leave these interesting extensions to future research.

Appendix

Filter Rules (FR)

x: increase in the log return required to generate a "buy" signal

y: decrease in the log return required to generate a "sell" signal

e: the number of the most recent days needed to define a low (high) based on which the filters are applied to generate a "long" ("short") signal

c: number of days a position is held during which all other signals are ignored

$x = 0.0005, 0.001, 0.005, 0.01, 0.05, 0.10$ (6 values)

$y = 0.0005, 0.001, 0.005, 0.01, 0.05$ (5 values)

$e = 1, 2, 5, 10, 20$ (5 values)

$c = 1, 5, 10, 25$ (4 values)

Note that y must be less than x, there are 15 (x,y) combinations

Number of rules in FR class $= x \times c + x \times e + x \times y +$ ((x,y) combinations)) $= 24 + 30 + 15 = 69$

Moving Average Rules (MA)

n: number of days in a moving average

m: number of fast-slow combinations of n

b: fixed band multiplicative value

d: number of days for the time delay filter

c: number of days a position is held, ignoring all other signals during that time

$n = 2, 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ (11 values)

$m = \sum_{i=1}^{n-1} i = 55$

$b = 0, 0.0005, 0.001, 0.005, 0.01, 0.05$ (6 values)

$d = 2, 3, 4, 5$ (4 values)

$c = 5, 10, 25$ (3 values)

Number of rules in MA class: $= b \times (n+m) + d \times (n+m) + c \times (n+m) = 396 + 264 + 198 = 858$

Support and Resistance (SR, or Trading Range Break) Rules

n: number of days in the support and resistance range;

e: used for an alternative definition of extrema where a low (high) can be defined as the most recent closing price that is less (greater) than the n previous closing prices;

b: fixed band multiplicative value;

d: number of days for the time delay filter;

c: number of days a position is held, ignoring all other signals during that time

$n = 5, 10, 15, 20, 25, 50, 100$ (7 values);

$e = 2, 3, 4, 5, 10, 25, 50$ (7 values);

$b = 0.0005, 0.001, 0.005, 0.01, 0.05$ (5 values);

$d = 2, 3, 4, 5$ (4 values);

$c = 1, 5, 10, 25$ (4 values);

Number of rules in SR class $= c \times (n+e) + b \times (n+e) \times c + d \times c \times (n+e) = 100 + 800 + 320 = 1220$

Channel Breakout Rules (CBO)

n: number of days for a channel

x: difference between the high price and the low price ($x\times$low price) required to form a channel

b: fixed band multiplicative value ($b < x$)

c: number of days a position is held, ignoring all other signals during that time

$n = 5, 10, 15, 20, 25, 50, 100, 200$ (8 values)

$x = 0.001, 0.005, 0.01, 0.05, 0.10$ (5 values)

$b = 0.0005, 0.001, 0.005, 0.01, 0.05$ (5 values)

$c = 1, 5, 10, 25$ (4 values)

Note that b must be less than x. There are 15 (x,b) combinations.

Number of rules in CBO class $= n \times x \times c + n \times c \times$((x,b) combinations)$= 160 + 480 = 640$

Total number of trading rules $= 2127$

CHAPTER 2

HOW ILLUSORY IS THE PROFITABILITY OF TECHNICAL ANALYSIS ?*

ABSTRACT

This paper quantifies the extent of data snooping bias in the commonly used technical trading strategies. Studying 25988 trading strategies for emerging foreign exchange markets, we find that the best rules can sometimes generate an annually mean excess return of more than 30%. Applying standard tests, we find hundreds to thousands of seemingly significant profitable strategies. Almost all these profits vanish once the data snooping bias is taken into account. We also show that out-of-sample tests, which are often proposed as an effective way of reducing the bias, can have a severe data mining bias. Overall, we show that the profitability of technical analysis is illusory.

*This chapter is based on a joint paper with Pei Kuang (Goethe University Frankfurt) and Michael Schröder (Centre for European Economic Research).

2.1 Introduction

After decades of debate, the profitability of technical trading rules (hereafter "TTR") remains a conundrum. On the one hand, technical analysis only needs public information such as asset prices. If the market is (weakly) efficient, as widely believed among financial economists, rational investors should quickly arbitrage away the profits, implying that technical analysis is useless. On the other hand, if TTR cannot generate persistent profits, why do at least 90% percent of experienced traders place some weight on it in costly trading activity [Taylor and Allen (1992)]?

In this paper, we show that the profitability of technical analysis is illusory. Studying 25988 trading strategies for emerging foreign exchange markets, we find that best rules can sometimes generate an annually mean excess return of more than 30%. Applying standard tests, we find hundreds to thousands of seemingly significant profitable strategies. Almost all these profits become insignificant once the data snooping bias is taken into account. We also show that out-of-sample tests, which are often proposed as an effective way of reducing the bias, can have a severe data mining bias. We do not find that certain rules dominate all others, either. It implies that, even if one could find ex post profitable rules, ex ante the exploitation of these rules is extremely difficult if not impossible, since we do not know which one to implement.

Economists have long acknowledged the data snooping bias in this context. Indeed, mechanical trading rules have ex ante parameters, which are unknown at the beginning of the trading period. So investors usually need to search intensively among a potentially large universe of trading rules on a single historical data to find the desired trading rules. Trying hard and long enough makes it very likely to find seemingly profitable but in fact wholly spurious trading strategies. The classical statistical inference typically conducted in the previous literature is biased in this case.

Still, how much can the data snooping bias explain the identified profitability? This question carries obvious importance for both investors and researchers. Due to its statistical difficulty, however, few existing research is able to answer this question. Our paper quantify the extent of data snooping bias by applying two recently developed and more powerful methods: the StepM test [Romano and Wolf (2005)] and the SSPA test [Hsu and Hsu (2006)]. For a given universe of trading strategies, these tests are data-snooping free since they take the entire search process into account, and hence are able to detect the genuine profitable trading rules from the universe. We compare the number of profitable rules from these two tests to that from the classical tests, which do not control for data snooping bias. The comparison can show how many profits are genuine and how many are spurious. This provides richer information about the extent of data snooping bias than only knowing whether the best performing rule has a true predictive ability or not (as in Reality Check [White (2000)] and the SPA test [Hansen (2005)]). Therefore we are able to make the extent of data snooping bias more transparent.

From methodological perspective, we empirically show that data snooping bias can be substantial in out-of-sample tests too. Out-of-sample tests are usually seen as an effective way of reducing the bias. If the in-sample result is spurious, it is unlikely to survive in a new and independent sample. We show, however, that out-of-sample tests can also result in substantial biases in tests such as those for the trading rule performance. In this context, if in-sample data snooping is not controlled, many seemingly satisfactory results emerge. When a single new data is available for out-of-sample test, it is repeatedly used by one or more researchers and practitioners for checking whether some or all of those seemingly satisfactory results still hold. Therefore it is not free of data snooping bias.

Our study provides a comprehensive test of trading rule profitability across 25,988 trading rules. It consists of the simple rules (e.g., filter rules), the charting rules with kernel smoothing (e.g., Head and Shoulders, as studied by Lo, Mamaysky, and Wang (2000)), and complex trading rules. The majority of these rules are popular among professional traders, but have not been studied in the literature for emerging FX markets. Hence, our paper depicts a more complete picture regarding the performance of trading rules and market efficiency for these markets.

Although we consider a large universe of trading rules, we do not assume that any particular individual has searched across them. Rather, researchers and investors as a whole are likely to have done this. In fact, before investor communities start to speculate on an emerging FX market, they have no prior information on which trading rule will work, hence it appears natural for them to search among the trading rules known from other markets such as developed countries' FX markets. Given today's high computer capacity and the fact that those trading rules have been implemented for other markets, we feel it is realistic to assume investors have collectively and sequentially searched over a large universe of trading rules. Testing such a large universe has both pros and cons. On the one hand it depicts a much more complete picture of the trading rule profitability and market efficiency in emerging FX markets than the existing literature. On the other hand, it raises a concern about the power of tests. However, we show that our major conclusions continue to hold when testing with smaller universes.

A study of TTR profitability for emerging FX market is interesting in its own right. Factors such as spot exchange rate movements, interest rate differentials and transaction costs in emerging markets can contribute to TTR profitability differently from their developed FX market counterparts. For example, spot exchange rates in emerging market are usually more volatile, and are more likely to undergo a currency crisis. If the TTR captures the risk premium of a spot rate movement, it is likely to be more profitable than a developed market, ceteris paribus. On average, the emerging countries also have higher nominal interest rates, which directly affects the profitability of TTR. Furthermore, the bid-ask spreads are usually two to four times larger in emerging markets than in developed countries [Burnside, Eichenbaum, and Rebelo (2007)], indicating higher transaction costs. In addition, emerging markets have stricter regulation and capital control, which makes it more difficult for speculation to arbitrage away the profits.

The remainder of this paper is structured as follows. Section 2.2 provides a short literature review. Section 2.3 presents the universe of the trading rules. Section 2.4 discusses the methodology for accounting for the effect of data snooping bias and for identifying the profitable strategies. Section 2.5 briefly documents our data and then reports our main empirical findings. Section 2.6 concludes. A detailed documentation of the trading rules considered can be found in the appendix.

2.2 Literature Review

The FX market has substantial supportive evidence for the profitability of technical analysis. Examples are Sweeney (1986), Levich and Thomas (1993), Neely (1997), Chang and Osler (1999), LeBaron (1999), Osler (2000), and Qi and Wu (2006). There are also several studies showing the opposite [see Lee and Mathur (1996), and Neely and Weller (2003)]. Most of these studies, however, confined themselves to the currencies of developed economies. It is unclear whether their findings can be carried over to the emerging markets, which themselves are heterogeneous.

Among the few studies on the profitability of TTR for emerging FX markets, Martin (2001) investigates moving average rules using daily data between 1993 and 1995 for 12 emerging markets, and finds significant improvement in the mean excess return, but not in the Sharpe ratio; Lee, Gleason, and Mathur (2001) examine the profitability of moving average and channel rules in 13 Latin American countries over the period 1992-99 and find mixed results; Pojarliev (2005) documents profitability of moving average rule using monthly data for the period of 1999-2004. However, it would still be premature to conclude whether TTR can generate sizable profits based on these mixed results. Furthermore, these studies do not formally control for the effect of data snooping bias, which is a critical concern in this line of research. A notable exception is Qi and Wu (2006). They are the first to study the TTR profitability with formal data snooping check for developed countries' currencies. They consider a universe of 2,127 simple technical trading rules and find that data snooping biases do not change the conclusion of profitability of trading rules in full sample. But in the second half of the sample the data snooping bias is more serious.

Data snooping has long been considered by academic researchers [Jensen (1967), Jensen and Bennington (1970), Lo and MacKinlay (1990) and Brock, Lakonishok, and LeBaron (1992)]. However, a rigorously founded and generally applicable test remained unavailable until White (2000) introduced the Reality Check. The Reality Check can directly quantify the effects of data snooping when testing the best trading rule from the "full universe" of trading strategies. Since then, a couple of studies [e.g., Sullivan, Timmermann, and White (1999), White (2000), Hsu and Kuan (2005) and Qi and Wu (2006)] have applied the Reality Check and found mixed results.

Hansen (2005) points out that the power of the Reality Check can be reduced and even be driven to zero when too many poor and irrelevant rules are included in the set of alternatives. Simply excluding poor performing alternatives does not lead to valid inference in general either. To solve this problem, Hansen (2005) proposes a new test statistic for superior predictive ability (hereafter "SPA"), which invokes a sample-dependent distribution under the null hypothesis. This SPA test is more powerful compared to White's Reality Check and less sensitive to the inclusion of poor and irrelevant alternatives.

The question of interest for both the Reality Check and the SPA test is whether the best trading rule beats the benchmark. An investor might want to know whether a particular trading rule is profitable. A researcher may want to test whether a certain trading rule found profitable in the literature indeed outperforms the market. Furthermore, as pointed out by Timmermann (2006), choosing the forecast with the best track record is often a bad idea, and a combination of forecasts dominates the best individual forecast in out-of-sample forecasting experiments. So we may want to know all or some of the profitable trading strategies and combine them for decision making (this is what our complex trading rules do).[1] We may also want to avoid worse strategies when combining trading rules, since trimming them off often helps to improve forecasting performance as found in the forecasting literature [Timmermann (2006)]. Romano and Wolf (2005) modify the Reality Check and propose a stepwise multiple test (hereafter "StepM"), which can detect as many profitable trading rules as possible for a given

[1]Some evidence of unstable performance of the trading strategy is provided in Sullivan, Timmermann, and White (1999). They find that the best trading rule applying to DJIA for the period of 1897-1986 does not outperform the benchmark for the period of 1987-1996.

significance level. Hsu and Hsu (2006) propose a stepwise SPA (hereafter "SSPA") test to combine the SPA test and the StepM test, which is more powerful in identifying all good trading strategies than those two. These two stepwise tests enable us to separate the genuine profitable rules from spurious ones among all seemingly profitable rules obtained from classical tests, and hence provide a more complete picture of the extent of data snooping bias.

When the new data is available, some researchers [for example, Neely, Weller, and Ulrich (forthcoming)] conduct out-of-sample tests to check whether previously identified profitable trading rules still are still profitable in the new data, as a way to control for data snooping bias. This is generally regarded as an effective way. In the large scale tests, however, the out-of-sample data snooping bias can also occur.[2] In this paper we also investigate this potential bias.

2.3 Universe of Trading Rules

Defining the universe of trading rules is a key step for obtaining valid inference in a superior predictive ability test. On the one hand, having a too small universe may miss the data snooping bias incurred during the search for profitable trading rules. On the other hand, the power of test may be reduced when too many irrelevant trading rules are included. Furthermore, defining a complete and exhaustive universe of trading rules is neither possible nor viable. In balancing these considerations, we expand the trading rules to a large universe while keeping it computationally feasible. In total we have 25,988 trading rules, consisting of simple rules, charting rules with kernel smoothing and complex trading rules. Table 2.1 provides an overview of the universe of trading rules we considered. Since the class of simple trading rules has well been described in the literature, we leave its details to the appendix. In the following, we focus on describing the other two group of trading. The parameters used in these trading strategies are reported in the appendix.

[2] Inoue and Kilian (2004) point out that out-of-sample tests can have as serious data mining bias as in-sample test.

Table 2.1: The Universe of Trading Rules

Trading Rules	Number of rules
Simple trading rules	
Filter Rules (FR)	497
Moving Averages (MA)	2049
Support and Resistance (SR)	1220
Channel Break-Outs (CB)	2040
Momentum Strategies in Price (MSP)	1040
Chart rules with kernel smoothing	
Head and Shoulders (HS, IHS)	3384
Triangle (TA)	1128
Rectangle (RA)	3384
Double Tops and Bottoms (DTB)	3384
Broadening Tops and Bottoms (BTB)	1128
Complex trading rules	
Learning Strategies (LS)	1872
Voting Strategies (VS)	26
Fraction Position Strategies (FPS)	26
Voting by Learning Strategies (VLS)	4810
Total number of trading rules	25988

Charting Rules with Kernel Smoothing

Similarly to Lo, Mamaysky, and Wang (2000), we consider 5 pairs of technical patterns, applying each to kernel smoothed spot series. These strategies include Head and Shoulders (HS) and Inverse Head and Shoulders (IHS), Triangle (TA), Rectangle (RA), Double Tops and Bottoms (DTB) and Broadening Tops and Bottoms (BTB).

Using non-parametric smoothing does not mean that the rules we are considering are too complicated for investors to use. On the contrary, we do this because investors filter out noise when using these nonlinear rules. In this sense, they are carrying out smoothing in their trading practice. Kernel smoothing is an ideal way to smooth the spot rates. The critical issue in kernel smoothing is the choice of bandwidth. The aim here is to find which degree of smoothing can mimic the eyeball smoothing adopted by investors. Lo, Mamaysky, and Wang (2000) recommend a bandwidth of $0.3\times$*(optimal band width)*. The optimal band width is calculated from a cross validation method [Lo, Mamaysky, and Wang (2000)].[3] Lo, Mamaysky, and Wang (2000) advocate the

[3]For a detailed discussion on kernel smoothing and cross validation, see Härdle (1990).

use of this under-smoothed stock price series since they are neither too volatile nor too smooth, and the professional technical analysts they interviewed feel that such a smoothing is more acceptable than other versions of smoothed series. They admit, however, that such an approach is ad hoc. Commenting on that paper, Jegadeesh (2000) recommends the use of different choices of bandwidth to assess whether their results are sensitive to these choices. Gene Savin and Zvingelis (2007) consider the multiples 1, 1.5, 2 and 2.5 of the optimal bandwidth. We use three different bandwidths (scaling the optimal bandwidth with 0.3, 1 and 4), as well as the original series, in order to see the effect from under-smoothing, optimal smoothing, over-smoothing and no smoothing. Therefore, each pair of charting rules in this class is applied to four versions of spot rates, depending on the smoothing parameters used.

Head and Shoulders and Inverted Head and Shoulders

Head and Shoulders is one of the most popular and trusted chart patterns by practitioners in technical analysis. It occurs when the second (head) of three consecutive peaks exceeds the first (left shoulder) and the third (right shoulder). The minima between left (right) shoulder and the head are called left (right) troughs. We require the two shoulders (troughs) to be approximately equal such that their differences are no more than a differential rate x. The HS pattern is completed when the adjacent local minimum of the right shoulder penetrates the neckline and the band. Short position in foreign currency is taken exactly on the day when the price crosses the neckline and the band. Following Chang and Osler (1999), two exit rules for HS strategy are considered, namely endogenous and exogenous. For the endogenous one, we distinguish two kinds of situations. We define a cutoff as y times the standard deviation of the daily exchange rate change. If the price falls by d percent times difference between head and average trough(the difference is referred to as "measuring objective" or "price objective" in technical manual), we exit on the day when the price has risen above a local minimum by the cutoff percentage, which implies that the price has conclusively stopped moving in the predicted direction. In the second case, if the price does not fall by such an amount, we allow for a possible bounce or interruption, that is, the price may temporarily move back toward the neckline. When the second trough falls below the aforementioned d percent line, we are back to the former case. Otherwise, we liquidate the position at the second trough. In both situations, in order to limit the loss, a stop-loss line is incorporated whenever the price goes sufficiently far in the wrong direction. An exogenous exit rule means we close our position after holding for an exogenous specified number of days f. As the name tellingly reveals, the inverted Head and Shoulders is simply an inverse version of Head and Shoulders; once the IHS is completed, the speculator expects an upward trend in the future spot exchange rate. As a result he or she will borrow the British Pound to buy the foreign currency.

Triangle

Triangle is one of the reverse patterns which is also based on pricing movements showing five consecutive local extrema. Triangle tops (TTOP) are characterized by three descending local maxima and two ascending local minima. Triangle bottoms (TBOP) are characterized by three ascending local minima and two descending local maxima. Once a triangle is completed, it will constitute a signal for taking a long (short) position in foreign currency if the future closing spot exchange rate exceeds the latest top (or falls below the latest bottom) by a fixed proportion x, known as the "trend filter". We consider similar liquidation methods as for HS.

Rectangle

The rectangle pattern is also characterized by five consecutive local extrema. Rectangle tops (RTOP) require three tops and two bottoms to lie near an upper horizontal lines, that is, within x percent of their average, respectively. Moreover, we require the lowest top to be higher than the highest bottom. Similarly, the rectangle bottoms (RBOT) require two tops and three bottoms lie near the upper horizontal lines. Signals are generated in a similar way as the Triangle rule, so are the liquidation methods.

Double Tops and Bottoms

The double tops (bottoms) are characterized by two tops (bottoms) that lie near an upper horizontal lines, with one bottom (top) lies in between. Following Lo, Mamaysky, and Wang (2000), we require the two tops (bottoms) to occur at least a month, or with 22 intermediatory trading days. In addition, the second top (bottom) should be higher (lower) than all the local maxima (minima) in between.

Broadening Tops and Bottoms

Similarly to the Triangle class, the Broadening Tops and Bottoms are characterized by five consecutive local extrema. Broadening Tops (BTOP) requires 3 descending local maxima and two ascending local minima. Broadening Bottoms requires 3 descending local minima and two ascending local maxima. Compared to the Triangle class, this class presupposes a "divergence" shape, while the Triangle class requires a "convergence" shape.

Complex Trading Rules

Single rules can generate false signals when prices fluctuate in a broad sideways pattern. Relying on a single rule can be a dangerous practice, even for the historical best rule. Technical analysts can combine other trading rules to confirm the prediction of price direction. Following Hsu and Kuan (2005), we consider three classes of complex trading rules: the learning strategy (LS), the voting strategy (VS), and the fractional position strategy (FPS). Besides these complex trading rules, we also propose a new class of complex rules, namely the voting by learning strategy (VLS), which combines the voting and learning rules.

Learning Strategy (LS)

A learning strategy assumes that after an investor learns about the strategies' performances from the past m days (memory span) within the certain class, he can switch his position by following the best strategy found during the memory span. After his switch, he waits for r days (review span) and then reevaluates the strategies' performances of the past m days (memory span) to decide whether to switch. For the evaluation of the trading rules, we use the mean return and the Sharpe ratio during the memory span as the performance measure.

Voting Strategy (VS)

Voting strategy is a system of voting by trading rules within each class. Each rule is assigned one vote for recommending a ballot. We consider the three choice ballot, where either a long, a short or no position can be voted. The decision follows the recommendation of majority votes. To avoid the voting result being dominated by a class with a large number of rules, we consider every class of non-complex rules separately but not the one consisting of all rules.

Fractional Position Strategy (FPS)

Note that both the learning strategy and the voting strategy yield the signal as an integer. The fractional position strategy, in contrast, allows to take the position of a non-integer between -1 and 1. The fraction of a position is determined by an "evaluation index". In our case, it is only the fraction of winning votes that recommend the same positions relative to all votes within the same class.

Voting by Learning Strategy(VLS)

To implement this strategy, we consider the best n trading rules evaluated from the memory span within each class of non-complex rules, and then assign votes to these rules as in the voting strategy class. The decision follows the recommendation of majority votes from the best n trading rules. In addition, we consider the case in which the top n trading rules are selected from the set of all non-complex trading rules.

2.4 Methodology

In this section, we discuss the empirical tests used in this paper. The discussion is kept informal and heuristic, with more technical description provided in Appendix C.

2.4.1 Performance Measures

Suppose that our universe of trading rules includes m rules. Let δ_{t-1}^k be a "signal" function which generates a trading signal by the k^{th} trading rule using information up to $t-1$. This signal function can assume three values that instructs a trader to take a short position ($\delta_{t-1}^k = -1$), a long position ($\delta_{t-1}^k = 1$), or no position ($\delta_{t-1}^k = 0$) in a foreign currency at time $t-1$. The k^{th} trading rule yields the profit as:

$$R_t^k = (s_t - s_{t-1} + r_t^* - r_t)\delta_{t-1}^k - abs(\delta_{t-1}^k - \delta_{t-2}^k)g \tag{2.1}$$

where R_t^k is the excess return from trading on the currency in period t using k^{th} trading rule, s_t is the logarithm of the spot exchange rate (British Pound price of one unit foreign currency) at time t; r_t and r_t^* are domestic and foreign interest rates from time t-1 to t, respectively; g is a one way transaction cost.

We consider two performance measures: the mean excess return and the the Sharpe ratio.[4] We use a natural benchmark that the investor does not take position in the foreign exchange market and hence earns a zero excess return (alternatively, zero Sharpe ratio). Therefore our performance measure is in fact a relative performance

[4]The Sharpe ratio used here is in fact so called information ratio, defined as the mean excess return divided by the standard deviation. In practice, people do not make a strict distinction between these two measures. So we still use the name "Sharpe ratio" here. We calculate the robust Sharpe ratio with studentization and HAC standard errors. See Wolf (2007) for details.

measure [$d_{k,t}$ in the notation of Hansen (2005)] of k^{th} trading strategy compared to the benchmark.

2.4.2 Empirical Tests

Reality Check

White (2000) tests the null hypothesis that the benchmark is not inferior to any of the alternative trading rules. Suppose we search in 1000 trading rules to test whether the best trading rule from these 1000 trading rules is significantly better than a benchmark. Instead of comparing one trading rule to the benchmark as in the classical test, we look at a vector of 1000 relative performance measure at the same time. Each element in the vector represents the relative performance of one model in the universe. Testing whether the best rule is profitable is equivalent to ask whether the **maximal** value in this **vector** is significantly larger than zero. The distribution of the maximum in a vector of elements differs from the distribution of the element in the vector. The latter is used for inference in classical tests. Doing this gives consideration to the full set of models underlying the vector that led to the best performing trading rule. Rejecting the null hypothesis implies that at least one trading rule beats the benchmark.

Due to the complication of the true distribution of the test, White (2000) recommends the stationary bootstrap method of Politis and Romano (1994) to first obtain the empirical distribution of the test statistic, and obtain the p-value by comparing it with the quantiles of the empirical distribution from bootstrap re-sampling. If the p-value is smaller than a given significance level, the null hypothesis is rejected.

SPA test

The null hypothesis of SPA test [Hansen (2005)] is the same as in White's Reality Check. Unlike the White's Reality Check, the SPA test uses the studentized test statistic, which will typically improve the power. Hansen (2005) provides a concrete example for highlighting the advantage of studentizing the individual statistics, since it avoids a comparison of the performance measured in different "units of standard deviation". Furthermore, the SPA test invokes a sample-dependent distribution under the null hypothesis, which can discard the poor models asymptotically. Therefore, the new test is more powerful and less sensitive to the inclusion of poor and irrelevant alternatives. The improvement of the power of the SPA test over the Reality Check is further confirmed by the simulation experiment conducted in Hansen (2005). The p-value of SPA test is calculated by bootstrapping its empirical distribution.

StepM test

Both the Reality Check and the SPA test seek to answer whether the best trading strategy beats the benchmark. As discussed in section 3.1, it is often more interesting to identify all outperforming trading rules, or to know whether a particular trading rule improves upon the benchmark. The Reality Check can be modified easily for identifying potential strategies that beat the benchmark, but Romano and Wolf (2005) show that this is only suboptimal, and only amounts to the first step of StepM test of Romano and Wolf (2005), which can detect more good strategies from the second step on. The StepM test is therefore more powerful than the Reality Check in detecting superior trading rules. The aim of the StepM test is to find as many profitable trading rules as possible. They test whether the individual trading rule is better than the benchmark by checking whether it falls into a joint confidence region constructed from all trading strategies. If not, that trading strategy is detected as profitable. When some trading strategies are detected, the remaining trading strategies can be used to construct a new joint confidence region. Then the first step is repeated to find whether there are remaining profitable trading strategies. One continues this way until no profitable rules can be detected.

Romano and Wolf (2005) also propose the use of studentization to improve the power and level properties of the StepM test.

Stepwise SPA test

The Stepwise SPA test of Hsu and Hsu (2006) aims at combining the advantages of both the SPA test and the StepM test, to improve upon the White's Reality Check in two different ways. In this setting, the null hypothesis is similar to the StepM test, though it invokes a sample-dependent distribution under the null hypothesis, such that poor and irrelevant trading rules will be discarded asymptotically. They provide a formal proof and simulation results to demonstrate that the SSPA test is more powerful than the Reality Check, SPA test, and StepM test.

2.5 Empirical Results

Data and Summary Statistics

We collect daily spot exchange rate data from Datastream, which covers the period from January 1994 to July 2007. The spot exchange rate is quoted as foreign currency units per British pound (hereafter GBP). We study the countries which have both a spot exchange rate and an overnight interest rate available (if no overnight interest rate is available, we use other daily short rate instead). In total, we include ten currencies from emerging markets, which represent various geographic regions. The data for most of them starts in the 1990s. All considered currencies and their data availability are reported in table 2.2.

Table 2.3 reports the summary statistics of daily returns (defined as the log difference of spot exchange rates). The mean return shows the average depreciation of individual emerging market currency against the GBP in the sample period. Almost all the emerging market currencies on average depreciate relatively to the GBP, with only one currency [Czech Republic (Koruny)] that appreciates. The highest daily depreciation happens to the Mexican Peso, with an amount of 11.4 percent. The highest appreciation happens to the Mexican Peso, too (15.56 percent). The standard deviations are usually higher than those commonly found in the exchange rates of developed countries [see

Table 2.2: Data Used for Testing the Profitability of Trading Rules

All data are from Datastream.

Currency	Symbol	Available data	Out-of-Sample test period
Brazil, Reais	BRL	01/2000-07/2007	08/2002-07/2007
Czech Republic, Koruny	CZK	01/1998-07/2007	08/2002-07/2007
Hungary, Forint	HUF	09/1995-07/2007	08/2002-07/2007
India, Rupees	INR	01/1994-07/2007	08/2002-07/2007
Indonesia, Rupiahs	IDR	01/1999-07/2007	08/2002-07/2007
Mexico, Pesos	MXN	01/1995-07/2007	08/2002-07/2007
Poland, Zlotych	PLN	01/1995-07/2007	08/2002-07/2007
South Africa, Rand	ZAR	01/1996-07/2007	08/2002-07/2007
Thailand, Baht	THB	01/1999-07/2007	08/2002-07/2007
Turkey, New Lira	TRY	08/2002-07/2007	n.a.

Qi and Wu (2006) for comparison]. The skewness is relatively small for all currencies. The large kurtosis found in most of these currencies shows that extreme depreciation/appreciation occurred. Lee, Gleason, and Mathur (2001) also report extremely high kurtosis, consistent with our findings. Overall the summary statistics illustrate more volatile exchange rates with the emerging market than with their developed countries' counterparts.

Table 2.3: Summary Statistics for Daily Changes in the Logarithm of Exchange Rates

This table reports the daily returns, defined as the log difference of spot exchange rates

Currency	Observation	Mean (%)	Std (%)	Min (%)	MAX (%)	Skewness	Kurtosis
BRL	1975	0.0119	1.09	-10.87	9.08	-0.11	13.57
CZK	2498	-0.0127	0.55	-2.49	2.66	-0.01	4.52
HUF	3107	0.0198	0.54	-2.42	4.89	0.36	6.92
INR	3541	0.0161	0.52	-2.29	3.99	0.23	5.49
IDR	2237	0.0156	1.05	-8.70	7.96	-0.20	12.62
MXN	3281	0.0323	1.02	-15.56	11.39	-0.49	45.37
PLN	3281	0.0119	0.60	-3.02	4.19	0.42	6.70
THB	2237	0.0057	0.62	-7.50	8.92	0.59	32.76
TRY	1303	0.0002	0.88	-5.61	6.34	0.64	8.67
ZAR	3021	0.0310	0.93	-9.10	7.02	0.14	9.92

2.5.1 Full Sample Results

In this section we report the performance of trading rules using the full sample, with the whole universe of (25,988) trading rules, and with or without taking the transaction cost into account. We first present the data snooping bias in the best performing rules, as those reported in the literature. We then report the bias in all seemingly profitable rules.

Biases in the most profitable trading rules

We first assume no transaction costs. Table 2.4 shows the performance of the best trading rules (in terms of best past performance) under both the mean excess return (panel A) and Sharpe Ratio (panel B) criteria for each currency. According to the mean excess return, the best trading rule of each currency yields all positive but highly heterogeneous mean excess returns across currencies, ranging from a low of 6% (Thai Baht) to a high of 32% (Brazilian Real).

Interestingly, four out of ten best trading rules reported are charting rules, which are all based on an under-smoothed version of spot rate series.[5] These spot rates are first smoothed using the kernel smoothing method, and then the charting rules are applied to obtain the trading signals. The two best trading rules are from the class of Head and Shoulders and the other two are from Rectangle Rules. Filter rules are found to be the most profitable for all three currencies.

Panel A also reports four p-values (one p-value from the Reality Check and three p-values from the SPA test, including the upper bound and lower bound p-values).[6] We also calculate the nominal p-value by applying the Reality Check to each trading rule. Since the nominal p-value does not account for the search among all the trading rules, it is subject to data snooping bias. We find the smallest nominal p-values very close to zero for every currency, we do not enlist it in the table. Note that the best performing rule from the original sample is not necessarily the best performing rule from bootstrap

[5]In the table, the number following directly after the abbreviation of the charting rules with the kernel smoothing indicates the degree of smoothing. "1" stands for smoothing with 0.3 multiple of optimal bandwidth. Similarly, "2" and "3" stand for 1 and 4 multiple of optimal bandwidth, respectively. "4" stands for no smoothing.

[6]To calculate the p-values, we apply circular block bootstrap with a block length of 2 and 500 bootstrap replications. Our results hardly change when we use a different block length or other bootstrap procedures such as the stationary bootstrap or the moving block bootstrap.

samples, which has the smallest nominal p-value calculated. Its past performance may simply stem from pure luck. The reported four p-values can be compared to the smallest nominal p-value, which is an easy but rigorous way of quantifying the effect of data-snooping bias. Not surprisingly, all the p-values with data snooping check are bigger than the smallest nominal p-values, which ignore the data snooping bias. Two out of ten currencies (Hungarian Forint and Mexican Peso) have a p value (SPA_c) which is less than 0.05, indicating that at least one trading rule is found to be significantly profitable (at 5% significance level) if one ignores transaction costs. For the remaining eight currencies, no trading rules are found to be profitable even in the absence of transaction costs.

The Sharpe ratios of the best rules reported in panel B are all positive and large in the absence of transaction costs, ranging from 0.95 to 1.81. Five out of ten currencies are detected to have profitable rules. The majority of the best performing rules (six out of ten) are charting rules with kernel smoothing, among which five are Head and Shoulders rules and one is the Rectangle rule. Two complex rules are also among the best performing rules, both of which are voting by learning rules based on the Rectangle rules class. Still, the smallest nominal p-value for every currency is close to zero. When comparing them to the p-values of the Reality Check and the SPA test, they are much smaller in most cases, indicating that the data snooping bias is large. Although the results so far show that some currencies have at least one profitable trading rule in the absence of transaction costs, it is still premature to conclude that there are profitable rules in practice where non-zero transaction costs are involved.

Transaction costs are an important concern when testing TTR profitability. Lee, Gleason, and Mathur (2001) uses one-way transaction costs of 0.1% for Latin American Countries' currencies. Burnside, Eichenbaum, and Rebelo (2007) mention that the emerging countries' bid-ask spreads are usually 2-4 times higher than the developed counties' currencies. Qi and Wu (2006) use transaction costs of 0.04% for developed FX markets. So we choose 0.1% as one-way transaction costs. We have also considered a higher transaction cost of 0.3% and a lower one of 0.04%. We report results with a one-way transaction cost of 0.1%. Results with other transaction costs are qualitatively

Table 2.4: Performance of the Best FX Trading Rules in Emerging Market (without Transaction Cost)

Panel A: Mean Return Criteria

Currency	Best Trading Rule	Nr. of Trades	Mean Excess Return	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	RA1(y= 4.50, k=0.000, r=0.0100, d= 0.50, x=0.030)	126	0.32	0.18	0.22	0.24	0.26
CZK	FR(x=0.0005, e=2)	1151	0.07	1.00	0.50	0.54	0.55
HUF	FR(x=0.0005, e=2)	1477	0.08	1.00	0.04	0.04	0.04
INR	SR(e=2, c=10)	309	0.08	0.98	0.42	0.43	0.45
IDR	MA(n=10, m=5, c=5)	311	0.17	0.99	1.00	1.00	1.00
MXN	HS1(y= 3.50, k=0.000, r=0.0100, d= 0.50, x=0.050)	188	0.16	0.03	0.03	0.04	0.04
PLN	FR(x=0.0250, c=25)	79	0.08	1.00	1.00	1.00	1.00
TRY	RA1(y= 1.25, k=0.000, r=0.0075, d= 0.75, x=0.030)	72	0.20	1.00	0.98	0.99	1.00
THB	HS1(y= 4.50, k=0.000, r=0.0100, d= 0.75, x=0.050)	146	0.06	1.00	0.70	0.72	0.73
ZAR	SR(e=2, d=4, c=10)	235	0.20	0.10	0.41	0.46	0.51

Panel B: Sharpe Ratio Criteria

Currency	Best Trading Rule	Nr. of Trades	Sharpe Ratio	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	VLS_RA(x=2, m=250, r=250)	82	1.81	0.06	0.02	0.02	0.05
CZK	HS1(y= 1.00, k=0.000, r=0.0100, d= 0.75, x=0.030)	140	1.46	0.11	0.62	0.62	0.72
HUF	HS1(y= 1.50, k=0.000, r=0.0050, d= 0.75, x=0.015)	166	1.59	0.00	0.00	0.00	0.00
INR	HS1(y= 1.50, k=0.000, r=0.0100, d= 0.75, x=0.050)	242	1.21	0.13	0.04	0.04	0.35
IDR	RA1(y= 1.75, k=0.000, r=0.0050, d= 0.75, x=0.015)	158	1.34	0.80	0.21	0.21	0.77
MXN	HS1(y= 1.50, k=0.000, r=0.0100, d= 0.50, x=0.030)	190	1.45	0.00	0.00	0.00	0.01
PLN	MSP(m=2, f=5, k= 0.01)	119	0.95	1.00	0.99	0.99	1.00
TRY	VLS_RA(x=5, m=125, r=40)	54	1.66	1.00	0.33	0.33	0.94
THB	HS1(y= 1.75, k=0.000, r=0.0100, d= 0.75, x=0.050)	148	1.59	0.08	0.04	0.04	0.54
ZAR	SR(e=2, d=4, c=10)	235	1.25	0.30	0.22	0.22	0.41

Note: Results are based on the whole universe containing 25988 trading rules. A detailed description of the parameters listed in the best trading rule can be found in the appendix A. The mean excess return and Sharpe ratio are both annualized.

similar.

Table 2.5 documents the performance of the best trading rules according to both the mean excess return (panel A) and Sharpe ratio (panel B).

Not surprisingly, both the mean excess returns and the Sharpe ratio are smaller than those in the previous case. They are still positive and large for all currencies, however. The annual mean excess return ranges from 0.05 (Thai Baht) to 0.30 (Brazilian Real). The annualized Sharpe ratio lie between 0.78 (Indian Rupee) and 1.72 (Brazilian Real). The number of transactions is also smaller. For the mean excess return criteria, seven of the best rules are simple rules, two are charting rules with under-smoothing and one is a learning rule based on the filter rule class. Under the Sharpe ratio criteria, four of the best performing rules are Head and Shoulders applied to under-smoothed spot rates, and two are learning rule and voting by learning rule. The remaining four are the momentum rules in price, the filter rule and the support and resistance rule.

A close look at p-values of the Reality Check and the SPA tests shows no profitable rules according to the mean excess return criteria. The smallest p-value with data-snooping check is 0.17 (Mexican Peso). For the Sharpe ratio, we find that two out of ten currencies (Brazilian Real and Mexican Peso) have profitable rules, even after the data-snooping bias is controlled for[7]. In general, since we still have the smallest nominal p-value close to zero for every currency, we will find profitable rules if the data snooping bias is not considered. Therefore, ignoring the data snooping effect leads to a significant bias.

Bias in all "profitable" trading rules

It is often desirable to know all profitable trading rules, as discussed in the previous sections. We detect all the profitable rules according to the individual nominal p-value, the StepM test and the SSPA test. We report the total number of profitable trading rules detected by the nominal p-value and compare them to those found by the StepM

[7]In the case of a 0.04% one-way transaction cost, we find no profitable rules according to the mean excess return criteria, but find three currencies have profitable rules according to the Sharpe ratio criteria. Assuming a transaction cost of 0.3%, we find no profitable rules under either criteria.

Table 2.5: Performance of the Best FX Trading Rules in Emerging Market (with a 0.1% one-way Transaction Cost)

Panel A: Mean Return Criteria

Currency	Best Trading Rule	Nr. of Trades	Mean Excess Return	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	RA1(y= 4.50, k=0.000, r=0.0100, d= 0.75, x=0.030)	114	0.30	0.43	0.27	0.30	0.37
CZK	SR(n=5, c=25)	75	0.05	1.00	1.00	1.00	1.00
HUF	MA(n=200, m=125)	15	0.05	1.00	0.81	0.89	0.94
INR	SR(e=2, c=10)	309	0.05	1.00	1.00	1.00	1.00
IDR	MA(n=15, m=2, b=0.030)	23	0.14	1.00	1.00	1.00	1.00
MXN	HS1(y= 3.50, k=0.000, r=0.0100, d= 0.50, x=0.030)	188	0.14	0.21	0.14	0.17	0.21
PLN	FR(x=0.0250, c=25)	79	0.08	1.00	1.00	1.00	1.00
TRY	LS_FR1(m=125, r=40)	57	0.19	1.00	1.00	1.00	1.00
THB	CB(n= 100, x=0.100, c= 50)	3	0.05	1.00	1.00	1.00	1.00
ZAR	SR(e=2, d=4, c=10)	235	0.18	0.49	0.89	0.93	0.99

Panel B: Sharpe Ratio Criteria

Currency	Best Trading Rule	Nr. of Trades	Sharpe Ratio	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	VLS_RA(x=2, m=250, r=250)	60	1.72	0.12	0.04	0.04	0.12
CZK	HS1(y= 1.00, k=0.000, r=0.0100, d= 0.75, x=0.050)	140	1.02	1.00	1.00	1.00	1.00
HUF	HS1(y= 1.25, k=0.000, r=0.0050, d= 0.75, x=0.030)	168	1.12	0.75	0.16	0.16	0.64
INR	MSP(m=250, f=25, k= 0.01)	4	0.73	1.00	1.00	1.00	1.00
IDR	MSP(m=50, f=5, w=50, k=0.01)	14	1.18	1.00	0.48	0.49	0.98
MXN	HS1(y= 1.50, k=0.000, r=0.0100, d= 0.75, x=0.015)	154	1.29	0.07	0.02	0.02	0.09
PLN	FR(x=0.0250, c=25)	79	0.78	1.00	1.00	1.00	1.00
TRY	VLS_RA(x=5, m=125, r=40)	50	1.51	1.00	0.92	0.92	0.99
THB	HS1(y= 1.75, k=0.000, r=0.0100, d= 0.75, x=0.030)	148	1.08	1.00	1.00	1.00	1.00
ZAR	SR(e=2, d=4, c=10)	235	1.11	0.95	0.70	0.70	0.96

Note: Results are based on the whole universe containing 25988 trading rules. Detailed description of the parameters listed in the best trading rule can be found in the appendix A. The mean excess return and Sharpe ratio are both annualized.

or the SSPA test. This comparison can serve as a way of summarizing the danger of data snooping for *all* profitable rules according to the nominal p-value.

Table 2.6 reports the number of profitable trading rules without transaction costs (panel A), with one-way transaction costs of 0.04% (Panel B) and with one-way transaction costs of 0.1% (panel C) under both the mean excess return and the Sharpe ratio criteria. In the absence of transaction costs, and ignoring data snooping bias (nominal p-value), we find for almost every currency more than 1,000 profitable trading rules for each criteria. When the data snooping bias is taken into account, we find that only two out of ten currencies have profitable rules according to the mean excess return, and five out of ten according to the Sharpe ratio criteria. These findings are consistent with the results in table 2.4. Furthermore, the total number of trading rules detected to be profitable is quite small compared to those found by individual nominal p-values. As to the mean excess return criteria, the largest number of profitable trading rules found is 41 (Hungarian Forint by the SSPA test) while a test based on nominal p-value finds 1,266 profitable rules. For the Sharpe criteria, the detected number of profitable rules is at most 134 (Mexican Peso by the SSPA test) and a test based on nominal p-value finds 2,117 profitable rules. The Brazilian Real is found to have the largest number (more than 3,000) of profitable trading rules under both criteria if ignoring data snooping bias, though no more than one rule remains significantly profitable according to StepM and SSPA. This means that for about every 8 trading rules we can find one profitable rule, but none of them are indeed profitable. These results are striking, highlighting the enormous danger of data snooping.

When we have one-way transaction costs of 0.04%, we find fewer profitable trading rules (Panel B). They range from 242 to 3016 according to mean excess return, and from 468 to 3174 rules according to the Sharpe ratio criteria. StepM and SSPA test find only three currencies with profitable rules in the case of the Sharpe ratio. Even in the case of a 0.1% transaction cost, we can still find hundreds or thousands of trading rules for every currency according to both criteria when we consider the individual nominal p-value. The StepM test on the contrary, finds no profitable rules. Only under the Sharpe ratio criteria, does the SSPA test find one profitable rule for the Brazilian

Real and 24 profitable rules for the Mexican Peso. Again, these striking results indicate that the data snooping bias is substantial.

All three panels also show that the SSPA test can detect more, but not less profitable trading rules than the StepM test, confirming that the SSPA test does have higher power than the StepM test.

2.5.2 Sub-sample Analysis

Trading rule performance is often not stable. Today's profitable rule does not guarantee sure profits for tomorrow. For example, Sullivan, Timmermann, and White (1999) find that the best trading rule applying to DJIA for the period of 1897-1986 does not outperform the benchmark for the period of 1987-1996. Qi and Wu (2006) find the profitability has declined in recent periods. For this reason, we document the trading rule profitability in two subperiods of our sample. The second subperiod is between 08/2002 and 07/2007 for each currency. Having the same length of second subperiods facilitates a comparison across currencies. We report the results which assume one-way transaction costs of 0.1%. The results with transaction costs of 0.04% are qualitatively similar.

Table 2.7 reports the results for the first subperiod. Most of the best performing rules (eight out of ten) are found to be simple rules like channel breakout rules, and the remaining two are Head and Shoulders rules with under smoothing for mean excess return (panel A). Under the Sharpe ratio criteria (panel B) the best trading rules disperse over the simple rules, charting rules with kernel smoothing and complex rules. For both criteria, we find extremely high performance for Brazilian Real (annual mean excess return of 0.76 and Sharpe ratio of 2.97). We still find thousands of profitable trading rules according to an individual nominal p-value (results not reported), but we find no profitable trading rules according to the SPA test.

Table 2.6: Number of Profitable Trading Rules

All profitable trading rules are detected at 5% level. The universe of trading rules contains 25988 rules. The whole sample data are used for detecting the profitable rules.

Panel A: without transaction cost

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	3393	0	0	3556	1	1
CZK	442	0	0	704	0	0
HUF	1266	32	41	1725	103	121
INR	1286	0	0	1844	0	16
IDR	1470	0	0	2370	0	0
MXN	1822	18	22	2117	102	134
PLN	1115	0	0	1263	0	0
TRY	1116	0	0	1554	0	0
THB	944	0	0	2234	0	5
ZAR	1949	0	0	2164	0	0

Panel B: with one-way transaction cost of 0.04%

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	3016	0	0	3174	0	1
CZK	242	0	0	468	0	0
HUF	822	0	0	1134	75	95
INR	574	0	0	917	0	0
IDR	909	0	0	1713	0	0
MXN	1471	0	0	1648	32	113
PLN	578	0	0	606	0	0
TRY	880	0	0	1158	0	0
THB	562	0	0	1812	0	0
ZAR	1545	0	0	1784	0	0

Panel C: with one-way transaction cost of 0.1%

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	2599	0	0	2744	0	1
CZK	117	0	0	192	0	0
HUF	338	0	0	417	0	0
INR	198	0	0	419	0	0
IDR	396	0	0	657	0	0
MXN	1139	0	0	1316	0	24
PLN	323	0	0	358	0	0
TRY	680	0	0	871	0	0
THB	264	0	0	1224	0	0
ZAR	1166	0	0	1361	0	0

Table 2.7: Performance of the Best FX Trading Rules in First Sub-period (with a 0.1% one-way Transaction Cost)

Panel A: Mean Return Criteria

Currency	Best Trading Rule	Nr. of Trades	Mean Excess Return	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	CB(n=5, x=0.025, c=5, b=0.0050)	13	0.76	0.92	0.99	0.99	0.99
CZK	MA(n=15, m=5, b=0.015)	5	0.12	1.00	1.00	1.00	1.00
HUF	SR(e=4, d=2, c=25)	49	0.07	1.00	1.00	1.00	1.00
INR	CB(n=15, x=0.010, c=25, b=0.0010)	11	0.05	1.00	1.00	1.00	1.00
IDR	CB(n=10, x=0.050, c=10, b=0.0010)	17	0.36	1.00	1.00	1.00	1.00
MXN	HS1(y= 4.00, k=0.000, r=0.0075, d= 0.50, x=0.030)	108	0.15	1.00	0.61	0.69	0.83
PLN	FR(x=0.0250, c=25)	39	0.12	1.00	1.00	1.00	1.00
THB	MA(n=10, m=5, c=50)	13	0.11	1.00	1.00	1.00	1.00
ZAR	HS1(y= 3.00, k=0.000, r=0.0100, d= 0.75, x=0.050)	74	0.18	1.00	1.00	1.00	1.00

Panel B: Sharpe Ratio Criteria

Currency	Best Trading Rule	Nr. of Trades	Sharpe Ratio	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	CB(n=5, x=0.025, c=5, b=0.0050)	13	2.97	0.72	0.28	0.28	0.47
CZK	MA(n=15, m=5, b=0.015)	5	1.27	1.00	1.00	1.00	1.00
HUF	HS1(y= 2.00, k=0.000, r=0.0050, d= 0.75, x=0.015)	76	1.07	1.00	1.00	1.00	1.00
INR	VLS_TA(x=2, m=125, r=125)	28	1.15	1.00	0.98	0.98	1.00
IDR	MSP(m=50, f=5, k= 0.01)	8	1.76	1.00	0.95	0.95	1.00
MXN	VLS_HS(x=50, m=60, r=40)	78	1.51	0.47	0.09	0.09	0.42
PLN	MSP(m=5, f=10, w=5, k=0.02)	6	1.17	1.00	0.93	0.93	1.00
THB	DTB2(f= 5, k=0.005, n= 22, x=0.030)	18	1.39	1.00	1.00	1.00	1.00
ZAR	HS1(y= 1.25, k=0.000, r=0.0100, d= 0.75, x=0.030)	76	1.28	1.00	1.00	1.00	1.00

Note: Results are based on the whole universe containing 25988 trading rules. A detailed description of the parameters listed in the best trading rule can be found in the appendix A. The mean excess return and Sharpe ratio are both annualized. The first sub-sample periods for each currencies are defined in the table 2.2

Table 2.8: Performance of the Best FX Trading Rules in Second Sub-period (with a 0.1% one-way Transaction Cost)

Panel A: Mean Return Criteria

Currency	Best Trading Rule	Nr. of Trades	Mean Excess Return	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	RA1(y= 4.00, k=0.000, r=0.0075, d= 0.75, x=0.030)	78	0.23	1.00	1.00	1.00	1.00
CZK	MA(n=10, c=25)	39	0.06	1.00	1.00	1.00	1.00
HUF	SR(e=5, b=0.0050, c=50)	15	0.11	1.00	1.00	1.00	1.00
INR	LS_SR1(m=60, r=60)	63	0.12	1.00	1.00	1.00	1.00
IDR	MA(n=2, b=0.015)	5	0.12	1.00	1.00	1.00	1.00
MXN	SR(n=100, b=0.0050, c=50)	5	0.13	1.00	1.00	1.00	1.00
PLN	CB(n=5, x=0.025, c=10, b=0.0010)	15	0.10	1.00	1.00	1.00	1.00
THB	FR(x=0.0100, c=50)	9	0.10	1.00	1.00	1.00	1.00
ZAR	SR(e=2, d=4, c=10)	91	0.17	1.00	1.00	1.00	1.00

Panel B: Sharpe Ratio Criteria

Currency	Best Trading Rule	Nr. of Trades	Sharpe Ratio	P-value (RC)	P-value (SPA_l)	P-value (SPA_c)	P-value (SPA_u)
BRL	RA1(y= 3.00, k=0.000, r=0.0075, d= 0.75, x=0.050)	78	1.57	1.00	1.00	1.00	1.00
CZK	HS1(y= 2.00, k=0.000, r=0.0100, d= 0.75, x=0.030)	76	1.43	1.00	1.00	1.00	1.00
HUF	SR(e=5, b=0.0050, c=50)	15	1.24	1.00	1.00	1.00	1.00
INR	MSP(m=250, f=10, w=20, k=0.05)	4	1.47	1.00	0.96	0.96	1.00
IDR	VLS_RA(x=5, m=125, r=20)	58	1.45	1.00	1.00	1.00	1.00
MXN	MSP(m=40, f=50, w=20, k=0.01)	2	1.41	1.00	1.00	1.00	1.00
PLN	MSP(m=250, f=5, k= 0.02)	10	1.46	1.00	0.89	0.89	1.00
THB	HS1(y= 1.00, k=0.000, r=0.0100, d= 0.75, x=0.030)	98	1.14	1.00	1.00	1.00	1.00
ZAR	MSP(m=60, f=50, w=40, k=0.01)	2	1.19	1.00	1.00	1.00	1.00

Note: Results are based on the whole universe containing 25988 trading rules. A detailed description of the parameters listed in the best trading rule can be found in the appendix A. The mean excess return and Sharpe ratio are both annualized. The sample period is 08/2002-07/2007.

We report the results for the second sample period in table 2.8. Unlike Qi and Wu (2006), we do not find a uniform decline of mean excess returns and the Sharpe ratio relative to the first sub-sample. Indeed, we find that two currencies under mean excess return criteria and four currencies under Sharpe ratio criteria have higher profits than their first subperiod counterparts. We also find that the performance of the best rule for Brazilian Real is not that extreme (annual mean excess return of 0.23 and Sharpe ratio of 1.57). The best performing rules are most frequently found to be simple rules such as support and resistance and momentum strategy in price, though there are also charting rules applied to under smoothed spot rates and complex rules. When looking at the p-values with data snooping check, we again find no profitable rules.

2.5.3 Out-of-sample Data Snooping Bias

An out-of-sample test is an effective way of detecting a data snooping bias. Still, caution has to be taken, when the same new data is used repeatedly for testing various models in out-of-sample experiments. In this case, a data snooping bias can still emerge. This is particularly relevant when the number of models to be tested out-of-sample is really large. Unfortunately, this potential pitfall has not been well acknowledged in the literature. In this section we show that the out-of-sample data snooping bias can be substantial in testing a large number of trading rules.

We conduct the out-of-sample experiment as follows. First, we take the first sub-sample as the in-sample data, and detect all profitable trading rules according to the nominal p-value. That is, we ignore the data snooping bias. The detected profitable rules are then taken as the universe of trading rules for our out-of-sample analysis. The out-of-sample test uses the second sub-sample and detects all profitable trading rules using tests with and without a data snooping check.

Table 2.9: Number of Profitable Trading Rules in Out-of-Sample Experiment

Panel A: with one-way transaction cost of 0.04%

	Mean Excess Return Criteria					Sharpe Ratio Criteria				
Currency	$NP_{Sample1}$	$NP_{Sample2}$	NP_{OOS}	$StepM_{OOS}$	$SSPA_{OOS}$	$NP_{Sample1}$	$NP_{Sample2}$	NP_{OOS}	$StepM_{OOS}$	$SSPA_{OOS}$
BRL	5123	2326	630	0	0	6713	2663	939	0	0
CZK	245	442	120	0	0	545	754	136	6	6
HUF	426	269	170	0	0	914	767	251	1	1
INR	632	415	110	0	0	1133	696	211	1	1
IDR	410	710	31	0	0	914	2502	55	4	4
MXN	1498	548	205	0	0	1673	728	231	0	0
PLN	836	562	3	0	0	920	641	23	1	1
THB	142	385	52	0	0	677	1633	162	0	0
ZAR	1310	1402	85	0	0	1470	1784	110	1	1

Panel B: with one-way transaction cost of 0.1%

	Mean Excess Return Criteria					Sharpe Ratio Criteria				
Currency	$NP_{Sample1}$	$NP_{Sample2}$	NP_{OOS}	$StepM_{OOS}$	$SSPA_{OOS}$	$NP_{Sample1}$	$NP_{Sample2}$	NP_{OOS}	$StepM_{OOS}$	$SSPA_{OOS}$
BRL	4665	1903	427	0	0	6320	2188	704	0	0
CZK	131	210	52	0	0	311	328	75	0	0
HUF	159	150	97	0	0	485	205	106	0	0
INR	219	102	1	0	0	572	212	9	0	0
IDR	296	285	12	0	0	753	1395	22	0	0
MXN	1234	300	72	0	0	1400	430	83	0	0
PLN	438	296	2	0	0	501	372	12	1	1
THB	44	119	1	0	0	481	1030	44	0	0
ZAR	980	97	10	0	0	1094	232	21	0	0

Note: All profitable trading rules are detected at the 5% level. The TTR universe for sub-sample analysis is the whole universe of 25988 rules. The TTR universe for the out-of-sample test consists of the profitable rules found in the first sample according to nominal p-values. “NP” refers to the number of profitable rules according to nominal p-values. “OOS” refers to out-of-sample test

Table 2.9 reports the number of trading rules in the out-of-sample tests. "NP$_{Sample1}$" refers to the number of profitable trading rules found in the first sub-sample when the data snooping bias is ignored. "NP$_{OOS}$" stands for the number of profitable trading rules found in the out-of-sample test without controlling for a data snooping bias. This implies that all these rules are found to be profitable both in-sample and out-of-sample according to the nominal p-value. "StepM$_{OOS}$" and "SSPA$_{OOS}$" report the number of profitable rules detected out-of-sample according to the StepM test and the SSPA test, respectively. The difference between NP$_{OOS}$ and StepM$_{OOS}$ or SSPA$_{OOS}$ provides a quantification of the extent of the out-of-sample data snooping bias. In panel A we consider a one-way transaction cost of 0.04%. According to the mean excess criteria, ignoring the data snooping bias, we find seemingly profitable trading rules for every country. They range from 3 (Polish Zloty) to 630 (Brazilian Real). 5 out of 10 currencies have more than 100 profitable rules. The percentage of rules which are profitable both in sample and out-of-sample differs a lot across currencies. For example, almost 50% of in-sample profitable rules are also out-of-sample profitable for Czech Koruna, but for Polish Zloty, this percentage is almost zero. When we control for data snooping bias, both the out-of-sample StepM test and SPA test find no profitable rules. If we consider the Sharpe ratio criteria, we find similar substantial out-of-sample data snooping bias, though in this case the StepM test and SSPA test sometimes identify a few (less than 7) profitable rules out-of-sample. Imposing a higher one-way transaction cost of 0.1% (panel B) reduces the number of profitable rules both in-sample and out-of-sample. Still, the difference between NP$_{OOS}$ and StepM$_{OOS}$ or SSPA$_{OOS}$ shows that ignoring the data snooping is a danger practice. In the case of the Czech Koruna under mean excess return criteria, 52 out of 131 in-sample detected profitable rules will be wrongly discovered as profitable in the absence of a data snooping check. With a one-way cost as high as 0.1%, there are still 704 trading rules incorrectly detected as profitable for Brazilian Real under Sharpe ratio criteria. Both the StepM test and SSPA test only find one rule to be profitable out-of-sample, which is free of data snooping bias. This indicates that, on average out-of-sample tests do alleviate the data snooping bias to a large extent, but ignoring the repeating testing in out-of-sample experiments can bias the results substantially.

2.5.4 Reduced Universe of Trading Rules

Our universe of trading rules is the largest in number in the FX trading rule literature so far. This, however, raises the concern about the power of the test, as the number of trading rules is large and is usually five times higher than the number of the observations. To address this issue, we consider a reduced universe, which contains the trading rules in each class.[8] The size of these universes ranges from 497 (filter rule) to 3,384 (Head and Shoulders for example). By doing this, we assume that the professional traders and researchers only search for profitable rules within each class. We have also considered a universe which either contains all simple trading rules, or all charting rules with kernel smoothing.

We report the results for the filter rule and the moving average rule, which are the most frequently studied classes in the FX trading rules literature. Results based on other classes of trading rules are similar, and are available from the authors upon request.

Panel A of table 2.10 shows the results for the class of filter rules. Even when assuming no transaction costs, and in a universe as small as the one of only 497 trading rules, we detect no profitable trading rules according to the StepM and the SSPA test for all currencies. Ignoring the data snooping bias involved, however, one can find much more profitable trading rules based on the nominal p-value. The extent of data snooping bias is substantial. Similarly, for the moving average rule in panel B, we find some profitable trading rules according to the nominal p-value. Once the data snooping bias is controlled for, we find no profitable rules in the moving average class, even when we assume zero transaction costs. Given the results in these two examples and similar results for unreported universes, we conclude that our major results, namely that the data snooping bias is substantial and that evidence supporting the TTR profitability is rare, are robust to a smaller universe of trading rules.

[8]We do not consider complex rules here, since they are based on simple rules and charting rules with kernel smoothing, which search implicitly in large universes.

Table 2.10: Number of Profitable Trading Rules (Reduced Universe)

All profitable trading rules are detected at the 5% significance level. We assume that there is no transaction cost involved.

Panel A: Filter Rule class (total number of trading rules: 497)

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	80	0	0	88	0	0
CZK	6	0	0	8	0	0
HUF	17	0	0	20	0	0
INR	19	0	0	20	0	0
IDR	18	0	0	19	0	0
MXN	6	0	0	6	0	0
PLN	32	0	0	33	0	0
TRY	32	0	0	27	0	0
THB	0	0	0	0	0	0
ZAR	25	0	0	24	0	0

Panel B: Moving Average class (total number of trading rules: 2049)

	Mean Excess Return Criteria			Sharpe Ratio Criteria		
Currency	Nominal P-value	StepM	SSPA	Nominal P-value	StepM	SSPA
BRL	476	0	0	438	0	0
CZK	4	0	0	3	0	0
HUF	16	0	0	16	0	0
INR	16	0	0	16	0	0
IDR	107	0	0	114	0	0
MXN	13	0	0	16	0	0
PLN	90	0	0	96	0	0
TRY	11	0	0	14	0	0
THB	0	0	0	0	0	0
ZAR	102	0	0	96	0	0

2.6 Conclusion

There is substantial empirical evidence documenting the success of the technical analysis in various markets, especially in FX markets. Critics, however, cite data snooping as one major source of this anomaly.

The purpose of this paper is to quantifying the extent of data snooping bias in technical trading rule profitability. We test trading rule performances across 25,988 trading rules for the emerging foreign exchange markets. We find that nearly all seemingly successes of technical analysis in our trading rule universe are in fact the result of data mining bias. Even the out-of-sample tests, when not controlling for data snooping bias, sometimes lead to many seemingly profitable but in fact spurious rules. These striking results show that data snooping bias can be enormous both in-sample and out-of-sample. Our findings continue to hold when we consider different transaction costs, sub-sample analysis and various reduced universes of trading rules.

Our empirical results also provide a first comprehensive evidence of the technical trading rule profitability for 10 emerging countries' exchange rates. The universe of our trading rules includes simple rules, charting rules with kernel smoothing and complex trading rules. Although popular among professional traders, the majority of these trading rules have not been studied in the literature for emerging FX markets. Our paper depicts a more complete picture of the performance of trading rules and market efficiency for emerging foreign exchange markets. Overall we find rare evidence against the efficiency of emerging FX markets.

Investors have to be careful when applying technical analysis to these markets, given the substantial data snooping bias that is likely to be involved. Further research, such as the investigation of economic explanations for trading rules' profitability or determinants of its cross country variation, also needs to account for the danger of data snooping.

Appendix

Appendix A: Description of Simple Trading Rules

Our descriptions of simple trading rules draw heavily on Sullivan, Timmermann, and White (1999), Qi and Wu (2006) and Hsu and Kuan (2005), though we have made some modifications to avoid ambiguities.

Filter Rules

The filter rule strategy for generating a trading signal follows Sullivan, Timmermann, and White (1999) and Qi and Wu (2006). The basic filter rule could be stated as follows: if the daily closing price (in British Pound) of a foreign currency moves up by $x\%$ or more from its most recent low, the speculator borrows the British Pound and uses the proceeds to buy and hold the foreign currency until its price moves down at least $x\%$ from a subsequent high, at which time the speculator short sells the foreign currency and uses the proceeds to buy the British Pound. Two definitions of the subsequent high (low) are considered. One is the highest (lowest) closing price over the period of holding a particular long (short) position. The alternative high (low) refers to the highest (lowest) price over the e most recent days. We also consider that a given long or short position is held for prespecified c days during which period all other signals are ignored.

Moving Average Rules

Moving averages are among the oldest trading rules used by chartist. The (equally weighted) moving average of a currency price for a given day t over the n days is $\frac{1}{n}\sum_{i=0}^{n-1} s_{t-i}$. Under a simple single moving average rule, when the current price is above the moving average by an amount larger than the band with $b\%$, the speculator borrows the British Pound to buy the foreign currency. Similarly, when the current price is below the moving average by $b\%$, the speculator short sells the foreign currency to buy the British Pound. Under dual moving average rule, when the short moving average of a foreign currency price is above the long moving average by an amount larger than the band with $b\%$, the speculator borrows the British Pound to buy the foreign currency. If the short moving average of a foreign currency price penetrates the long moving average from above, the speculator short sells the foreign currency to buy the British Pound. Following Sullivan, Timmermann, and White (1999) and Qi and Wu (2006), we implement the moving average rules with a time delay filter in addition to the fixed percentage band filter as described above. The time delay filter requires that the long or short signals remain valid for d days before action is taken. Similar to the filter rule case, we also consider holding a given long or short position for c days during which period all other signals are ignored.

Trading Range Break (or Support and Resistance) Rules

The support and resistance level refers to certain price levels acting as barriers to prevent traders from pushing the price of an underlying asset in a certain direction. Under a trading range break rule, when the price of a foreign currency moves above the maximum price (resistance level) over the previous n days by $b\%$, the speculator borrows the British Pounds to buy the foreign currency. When the price falls below the minimum price over the previous n days by $b\%$, the speculator sells short the foreign currency and buys the British Pound. Alternatively, we use the local maximum (minimum), which refers to the highest (lowest) price over the e most recent days, as the definition for the resistance level. Here we also allow a time delay filter, d, as well as the holding period of c days to be included, as in the case of moving average rules.

Channel Breakout Rules

A channel occurs when the high price of a foreign currency over the previous n days is within $x\%$ of the low over the previous n days. Under a channel breakout rule, when the closing price of the foreign currency exceeds the channel by $b\%$, a signal is generated for the speculator to borrow the British Pound and buy the foreign currency. Likewise, when the closing price of the foreign currency drops below the channel by $b\%$, a signal is generated for the speculator to short sell the foreign currency and buy the British Pound. Again, we consider a holding period of c days.

Momentum strategies

A momentum strategy attempts to predict the strength or weakness of the current market based on an "oscillator" constructed from a momentum measure. We follow Hsu and Kuan (2005) and use the rate of change (ROC) as the momentum measure. The m-day ROC at time t is defined as the change of spot exchange rate divided by the closing spot exchange rate at time $t - m$. Two oscillators are considered: the simple oscillator (which is just m-day ROC), and the moving average oscillator (which is the w-day moving average of m-day ROC with $w \leq m$). An overbought/oversold level k (say 5% or 10%) is needed to determine whether a position should be initiated. When the oscillator penetrates the overbought level from below, the speculator borrows the British Pound to buy the foreign currency. If the oscillator crosses the oversold level from above, the speculator sells short the foreign currency and buys the British Pound. Once again, we consider a holding period of c days.

Appendix B: Documentation of Trading Rules Parameters

Simple Rules

<u>Filter Rules (FR)</u>

x: increase in the log pound value of foreign currency required to generate a "buy" signal

y: decrease in the log pound value of foreign currency required to generate a "sell" signal

e: the number of the most recent days needed to define a low (high) based on which the filters are applied to generate "long" ("short") signal

c: number of days a position is held during which all other signals are ignored

$x = 0.0005, 0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.25, 0.3$ (24 values)

$y = 0.0005, 0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, 0.075, 0.1$ (12 values)

$e = 1, 2, 3, 4, 5, 10, 15, 20$ (8 values)

$c = 5, 10, 25, 50$ (4 values)

Noting that y must be less than x, there are 185 (x,y) combinations

Number of rules in FR class $= x + x \times e + x \times c +$ ((x,y) combinations) $= 24 + 192 + 96 + 185 = 497$

Moving Average Rules (MA)

n: number of days in a moving average

m: number of fast-slow combinations of n

b: fixed band multiplicative value

d: number of days for the time delay filter

c: number of days a position is held, ignoring all other signals during that time

$n = 2, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 200, 250$ (15 values)

$m = \sum_{i=1}^{n-1} i = 105$

$b = 0.0005, 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$ (8 values)

$d = 2, 3, 4, 5$ (4 values)

$c = 5, 10, 25, 50$ (4 values)

Number of rules in MA class: $= n + m + b \times (n + m) + (d \times (n + m) + c \times (n + m) + 9$

$= 15 + 105 + 960 + 480 + 480 + 9 = 2049$

Support and Resistance (SR, or Trading Range Break) Rules

n: number of days in the support and resistance range;

e: used for an alternative definition of extrema where a low (high) can be defined as the most recent closing price that is less (greater) than the n previous closing prices;

b: fixed band multiplicative value;

d: number of days for the time delay filter;

c: number of days a position is held, ignoring all other signals during that time

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ (10 values);

$e = 2, 3, 4, 5, 10, 20, 25, 50, 100, 200$ (10 values);

$b = 0.0005, 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$ (8 values);

$d = 2, 3, 4, 5$ (4 values);

$c = 5, 10, 25, 50$ (4 values);

Number of rules in SR class $= [(1+c) \times (n+e)] + [b \times (n+e) \times (1+c)]$ $+[d \times c \times (n+e)]=$ $100 + 800 + 320 = 1220$

Channel Breakout Rules (CBO)

n: number of days for a channel

x: difference between the high price and the low price ($x\times$low price) required to form a channel

b: fixed band multiplicative value ($b < x$)

c: number of days a position is held, ignoring all other signals during that time

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ (10 values);

$x = 0.001, 0.005, 0.01, 0.02, 0.03, 0.05, 0.075, 0.10$ (8 values)

$b = 0.0005, 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$ (8 values)

$c = 1, 5, 10, 25$ (4 values)

Noting that b must be less than x. There are 43 (x,b) combinations.

Number of rules in CBO class $= n \times x \times c + n \times c \times$((x,b) combinations)$= 320 + 1720 = 2040$

Momentum Strategies in Price (MSP)

m: number of days rate of change in price;

w: number of days in moving average;

k: overbought/oversold level;

f: fixed holding days;

$m = 2, 5, 10, 20, 30, 40, 50, 60, 125, 250$ (10 values);

$w = 2, 5, 10, 20, 30, 40, 50, 60, 125, 250$ (10 values);

$k = 0.05, 0.10, 0.15, 0.2$ (4 values);

$f = 5, 10, 25, 50$ (4 values);

Noting that w must be less than or equal to m, there are 55 $w - m$ combinations.

Number of rules in MSP class $= m \times k \times f +$ ((m,w) combinations)$\times k \times f = 160 + 880$ $= 1040$

Charting Rules with Kernel Smoothing

The Charting rules share common parameters, so we first describe them here.

k: fixed band multiplicative value;

f: fixed holding days;

r: stop loss rate;

d: parameter for fixed liquidation price;

y: multiple of standard deviation of daily exchange-rate changes used to liquidate the position;

b: multiple of optimal bandwidth of kernel regression;

$k = 0, 0.005, 0.01;$ (3 values)

$f = 1, 5, 10, 25$ (4 values)

$r = 0.005, 0.0075, 0.01;$ (3 values)

$d = 0.25, 0.5, 0.75;$ (3 values)

$y = 1, 1.25, 1.5, 1.75, 2.00, 2.5, 3.00, 3.50, 4.00, 4.50$ (10 values)

$b = 0.3, 1, 4$ (3 values)

Head-and-Shoulders (HS)

x: differential rate of shoulders or troughs

$x = 0.015, 0.03, 0.05;$ (3 values)

Number of rules in HS class $= (x \times k \times r \times d \times y + x \times k \times f) \times (1 + b)$= $(810 + 36) \times 4$ $= 3384$

Triangle (TA)

Number of rules in TA class: $= (k \times r \times d \times y + k \times f) \times (1 + b)$= $(270 + 12) \times 4 = 1128$

Rectangle (RA)

Number of rules in RA class $= (x \times k \times r \times d \times y + x \times k \times f) \times (1 + b)$= $(810 + 36) \times 4$ $= 3384$

Double Tops and Bottoms (STB)

x: differential rate of shoulders or troughs

n: least day differential between two tops/bottoms;

$x = 0.015, 0.03, 0.05$; (3 values)

$n = 22$ (1 value);

Number of rules in DTB class: $= (x \times k \times r \times d \times y \times n + x \times k \times f \times n) \times (1 + b)$

$= (810 + 36) \times 4 = 3384$

Broadening Tops and Bottoms (BTB)

Number of rules in BTB class: $= (k \times r \times d \times y + k \times f) \times (1 + b) = (270 + 12) \times 4 = 1128$

Complex Rules

Learning strategies (LS)

m: memory span;

r: review span;

$m = 2, 5, 10, 20, 40, 60, 125, 250$ (8 values);

$r = 1, 5, 10, 20, 40, 60, 125, 250$ (8 values);

Noting that $r \leq m$, there are 36 (m,r) combinations.

We have 2 performance measures: the sum of m daily returns, and the sharpe ratio.

In addition, including class with kernel smoothed series and learning on all non-complex rules, there are $5 + 5 \times 4 + 1 = 26$ classes of trading rules.

Number of rules in LS class $= 36 \times 2 \times 26 = 1872$

Voting Strategies (VS)

Number of rules in VS class $= 26$

Fraction Position Strategies (FRS)

Number of rules in FRS class $= 25 + 1 = 26$

Voting by learning strategies (VLS)

m: memory span;

r: review span;

n: number of top trading rules chosed within a class;

$m = 1, 2, 5, 10, 20, 40, 60, 125, 250$ (9 values);

$r = 1, 5, 10, 20, 40, 60, 125, 250$ (8 values);

$n = 2, 3, 5, 10, 50$ (5 values)

Number of rules in VLS class: $= 37 \times 26 \times 5 = 4810$

Total number of trading rules $= 497 + 2049 + 1220 + 2040 + 1040 + 3384 + 1128 + 3384 + 3384 + 1128 + 1872 + 26 + 26 + 4810 = 25988$

Appendix C: Tests without Data Snooping Bias

In this section, we discuss the two stepwise bootstrap tests (StepM and SSPA) mentioned in the previous section. However, we start with the Reality Check and SPA test, since StepM is a stepwise version of the former and the SSPA test is a stepwise version of the latter. Our notations are similar to Hansen (2005).

Reality check

White (2000) tests the null hypothesis that the benchmark is not inferior to any of the m alternative trading rules:

$$\mathbf{H_0} : \mu \leq \mathbf{0}, \tag{A.1}$$

where $\mu = E(\mathbf{d_t})$ and is a $m \times 1$ vector ($\mu \in \mathbb{R}^m$), while $\mathbf{d_t} = (d_{1,t}, \cdots, d_{m,t})$ is the $m \times 1$ vector of relative performance measures. Look at the vector of relative performance measure gives consideration to the full set of models underlying the vector that led to the best performing trading rule. Rejecting equation A.1 implies that at least one trading rule beats the benchmark. White proceeds to construct the reality check from the test statistics,

$$T_n \equiv max(n^{1/2}\overline{d}_1, \cdots, n^{1/2}\overline{d}_m), \tag{A.2}$$

$\overline{d}_m$ is the average performance of model m across n observations.

To calculate the p-value for the null hypothesis, White (2000) recommends the stationary bootstrap method of Politis and Romano (1994) to first obtain the empirical distribution of T_n^*:

$$T_n^*(b) = max[n^{1/2}(\overline{d}_1(b) - \overline{d}_1), \cdots, (n^{1/2}\overline{d}_m(b) - \overline{d}_m)], \qquad b = 1, \cdots, B \qquad \textbf{(A.3)}$$

The p-value is obtained by comparing T_n with the quantiles of the empirical distribution of $T_n^*(b)$. If the p-value is smaller than a given significance level, the null hypothesis is rejected.

SPA test

The null hypothesis of SPA test [Hansen (2005)] is the same as in White's Reality Check. Unlike the White's Reality Check, the SPA test uses the studentized test statistic, which will typically improve the power. Hansen (2005) provides a concrete example for highlighting the advantage of studentizing the individual statistics, since it avoids a comparison of the performance measured in different "units of standard deviation". Furthermore, the SPA test invokes a sample-dependent distribution under the null hypothesis, which can discard the poor models asymptotically. Therefore, the new test is more powerful and less sensitive to the inclusion of poor and irrelevant alternatives. The test statistic for SPA is given as follows:

$$T_n^{SPA} = max[max_{k=1,\cdots,m} \frac{n^{1/2}\overline{d}_k}{\hat{\omega}_k}, 0], \qquad \textbf{(A.4)}$$

where $\hat{\omega}_k^2$ is some consistent estimator of $\omega_k^2 = var(n^{1/2}\overline{d}_k)$. Then an re-centered estimator $\hat{\mu}^c$ for μ is chosen such that it conforms with the null hypothesis based on $N_m(\hat{\mu}^c, \hat{\Omega})$. The particular $\hat{\mu}^c$ suggested by Hansen (2005) is $\hat{\mu}_k^c = \overline{d}_k \mathbb{I}_{\{n^{1/2}\overline{d}_k/\hat{\omega}_k \leq -\sqrt{2loglogn}\}}$ for $k = 1, \cdots, m$, where $\mathbb{I}$ is the indicator function. It can be shown that for a poor trading rule $\mu_k < 0$, it has little effect on the distribution, and for a sufficiently large n, it will be discarded eventually. The improvement of the power of the SPA test over the Reality Check is further confirmed by the simulation experiment conducted in Hansen (2005). The p-value of T_n^{SPA} is calculated by bootstrapping its empirical distribution and then comparing the T_n^{SPA} with the quantiles of the empirical distribution of $T_n^{SPA*}(b, n)$.

That is:

$$\hat{p}^{SPA} \equiv \sum_{b=1}^{B} \frac{\mathbb{I}_{\{T_n^{SPA*}(b,n) > T_n^{SPA}\}}}{B} \tag{A.5}$$

Furthermore, Hansen (2005) defines another two p-values, which are not consistent, but can serve as the upper and lower bound for the consistent p-value. It imposes the null by recentering the bootstrap variables at $\hat{\mu}^l$ or $\hat{\mu}^u$, instead of $\hat{\mu}^c$. That is:

$$Z^*_{k,b,t} \equiv d^*_{k,b,t} - g_i(\bar{d}_k), \tag{A.6}$$

where $i = l, c, u$, $g_l(x) = max(0, x)$, $g_u(x) = x$, and $g_c(x) = x \times \mathbb{I}_{\{x \geq -\sqrt{(\hat{\omega}_k^2/n)2loglogn}\}}$. One can show that $E(Z^*_{k,b,t}) = \hat{\mu}^i$ for $i = l, c, u$.

StepM test

Both the Reality Check and the SPA test seek to answer whether the best trading strategy beats the benchmark. As discussed in section 3.1, it is often more interesting to identify all outperforming trading rules, or to know whether a particular trading rule improves upon the benchmark. The Reality Check can be modified easily for identifying potential strategies that beat the benchmark, but Romano and Wolf (2005) show that this is only suboptimal, and only amounts to the first step of StepM test of Romano and Wolf (2005), which can detect more good strategies from the second step on. The StepM test is therefore more powerful than the Reality Check in detecting superior trading rules. The aim of the StepM test is to find as many profitable trading rules as possible. Their null hypothesis is considered as the individual hypothesis test:

$$H_0^k : \mu_k \leq 0, \tag{A.7}$$

The individual decisions are made in a manner that asymptotically controls for the familywise error rate (FWE) at the significance level α. The FWE is defined as the probability of incorrectly identifying at least one strategy as superior. The joint confidence region is constructed to have a nominal joint coverage probability of $1 - \alpha$ in each stage, which is fulfilled by choosing a parameter c_i at the stage i. In the first stage it assumes the form of

$$[d_1 - c_1, \infty) \times \cdots \times [d_m - c_1, \infty) \tag{A.8}$$

If $0 \notin [d_k - c_1, \infty)$, then the k^{th} trading rule is detected as profitable. This first step can detect more than one profitable rule. For the second step, all the profitable rules found in the first step are dropped from the universe and only remaining rules are used to form the new universe. Forming a similar confidence interval as in (A.8), though replacing c_1 by c_2, one can detect the profitable trading rules, if any, from the remaining rules by checking whether or not the individual confidence interval $[d_k - c_2, \infty)$ contains zero. If no profitable rules are found, one should stop; otherwise, one continues this way until no profitable rules can be detected. Since the individual confidence interval in the joint confidence interval typically shrinks with the increasing number of testing steps, more profitable trading rules can be detected than when only relying on the first step.

Romano and Wolf (2005) also propose the use of studentization to improve the power and level properties of the StepM test.

Stepwise SPA test

The Stepwise SPA test of Hsu and Hsu (2006) combines the SPA test and the StepM test to improve the power. It uses the studentized test statistic as in SPA test, and impose a null to center the bootstrap variables around $\hat{\mu}^c$ as in the SPA test, such that poor and irrelevant trading rules will be discarded asymptotically. They provide formal proof and simulation results to demonstrate that the SSPA test is more powerful than the Reality Check, SPA test, and StepM test.

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 Andreas Schrimpf (Aarhus University). A revised version of the paper is forthcoming 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 3.4 examines the role of other yield curve information beyond the term spread. 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.

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.

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:

$$y_{t+k} = \beta_0 + \beta_1' z_t + \beta_2' x_t + \epsilon_{t+k}. \qquad \textbf{(A.1)}$$

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 (A.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 (1987b), 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.

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

$$\hat{\sigma}_1^2\text{-adj} \;=\; n^{-1}\sum_{t=m+1}^{T-k}(y_{t+k}-\hat{f}^{(1)}_{t,t+k})^2 - n^{-1}\sum_{t=m+1}^{T-k}(\hat{f}^{(0)}_{t,t+k}-\hat{f}^{(1)}_{t,t+k})^2, \qquad \text{(A.2)}$$

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 (A.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 $\hat{\sigma}_0^2 - \hat{\sigma}_1^2 - \text{adj}$) 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

$$\mathrm{Net}-\mathrm{SSE}(\tau_0,\tau_1)=\sum_{t=\tau_0}^{\tau_1}[(y_{t+k}-\hat{f}^{(0)}_{t,t+k})^2-(y_{t+k}-\hat{f}^{(1)}_{t,t+k})^2], \qquad \text{(A.3)}$$

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).

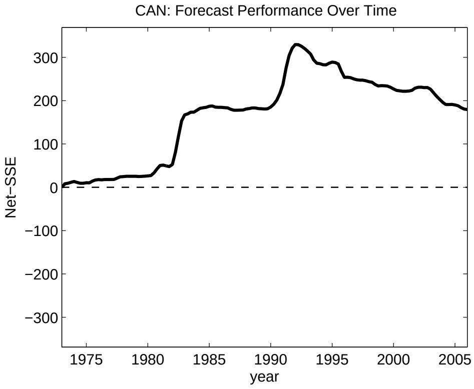
[11] In an extensive analysis for the US stock market, Goyal and Welch (2008) question the existence of stock return predictability based on their finding of poor OOS performance over time.

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.

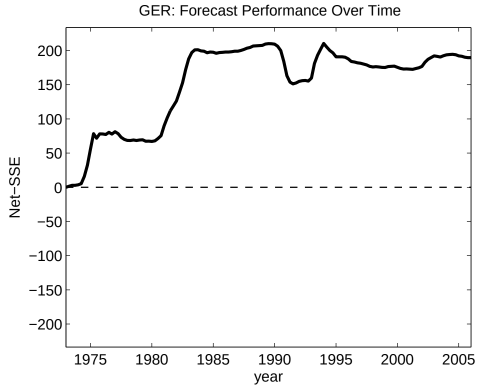
Figure 3.1 (based on a forecast horizon of 4 quarters) depicts how the OOS performance of prediction models using the term spread model evolves over time relative to the AR(1) benchmark. All four panels in Figure 3.1 indicate a strong time-variation of the forecast performance. More concretely, Canada, 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.

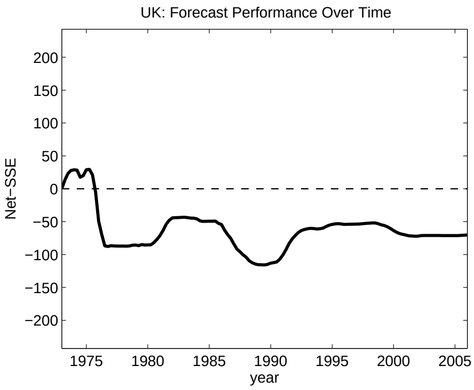
Figure 3.1: Time-varying Forecast Performance, Net-SSE, k=4



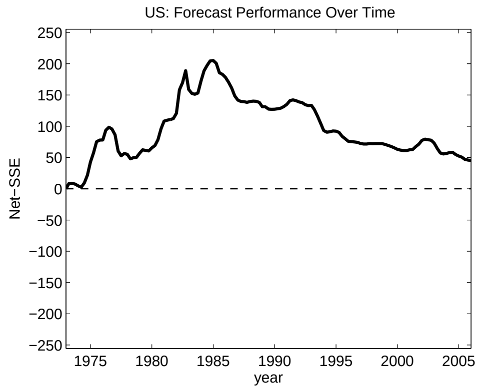
(a) Canada



(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.

Econometric Methods: Structural Break Tests and Window Selection for Forecasting

Predictive regressions for output growth using the yield spread as in Equation (A.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 (A.1) to vary across $r+1$ regimes, we consider a predictive regression of the following form:

$$y_{t+k} = \beta'_{0,j} + \beta'_{1,j} z_t + \beta'_{2,j} x_t + \epsilon_{t+k}. \qquad (t = T_{j-1}+1, ..., T_j) \tag{A.4}$$

where $j = 1, .., r+1$, and r is the number of breaks in the linear regression. Note that Equation (A.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

$$Dmax = max_{1 \leq r \leq R} \; a_r SupF_T(r). \tag{A.5}$$

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

$$\hat{f}_{t+k}^{weighted}(\underline{\omega}, \tilde{\omega}) = \frac{\sum_{w=1}^{t-\underline{\omega}-\tilde{\omega}} (\hat{f}_{t+k,w} MSFE(w|t, \tilde{\omega})^{-1})}{\sum_{w=1}^{t-\underline{\omega}-\tilde{\omega}} MSFE(w|t, \tilde{\omega})^{-1}}. \tag{A.6}$$

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).

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).

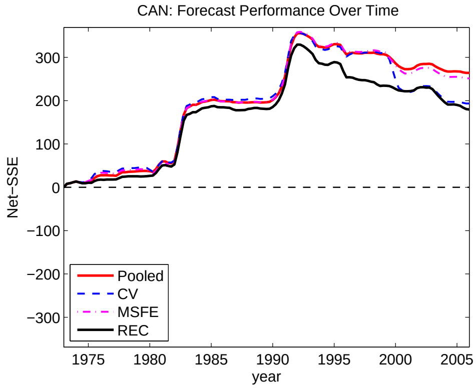
[21]This finding links to the general result in the forecasting literature that simple combination approaches often produce better forecasts compared to single forecasts or very sophisticated combination approaches (See e.g. Timmermann, 2006).

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

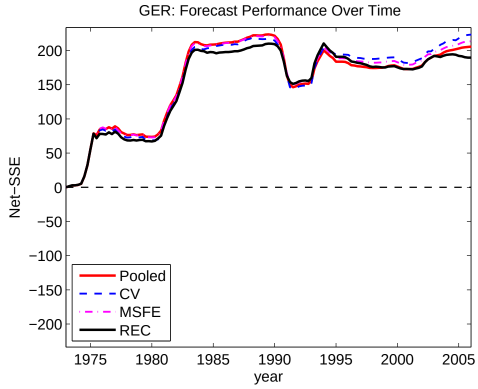
Panel A: CAN	**Recursive**		**MSFE-weighted**		**Cross-Validation**		**Pooled**	
Mean Forecast Error	-1.12	(0.25)	-0.55	(0.30)	-0.48	(0.35)	-0.66	(0.24)
Theil's U	0.87		0.81		0.86		0.80	
MSFE-adj	3.59***		4.37**		4.27***		4.25***	
Mincer-Zarnowitz: $\hat{a}$	-0.81	(0.81)	0.20	(0.81)	0.71	(0.87)	-0.17	(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.

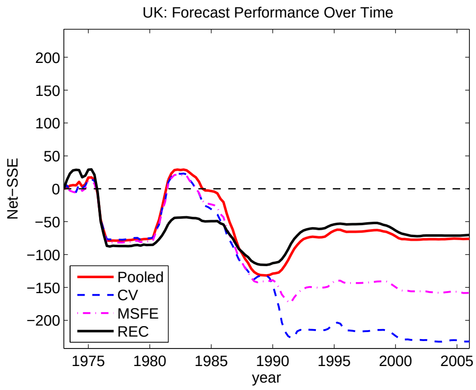
Figure 3.2: Time-varying Forecast Performance, Window Selection Methods



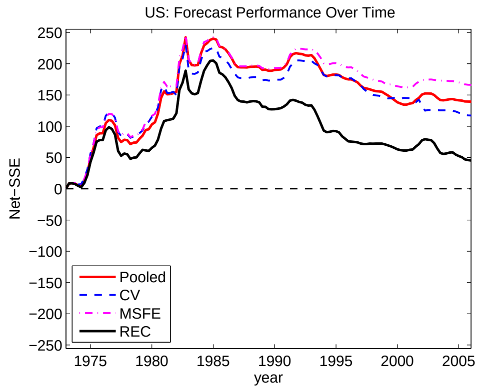
(a) Canada



(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 $f_t^{(0,1)}, \ldots, f_t^{(m-1,m)}$ is a significant predictor of (one-year) bond excess returns of bonds of all maturities ($n = 2, \ldots, m$):[25]

$$rx_{t+4}^{(n)} = \beta_0^{(n)} + \beta_1^{(n)} f_t^{(0,1)} + \ldots + \beta_m^{(n)} f_t^{(m-1,m)} + \epsilon_{t+4}^{(n)}, \tag{A.7}$$

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 ($n = 2, \ldots, m$) 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.

BIBLIOGRAPHY

ABREU, D., AND M. BRUNNERMEIER (2003): "Bubbles and crashes," *Econometrica*, pp. 173–204.

ANDREASSEN, P., AND S. KRAUS (1988): "Judgmental prediction by extrapolation," *unpublished paper, Department of Psychology, Harvard University*.

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.

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.

BAKER, M., AND J. WURGLER (2007): "Investor sentiment in the stock market," *Journal of Economic Perspectives*, 21(2), 129–151.

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.

BLACK, F. (1976): "Studies of stock price volatility changes," 177(181), 351–367.

BLANCHARD, O., AND M. WATSON (1982): *Crises in Economic and Financial Structure*chap. Bubbles, rational expectations, and financial markets, pp. 295–315. Lexington Books, Lexington, MA.

BROCK, W., J. LAKONISHOK, AND B. LEBARON (1992): "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," *Journal of Finance*, 47(5), 1731–64.

BROWN, G., AND M. CLIFF (2005): "Investor Sentiment and Asset Valuation," *The Journal of Business*, 78(2), 405–440.

BRUNNERMEIER, M., S. NAGEL, AND L. PEDERSEN (2008): "Carry trades and currency crashes," *NBER Macroeconomics Annual*.

BURNSIDE, C., M. EICHENBAUM, AND S. REBELO (2007): "The Returns to Currency Speculation in Emerging Markets," *American Economic Review Papers and Proceedings*, 97(2)(12916), 333–338.

CAMPBELL, J., AND L. HENTSCHEL (1992): "No news is good news: an asymmetric model of changing volatility in stock returns," *Journal of Financial Economics*, 31, 281–318.

CHANG, P. H. K., AND C. L. OSLER (1999): "Methodical Madness: Technical Analysis and the Irrationality of Exchange-Rate Forecasts," *Economic Journal*, 109(458), 636–61.

CHAROENROOK, A., AND H. DAOUK (2008): "Conditional Skewness of Aggregate Market Returns," .

CHEN, J., H. HONG, AND J. STEIN (2001): "Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices," *Journal of Financial Economics*, 61(3), 345–381.

CHRISTIE, A. A. (1982): "The stochastic behavior of common stock variances: Value, leverage and interest rate effects," *Journal of Financial Economics*, 10(4), 407–432.

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

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

DE LONG, J., A. SHLEIFER, L. SUMMERS, AND R. WALDMANN (1990): "Noise trader risk in financial markets," *The Journal of Political Economy*, 98(4), 703–738.

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

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.

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

FRANKEL, J., AND K. FROOT (1988): “Explaining the demand for dollars: International rates of return and the expectations of chartists and fundamentalists,” *Agriculture, Macroeconomics and the Exchange Rate*.

FRENCH, K., G. SCHWERT, AND R. STAMBAUGH (1987): “Expected stock returns and volatility,” *Journal of Financial Economics*, 19, 3–29.

GEHRIG, T., AND L. MENKHOFF (2004): “The use of flow analysis in foreign exchange: exploratory evidence,” *Journal of International Money and Finance*, 23(4), 573–594.

GENE SAVIN, P. W., AND J. ZVINGELIS (2007): “The Predictive Power of "Head-and-Shoulders" Price Patterns in the U.S. Stock Market,” *Journal of Financial Econometrics*, 5, 243 – 265.

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.

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

GRIFFIN, J., J. HARRIS, AND S. TOPALOGLU (2006): “Who Drove and Burst the Tech Bubble?,” *International Center for Finance at Yale School of Management, Research Paper Series*.

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., 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, P. R. (2005): “A Test for Superior Predictive Ability,” *Journal of Business & Economic Statistics*, 23, 365–380.

HARUVY, E., Y. LAHAV, AND C. NOUSSAIR (2007): “Traders’expectations in asset markets: experimental evidence,” *American Economic Review*, 97(5), 1901–1920.

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

HOMMES, C., J. SONNEMANS, J. TUINSTRA, AND H. VAN DE VELDEN (2005): “Coordination of expectations in asset pricing experiments,” *Review of Financial Studies*, 18(3), 955–980.

HONG, H., AND J. STEIN (2003): “Differences of Opinion, Short-Sales Constraints, and Market Crashes,” *Review of Financial Studies*, 16(2), 487–525.

HÄRDLE, W. (1990): *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, U.K.

HSU, P.-H., AND Y.-C. HSU (2006): “A Stepwise SPA Test for Data Snooping and its Application on Fund Performance Evaluation,” .

HSU, P.-H., AND C.-M. KUAN (2005): “Reexamining the Profitability of Technical Analysis with Data Snooping Checks,” *Journal of Financial Econometrics*, 3(4), 606–628.

HUENG, C., AND J. MCDONALD (2005): “Forecasting asymmetries in aggregate stock market returns: Evidence from conditional skewness,” *Journal of Empirical Finance*, 12(5), 666–685.

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.

JEGADEESH, N. (2000): "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation: Discussion," *The Journal of Finance*, 55(4), 1765–1770.

JENSEN, M. (1967): "Random Walks: reality or myth – comment," *Financial Analysts Journal*, 23, 77–85.

JENSEN, M. C., AND G. A. BENNINGTON (1970): "Random Walks and Technical Theories: Some Additional Evidence," *Journal of Finance*, 25(2), 469–482.

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

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

LEBARON, B. (1999): "Technical trading rule profitability and foreign exchange intervention," *Journal of International Economics*, 49, 125–143.

LEE, C., A. SHLEIFER, AND R. THALER (1991): "Investor sentiment and the closed-end fund puzzle," *Journal of Finance*, 46(1), 75–109.

LEE, C. I., K. C. GLEASON, AND I. MATHUR (2001): "Trading rule profits in Latin American currency spot rates," *International Review of Financial Analysis*, 10(2), 135–156.

LEE, C. I., AND I. MATHUR (1996): "Trading rule profits in european currency spot cross-rates," *Journal of Banking & Finance*, 20(5), 949–962.

LEMMON, M., AND E. PORTNIAGUINA (2006): "Consumer confidence and asset prices: Some empirical evidence," *Review of Financial Studies*, 19(4), 1499.

LEVICH, R. M., AND L. I. THOMAS (1993): "The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach," *Journal of International Money and Finance*, 12(5), 451–474.

LO, A. W., AND A. C. MACKINLAY (1990): "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3(3), 431–67.

LO, A. W., H. MAMAYSKY, AND J. WANG (2000): "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation," *Journal of Finance*, 55, 1705–1765.

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

MARTIN, A. D. (2001): "Technical trading rules in the spot foreign exchange markets of developing countries," *Journal of Multinational Financial Management*, 11(1), 59–68.

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

NAGEL, S. (2004): "Trading rules and trading volume," *working paper, Harvard University*.

NEELY, C. J. (1997): "Technical analysis in the foreign exchange market: a layman's guide," *Review*, (Sep), 23–38.

NEELY, C. J., AND P. A. WELLER (2003): "Intraday technical trading in the foreign exchange market," *Journal of International Money and Finance*, 22(2), 223–237.

NEELY, C. J., P. A. WELLER, AND J. M. ULRICH (forthcoming): "The adaptive markets hypothesis: evidence from the foreign exchange market," *Journal of Financial and Quantitative Analysis*, (2006-046).

NEILL, H. (1997): *The art of contrary thinking*. Caxton Press.

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

——— (1987b): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

OSLER, C. (2000): "Support for resistance: technical analysis and intraday exchange rates," *Economic Policy Review*, (Jul), 53–68.

PARK, C., AND S. IRWIN (2007): "What do we know about the profitability of technical analysis?," *Journal of Economic Surveys*, 21(4), 786–826.

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

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

PINDYCK, R. (1984): "Risk, inflation, and the stock market," *The American Economic Review*, pp. 335–351.

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

POJARLIEV, M. (2005): "Performance of Currency Trading Strategies in Developed and Emerging Markets: Some Striking Differences," *Financial Markets and Portfolio Management*, 19(3), 297–311.

POLITIS, D. N., AND J. P. ROMANO (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

QI, M., AND Y. WU (2006): "Technical Trading-Rule Profitability, Data Snooping, and Reality Check: Evidence from the Foreign Exchange Market," *Journal of Money, Credit and Banking*, 30, 2135–2158.

QIU, L., AND I. WELCH (2006): "Investor sentiment measures," *SSRN: http://ssrn.com/abstract=589641*.

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

ROMANO, J. P., AND M. WOLF (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73(4), 1237–1282.

SHILLER, R. (2005): *Irrational Exuberance*. Princeton University Press.

SHLEIFER, A., AND L. SUMMERS (1990): "The noise trader approach to finance," *The Journal of Economic Perspectives*, pp. 19–33.

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.

SULLIVAN, R., A. TIMMERMANN, AND H. WHITE (1999): "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap," *Journal of Finance*, 54(5), 1647–1691.

SWEENEY, R. J. (1986): "Beating the Foreign Exchange Market," *Journal of Finance*, 41(1), 163–82.

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

TAYLOR, M. P., AND H. ALLEN (1992): "The use of technical analysis in the foreign exchange market," *Journal of International Money and Finance*, 11(3), 304–314.

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

WHITE, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68(5), 1097–1126.

WOLF, M. (2007): "Robust Performance Hypothesis Testing With the Sharpe Ratio," *ZEW - Center for European Economic Research Paper No. 320*.

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

XU, J. (2007): "Price Convexity and Skewness," *The Journal of Finance*, 62(5), 2521–2552.

Qingwei Wang

March 2010

Centre for European Economic Research
L 7, 1, D-68161 Mannheim
Germany
E-mail: wang@zew.de

Office: +49 (0)621 1235-233
Phone: +49 (0)179 7430-333
Fax: +49 (0)621 1235-223

Education

Ph.D. candidate in Finance, University of Mannheim, 2006-present
Dissertation topic: “Investor Beliefs and Forecast Evaluation”
Supervisor: Professor Ernst Maug, Ph.D.

M.A. in Economics and Management Science, Humboldt University at Berlin, 2005
LL.B. (with distinction), Fudan University, 2001
B.A. in Economics, Anhui Normal University, 1999

Working Experiences

Researcher, Centre for European Economic Research (ZEW), 2005 – 2010
Research assistant for Prof. Bernd Fitzenberger, Goethe University, 2005
Research assistant for Prof. Wolfgang Härdle, Humboldt University at Berlin, 2003 – 2005
Research assistant for Prof. Bengt-Arne Wickström, Humboldt University at Berlin, 2003 – 2005

Publications

“A Reappraisal of the Leading Indicator Properties of the Yield Curve under Structural Instability” (with Andreas Schrimpf), forthcoming in ***International Journal of Forecasting***

“Union Membership in Germany: Determinants, Densities and Decompositions” (with Bernd Fitzenberger and Karsten Kohn), accepted in ***Journal of Population Economics***

“The Economic Impact of Olympic Games: Evidence from Stock Markets” (with Christian Dick), to appear in ***Applied Economics Letters***

“Alternative Methods for a Forward-Looking Assessment of Potential GDP Growth”, 2009 chapter in ***Projecting Potential Output: Methods and Problems*** (with Andreas Schrimpf), 141-160, Physica-Verlag.