An Analysis of Value Investing Determinants under the Behavioural Finance Approach

by

René-Christian Kumsta

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of

Doctor of Philosophy of Loughborough University

September 2015

Certificate of Originality
Thesis Access Conditions and Deposit Agreement

Students should consult the guidance notes on the electronic thesis deposit and the access conditions in the University's Code of Practice on Research Degree Programmes

Author: René-Christian Kumsta…………………………………………………………..…….

Title: An Analysis of Value Investing Determinants under the Behavioural Finance Approach………..

I, René-Christian Kumsta, Straßbergerstr. 11, 80809 Munich, Germany, "the Depositor",
would like to deposit **An Analysis of Value Investing Determinants under the Behavioural Finance Approach**, hereafter referred to as the "Work", once it has successfully been examined in Loughborough University Institutional Repository

Status of access OPEN / ~~RESTRICTED~~ / ~~CONFIDENTIAL~~

Moratorium Period…………………………………years, ending…………../…………20……………………….

Status of access approved by (CAPITALS):…………………………………………………………………

Supervisor (Signature)……………………………………………...…………………………………...

School of…………………………………………………………………...…………………………………

Author's Declaration *I confirm the following :*

CERTIFICATE OF ORIGINALITY
This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgements or in footnotes, and that neither the thesis nor the original work therein has been submitted to this or any other institution for a degree

NON-EXCLUSIVE RIGHTS
The licence rights granted to Loughborough University Institutional Repository through this agreement are entirely non-exclusive and royalty free. I am free to publish the Work in its present version or future versions elsewhere. I agree that Loughborough University Institutional Repository administrators or any third party with whom Loughborough University Institutional Repository has an agreement to do so may, without changing content, convert the Work to any medium or format for the purpose of future preservation and accessibility.

DEPOSIT IN LOUGHBOROUGH UNIVERSITY INSTITUTIONAL REPOSITORY
I understand that open access work deposited in Loughborough University Institutional Repository will be accessible to a wide variety of people and institutions - including automated agents - via the World Wide Web. An electronic copy of my thesis may also be included in the British Library Electronic Theses On-line System (EThOS). I understand that once the Work is deposited, a citation to the Work will always remain visible. Removal of the Work can be made after discussion with Loughborough University Institutional Repository, who shall make best efforts to ensure removal of the Work from any third party with whom Loughborough University Institutional Repository has an agreement. Restricted or Confidential access material will not be available on the World Wide Web until the moratorium period has expired.

- That I am the author of the Work and have the authority to make this agreement and to hereby give Loughborough University Institutional Repository administrators the right to make available the Work in the way described above.

user's declaration: for signature during any Moratorium period (Not Open work): ***I undertake to uphold the above conditions:***			
Date	Name (CAPITALS)	Signature	Address

Abstract

WHAT WAS DONE? This study researches the success of several value investment strategies in the stock markets of the United Kingdom and Germany based on nine firm fundamentals that are extracted from listed firms' annual financial statements. In this regard, we first examine alternative forecast combination methods in a novel way to utilise fully the financial information at hand. Second, we examine the drivers of investment returns, particularly the role of information uncertainty, for which a new direct measure is developed. Finally, we evaluate the performance of these financial health investment strategies in alternative institutional environments by focusing on the differences between the two markets regarding both their corporate culture and their legal environment.

WHY WAS IT DONE? Similar to economics, the discipline of finance is a social science because its observations emanate from economic transactions between humans. Nevertheless, a significant part of the research in this area is undertaken by means that are almost exclusively applied to the natural sciences, such as mathematics or physics. Although the reasons seem manifold, an increased form of scientificity, in conjunction with greater credibility of the research process and results, is deemed to be of primary importance. However, the benchmark for evaluating these research outcomes differs from those used in the natural sciences. From the example of the efficient market hypothesis one can see that alternative research results that cast serious doubt upon efficiency per se are disregarded as aberrations, leading to the assumption that the hypothesis in its entirety is more or less valid. This study assumes that inefficiencies in the stock market do exist for prolonged periods of time and investors are actually able to benefit from them.

HOW WAS IT DONE? Secondary financial statement data of listed companies in the United Kingdom and Germany were downloaded from Datastream for the period between 1992 and 2010. A quantitative analysis of the significance of the correlation between groups of firms with similar financial characteristics and their one-year-ahead stock returns was subsequently performed. Various combination methods for differential weighting of individual financial statement items were conducted. The aim was to increase the profitability of the investment strategy.

WHAT WAS FOUND? In general, a classification of stocks according to certain internal criteria of financial health is capable of separating future winners from losers and at the same time confirms the results of a previous US study. More specifically, we first show that a wide range of combination methods generate profitable investment strategies whereby especially measures of profitability are the central indicator of a firm's future performance. Secondly, the more complex methods neither consistently nor substantively outperform the simpler methods. Thirdly, information uncertainty does not seem to be the prime driver of the profitability of an investment strategy. Lastly, we show that financial health investment strategies are profitable both in market-oriented, common law settings and in bank-oriented, code law settings.

Keywords: Capital markets; market efficiency; behavioural finance; financial statement analysis; valuation; UK stock market; German stock market; value investing; information uncertainty; liquidity.

Acknowledgments

I would like to express my deepest gratitude to my supervisors, Dr Andrew Vivian and Dr Ali Ataullah, for their continuous advice, time, dedication, patience and encouragement during the course of my PhD. I am truly honoured to have had both of them with me as distinguished companions on this academic journey.

I would also like to thank my internal examiner, Professor Mark Freeman, for his valuable suggestions and comments which helped me to improve the quality of my research.

I would like to give a thank you to my friends and colleagues within Loughborough's PhD community who made this journey a once in a lifetime never to be missed experience. A special thank you goes to I-Chen, Andrei, Hitesh and João.

I am also indebted and grateful to the School of Business and Economics at Loughborough University for having funded my PhD. I would also like to thank the staff at the University for their support in various ways.

Finally, I would like to give a big thank you to my family for their kindness, openness and encouragement. Without you, this journey would have been impossible!

Table of Contents

THESIS ACCESS CONDITIONS AND DEPOSIT AGREEMENT ERROR! BOOKMARK NOT DEFINED.

ABSTRACT .... **I**

ACKNOWLEDGMENTS .... **III**

TABLE OF CONTENTS .... **IV**

LIST OF TABLES .... **VIII**

LIST OF FIGURES .... **X**

LIST OF ABBREVIATIONS .... **XI**

CHAPTER 1 INTRODUCTION **1**

1.1 INTRODUCTION 1
1.2 MOTIVATION OF THE STUDY 1
1.3 RESEARCH OBJECTIVES AND CONTRIBUTION OF THE STUDY 3
1.4 THE STRUCTURE OF THE THESIS 5

CHAPTER 2 A REVIEW OF THE LITERATURE **8**

2.1 INTRODUCTION 8
2.2 THEORETICAL FOUNDATIONS OF MODERN FINANCE 9
2.2.1 The evolution of asset pricing models 9
2.2.2 The efficient market hypothesis (EMH) 10
2.2.3 Problems with the efficient market hypothesis 11
2.2.4 Arguments in favour of the efficient market hypothesis 18
2.3 BEHAVIOURAL FINANCE 22
2.3.1 Problems with the behavioural finance approach 25
2.4 PRACTICAL EXPLANATIONS OF ANOMALIES AND THEIR PERSISTENCE 27
2.5 SUMMARY 28
2.6 MARKET ANOMALIES AND THE BEHAVIOURAL FINANCE APPROACH 28
2.6.1 Links to psychology 28
2.6.2 Information processing 29
2.6.3 Heuristic simplification 30
2.6.4 Self-deception 31
2.6.5 Risk and uncertainty 31
2.6.6 Social interactions 32
2.6.7 Modelling alternatives to expected utility and to Bayesian updating 32
2.6.8 Summary 33
2.7 REVIEW OF EMPIRICAL RESEARCH ON FUNDAMENTALS-BASED INVESTING 33
2.7.1 A preliminary note 34
2.7.2 The US analysis 34
2.7.3 The international analysis 36
2.7.4 A simple, fundamentals-based investment strategy 37
2.7.5 Gaps in the literature 39
2.8 CONCLUSIONS 40

CHAPTER 3 METHODOLOGY AND METHODS **42**

3.1 INTRODUCTION 42
3.2 PHILOSOPHICAL PARADIGMS: POSITIVISM AND POSTPOSITIVISM 42

3.2.1 Methodology in the finance literature ... *44*
3.3 The UK stock market ... 45
3.4 The German stock market ... 46
3.5 Research methods ... 46
3.5.1 Replication and extension ... *47*
3.5.1.1 Firm selection ... 47
3.5.1.2 F-Score calculation ... 48
3.5.1.3 Benchmark definition and calculation of returns ... 50
3.5.1.4 Portfolio construction with the original F-Score ... 52
3.5.1.5 Portfolio construction with an alternative to the F-Score ... 55
3.5.1.6 Alternative combination measures of financial strength: The weighted DMSFE ... 56
3.5.1.7 Alternative combination measures of financial strength: Clusters ... 59
3.5.1.8 Inclusion of a second dimension: The book-to-market ratio ... 60
3.5.2 Driving forces behind stock returns ... *61*
3.5.2.1 Information uncertainty proxy: Size ... 61
3.5.2.2 Information uncertainty proxy: The SDF and MADF ... 62
3.5.2.3 Causality of investment performance: The Amihud liquidity measure ... 64
3.5.2.4 Causality of investment performance: The ToR liquidity measure ... 65
3.5.3 F-Score and F-rank: The case of Germany ... *67*
3.6 Empirical tests ... 69
3.6.1 The rationale behind the choice of statistical methods ... *69*
3.6.2 Visualisation of the test methods ... *70*
3.6.3 Parametric and non-parametric tests ... *71*
3.7 Conclusions ... 72

CHAPTER 4 EMPIRICAL ANALYSIS AND RESULTS – THE UK ... 73

4.1 Introduction ... 73
4.1.1 Hypothesis development ... *74*
4.2 An overview of the UK data set ... 77
4.2.1 Descriptive statistics of UK firms ... *79*
4.2.2 The correlation matrix and basic F-Score strategy ... *81*
4.2.3 One-way portfolios with the original F-Score ... *87*
4.2.4 One-way portfolios with an alternative: The F-rank approach ... *88*
4.3 The first set of alternative combinations of F-ranks: An overview ... 90
4.3.1 Alternative combinations of F-ranks: The weighted DMSFE ... *90*
4.3.2 Discussion and links to the literature: DMSFE ... *93*
4.4 The second set of alternative combinations of F-ranks: An overview ... 95
4.4.1 Alternative combinations of F-ranks: Clustering (k = 2) ... *95*
4.4.2 Alternative combinations of F-ranks: Clustering (k = 3) ... *96*
4.4.3 Discussion and links to the literature: Clustering ... *97*
4.5 Empirical results in the two-way environment ... 98
4.5.1 The F-Score and the book-to-market ratio ... *98*
4.5.2 The F-rank and the book-to-market ratio: The mean F-rank ... *100*
4.5.3 The F-rank and the book-to-market ratio: The median F-rank ... *102*
4.6 Two-way alternative F-rank combinations: An overview ... 104
4.6.1 Two-way alternative F-rank combinations: The weighted DMSFE ... *104*
4.6.2 Two-way alternative combination measures: Clustering ... *106*
4.7 Relation and relevance to the empirical finance literature ... 108
4.8 Chapter summary ... 111

CHAPTER 5 INFORMATION UNCERTAINTY AND LIQUIDITY – THE UK ... 115

5.1 Introduction ... 115
5.2 Hypothesis development: Firm size as an indirect information uncertainty proxy ... 117
5.2.1 Discussion of the results and links to the literature ... *119*
5.3 Size and alternative combinations of the F-rank: An overview ... 124

5.3.1 The weighted DMSFE in the UK *124*
5.3.2 Clusters *126*
5.4 MEASURES OF FUNDAMENTAL VALUE VOLATILITY: AN OVERVIEW 128
5.4.1 Hypothesis development: Direct information uncertainty measures *128*
5.4.2 Direct measures of fundamental value volatility: The SDF *130*
5.4.3 Direct measures of fundamental value volatility: MADF *133*
5.5 SIZE, SDF AND MADF: A DISCUSSION AND LINKS TO THE LITERATURE 135
5.6 LIQUIDITY AND STOCK RETURNS 140
5.6.1 Hypothesis development: Liquidity as a key driver of investment performance *140*
5.6.2 The first liquidity proxy: The Amihud measure *143*
5.6.2 The second liquidity proxy: The ToR measure *147*
5.7 DISCUSSION OF THE RESEARCH FINDINGS AND LINKS TO THE LITERATURE 150
5.8 CHAPTER SUMMARY 154

CHAPTER 6 EMPIRICAL ANALYSIS AND RESULTS – GERMANY 157

6.1 INTRODUCTION 157
6.2 CHARACTERISTICS OF THE GERMAN ECONOMY 158
6.2.1 The German Mittelstand *159*
6.2.2 Investor protection in common law and code law countries *161*
6.2.3 Hypothesis development *162*
6.3 AN OVERVIEW OF THE GERMAN DATA SET 166
6.3.1 Descriptive statistics of German firms *166*
6.3.2 Correlation matrix and basic F-Score strategy in Germany *170*
6.3.3 Comparison of correlation matrices between Germany, the UK and the US *171*
6.3.4 Analysis of one- and two-year buy-and-hold returns in Germany *175*
6.3.5 One-way portfolios with the original F-Score *178*
6.3.6 One-way portfolios with an alternative: The F-rank approach *179*
6.4 THE FIRST ALTERNATIVE COMBINATION MEASURES FOR THE F-RANK: AN OVERVIEW 181
6.4.2 Combination measures for the F-rank: The weighted DMSFE *181*
6.5 THE SECOND ALTERNATIVE COMBINATION MEASURES FOR THE F-RANK: AN OVERVIEW 183
6.5.1 Combination measures for the F-rank: Clustering with k = 2 and k = 3 .... *183*
6.5.2 A comparison between one-way strategies in Germany and the UK *184*
6.6 EMPIRICAL RESULTS IN THE TWO-WAY ENVIRONMENT 188
6.6.1 The F-Score and the book-to-market ratio *188*
6.6.2 The F-rank and the book-to-market ratio *189*
6.7 TWO-WAY ALTERNATIVE COMBINATION MEASURES: AN OVERVIEW 191
6.7.1 Two-way alternative combination measures: The weighted DMSFE *191*
6.7.2 Two-way alternative combination measures: Clustering *192*
6.7.3 Summary and comparison of the results *193*
6.8 SIZE AS AN INFORMATION UNCERTAINTY PROXY: OVERVIEW 196
6.8.1 Size, the F-Score/F-rank and combinations of the F-rank *196*
6.8.2 Size and combinations of DMSFE *198*
6.8.3 Size and cluster combinations *199*
6.9 ALTERNATIVE IU MEASURES: THE SDF AND MDF 200
6.9.1 The SDF and the German stock market *201*
6.9.2 The MADF and the German stock market *203*
6.9.3 Summary and comparison of IU strategies in Germany *204*
6.10 CHAPTER SUMMARY AND CONCLUSION 207

CHAPTER 7 CONCLUSION AND LIMITATIONS 211

7.1 INTRODUCTION 211
7.2 RESEARCH SCOPE 212
7.3 RESEARCH DESIGN, IMPLEMENTATION AND BOUNDARIES 214

7.4 Research questions and factual conclusions 215
7.5 Conceptual conclusions and the contribution to knowledge 217
7.6 Limitations of the study 223
7.7 Recommendations for future research 223

APPENDIX 225

BIBLIOGRAPHY XIV

List of Tables

Table 2.1 Comparison of psychology and economics 33
Table 3.1 Basic Beliefs of Alternative Inquiry Paradigms 43
Table 3.2 Datastream items and description 48
Table 3.3 F-Score computation 49
Table 3.4 Firm years of the F-Score 54
Table 3.5 Parametric (p) and non-parametric (np) tests on portfolio returns 71
Table 4.1 Firms with complete F-Scores 78
Table 4.2 Financial characteristics of UK firms 80
Table 4.3 Comparison of buy-and-hold returns between UK and US firms 81
Table 4.4 Spearman correlation matrix between stock returns, F-components and F-Score[a] 83
Table 4.5 One-year and two-year buy-and-hold returns of F-Score portfolios 85
Table 4.6 One-way F-Score portfolios 88
Table 4.7 One-way F-rank portfolios 89
Table 4.8 An overview of DMSFE portfolios by year 91
Table 4.9 One-way weighted DMSFE portfolios 92
Table 4.10 One-way clustered portfolios ($k = 2$) 95
Table 4.11 One-way clustered portfolios ($k = 3$) 96
Table 4.12 Two-way F-Score portfolios 99
Table 4.13 Two-way mean F-rank portfolios 101
Table 4.14 Two-way median F-rank portfolios 103
Table 4.15 Two-way weighted DMSFE portfolios 105
Table 4.16 Two-way clustered portfolios 107
Table 5.1 Two-way F-Score/F-rank-size portfolios 120
Table 5.2 Two-way weighted DMSFE–size portfolios 125
Table 5.3 Two-way clustered-size portfolios 126
Table 5.4 Two-way SDF portfolios 130
Table 5.5 Two-way MADF portfolios 134
Table 5.6 Two-way Amihud portfolios 144
Table 5.7 Two-way ToR portfolios 148
Table 6.1 Financial characteristics of German firms 167
Table 6.2 Comparison of financial characteristics 168
Table 6.3 Comparison of buy-and-hold returns between German, UK and US firms 170
Table 6.4 Spearman correlation matrix between stock returns, F-components and F-Score[a] 173
Table 6.5 Correlation between F-components and one-year winsorised returns 174
Table 6.6 One-year and two-year buy-and-hold returns of F-Score portfolios[a] 177
Table 6.7 One-way F-Score portfolios 179
Table 6.8 One-way F-rank portfolios 180
Table 6.9 Overview of DMSFE portfolios by year 182
Table 6.10 One-way weighted DMSFE portfolios 182
Table 6.11 One-way clustered portfolios ($k = 2$ and $k = 3$) 184
Table 6.12 One-way strategies in Germany and the UK 185
Table 6.13 Test results of return differences between Germany and the UK 187
Table 6.14 Two-way F-Score portfolios 189
Table 6.15 Two-way mean/median F-rank portfolios 190

Table 6.16 Two-way weighted DMSFE portfolios 192
Table 6.17 Two-way clustered portfolios 193
Table 6.18 Comparison of two-way strategies in Germany and the UK (horizontally) 195
Table 6.19 Comparison of two-way strategies in Germany and the UK (vertically) 195
Table 6.20 Two-way F-Score/F-rank-size portfolios 197
Table 6.21 Two-way weighted DMSFE portfolios 198
Table 6.22 Two-way clustered-size portfolios 199
Table 6.23 Two-way SDF portfolios 201
Table 6.24 Two-way MADF portfolios 204
Table 6.25 Comparison of two-way strategies in Germany and the UK (horizontally) 205
Table 6.26 Comparison of two-way strategies in Germany and the UK (vertically) 206

List of Figures

Figure 3.1 Visualisation of F-Score construction 52
Figure 3.2 Illustration of $DMSFE_i$ computation 57
Figure 3.3 Visualisation of DMSFE portfolio construction 58
Figure 3.4 Visualisation of cluster portfolio construction 60
Figure 3.5 Visualisation of the statistical analyses 70
Figure 4.1 Histogram of firms with complete F-Scores 78

List of Abbreviations

ADR	American Depository Receipt
AIM	Alternative Investment Market
APT	Arbitrage Pricing Theory
ARCH	Autoregressive Conditional Heteroscedasticity
BHAR	Buy-and-Hold Abnormal Returns
BM	Book-to-market ratio
BMW	Bayerische Motoren Werke
CAPM	Capital Asset Pricing Model
CAR	Cumulative Abnormal Return
CFO	Chief Financial Officer
cp.	Compare
DAX	Deutscher Aktienindex
DE	Germany
DMSFE	Discounted Mean Square Forecast Error
DS	Datastream
e.g.	exempli gratia
EMH	Efficient Market Hypothesis
EP	Earnings-to-price ratio
ES	Entry Standard
ETF	Exchange Traded Fund
EU	European Union
FCA	Financial Conduct Authority
GS	General Standard
GBP	British pound
HGB	Handelsgesetzbuch

IAS	International Accounting Standards
IfM	Institut für Mittelstandsforschung
IFRS	International Financial Reporting Standards
IPO	Initial Public Offering
IU	Information Uncertainty
LSE	London Stock Exchange
LTCM	Long-Term Capital Management
MADF	Mean Absolute Deviation of F-rank
MD	Mean Deviation
MPT	Modern Portfolio Theory
MM	Main Market
MWW	Mann-Whitney-Wilcoxon
NASDAQ	National Association of Securities Dealers Automated Quotations
NYSE	New York Stock Exchange
OOS	Out-Of-Sample
PS	Prime Standard
PSM	Professional Securities Market
S&P500	Standard & Poor's 500
SD	Standard Deviation
SDF	Standard Deviation of F-rank
SFM	Specialist Fund Market
SOX	Sarbanes-Oxley Act
ToR	Turnover Ratio
UK	United Kingdom
US	United States
US-GAAP	United States General Accepted Accounting Principles
VW	Volkswagen

Winsd.	Winsorised

Chapter 1 Introduction

1.1 Introduction

The first chapter of the thesis presents an overview of the rationale for undertaking this research, which analyses the performance of financial statement-based investment strategies in the stock markets of the United Kingdom and Germany. It also highlights the research objectives and how the study aims to contribute to the literature by answering the relevant research questions. Finally, a brief summary of the chapters in this thesis is provided.

1.2 Motivation of the study

The basic principle of a fundamentals-based investment strategy is to provide investors with an easily implementable tool that generates reliable stock returns both consistently and reliably. However, in academia, this very principle is controversial because, inter alia, it assumes that market participants have different perceptions of financial statement information and the way in which this information is actually incorporated into stock prices. This lack of unity amongst investors should lead to profit opportunities that can be benefited from with an adequate investment strategy.

Naturally, two conditions need to be assumed. On the one hand, stock markets are required to be actually informationally inefficient, which means that stock prices do not move randomly and can therefore be forecasted. This stands in contrast to the so-called efficient market hypothesis (EMH), which has been the dominating tenet in modern finance for more than three decades (Lo, 2004). According to Fama (1970), prices fully reflect all the available information. Ever since its inception, however, the EMH has received considerable criticism from various quarters. For instance, Grossman and Stiglitz (1980) excluded the possibility of efficient markets on logical grounds. According to their argument, if all the information is already priced in, investors would stop the costly activity of collecting information and trading on it, with the consequence that markets would cease to exist. As a result of this academic debate, an alternative school of thought has emerged more recently that relaxes some of the strict assumptions made by the EMH. Adherents to the so-called school of behavioural finance believe that investors should be viewed in the financial markets as individual rather than aggregate actors, who do not always exhibit rational behaviour during economic decision making but are rather subject to their feelings and emotions. Consequently, their actions might not always be predictable, which would then lead to temporary inefficiencies in stock markets. Due to its mathematical approach to understanding financial markets, that is, prices follow martingales (Lo, 2004), the EMH has been criticised by behavioural finance advocates for being the outcome of so-called “physics envy”. Bennis and O’Toole’s (2005) critique was that the social sciences, such as economics and finance, borrow the scientific approach from the hard sciences, such as physics or mathematics, and that research outcomes are hardly evaluated regarding their practical use. Behavioural finance is not free from similar criticism. There is considerable doubt in the academic literature whether the test findings of psychological experiments, which are regarded as one of the foundations of behavioural finance, are as easily transferable to financial markets as their followers assume (Barberis and Thaler, 2005).

On the other hand, even in the case in which the above-mentioned issues have been addressed, the questions of which one of a plethora of investment strategies is most suitable for benefiting from market irregularities and on what grounds remain unanswered. As summarised by Ke and Ramalingegowda (2005), market participants can be classified generally as transient, dedicated or quasi-indexing investors, the transient ones being characterised by active trading whereas the quasi-indexing ones aim for a low portfolio turnover. Each of these categories, in turn, requires different approaches to or strategies for investing. However, the overarching aim to realise adequate returns depending on the chosen strategy is common to all approaches. While the search for the optimal strategy might be criticised as an exercise of data mining, this problem can be moderated by using theory as a guide (Richardson et al., 2010). According to the broad definition provided by Ke and Ramalingegowda (2005), this study adds to the literature in the field of dedicated investors.

As one result of physics envy, theorists in finance are faced with the problem of not always being able to apply rigorously the scientific method that is closely intertwined with a positivist research methodology in the natural sciences. An example of this is the Black–Scholes formula, which is intended to price options on the basis of observable input variables. As Ferraro et al. (2005) described, the divergence between the actual and the predicted option price from the formula was as high as 40% in the months following its presentation.

Subsequently, the accuracy of the formula increased substantially. Because the formula remained the same over time, there is a presumption that the options market changed in such a way that it now corresponds to the formula's prediction (MacKenzie and Millo, 2003). As a consequence, the options market could be regarded as efficient because the prices reflect all the available information, which is postulated by Fama's (1970) efficient market hypothesis. Again, however, this efficiency only comes into existence because the market participants are using the Black–Scholes model, not because it has uncovered an innate pricing mechanism in the options market. As it is widely used by investors because of its accuracy, this efficiency is not likely to change in the foreseeable future. By contrast, less accurate models exist in the equity markets, such as the capital asset pricing model, the arbitrage pricing theory or the Fama–French three-factor model. The reason for this might be that financial theory and practice do not coincide as expected. Evidence for this is provided, for instance, by Graham and Harvey (2001), who found differences in the application of project evaluation methods by CFOs between small and large firms that, in turn, might have a significant influence on equity prices. Therefore, no single pricing formula occupies such a dominant position in the equity markets as Black–Scholes. This should not be regarded as a drawback, because once a reliable investment strategy has been found, the user, the investor in this case, should be reasonably assured that an actual relationship between firm fundamentals and stock returns is evident. Naturally, it is conceivable that other strategies lead to similar results.

As a consequence, this empirical study is motivated to research the determinants of successful value investing with regard to behavioural finance and specifically i) to test various fundamentals-based investment strategies in two core European stock markets, the UK and Germany; ii) to extend these strategies with a view to improving them by better combining information from different indicators; iii) to focus on the practising investor by providing an easily implementable investment tool; and iv) to reduce data mining by emphasising the economic theory of the input variables and the construction of alternative strategies.

1.3 Research objectives and contribution of the study

The present study aims to research empirically and understand the relationship between a firm's fundamentals that can be derived from its financial statements and its subsequent stock return. As mentioned before, this is associated with the assumption that stock markets are either entirely or at least partly inefficient. In this regard, the following research objectives are set to be achieved:

i. To test empirically an earlier successful investment strategy in the European context.
ii. To extend this strategy by borrowing concepts from the forecasting literature with the objective of improving the results of the original investment strategy.
iii. To examine the drivers of the investment returns, particularly the role of information uncertainty and liquidity.
iv. To evaluate the results against the background of differing legal, accounting, country-, stock market- and firm-specific characteristics.
v. In general, to contribute to the literature on capital markets and more specifically to the literature on fundamental analysis and valuation.

In terms of theory, the completion of these research objectives extends the current state of the aforementioned part of the literature by adding another dimension. This dimension consists of two parts. First, the research is undertaken in another geographical location with differing characteristics from the original study and intends to validate the original results. Second, the original idea is adjusted and subsequently tested for improvement. At the same time, practical relevance is ensured. To achieve completion, the following research questions are posed.

i. To what extent can the success of a simple fundamentals-based investment strategy be replicated in the stock markets of the UK and Germany?
ii. Are there alternative strategies that can be used by investors to achieve better results and, if so, which one of them is the best?
iii. How do a firm’s book-to-market ratio, different kinds of information uncertainty and stock liquidity drive the success of those strategies?
iv. Does the country-specific legal system have an influence on the success of the various investment strategies?

1.4 The structure of the thesis

This subsection provides an overview of the structure of the thesis and a summary of each of the following six chapters.

Chapter 2: A Review of the Literature

The purpose of the literature review chapter is to highlight the evolution of the contributions to the literature in the field of finance and their relevance to the research objectives of the present study. In this context, the chapter starts with an introduction to the efficient market hypothesis, which has dominated the academic dispute in the area of modern finance ever since its revelation in the early 1970s. As a solution to the continuing controversy between scholars, the emerging area of behavioural finance has been proposed as an alternative theory. Having presented these two theoretical concepts, the chapter continues by scrutinising critically some of the past empirical studies and elaborating how and the extent to which theory and practice correspond. In part, the deviations between some of the research results and the theory have been consistently so large that academics of both schools of thought have usually agreed on their existence. The chapter describes how these anomalies are assessed by advocates of the two schools and range from simple to more complex attempts to explain them. Finally, the chapter condenses the literature into those parts that are relevant to the present study. It thus helps to place it within the body of academic literature and demonstrates how it contributes to it. A summary concludes the chapter.

Chapter 3: Methodology and Methods

The third chapter describes the philosophical underpinnings on which this study is based and explains why it was chosen and what it makes different from an alternative philosophical paradigm. It also includes a brief discussion about the relevance of this specific paradigm regarding the finance literature. In the further course, the characteristics of the two stock markets in the UK and Germany on which this study focuses are presented. Following that, the attention turns to the detailed presentation of the research methods. In the first step, a previously successful investment strategy is identified as a starting point on which both the empirical analyses and the newly introduced strategies are based. Alongside that, the procedure for collecting and processing the secondary data is described. In the second step, the reasons for the choice of specific statistical test methods are explained and depicted. Finally, a short summary is provided.

Chapter 4: Empirical Analysis and Results – The UK

The key contribution of this chapter is an analysis of whether alternative methods of combining financial statement information can enhance the profitability of trading strategies. In addition, the analysis is conducted in a more recent time period and considers firms with various book-to-market ratios. Beginning with the UK, the chapter first reports the descriptive results of publicly listed firms, then replicates and extends this strategy in various ways. To test whether those strategies work better for firms with certain characteristics, the analysis then takes into account a firm's book-to-market ratio. A discussion about the relevance of the UK results to the existing empirical finance literature concludes the chapter.

Chapter 5: Information Uncertainty and Liquidity – The UK

The fifth chapter aims to answer the question about the key drivers of the investment strategies under scrutiny. The main contribution of this chapter is the identification of the prime force behind the one-year investment returns, thereby developing a novel measure of information uncertainty. Possible other drivers, such as firm size and stock liquidity, are used. Size, in this regard, serves as a surrogate for the overarching wider concept of information uncertainty and is complemented by two more specific measures. Liquidity, in turn, is represented by proxies that capture either the price impact dimension or the trading quantity dimension while they are used to test reciprocally the robustness of the results at the same time. Finally, the empirical results of this chapter are discussed against the background of the existing body of academic literature.

Chapter 6: Empirical Analysis and Results – Germany

The last empirical chapter of the research deals with the application of the financial health investment strategy and its respective variants in the German stock market. Its purpose is to provide further insights into the success of the strategies in a different setting, helping us to examine whether the profitability of the strategies is influenced by the institutional or legal setting. This is of interest because, in a bank-oriented economy such as Germany, published financial statement information is not the prime method of mitigating information asymmetries. In this regard, this chapter particularly emphasises the peculiarities of the German economy and legal system and contrasts them with those prevalent in the UK.

Therefore, the testable hypotheses are adjusted accordingly to predict whether a country characteristic is causal for either increasing or diminishing trading profitability. In conclusion, the results are related to the literature with a conclusive chapter summary.

Chapter 7: Conclusion and Limitations

The concluding chapter highlights the main contributions of the present research and how the gaps in the literature were narrowed. It then continues by pointing out the limitations of the research and proposes potential avenues for further research.

Chapter 2 A Review of the Literature

2.1 Introduction

The purpose of the following literature review is to provide an overview of the evolution of modern finance to explore and legitimate the research gap that this thesis intends to narrow. This source is then used to derive research questions and to formalise the topic under scrutiny in each of the three empirical chapters. Moreover, the literature review highlights the research that has already been conducted in the past, the contemporary state of academic knowledge in this area and how this thesis is relevant to it. A further important aspect is the critical evaluation of research frameworks used by scholars that will play a leading part in the subsequent chapters, particularly concerning the research design and methodology. The general rationale is therefore to present a guideline for the reader to follow the research process, the author's line of argument and what can be expected from this thesis (Blumberg et al., 2008).

The literature review is structured as presented below:

- The first part deals with the two prevalent schools of thought in modern finance and lays the theoretical foundation for the necessity of the research at hand. It includes an outline of their fundamental assumptions and shortcomings and answers the question of why two schools exist at all. Finally, recent developments in this area are mentioned.
- After presenting the general theoretical concepts of mainstream finance, the second part of the review emphasises the practical ramifications of their assumptions. This includes a summary, interpretation and critical evaluation of empirical findings that were obtained in stock markets, mostly in the US. The research results indicate the existence of anomalous stock return patterns with regard to firm valuation and firm size, known as the value effect and the size effect, respectively.
- Since the existence of these anomalies is widely acknowledged by academics, parts three and four seek to clarify both doctrines' stance on the issue and the explanations that they state in each case. The possible reasons for those anomalies range from simple and logical to more complex explanations with no definite, unambiguous final answer.
- The final part of this chapter relates the existing body of knowledge to the current research. Testable hypotheses are established in subsequent chapters. It therefore locates the work within the academic literature and demonstrates how it extends,

differs from and ultimately contributes to the knowledge. Some concluding remarks are made.

2.2 Theoretical foundations of modern finance

2.2.1 The evolution of asset pricing models

In the 1950s, Harry Markowitz (1959) presented the conceptual framework for what is now known as modern portfolio theory (MPT). The basic idea was that a certain asset return is only obtainable by taking on a certain amount of risk and therefore leaves the investor with a given set of investment portfolios from which to choose. In this context, the element of risk is tantamount to an uncertain outcome, whereas the asset price is a representation of the present value of its discounted future payoffs. Based on Markowitz's work, Sharpe (1964), Lintner (1965) and Mossin (1966) independently extended the basic concepts of MPT by presenting the capital asset pricing model (CAPM). According to this, the expected return of an asset can be presented as the risk-free rate of interest plus the asset's market covariance times the difference between the market's expected rate of return and the risk-free rate divided by the variance of the market return. In contrast to MPT, the CAPM assumes that overall risk can be split up into diversifiable and market risk and thus it provides a means of valuing assets subject to market risk only.

However, questioning the practicality of the model due to its strict assumptions, Ross (1976) proposed arbitrage pricing theory (APT) as an alternative to the widely used CAPM. The main point of critique was the inflexibility of the CAPM, in which the future return expectations and the risk assessment of investors are assumed to be identical or at least homogeneous. Further, the CAPM assumes a mean-variance utility-optimising investor. Relaxing these conditions, the alternative model is still able to provide better estimates of the expected stock returns than the CAPM (e.g. Bower et al., 1984; Fama and French, 1997). According to APT, a financial asset's expected return can be written as a linear function of a range of macro-economic factors or theoretical market indices with a beta coefficient for each factor. The obtained rate of return can then be used to discount the future payoffs to arrive at the appropriate asset price. Differences between computed and actual asset prices are not supposed to exist for a long time because they are arbitraged away. This logic implies that not all investors need to be well-informed or rational to ensure that stock prices are priced according to their underlying fundamental values.

Black and Scholes (1973) and Merton (1973) also based their work on the arbitrage argument to derive a formula for pricing options. A summary of the numerous developments in this area up to this point can be tracked in Smith (1976). Finally, Fama and French (1993) extended the CAPM by expanding the set of variables that are deemed to explain stock returns. By employing time series regressions, they showed that firm size and book-to-market equity play a vital part in explaining the differences in average stock returns. In general, it has been believed since then that this so-called three-factor model provides a more reliable tool for explaining asset returns against which anomalies should be measured (Malkiel, 2003).

The different asset pricing models presented up to this point are part of financial economics and started to gain scientifically relevant status in the 1960s (Jovanovich, 2008). Although they offer various approaches to asset valuation, they commonly hold the same assumptions as the efficient market hypothesis. The following section will present the underlying principles of this theory and point out its disadvantages.

2.2.2 The efficient market hypothesis (EMH)

The purpose of the following paragraph is to elaborate and review critically the basic theory on which all of the aforementioned asset pricing models are built. Since these models are commonly used to mimic reality to comprehend its complexity fully, certain assumptions need to be defined first. Although not all aspects of reality are essential for its understanding, a researcher has to ponder the extent to which each of those is both vital and measurable in the process of abstraction. In general, the EMH assumes that all information is rapidly and fully incorporated into prices. More informally, there is no “free lunch” because higher returns are associated with higher risk.

The EMH is the dominant assumption of many asset pricing theories and was first introduced by Fama in 1970. The theory is based on the concept that stock prices follow a random walk, which means that they have no memory and therefore cannot be used as a means to predict future returns either reasonably well or reliably. However, the central idea goes back to Samuelson (1965), who originally presented the random walk hypothesis and outlined the reasons behind it. According to the theory, randomness is a result of the insatiable greed of investors, who constantly attempt to secure a financial advantage by trading on any kind of information. Because stock prices are a result of these cumulative transactions, the available information is then reflected in the stock market prices. For that reason, arbitrage opportunities will fade quickly and above-average returns are practically impossible (cp. APT in section 2.2). In short, the EMH is "the simple statement that security prices fully reflect all available information" (Fama, 1991). Security prices in this respect represent the present value of the future payoffs of stocks and "fully reflect" the available information. The term information is split into three sub-categories: i) weak form, ii) semi-strong form and iii) strong form. Category i) assumes that past stock returns are useless in forecasting future stock returns and therefore render technical analysis futile. In other words, weak form efficient markets are characterised by the random walk hypothesis and therefore past prices play no role in the prediction of future stock returns (Smith, 2012). Testing for weak form market efficiency is usually undertaken by variance ratio tests based on Lo and MacKinlay (1988) and newer version of these such as in Kim and Shamsuddin (2008). To meet the requirements of category ii), prices should additionally incorporate new information, such as financial statements or ad hoc reports, in a timely manner. In order to test the effect of the arrival of new information on stock prices, event studies are usually conducted. For instance, Chordia et al. (2005) analyse stock price movement of up to thirty minutes to provide insight on how fast prices converge to market efficiency. The last case iii) holds true if even investors with access to confidential information are unable to benefit from it. The prevailing view amongst academics is that financial markets are at least the semi-strong form of efficient.

2.2.3 Problems with the efficient market hypothesis

Fama (1970) himself acknowledged the vagueness of the EMH and especially the exact meaning of the statement of full reflection of information. Due to the joint hypothesis problem, the theory cannot be tested empirically unless the underlying asset pricing model is known. In other words, if the null hypothesis of market efficiency is falsified, this would not necessarily mean that markets are inefficient. It may also be due to the deficiency of the underlying valuation model that was used in falsifying the null hypothesis. Hence, the results would still be open to interpretation. However, the benefit of the theory lies in its ability to isolate risk as the only distinctive characteristic amongst the theoretical models, because the definition of risk is model-dependent. This is possible because the expected return is assumed to be a function of risk alone on condition of a given set of information. The EMH therefore seamlessly follows on from both the APT and the CAPM and is formalised as follows:

$$E\left(\tilde{p}_{j,t+1}|\Phi_t\right) = [1 + E(\tilde{r}_{j,t+1}|\Phi_t)]p_{jt} \quad (1)$$

The equation states that the price $\tilde{p}$ of asset j at time t + 1 corresponds to the percentage return of that asset after one period times its price p at time t. The Greek letter Φ represents a vector of all the available information at time t. Although this equation suggests a veneer of mathematical precision, its interpretation is rather fuzzy. Markets are efficient if all the available information is “fully reflected” in the asset prices. However, it is still unclear how the term “fully reflected” should be interpreted. Fama (1970) stated that markets are already efficient if investors “utilise” information and thus moderate the strict requirement of “full” reflection. This could be regarded as another reason why the EMH cannot be falsified as the scale on which information is utilised is irrelevant to market efficiency.

Fama (1970) proceeded by providing three basic prerequisites that an efficient market has to satisfy sufficiently to be able to reflect fully all the available information: transaction costs are assumed to be zero, all the available information is costless and there is agreement on the pricing implications of the information amongst the investors.

The first assumption clearly does not resemble the reality in the financial markets of the 1970s, although transaction costs have fallen dramatically since Fama’s (1970) work was published. This can be largely ascribed to the computerisation of stock markets and the associated reductions in overheads as well as vast increases in the trading volumes of many financial markets. Although noted, this assumption is not considered in the further course of this thesis.

The second assumption, however, deserves further attention. The costs of accessing information have decreased in a similar way to transaction costs, but another dimension exists. Grossman (1976) and Grossman and Stiglitz (1980) logically proved the impossibility of perfectly efficient financial markets for this assumption. They reasoned that information gathering would become an unprofitable activity due to the implied costs and therefore would not be undertaken by any investor. Consequently, trading would come to a standstill and markets would cease to exist. However, if all investors stopped paying for information, the markets would be in disequilibrium, because every single investor assumes that he would be better off once he resumes paying for information. This rationale was taken up by Fama (1991) in his literature review of 20 years of EMH research. From the outset of his paper, he referred to this issue and lessened this assumption to the point that investors will only pay for information if the marginal benefits exceed the marginal costs. By rendering the strict version of the assumptions false, he admitted to bypassing the first assumption as well. Therefore, the focus rests mainly on the last assumption.

Finally, the third assumption leaves some room for interpretation. In Fama (1970), there is no explicit statement about whether it is sufficient for prices to reflect investors' aggregate agreement on the pricing implications of information or whether all investors reach the same conclusion independently. The literature on this point tends towards the latter explanation, as Clement (1999) identified systematic and time-persistent analyst forecast variability and Gleason and Lee (2003) documented differences in the pricing process dependent on both analyst coverage and prominence.

As mentioned above, investors are expected to utilise all the available information, which is categorised as i) historic (past prices), ii) current and public (company announcements, etc.) or iii) private and non-public (insider knowledge), in exactly the same way. It can therefore be concluded that such a restrictive definition of "information" accompanied by the catch-all term of "utilisation" is another reason why the EMH cannot be falsified. The reason is presented in the next paragraph.

For instance, in a semi-strong-form efficient market, investors would be able to differentiate between different kinds of information, such as company-dependent (internal) or independent (external) information. As no explicit distinction is made, Fama's (1970) definition of information in the semi-strong-form case leaves the reader with room for interpretation. Both kinds of information also have consequences with regard to the amount of time it takes for the asset price to adjust. This is because in reality different levels of agreement on how to interpret information exist amongst investors. A profit-warning company statement (internal) is likely to be reflected more quickly in the asset price than a rise in interest rates (external). If profits slump, this cannot be a good sign for the health of any company, regardless of the industry in which it operates. A rise in interest rates, however, is not a bad sign per se. Companies in the financial sector, for instance, might even profit from it in the future. There should also be a difference between companies with different levels, types and durations of corporate debt in this scenario. Fama (1970) took up this thought and presented evidence from Waud (1970), who found little proof that prices adjust markedly after interest rate announcements because the external information has already been anticipated. However, it is questionable whether investors always have the ability to comprehend fully the complexity that some types of information entail and whether they can predict the effect on any single asset within their portfolios at the same time. However, even assuming that investors are incapable of seeing through all the layers of complexity, this point of criticism is undermined by the catch-all term. It suffices that investors just utilise information to make the EMH work.

Although the EMH approximately held true at that time (e.g. Ball and Brown, 1968), it is not surprising that since then it has earned a considerable amount of criticism because of these somewhat unrealistic assumptions. LeRoy (1976) equated the entire theory with a tautology because prices are conditional on Φ in (1) to make them efficient. However, prices are efficient by definition as they are regarded as having incorporated all the available information. Thus, the theory is always true and cannot be rejected. In an attempt to dispel the ambiguities of the EMH, Beaver (1981) relaxed the assumption of investors' full agreement on information. According to his definition of efficient markets, it is sufficient for asset prices to behave as if every investor has interpreted the information in the same way. Although this definition shifts the EMH towards more realistic assumptions in that it acknowledges heterogeneity amongst investors, it raises another question: how can the market in aggregate be efficient if each single investor has different views on how to interpret information? A logical answer was given by Ryan (1982), who interpreted Beaver's (1981) assumption as a statement that "[…] market prices reflect the rational component of individual behaviour […] whilst unsystematic (irrational) behaviour is diversified away in the pricing process". However, in his view, this answer is unsatisfactory as it is also possible that the situation could be vice versa in that the rational component is cancelled out. Nevertheless, Ryan (1982) acknowledged the possibility of bypassing his alternative answer by citing Friedman's (1953) analogy of billiard players.

"Consider the problem of predicting the shots made by an expert billiard player. It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas that would give the optimal directions of travel … It is only a short step from these examples to the economic hypothesis that under a wide range of circumstances individual firms behave *as if* they were behaving rationally to maximise their expected returns …"

Although this explanation can serve as defence for the EMH as a whole, it is still unclear how the transition process from irrationality to rationality actually works (Ryan, 1982). However, this criticism could be regarded as unjustified from a practical point of view. As regards the development of uniform accounting and financial regulation standards in an increasingly globalised world, the assumption of inefficient markets would not be helpful (Kothari et al., 2010). As the authors pointed out, efficient markets are based on an equilibrium theory and therefore describe a state from which accounting principles, for instance, can be derived. Market inefficiency theories in turn describe transitory pricing only but more importantly do not provide a framework for the restitution of market efficiency. Therefore, despite the shortcomings and ambiguities of the EMH, it remains an integral part of the academic literature and policy makers' decision process until such time as advocates of market inefficiency present an adequate substitute. This statement, of course, is easy to make with the benefit of hindsight and because the implications of the EMH for policy making are now apparent.

However, the ongoing academic dispute and exchange of ideas should stimulate the search for such better theory. At the end of the 1970s, though, EMH proponents were still pointing to the substantial empirical evidence in support of this theory (e.g. Jensen, 1978). In a statement on the possibilities of falsifying a theory in general, Friedman (1953) stated that it is impossible both to refute theoretical hypotheses and to prove them empirically. Advocates of the EMH usually argue exactly on this basis. This means that by referring to the joint hypothesis problem, empirical evidence against the validity of the EMH is refuted. On the basis of Friedman's (1953) statement, it may seem rather surprising that the validity of empirical evidence (e.g. Jensen, 1978) is approved in support of the EMH but rejected in cases against it (e.g. Fama, 1991).

Nevertheless, since the end of the 1970s, a vast amount of literature has been gathered on anomalies that do not conform to the idea of efficiency in financial markets. Banz (1981) was supposedly amongst the first to spur this development. Studying risk-adjusted stock returns depending on firm size, he discovered unusually high returns for small firms that were out of line with expected return models, such as the CAPM. Although this so-called size effect fades with increasing firm size, he concluded that the CAPM is error-prone. Building on these results, Basu (1983) documented that, additional to size, firms with higher earnings' yields tend to show higher subsequent risk-adjusted returns in the cross-section, thus calling market efficiency into question.

A further common anomaly that survived until recent times has been covered by a vast majority of research in the area of what is known as calendar effects. At certain times in the year, stock returns are either below or above their expected level. Studies have shown that mean returns on Fridays are above the average, whereas mean returns on Mondays are below the average (Siegel, 2002; Cho et al., 2007). A subset of this conundrum was first found in a study conducted by Keim (1983), who examined stock returns over one calendar year. Now known as the January effect, he found a higher negative correlation between size and stock returns in January than in the remaining months. A further challenge to the semi-strong form of the EMH is the results presented by Copeland and Mayers (1982). They found significant abnormal risk-adjusted returns for stocks that are part of the first quintile and are assumed to perform best. Being aware of the shortcomings of the CAPM, they implemented a holdout period approach to create a valid benchmark against which stock returns are measured. Although there were problems with this method, such as non-stationarity of mean returns during the benchmark creation process, they proved that it did not undermine the validity of their results.

In his update and review of 20 years of research after the EMH was first published, Fama (1991) acknowledged ambiguities in the EMH and left the choice of accepting it as a useful simplification of reality to the reader. A decade later, Kothari (2001) provided an insightful review of empirical research on the relationship between capital markets and financial statements. According to his classification, the literature can be divided into five main areas: i) methodological capital market research, ii) evaluation of alternative accounting performance measures, iii) valuation and fundamental analysis research, iv) tests of market efficiency and v) value relevance of disclosures according to various financial accounting standards and the economic consequences of new accounting standards. With the assistance of this classification, it should be noted that the topic of the present thesis is clearly located in the area of bullet point iii). Based on this, market inefficiency, more precisely semi-strong-form inefficiency, is assumed from the outset. As described earlier, this implies that market prices do not reflect all the available information of financial statements and information is not anticipated immediately in the pricing process.

Although the classification by Kothari (2001) is a helpful starting point, it needs to be remarked that some overlap exists between point iii) and point v) for two reasons. On the one hand, researchers test alongside the evidence against the EMH whether trading strategies based on the anomaly are viable. On the other hand, tests of trading strategies based on those anomalies are a logical consequence of finding them first. For this reason, the literature can be further divided into the following three categories. Firstly, confirmation of already-documented and revelation of new anomalies in stock markets. Into this category, for instance, fall studies on short-term momentum and under-reaction to new information. According to the claims of the random walk hypothesis, stock prices have no memory and therefore the serial correlation should be zero. Research by MacKinlay (1997) revealed the opposite in that stock prices drifted too many times in one direction, which caused the authors to reject the hypothesis of random walks. Jegadeesh and Titman (1993) reached the same conclusion earlier and found outperformance of past losers compared with past winners over the following three to twelve months. It can be reasonably inferred that those results are in violation of the EMH, because information does not seem to trickle down to stock prices instantly. This work is based on findings in the long-run scenario insofar as stock market returns substantially tend to mean revert over time. Negative serial correlation exists amongst those stocks that have been former winners and turn out to be losers in the subsequent one to three years (De Bondt and Thaler, 1985; Fama and French, 1988; Poterba and Summers, 1988).

Newer findings document anomalies based on sales growth (Lakonishok et al., 1994), on the momentum and trading volume effects (Lee et al., 1999) and on the industry-factor effects (Moskowitz and Grinblatt, 1999).

Secondly, research has focused on the implementability of trading strategies based on those market aberrations. The term implementability includes a number of further aspects, such as the cost–benefit analysis, reliability and flexible applicability of a particular trading strategy. Studies in this area mainly focus on fundamental analysis in an attempt to predict stock market returns in relation to some fundamental variables and financial statement ratios, such as Lev and Thiagarajan (1993) and Abarbanell and Bushee (1997).

Thirdly, research is concerned with the potential drivers of market anomalies and attempts to provide answers to why they exist and why some of them persist. This relatively new strand of literature is dealt with in the second part of this review and mainly focuses on the investor's mind and the effects on financial markets. It represents an alternative to the established ideas of the EMH by focusing not on the aggregate market outcome but on understanding the decision-making process that finally leads to it.

With regard to the further subdivision of the literature, the present thesis contributes to the second point mentioned above. In general, one can conclude that the doubts about the validity of market efficiency have not been dispelled entirely until today. This is particularly interesting in the light of the increasing evidence against the EMH compiled by its originator. Using the CAPM as an asset pricing model, Fama and French (2006) found considerable inconsistencies for the period 1963 to 2004. These are, among other findings, larger than the anticipated betas for growth stocks compared with value stocks, indicating exactly the opposite of what one would expect according to the risk-to-reward relationship of the CAPM. More recently, Fama and French (2012) documented a value premium across four major economic regions that is negatively correlated with size. For these reasons, it is intriguing that the implications for market efficiency are not mentioned in either paper. As noted by Dempsey (2013), it is also bewildering why the inefficiencies of the CAPM are pointed out in detail, but those of the EMH are not. The CAPM can be viewed as an elegant representation of the EMH by relating risk to expected return and thus modelling the investing process subject to investors' rationality. However, according to Fama and French (1992), β as a risk measure is inadequate for explaining cross-sectional expected returns. Consequently, it might be asked why in the three-factor model that was introduced in this paper β remained a factor in the model, next to size and book-to-market equity. Dempsey (2013a) ventured so far as to conjecture far-reaching consequences for the subject of finance because "[s]uch paradigm, after all, justified the status of finance as a subject worthy of 'scientific inquiry'".

The presentation of evidence in support of the EMH in the further course of the text will demonstrate the rationale of the academic dispute and its current situation.

2.2.4 Arguments in favour of the efficient market hypothesis

The EMH has attracted vehement criticism for many years, some of which might not be entirely justified. Although some of its prominent followers and even its originator seem to have undergone at least partly a change of mind, the EMH should not be ruled out as a means of describing financial markets. As mentioned before, the intent of the EMH is to describe markets in equilibrium, which can be regarded as an idealised state. It follows logically that, on this basis, some of the documented anomalies are nothing out of the ordinary. In the short run, markets tend to be a voting mechanism but convert to being a weighing mechanism in the long run (Graham, 2005). This analogy, created by the father of value investing, Benjamin Graham, in the 1960s, highlights that short-term market inconsistencies are expressly accepted in the pricing process. Supporters of the EMH have received prominent assistance from the same Graham, who was cited by Malkiel (2003) with the words: “I am no longer an advocate of elaborate techniques of security analysis in order to find superior value opportunities. [Today] I doubt whether such extensive efforts will generate sufficiently superior selections to justify their cost. I’m on the side of the ‘efficient market’ school of thought.” However, it is interesting to compare Graham’s utterance as a practising investor with the theoretical findings, which tell a different story. Fama and French (2006) showed that the CAPM is able to explain value premiums for the period of 1926 to 1963, but struggles to do so afterwards. Ang and Chen (2007) reached the same conclusion, at least for the pre-1963 period.

A reasonable explanation for this conundrum was provided by Malkiel (2003). On the one hand, anomalies occur because fundamental valuation measures might represent better proxies for risk than beta. Direct evidence for this was already presented by Fama and French (1993), who introduced the three-factor model based on their previous findings of market anomalies. Moreover, Malkiel (2003) made the point that some of these patterns are not robust and depend on the sample period researched. For instance, the period from the early 1960s to 1990 is regarded as unique in terms of generating abnormal returns. Besides, even if these patterns were actually present in the past, they would already have been exploited and therefore would have ceased to exist. Although this logic holds true in theory, the picture looks rather different when looking at the results of practising investors. It can indeed be argued that trading strategies that are accessible to the public are subject to diminishing returns. However, this does not automatically imply that markets are efficient in the sense that abnormal returns cannot be generated with trading strategies that are held secret. One famous example is long-term capital management (LTCM), which was founded after 1990. Generating close to 50% returns after fees, it is apparent that abnormal profits were still possible. However, as the LTCM strategies became increasingly apparent to other investors, the trading opportunities became “crowded” and the returns declined (MacKenzie, 2003). This example confirms

Malkiel's (2003) first statement but challenges the second one, namely the absence of market anomalies. He likewise referred to the substandard performance of professional fund managers and regarded it as one of the most convincing proofs of market efficiency. It should be noted, however, that no single fund is big enough to pose a systemic risk to global markets in the case of a total failure. The finance literature appears to be amazingly silent about these implications, with the risk that regulators focus too much on an idealised state or equilibrium.

The reasoning against the existence of a value premium can be rather straightforward. The higher expected return can only be achieved by higher risk exposure. Because it is assumed that value stocks exhibit higher risk, this premium is not an abnormal return or α, but compensation for bearing the risk of a distressed asset. Therefore, higher systematic risk is synonymous with higher returns. The question is why higher systematic risk should be associated with higher unsystematic risk or distress of a firm if this type of risk is diversifiable? To sidestep this inconsistency, Heaton and Lucas (1997) argued that value firms share certain characteristics, making them weak in the face of macro-economic factors such as recessions. Because growth firms are more robust against those imponderables, investors prefer them and accept lower returns. However, this logic is not backed empirically.

On the other hand, evidence against the EMH could be discarded as the result of data mining. Kothari et al. (2005) examined the data sets that other researchers used to find arguments against the EMH and for market anomalies such as the value premium. Due to non-random data deletion, the impression might arise that stock returns are in fact predictable. Accounting for truncations in the data sets used, there seemed to be a bias of up to 50% in the results. Malkiel (2003) partially blamed editors of academic journals as they prefer to publish new results than research that is confirmatory in nature or has negative outcomes. These two aspects weigh heavily in favour of the EMH advocates. In his review of cross-sectional return predictors, Subrahmanyam (2010), not a follower of either the EMH or the CAPM, cast doubt on the process of adding variables to a predictive model to explain data. He reasoned that this data-mining procedure comes at the expense of model robustness and ultimately reduces reliability. Black (1993) even ventured so far as to view the three-factor model as the result of data mining.

The criticism of the urge to publish in whichever new area of finance is arguable but does not seem to be completely unwarranted. In light of the publications of some abstract research findings in renowned journals in the past decade, it may be asked whether this development is conducive for policy makers and the finance discipline. A good example is Hirshleifer and Shumway (2003), who found a significant correlation between sunshine and stock returns after running tests for twenty-six cities. Although they argued that sunshine lifts investors' mood, their results could seriously be called into question for two reasons. Firstly, using a simple OLS regression, only four cities showed a significant correlation using a two-tailed 5% test. However, those markets are mostly of local importance. Employing a logit model, the only international market with a significant correlation was New York. Secondly, some cities with internationally important stock markets were ignored. The stated correlation might simply not be there because market participants in different time zones are not affected by the current weather conditions in a particular location. This also seems to be corroborated by the finding that mostly local markets are influenced by sunshine. Although the weather might have an indirect influence on the risk-taking behaviour of investors and challenge the market efficiency, as documented by Bassi et al. (2013), the added value is most likely to lie in the academic exercise itself. Weather is uncontrollable; therefore, new insights for regulators or policy makers are most likely to be of limited importance.

In his review, Malkiel (2003) cited a conversation between Roll and Shiller. Roll, an economist and fund manager, stated that he had tried to invest both his own and his clients' money in every conceivable way in the pursuit of successfully exploiting the anomalies presented by academics but that he could not find any kind of market inefficiency. As mentioned before, there is evidence in favour of the EMH because the majority of fund managers underperform the general market. A failed financial incentive scheme in the fund industry might be a reason for that. This would also go hand in hand with Friedman's (1990) statement about the recklessness of individuals in spending money, which varies depending on its origin and the investment objectives. According to this idea, the recklessness in spending money is lowest if one's own money is spent on oneself and highest if others' money is spent on others. Applied to the case of active fund managers, a commission is paid irrespective of their performance during the year. This means that others' money is spent not only on others but at least partly on fund managers themselves, which indicates a high level of recklessness.

The above section provided an overview of the origins of modern finance and a presentation of the ongoing debate in the current literature with regard to efficient markets. In most cases, the abnormal returns to fundamentals-based investment strategies are owed to the extreme deciles or quintiles, respectively. Hence, it appears reasonable to assume that the majority of firms still tend to be approximately correctly priced. However, as pointed out earlier, some scholars are unconvinced about the assumption of efficient markets and doubt that models such as the CAPM or the three-factor model fully capture the complexity of financial markets. Therefore, the next section will provide a review of the literature in an alternative branch of finance called behavioural finance.

2.3 Behavioural finance

Established in the early 1980s, the behavioural finance approach is trying to provide an alternative understanding of the functioning of financial markets and answers to the question of the causes of market anomalies given the limitations of the EMH described earlier. In this regard, behavioural finance (BF) is relaxing the assumptions of rational investors set by von Neumann and Morgenstern (1944) in that it admits limits to arbitrage and takes into account cognitive psychology (Ritter, 2003). Arbitrage in this regard means that rational investors are assumed to move stock prices back to their efficient levels without capital and risk, thus offsetting irrational behaviour on an aggregate basis. However, this true version of arbitrage is only possible theoretically. In reality, arbitrage involves both capital and risk. Additionally, investors might avoid volatile situations as their risk is further increased. This poses a limit to efficient asset pricing, which is not anticipated by the EMH (Shleifer and Vishny, 1997). Investor irrationality, on the contrary, is the outcome of a multitude of cognitive biases that ultimately distort the price-building process (Ritter, 2003). The interplay between financial decision making and psychology occupies a large part of the BF literature. Thaler (2000) predicted an evolution from the homo economicus to the homo sapiens, an individual who is more emotional and not as smart as his or her creators. Because it is believed that investors are not able to comprehend fully the intricacies of financial markets and foresee the consequences of their decisions in the future, so-called heuristics are applied (e.g. Gilovich et al., 2002; Benartzi and Thaler, 2007). These rules of thumb may be a suitable and easy replacement for more elaborate decision techniques but can likewise lead to so-called biases. According to Statman (2005), proponents of the EMH assume that i) investors are rational, ii) markets are efficient, iii) investors design their portfolios according to the rules of mean-variance portfolio theory and iv) the expected returns are a function of risk and risk only.

However, in the eyes of the behavioural finance advocates, investors do not fully adhere to these assumptions. Even if there are rational investors, Barberis and Thaler (2005) noted that arbitrage is limited because counteracting the force of irrational investors who are pushing stock prices away from their fundamental values is almost impossible. It might even be assumed that usually rational investors sometimes engage in irrational behaviour. O'Hara (2008) referred to rational investors who believe they can buy a stock and sell it on for a higher price despite knowing that the aggregate market is irrational. A further example is Keynes's analogy of a beauty contest. Investors buy stocks because they believe they are popular with the aggregate market, not on the basis of their intrinsic value. As Keynes stated, markets can remain irrational longer than investors can stay solvent. Professional investors, such as fund managers, are likely to engage in such behaviour if the competition for and return requirements of clients are high. In contrast, rational finance theory is based on the analogy of the homo economicus, which was defined by von Neumann and Morgenstern and which assumes a rational and utility-maximising investor. Rational investors offset profits and losses against each other to evaluate which stocks show the highest expected utility. Beyond that, rational decision makers are indifferent to the weight of gains and losses.

The literature on BF can be split into two complementary parts. Since the 1980s, the classical view of efficient markets has been challenged by an increasing number of studies documenting market anomalies. As described earlier, the first step at that time was to describe the different kinds of anomalies and to establish agreement on their existence within the academic community. Once those findings had been acknowledged and published, researchers could raise the question of why they occur in the next step. De Bondt and Thaler (1985, 1987) pioneered this work and were amongst the first to link investors' behaviour on financial markets to experimental psychology. Their results revealed signs of overreaction to unforeseen news as well as higher subsequent long-term returns for prior losers compared with prior winners. The anomaly was particularly pronounced in January but a function of neither risk nor size. This, however, contrasts with the earlier-mentioned results of Banz (1981), who reported unusually high returns for smaller-sized companies. In turn, Jegadeesh and Titman (1993) reported higher subsequent one-year returns on stocks with previous high six-month returns, thus finding exactly the opposite of De Bondt and Thaler (1985) in the short term. Since then, more anomalies have been found for which the EMH is unable to provide a coherent explanation. This is not surprising because both the EMH and the assumption of rational market participants are nothing more than metaphors for reality

(Shiller, 2003). This would also correspond to the finding that investors rely on heuristics to be able to deal with the complexities of financial markets.

The second part of BF addresses precisely this issue in that it is concerned with the question of how those anomalies emerge. In a seminal paper, Kahneman and Tversky (1979) presented their prospect theory, according to which investors make decisions dependent on the potential value of losses and gains and not only on the final outcomes. Besides, probabilities of outcomes are replaced by individual decision weights. Although prospect theory deals with the decision-making process of investors and likewise expected utility theory, there are differences in investors' choices in the end. This is because, in a two-stage process, investors first edit the offered prospects and simplify them. They then choose the prospect with the highest value after having evaluated and weighed the edited prospects by applying heuristics.

An essential part of the theory is the assumption that the carriers of values are changes and not final states. In more formal terms, this means that instead of determining utility as in rational finance, value is rather a function of a reference point and the extent of change in relation to this point. Interestingly, the value function is characterised by the observation that monetary losses weigh heavier than gains and investors avoid fair bets when the amount of gain and loss is equal. Therefore, the graphical presentation of the value function is S-shaped and concave on the right of the reference point and convex on the left of it. The convex segment features a much steeper slope than the concave one, representing the greater discomfort of investors in reaction to losses. In fact, experiments have shown that the discomfort of giving up an asset is twice as much as the benefit of retaining it (Tversky and Kahneman, 1991). This is closely linked to the observations of risk aversion examined by Shefrin and Statman (1985). Investors hold on to losers for too long and sell winners too quickly. Therefore, seeking risk in the hope that stock prices will recoup their losses to avoid discomfort is likely to be widespread behaviour. For this reason, the evaluation of risk is seemingly a very intuitive process and far from being objective. Kahneman and Tversky (1984) compared this behaviour with the sale of lottery tickets and insurance policies. As a consequence of shifts in the reference point, investors tend to become even more risk seeking and accept prospects that they normally would not. Reference shifts represent any kind of changes in personal wealth, such as an unforeseen tax payment.

2.3.1 Problems with the behavioural finance approach

The main argument of EMH adherents against BF is the impossibility to predict anomalies reliably and implement successful trading strategies on that basis. In other words, it would be difficult to outperform arbitrageurs consistently as all unjustified price differences would be equalled out almost immediately. Even though calendar effects can be observed, they should be anticipated and therefore stock prices would already have reacted accordingly before January. Because the effect does not disappear over time, it can be argued that there are limits to arbitrage as assumed in the BF literature. However, this in turn would be indicative that markets are indeed efficient because the anomalies are too small to profit from them after deducting the implementation costs (Malkiel, 2003). Hence, EMH advocates might argue that the concept of limits to arbitrage actually turns out to be evidence against the validity of BF.

Nevertheless, it should be noted that limits to arbitrage exist for two other reasons than implementation costs only. On the one hand, if rational investors realise that prices are deviating significantly from their efficient levels, they still might not engage in arbitrage because of the risk of being stopped out on their position before the prices move back again. Further, if the information is not perfect, rational investors might simply be wrong after all and this is likely to limit their actions (e.g. DeLong et al., 1990; Shleifer and Summers, 1990). Imperfect information might be caused by young and growing firms with limited information quality and availability or little analyst coverage. On the other hand, empirical research has found evidence for systematic trading across retail investors (e.g. Barber et al., 2009). This means that trading behaviour across investors is highly synchronised or correlated, but at the same time it is not predictable, a point of criticism by EMH advocates. It is also self-evident that for a market either to rise or to fall substantially, there has to be a period of highly correlated trading behaviour. For instance, markets that are rising sharply could still be regarded as efficient as long as rational investors believe that information is adequately reflected in the market prices and they have not yet moved too far away from their efficient levels. With the benefit of hindsight, it is easy for BF advocates to ascribe a stock market crash to irrational behaviour and its consequences. Dorn et al. (2008), for instance, documented a high correlation in trading behaviour amongst investors of a German discount broker for the period 1998–2000. Ex ante, however, a rational investor faces exactly the two problems described earlier. In 1998, the popularisation of new technologies, such as the Internet, might have induced rational investors to believe that their future potential is efficiently reflected in the rising stock prices at that time. Likewise, there would not be a need for the same investors to act as there would not be arbitrage opportunities. Even supposing that there would be opportunities, it would be nearly impossible to judge when they would actually materialise.

In general, BF challenges the idea of a homo economicus who behaves in a rational way throughout. The argument is that humans are influenced by feelings and moods, which in turn affect stock prices. Therefore, the notion of bounded rationality was introduced based on the work of Simon (1955). According to his model, humans neither have the ability to grasp fully all the aspects of their environment nor the necessary computational power to figure out the best course of action. He proceeded to claim a drastic revision of the assumption of irrationality from which the theory of either how investors should behave or how they actually do behave is derived. This point was taken up by Lo (2004) in his attempt to synthesise the schools of EMH and BF. The reason for the failure of the irrationality assumption in economic modelling is the difference between a satisfactory and an optimal outcome. In the latter case, agents are assumed to maximise their expected value. In the former case, however, agents subjectively set only an aspiration level that might not necessarily correspond to the optimal level. However, this idea was discarded because EMH supporters objected that the exact point at which agents stop optimising and set the aspiration level instead is unclear.

Another point of criticism is that the cognitive biases that were found in predefined testing environments are simply assumed to be present in financial markets as well. Although Barberis and Thaler (2005) argued that investment experience does not help in reducing cognitive biases and might even exacerbate them, this does not necessarily contradict the assumption of an efficient market. The knowledge of being part of a psychological experiment may also adversely affect participants in that they do not feature their normal behaviour. Besides, test participants have limited means, for example time restrictions, to make decisions and the results are potentially biased as examiners decide which answers are right or wrong. Although Hirshleifer (2001) argued that efforts have been made to address these issues, it remains a challenge to emulate real situations in which financial decisions have to be made. Here again, Friedman’s (1990) argument comes into play: it is at least doubtful whether participants in an experiment would behave in the same way as they would when investing either their own money or on behalf of a client. However, the first step in this direction was undertaken by Lo and Repin (2002), who studied the trading success of ten professional securities traders in practice. The results showed that traders who exhibited stronger emotional reactions to both gains and losses had a significantly worse trading performance. These findings were later confirmed with a group of 80 day-traders who subscribed to an online training course (Lo et al., 2005).

Finally, the interpretation and application of a range of cognitive biases is deemed to be somewhat arbitrary. In the investment universe, nearly every bias can be used ex post to explain anomalies in the market, but it is unclear which of those biases is indeed prevalent. Hirshleifer (2001) acknowledged this drawback and noted that model mining is a serious issue to be considered. However, it can also be regarded as an advantage because a range of different explanations for anomalies can be provided as long as they are not contradictory.

2.4 Practical explanations of anomalies and their persistence

It is reasonable to assume that anomalies might also be the result of regulatory constraints or self-imposed investment restrictions. The best examples of a legal constraint are the two types of shares traded on Chinese stock exchanges. Foreign investors are only allowed to trade B-shares and do not have access to A-shares.[1] For this reason, arbitrage opportunities between A-shares and B-shares cannot be traded on and mispricing is likely to persist. In the other case, a fund may have self-imposed investment restrictions to attract a certain clientele. These can prevent fund managers from investing in companies below a certain market capitalisation and/or prohibit shorting them. Lastly, Chevalier and Ellison (1999) examined the labour market for mutual fund managers and found that younger mangers tend to follow their peers and hold conventional portfolios. The reason for this is that those young managers have to prove their investment abilities and try not to be laid off during their trial period. It may thus be concluded that anomalies could be enhanced and drive stock prices further away from a rational level.

[1] Source: Shanghai Stock Exchange (http://www.sse.com.cn) and Shenzhen Stock Exchange (http://www.szse.cn)

2.5 Summary

The two schools of thought agree on one price that is achieved by the forces of supply and demand. The difference is that EMH supporters assume that this price is always efficient because it fully reflects all the available information. On the contrary, BF advocates that, due to investor irrationality, this price cannot be achieved most of the time. It can therefore be derived that the EMH represents equilibrium and BF disequilibrium, as mentioned earlier (Kothari et al., 2010). Thus, rather than being antagonists, the two theories could even be seen as complementary parts of a holistic theory to describe financial markets that has yet to be discovered. It seems noteworthy that in his original work with regard to the EMH, Fama (1970) did not mention the term “irrationality”, which could mean that BF initially raised doubts against the assumptions of the EMH that it did not actually make. However, given that Fama (1991) elaborated on the term “rationality” in his review (Fama, 1991), it is reasonable to assume that advocates of the EMH have now incorporated investor rationality into their theory.

2.6 Market anomalies and the behavioural finance approach

After presenting the theoretical foundations of behavioural finance, this section gives a short insight into the main experimental research findings underpinning the theory. As a matter of fact, behavioural finance builds upon irrational behaviour documented in the area of psychology and transfers it to a financial setting. As mentioned before, although this method is not free from criticism, it has helped to provide an alternative to the EMH in explaining the functioning of financial markets.

2.6.1 Links to psychology

Hirshleifer (2001) listed five biases that are deemed to be causal for the existence of market anomalies. These are i) heuristic simplification, ii) self-deception, iii) emotion and self-control, iv) social interactions and v) modelling alternatives to expected utility and to Bayesian updating. He made the point that biases are not cancelled out in equilibrium as proposed by economists because they are innate from an evolutionary point of view. Thus, they should also be reflected in stock prices. Biases are the result of bounded rationality. In other words, humans employ heuristics to reduce the complexity of their environment and to arrive at a decision. These so-called cognitive biases are just one part. Moods and emotional condition can also lead to biases, which are classified as affective biases.

2.6.2 Information processing

Information processing is the preliminary stage before decisions are made under uncertainty. Tversky and Kahneman (1974) summarised this stage as the application of certain heuristics, which are applied to subsume complex problems. In general, this kind of complexity reduction suffices for most of the tasks in everyday life but can lead to significant errors with increasing complexity of tasks. People neglect base-rate information and make a poor job of estimating probabilities and forecasting values. Due to mostly superficial information processing, investors are regarded as non-rational by advocates of behavioural finance. It is also the case that some of the cognitive and affective biases are present at the same time.

A practical example is the fall and demise of the long-term capital management (LTCM) illustrated by Lowenstein (2002), which shows that reliance on highly sophisticated models can end disastrously if the human aspect is left out of the equation. In this case, all the biases from i) to v) can reasonably well be applied to explain the failure ex post. Although such a broad justification appears not to be meaningful, it still provides another perspective and thus contributes to a better understanding of the issue. With this explanation, it seems likely that the huge amount of daily trades from noise traders who caused erratic price behaviour was not anticipated by LTCM’s models.

The term “noise trader” dates back to Black (1986) and generally describes investors who either enjoy trading for the sake of trading itself or believe that they are trading on useful information, even if it would be better for them not to do so. Even though rational investors should be able to make profits quickly from arbitrage, as assumed by rational finance theory, noise traders are not likely to be discouraged from further trading. In fact, they will create their own space at some point and prevent rational traders from dominating the market. Therefore, markets will remain inefficient because noise traders bear a disproportionately high risk against which rational investors are not willing to bet. For this reason, noise traders are able to earn even higher returns than informed traders (De Long et al., 1990). Noise traders are a vital part of financial markets as they provide liquidity at all times. It may even be argued that they attenuate both upward and downward cycles. In the case of rising markets, noise traders are likely to sell their portfolios too early, as identified by Shefrin and Statman (1985). However, in falling markets, some of these investors might buy too early in the hope of gaining a bargain or averaging down their losses because they have a reference point in mind (Tversky and Kahneman, 1991).

Given these anomalies, Thaler (2005) conjectured that it is hardly conceivable that the stock prices at the NASDAQ were reasonably priced at index levels of 5,000 and 1,300 within a few years' time and were compliant with rational finance theory at the same time. This agrees with Shiller's (1981) study, which found much higher stock price volatility than predicted by the efficient markets model. He concluded that volatility is up to thirteen times higher than justified and cannot be ascribed to the arrival of new information only. As it seems, this issue is not rooted solely in the amount of information but rather in how market participants process it to make their final investment decisions. Behavioural finance acknowledges anomalies in financial markets and tries to find reasonable explanations for them by studying the human component of the homo economicus or rather the homo sapiens. For this reason, behavioural academics have incorporated concepts from the literature on psychology and sociology. This approach is not new, as can be seen in the transfer of models used in physics to finance in the 1970s.

The following section will provide a brief outline of the above-mentioned biases i)–v) and how they relate to the recent empirical results in the literature.

2.6.3 Heuristic simplification

This term summarises heuristics that are deemed to reduce complexity during the decision-making process. The availability heuristic is probably one of the best known and was documented by Tversky and Kahneman (1973). It suggests that people ascribe more attention to certain events once information is available. In other words, the probability of a future event is deemed to be higher dependent on how many instances of a similar event enter one's mind. Of course, this probability is highly arbitrary and personal and therefore can lead to substantial losses in financial markets. Paradoxically, the availability heuristic can explain biases in terms of both over- and under-reaction. Investors rate tail events depending on their media coverage and the time elapsed since the occurrence. Thus, the probabilities of recent tail events are overestimated and vice versa. As Barberis (2013) argued, this could be a reason for the 2008 crisis because most investors attached too low a probability to this kind of tail event.

2.6.4 Self-deception

The research on self-deception is extensive and mostly focuses on humans' overconfidence. Griffin and Tversky (1992) provided a summary in a non-financial context, while Odean (1998) found that people are likewise overconfident in financial markets. The underlying hypothesis is that people turn their attention more to the extremeness of the available information and disregard sample sizes. The law of small numbers is closely linked to this kind of bias and was first observed by Tversky and Kahneman (1971). It describes the tendency of people to underestimate the sample size during the evaluation process. In other words, most people attribute the same weight to smaller samples as to larger ones. This may result in the so-called gambler's fallacy, whereby a fair coin is flipped and after a streak of heads many people feel certain that tails will show up next because they are now due. Barberis and Thaler (2005) linked this finding to analysts, who can sometimes reach "star analyst" status by backing the right horse a couple of times simply by chance. Further empirical research has shown that overconfidence induces active trading and in fact leads to substandard performance (Barber and Odean, 2000). Based on these findings, Barber and Odean (2001) also found gender-specific differences in trading, which indicated that male investors are more overconfident than their female counterparts. More recently, Barber et al. (2009b) found that excessive trading results in systematic and substantial losses for retail traders in Taiwan. Interestingly, this was not the case for institutional investors, whose performance was positively related to aggressive trading. This suggests that private investors seem to be more likely to succumb partly to their emotions.

2.6.5 Risk and uncertainty

Ellsberg (1961) showed that individuals prefer known risks to unknown risks or uncertainty, despite the same probability distributions of return results. Following Knight (1921) and Keynes (1937), Anderson et al. (2009) defined an event as risky (uncertain) if the outcome is unknown (unknown) and its probability distribution is known (unknown). Noting the inconsistent empirical results regarding the risk–return relationship, they focused on the understudied uncertainty–return relationship. Their results indicated a positive correlation between phases of high levels of uncertainty and high subsequent quarterly excess returns. Thus, mispricing due to uncertainty aversion is likely to benefit those investors who either are willing to accept uncertainty or are actually able to attach some probability distribution to the outcome due to their familiarity or experience with the matter.

2.6.6 Social interactions

In their empirical study about herding behaviour, Ivković and Weisbenner (2007) reported conformity in stock purchases amongst US households. Within a radius of 50 miles around the individual household, they observed an increase in the same household's purchases after neighbouring households had increased their position. The results were believed to be related to word-of-mouth communication in the more social states of the US. Similar findings were presented by Ng and Wu (2010) for China. Investors placing their orders in person at a brokerage branch tended to be influenced by word of mouth, with the effect being particularly present in buying decisions. The authors emphasised that this might be the reason for the pricing differences between the A-shares and the B-shares because the former may be more subject to speculation than liquidity. Due to the legal restrictions mentioned in section 2.4, arbitrage opportunities are likely to persist. Although both studies consider direct contact between investors to be necessary, Sabherwal et al. (2011) found consistent results when studying online stock message boards. The authors mostly monitored the price behaviour of small stocks for which no new information was published and therefore the price action was solely dependent on word-of-mouth recommendations. Abnormal returns over a two-day period were found to be significantly positive.

2.6.7 Modelling alternatives to expected utility and to Bayesian updating

As outlined earlier, Kahneman and Tversky (1979) documented deviations from the behaviour that is usually expected from rational investors. Their alternative, prospect theory, is able to model decision making under risk better and thus to explain some of the irregularities of the more traditional approaches. Prospect theory, for instance, is a helpful means to shed light on the well-documented disposition effect, which states that investors adhere to their losing positions and sell winning positions too early. Barberis and Xiong (2009) found that using a realised gains and losses model is able to predict this effect reliably. However, this was not the case for paper gains and losses. More recently, Ben-David and Hirshleifer (2012) presented evidence in favour of the von Neumann–Morgenstern preference theory and in fact found what they called a reverse disposition effect. They argued that despite affective biases, investors are still able to question their prior investment decision and trade from a losing position if necessary.

2.6.8 Summary

As the literature review has shown up to now, the discussion about the correct approach to the description of financial markets is far from settled. However, it appears that the different approaches of the two schools of thought stimulate the academic debate and thus help to promote the understanding of this complex matter. According to Lo (2004), the controversy is rooted in the fundamentally different points of departure of each theory, because BF refers to psychology and the EMH adopts economic concepts, as shown in table 2.1. For this reason, BF and the EMH should be regarded as equivalent contributors to the literature in understanding the mechanics of financial markets.

Table 2.1 Comparison of psychology and economics

Psychology	***Economics***
• Based primarily on observation and experimentation.	• Based primarily on theory and abstraction.
• Field experiments are common.	• Field experiments are not common.
• Empirical analysis leads to new theories.	• Theories lead to empirical analysis.
• There are multiple theories of behaviour.	• There are few theories of behaviour.
• Mutual consistency among theories is not critical.	• Mutual consistency is highly priced.

2.7 Review of empirical research on fundamentals-based investing

The following section provides a review of the literature related to value investing using financial statement ratios. This strand of literature is particularly concerned with the question of whether value investing strategies are able to outperform the market on a consistent basis and whether they can be implemented economically. It therefore implies market inefficiencies and scrutinises the interplay between stock prices and fundamental value represented in the form of accounting numbers. This section also highlights the main research findings classified according to geographic region.

2.7.1 A preliminary note

Although the literature review is deemed to summarise the past and ongoing discussion amongst academics, it seems to be relevant at this point to shift the focus briefly from theory to practice. As one of the best-known investors, Warren Buffett significantly outperformed the market in the period from 1976 to 2006 (Martin and Puthenpurackal, 2008). Although he opposed the theory of efficient markets in his 1985 shareholder letter,[2] his opinion seems to have changed since then. In his 1996 shareholder letter, he argued that most investors would be better off if they invested in a low-cost index fund.[3] In fact, he even set up a wager with a hedge fund, betting on the outperformance of an index fund over a hedge fund of funds (Moy, 2014). Given the evidence against the assumption of fully rational markets (BF) and the evidence against the reliability of profit opportunities (EMH), Warren Buffett seems to feature traits that allow him to be rational and profit from mispricing. According to the literature, the reason for that might be that Mr Buffett actively insulates himself from biases i)–v) as presented previously. The so-called "oracle of Omaha" does not apply simplifying heuristics because he invests in easy-to-understand businesses (Buckley, 1999). He also stays within his circle of competence and thus avoids being subject to overconfidence (Døskeland and Hvide, 2011). Due to his critical examination of companies and his focus on only those with a stable earnings history, he ensures that the uncertainty is reduced (Buckley, 1999). As his nickname suggests, he lives detached from any financial centre and is therefore not influenced by word of mouth or the conformity effect (Døskeland and Hvide, 2011). Lastly, all these traits ensure that he avoids entering losing positions, which in turn guarantees that he will not fall into the trap of the disposition effect.

2.7.2 The US analysis

Ou and Penman (1989) were among the first to examine thoroughly the predictive power of a set of 68 individual accounting ratios with regard to subsequent earnings between 1965 and 1985. They used a composite measure called $\hat{Pr}$, which estimates the probability of an

[2] Source: Berkshire Hathaway (http://www.berkshirehathaway.com/letters/1985.html)
[3] Source: Berkshire Hathaway (http://www.berkshirehathaway.com/letters/1996.html)

increase in earnings in the following year. Because future dividends are dependent on earnings, an increase in earnings is associated with an increase in the firm's stock price. To reduce the overall amount of accounting ratios, the sample was divided into two data sets and the most significant descriptors were filtered out. This process left the authors with 16 and 18 descriptors, respectively, for the first and the second estimation period. Following that, an investment strategy was applied whereby stocks were assigned to 10 portfolios according to their $\hat{Pr}$ value with a holding period of 2 years in a theoretical zero-investment long-short portfolio. The returns were computed either monthly as cumulative returns or as buy-and-hold returns over a 2-year period. The former method was necessary to account for firms that stopped trading during this period. The results showed that the use of $\hat{Pr}$ as a valuation measure reliably predicted returns of up to 3 years after portfolio formation, whereby any mispricing disappeared. Buy-and-hold returns amounted to 12.5% and to 7.0% if adjusted for size effects. Although the authors expressly pointed out that the theory was guided by the data, fellow academics raised concerns about heavy in-sample fitting (e.g. Greig, 1992; Richardson et al., 2010). To overcome this problem and avoid the need for a holdout period, Lev and Thiagarajan (1993) identified a set of 12 fundamentals that investment professionals believe are key drivers of security valuation. They found that 10 out of 12 coefficients are statistically significant and the adjusted R^2 of the fundamentals model is substantially higher than that of the conventional returns–earnings model used as a benchmark.

In the literature, a controversial point is the question of the appropriate dependent variable in estimating coefficients. Ou and Penman (1989) used earnings but received criticism because earnings are only indirectly related to stock prices. Larcker (1989) rather suggested the use of direct measures, such as dividends or cash flows. In contrast, Abarbanell and Bushee (1997) agreed with the view of Ou and Penman (1989) and reasoned that using earnings corresponds better to the economic intuition that drives the construction of the descriptors. They used the 12 accounting signals to predict changes in future earnings successfully. Based on this study, Abarbanell and Bushee (1998) employed a reduced set of 9 accounting-based signals to implement an investment strategy that was able to generate significant abnormal returns of 13.2% within the first year. Similarly to Lev and Thiagarajan (1993), they emphasised that variable selection should be motivated by economic arguments rather than data mining.

2.7.3 The international analysis

Compared with the literature on the interplay between accounting ratios and stock returns in the US, the international evidence is scant. As the world's biggest economy, the research interest in the US stock markets is particularly marked as they are deemed to influence smaller markets and thus a "one-size-fits-all approach" seems justifiable. Another reason may be the greater availability and accessibility of stock market data to conduct research. The results from US studies can therefore be transferred to international markets, which appears to be particularly interesting for regulators in emerging markets or frontier markets where historical data are simply not available. However, with more data available over time, a better understanding of the functioning of individual markets outside the US should be beneficial in two ways. Firstly, it is of interest whether US results can be replicated and confirmed in other stock markets. If they can, this would be tantamount to a positive robustness check. For instance, if an investment strategy (e.g. based on the F-Score) also works outside the US, this would reduce the implementation costs for practitioners in individual markets. Secondly, regulators would be able to adjust their legal frameworks to satisfy the needs of the local market.

An early study by Levis (1989) found irregularities in the stock price behaviour on the London Stock Exchange (LSE) for the period from 1956 to 1985. Apart from the size effect, trading strategies based on the dividend yield, price–earnings ratios and market capitalisation appeared to be equally profitable in the following year. Focusing on the tax treatment of dividends in the UK, Morgan and Thomas (1998) showed that stocks with high (low) dividend yields generate higher (lower) risk-adjusted returns. The results were representative for the period from 1975 to 1993 and the total returns were computed monthly. However, no correlation with the tax treatment was discernible. Dimson et al. (2003) presented evidence for the existence of a value premium in the UK stock market for the period 1955–2001 irrespective of size. Again, the dividend yield was a reliable predictor of the value that materialised in the following year. In a more recent study, Siganos (2012) examined the profitability of various investment strategies based on seven individual return predictors between 1988 and 2009. The results indicated that only the earnings-to-price strategy was profitable after taking the transaction costs into account. Finally, Gregory et al. (2013) showed that UK directors enjoyed long-term abnormal returns compared with a simple value-glamour strategy between 1986 and 2003. The results were independent of the definition of value, indicating that directors are contrarians who sell growth and buy value stocks. Further, the authors found that abnormal returns last for more than two years after directors' trading.

Empirical research on the UK stock market is primarily based on market-based signals. To the best of my knowledge, studies that employ a composite, which in turn is derived from accounting-based ratios, are largely lacking. This indicates a gap in the present literature that this thesis aims to close.

Chan and Lakonishok (2004) examined the value premium in an international setting, including the UK and Germany. Using four value indicators – book value/market value, earnings/price, cash flow/price and dividend/price – the authors proved the outperformance of the value over the growth strategy for most of the time between 1975 and 1995.

In the case of Germany, the literature on the relationship between accounting numbers and stock returns is rather limited and, in general, the German stock market appears to have been neglected in academic research (Amel-Zadeh, 2011). Schiereck et al. (1999) back-tested a naïve buy-loser-sell-winner strategy and documented abnormal returns in the short and long term between 1961 and 1991. However, most of the literature is concerned with the testing of asset pricing models. For instance, Schrimpf et al. (2007) and Artmann et al. (2012) presented evidence that the CAPM and a three- and four-factor model are unable to explain the cross-section of returns. Whitaker (2003) used a combination of four variables to implement a value investing strategy across a combination of a range of European markets and found good performance. Based on these findings, they tested a more complex strategy by combining value and momentum investing, resulting in even better results (Whitaker, 2004). However, the German stock market was not researched on its own. Therefore, similar to the UK case, a gap in the literature is apparent.

2.7.4 A simple, fundamentals-based investment strategy

In a seminal paper, Piotroski (2000) contributed greatly to this strand of the literature by focusing on high book-to-market (BM) firms only with the reasoning that mispricing is more likely in those stocks. Instead of examining the relationships between each signal and the subsequent returns, he condensed nine accounting ratios into one signal on which portfolio construction was based. Using this so-called F-Score in the stock selection process, it was possible to increase the mean return by 7.5% p.a. and to yield 23% p.a. in a long-short portfolio in the period 1976 to 1996. This paper has garnered some interest on the part of academics because it applied procedural principles that can be implemented without the need for huge data sets and it required no holdout period. There has been criticism, inter alia, that such a strategy omits low-BM firms and similar returns in those were deemed impossible (Guay, 2000). In response to this critique, Mohanram (2005) presented a composite score that summarised eight accounting signals, which were specifically tailored to grasp the financial condition of low-BM firms. This so-called G-Score was able to generate a 20% annual return in a long-short portfolio for the period between 1978 and 2001. In derogation from the F-Score methodology, the G-Score measure is harder to implement with regard to time and costs, because, apart from company data, investors also need industry metrics. By nature, the G-Score is a relative measure and therefore might be inappropriate in the case in which one specific industry is overvalued. Moreover, the question arises of which peer group should be used as a holistic representation of an industry (Piotroski, 2005). Generally, Piotroski's (2000) study is consistent with the behavioural assumption that accounting information is not immediately incorporated into stock prices. Further, it has been given considerable attention by investment practitioners.

It should be noted that this thesis contributes to the research on valuation and fundamental analysis. It also extends the work undertaken on the F-Score and applies this approach to the stock markets of the UK and Germany.

Similar to Piotroski (2000), Beneish et al. (2001) applied a two-stage approach to predict extreme price movement in either direction in the next quarter. They motivated their study with the assumption that most analysts focus on a subset of firms in the stock market. By separating extreme performers first and applying financial statement analysis next, they attempted to crystallise common characteristics in the accounting numbers that ultimately help to forecast extreme volatility. Combining market-related and firm-based variables, they showed that extreme price movements can be anticipated four to six months in advance. In particular, the observations indicated that firm-based variables are more helpful in separating future winners from losers.

Attention has also been devoted to the question of how investment professionals interpret financial statement data. The results of prior studies of analysts' recommendations are not consistent, which might be due to a conformity bias. For instance, the results of Barber et al. (2003) and Drake et al. (2009) indicated a significantly negative relationship between buy recommendations and subsequent stock returns. Bradshaw (2004) instead reported no correlation between one-year returns and analyst recommendations, whereas Barber et al. (2001) and Jegadeesh et al. (2004) found a spurious but significant relationship in the near term. More recently, Wieland (2011) found that analysts are too optimistic in their recommendations and developed a scoring tool to separate out incorrect return forecasts. This tool should be particularly helpful for practitioners as there is evidence that investors follow recommendations unconditionally. The abnormal returns were reported to be 14.8% in the subsequent year.

In a similar paper, Wahlen and Wieland (2011) presented evidence that analyst consensus recommendations do not fully reflect the information from financial statements. They used a so-called predicted earnings increase score (PEIS), which represents the probability of an earnings increase in the following year. The measure consists of six accounting-based ratios similar to the F-Score and G-Score, and hedged portfolios built from high- and low-scoring stocks generated abnormal returns of 10.8% in the period from 1994 to 2005.

2.7.5 Gaps in the literature

In general, the research on accounting-based investment strategies is highly concentrated in the US stock markets. In this regard, the F-Score strategy might be the outcome of data mining, despite its robustness having been confirmed ever since its publication (Richardson et al., 2010). Further research outside the US, however, is likely to provide a more accurate picture of its overall success. Based on this groundwork, three main gaps have been identified in the current body of academic literature, which this study aims to narrow.

Firstly, the F-Score is based on a binary evaluation of accounting data, which makes it easily implementable. However, a vast amount of potentially useful data might be disregarded, thus constraining its true potential. Moreover, the composition of the F-Score itself may be another source of improvements in its success. While the original investment method aggregates each of the nine components into one score with equal weights, variable weights might be a better choice.

Secondly, the strategy is only applied to a certain type of stocks in Piotroski's (2000) study. This limits the validity of research outcomes in two ways. On the one side, the restriction to only high-BM firms is contrary to the aim of generalisability, which should be the focus of studies that are deemed to be used in practice. On the other side, the determinants of its success are still unclear. In this respect, measures of liquidity and information uncertainty, using a novel method, are introduced and tested.

Thirdly, it is of interest to study the behaviour of such a strategy against the background of different institutional settings. While the UK is based on a tradition of common law similar to the US, this is not the case in Germany, with a code law legal system. The resulting differences in investor protection rights, for instance, may have an influence on investment strategies that have not been researched until now.

2.8 Conclusions

This chapter provided an overview of the development of the two main schools of thought, which represent the two fundamental pillars of the contemporary finance literature. The literature was narrowed down further to highlight and justify the research gap. The process can be summarised as follows.

- The EMH is made under the assumption that a sufficiently large number of investors behave rationally in the sense of the homo economicus to ensure efficient markets. It therefore focuses on the aggregate market behaviour. BF, however, presumes that investors apply heuristics in the decision-making process, which may lead to biases and therefore mispricing in aggregate markets. Consequently, the attention is centred on individual market participants.
- As the present thesis starts from the premise that stock markets may not be fully efficient, the focus of the review shifted to the presentation of empirical research based on value investing and fundamental analysis, which challenges the assumption of semi-strong-form efficiency. The thesis builds directly upon previous work undertaken in this area by using established investment strategies and applying them outside the US stock markets. It then combines these strategies with measures of behavioural biases.
- The literature on investment strategies using accounting-based data outside the US is very limited, particularly in the case of the UK and Germany. The aim of the thesis is to close this gap and provide an answer to the question of whether the same

investment approach is simply transferable to the European context. Therefore, the present thesis tests rather than develops theory by the replication and extension of previous work.

Chapter 3 Methodology and Methods

3.1 Introduction

The following chapter intends to outline the link between research questions and data (Punch, 2006). For this purpose, the chapter is divided into three sections. The first part deals with the theoretical assumptions and philosophical paradigm on which this thesis is based. Subsequently, the characteristics of the stock markets in the UK and Germany are briefly presented and contrasted. Finally, the practical process of data collection and analysis is illustrated. More generally, the chapter highlights both the methods and the steps involved in answering the research questions posed.

3.2 Philosophical paradigms: Positivism and postpositivism

According to Guba and Lincoln (2005), every piece of research is inevitably related to or emerges from the nature of a particular academic discipline. Although research paradigms can be reasonably divided into five branches, three of them are generally used in qualitative research. Since this dissertation is purely quantitative, this section discusses only the positivist and postpositivist research approaches on which most quantitative research is based. The positivist paradigm assumes that only one reality exists, which can be interpreted objectively by analysing the available data. To ensure objectivity, the researcher is detached from the object under scrutiny (Haneef, 2013). This so-called scientific approach represents a two-step process comprising the aims of the research on the one hand and the methods of investigation on the other hand. In the first step, scientists are interested in the understanding of properties and relationships of reality to produce theories, forecast results or validate existing findings. In the second step, scientists are required to apply rigorous and systematic methods that constitute verifiable theories (Denscombe, 2002). Regarding ontology, positivists assume a direct relationship between cause and effect. Postpositivists, on the contrary, attenuate this strictness by attaching a probability to the occurrence of the outcome. This implies that the researcher accepts the possibility of multiple realities depending on the different ways in which the research was conducted (Creswell, 2013). The two paradigms regarding their ontology, epistemology and methodology are compared in table 3.1 (Heron and Reason, 1997).

Table 3.1 Basic Beliefs of Alternative Inquiry Paradigms

Issue	***Positivism***	***Postpositivism***
Ontology	Naïve realism – “real” reality but apprehendible	Critical realism – “real” reality but only imperfectly and probabilistically apprehendible
Epistemology	Dualism; findings true	Modified dualism; findings probably true
Methodology	Verification of hypotheses	Falsification of hypotheses

Epistemology deals with the relationship between the researcher and the object under study. The researcher is assumed to be impartial and the analysis of data is required to be neutral (Denscombe, 2002). Again, postpositivists relax these assumptions to some extent in that they acknowledge that the process of research itself influences the findings. However, they strive to follow the stricter positivist approach as closely as possible. According to Ponterotto (2005), methodology is concerned with the process and procedures of the research. Starting at the top of table 3.1, both epistemology and methodology are a direct and logical consequence of the choice of ontology. In general, (post)positivists’ research procedures resemble those of the so-called “natural sciences”, such as physics and mathematics. Therefore, the aim is to find relationships between variables to forecast future events.

The combination of all three of these premises constitutes the research paradigm (Denzin and Lincoln, 2005). This paradigm is a “basic set of beliefs that guides action” (Guba, 1990). Therefore, the foundation of every piece of research is the choice of the appropriate methodology to answer the research questions satisfactorily. The two paradigms of positivism and postpositivism are regarded as mainly quantitative because of their general assumptions that reality is ordered and potential relationships can be measured sufficiently well. As mentioned before, another more interpretive equivalent is a qualitative approach, which is not relevant in the present thesis and thus has not been mentioned. Both approaches are also known as either deductive (quantitative) or inductive (qualitative), respectively (Burger, 2008).

This thesis takes a postpositivist stance and will therefore test theory as a logical result of choosing this particular research paradigm.

3.2.1 Methodology in the finance literature

Although not explicitly stated, it can be assumed that most of the EMH (BF) advocates articulate the view of a positivist (postpositivist), because it goes hand in hand with the assumption of investors' rationality (Burger, 2008). As outlined in the previous chapter, these basic convictions are not shared by all academics. In fact, the differences of opinion may be more deeply rooted than just criticisms about model fitting or statistical nuances. Applying Kuhn's view on evaluating two competing theories, it may take many decades to find out which theory is superior to the other. The reason lies in what he called topic incommensurability. To illustrate this concept, Wray (2011) reproduced Kuhn's example of Newton's examination of gravity. In the pre-Newtonian era, a mechanical explanation of gravity would have been sought, whereas in Newton's understanding, gravity was not a mechanical problem and therefore research under this assumption was not regarded as scientific. Although it is now known that Newton's theory is superior, the question arises of what happens if the subsequent theory is inferior to the preceding one. In other words, BF could turn out not to be as promising an alternative to the established EMH as anticipated by its supporters. To sidestep this problem, Kuhn suggested evaluating superiority not by using reality as a benchmark but by using all four of the so-called secondary criteria. However, even if the new theory should be i) more accurate, ii) consistent, iii) broader in scope and iv) simpler, it might not necessarily be truer. Therefore, a shift in beliefs from the old to the new theory would not be justified because truth as such would still remain arcane (Wray, 2011).

Kuhn's view on comparing a theory's superiority with the four criteria and not with reality seems not to be widely shared amongst BF supporters, though. One of the main points of criticism by BF adherents is precisely the incompatibility of the "rationalists'" assumptions. They argue that the EMH is an idealised assumption of efficient markets and does not reflect the real world. However, criticism from EMH advocates also seems not to be justified under this view. This is because a competing theory first needs to be established to assess whether it is better suited to explaining the truth. During this process, it is hardly surprising that the methods used in generating the former theory are criticised, partly the "mathematisation" of finance, which is accompanied by the negligence of human nature or behaviour, respectively. This trend was observable in the emerging finance discipline starting in the 1950s and has led to the creation of numerous asset pricing models, such as the CAPM or APT, underpinned by the hypothesis of efficient markets.

According to Lo and Mueller (2010), this so-called physics envy describes the transfer of research methods used in modern physics to finance and was first initiated by Samuelson in 1947. Although great advances have been made in physics due to a purely deductive approach, economic systems such as financial markets appear to be more complex and unstable and therefore the use of these methods should be less fruitful. This rationale also speaks for the postpositivist stance taken in this thesis.

3.3 The UK stock market

Trading in the UK takes place at the London Stock Exchange (LSE), which officially opened in 1801. Although there were regional exchanges across the UK, the remaining eleven of them merged with the LSE in 1973. At present, the LSE offers four market segments, of which two are suited to the general public and the other two serve a more specialist demand. The first two segments consist of the Main Market (MM) and the Alternative Investment Market (AIM). The Professional Securities Market (PSM) and the Specialist Fund Market (SFM) form the other part of the LSE. As defined by the LSE, the MM is the flagship market for established companies, whereas the AIM offers a platform for smaller and growing companies to raise equity capital. The PSM offers a flexible process to issue debt or depository receipts and involves less strict regulatory requirements. Lastly, the SFM offers a regulated market environment for accessing capital from highly sophisticated global investors and is particularly interesting for hedge funds or private equity funds. As of 31 March 2014, there were 2,455 listed companies from 70 countries, of which 1,301 traded on the MM, 1,094 on the AIM, 21 on the SFM and 39 on the PSM. The average daily equity trade volume amounted to roughly £4.0 billion and the market capitalisation of all the stocks was around £4,109 billion. The amount of capital raised on the AIM since its inception in 1995 exceeds £60 billion for a total of 3,000 companies.[4] Due to the specialised nature of the SFM and PSM markets and their small size, this thesis only considers stocks that trade in either the MM or the AIM market.

[4] Source: London Stock Exchange (http://www.londonstockexchange.com)

3.4 The German stock market

There are a total of eight stock exchanges in Germany, with Frankfurt being the one of prominent importance. Most of the seven regional exchanges either serve a bespoke customer demand with regard to lower listing costs than in Frankfurt or specialise in derivatives or exchange traded funds (ETF) trading. Currently, the exchange is operated by the Deutsche Börse Group and offers two different market divisions for companies to raise equity capital. The first consists of the “Prime Standard” (PS) and the “General Standard” (GS), both of which are regulated by laws of the European Union (EU). Compared with listings in the GS, companies listed on the PS have to meet stricter transparency requirements to be included in the main indices, such as the DAX (Deutscher Aktienindex), the German stock market index that comprises the 30 largest public listed companies. The other division is represented by the “Entry Standard” (ES) and is an open market that is regulated by the exchange itself and therefore has lower listing requirements. As of 31 March 2014, the total amount of listed companies was 721, broken down as follows: 325 (PS), 214 (GS) and 182 (ES). On a daily basis, the average equity trade volume was approximately €3.83 billion with a total market capitalisation of nearly €1,260 billion.[5] Regarding the relatively smaller size of the German stock market in terms of the amount of listed companies, this research considers data from companies listed on the PS, GS and ES.

3.5 Research methods

To answer the research questions, this thesis consists of three empirical chapters and their respective structure will be outlined subsequently. In general, the F-Score investment strategy presented in Piotroski’s (2000) key paper is replicated first and then extended using UK stock market data[6]. In the next step, the analysis deals with the question of the particular drivers of this strategy. Finally, the strategy is applied to the German stock market and the results for the two markets are compared. The process of data collection for both the UK and Germany is performed analogously and all of the empirical chapters draw on their respective data set. All the necessary information is sourced from the Datastream (DS) database. The main part of the data is used to calculate the nine accounting ratios that form the F-Score composite measure,

[5] Source: Frankfurt Stock Exchange (http://www.xetra.com)
[6] A summary of the research methods is provided in the appendix.

unless this ratio is readily available. The remaining information is used to extend the previous research results. In total, 3,089 UK and 973 German firms are included in the data collection process for the period 1992 to 2010.

3.5.1 Replication and extension

As described before, the key paper on which this thesis is based was published by Piotroski (2000) and attracted sizeable attention within the academic community. In short, an investment strategy based on a composite measure of nine financial statement items, called the F-Score, was found to provide abnormal annual stock returns. In chapter 4, the thesis seeks to test whether the results of the same strategy applied to the UK stock market are as promising as those in the US. However, in contrast to the US study, no distinction is made between high and low book-to-market firms due to the relatively smaller UK stock market. Section 3.5.1 details the contents of chapter 4 and is mainly concerned with the UK data. Chapters 5 and 6 apply the methods presented here and differences are highlighted in the respective later sections of this chapter.

3.5.1.1 Firm selection

To be included in the analysis, firms need to have a stock market listing in the UK. Further, only firms with their reporting currency denominated in British pounds (GBP) and WorldScope data are considered. After the application of this filter, 4,481 firms remain and are eligible for further analysis. In the next step, all the necessary data for the F-Score calculation are obtained. As the F-Score is based on annual financial statement data, the following yearly accounting data from UK firms is downloaded for the period from 31 December 1990 to 31 December 2011:

Table 3.2 Datastream items and description

Item		***Description***
FYE (D)	→	financial year end [DS code: WC05350]
MP-FYE (D)	→	market price at the end of June [DS code: RI] (30/06/1991–30/06/2012)
BV-FYE (D)	→	book value at the financial year end [DS code: WC05491]
BM-FYE (C)	→	book value of outstanding shares (= BV-FYE / MP-FYE)
PB-FYE (D)	→	price to book value at the financial year end [DS code: PTBV]
NIN (D)	→	net income before extra items and dividends [DS code: WC01551]
TAS (D)	→	total assets [DS code: WC02999]
ATAS (C)	→	average total assets ($ATAS_t = (TAS_t + TAS_{t-1}) / 2$))
CFO (D)	→	cash flow from operating activities [DS code: WC04860]
LTD (D)	→	long-term debt [DS code: WC03251]
CRA (D)	→	current ratio [DS code: WC08106]
EQO (D)	→	net proceeds from equity offerings [DS code: WC04251]
GPM (D)	→	gross profit margin [DS code: WC08306]
TSA (D)	→	net sales [DS code: WC01001]
CAP (D)	→	market capitalisation [DS code: WC08002]
		(D) = downloaded; (C) = computed

Table 3.2 lists more than nine F-Score items. This is because some accounting ratios are not readily available via Datastream and therefore need to be computed manually, such as the book-to-market and average total asset ratios. A distinction between these items is made by the letters “C” (manual calculation) and “D” (available in Datastream). Another point that needs clarification is the earlier start date of the data, which is 31 December 1990. Although the analysis starts in the year 1992 and ends in 2010, some financial statement information needed for the construction of the F-Score necessitates older accounting information, such as average total assets (ATAS). The same applies to the extended end date of 31 December 2011, which is especially required for the market prices at the end of June each year.

3.5.1.2 F-Score calculation

The next step in preparing the data for analysis is to exclude firms with erroneous data, specifically firms for which no data are available for at least one of the nine F-Score items (Datastream error code: #ERROR). It became apparent that most errors were contained in the item representing the gross profit margin (GPM) – 722 in this case. All of these errors were handled successively with the aim of keeping the loss of firm data to a minimum. This was achieved by dealing with the next highest amount of errors after one loop of erasure was performed. After this process, the final data set consists of 3,089 firms for the F-Score calculation, which means that a total of 1,392 firms are excluded from the analysis. In the final step, the F-component for each firm is computed as presented in table 3.3.

Table 3.3 F-Score computation

No.	*F-component*	*Section*	*Description*	*Computation*
1	ROA	P	return on assets	NIN_t / TAS_{t-1}
2	CFO	P	cash flow from operations	CFO_t / TAS_{t-1}
3	ΔROA	P	change in return on assets	ROA_t – ROA_{t-1}
4	ACCRUAL	P	accruals	(NIN_t – CFO_t) / TAS_{t-1}
5	ΔLEVER	LLSF	change in leverage	LTD_t / $ATAS_t$
6	ΔLIQUID	LLSF	change in liquidity	CRA_t – CRA_{t-1}
7	EQ_OFF	LLSF	seasoned equity offering	N/A
8	ΔMARGIN	OE	change in gross profit margin	(GPM_t / TSA_t) – (GPM_{t-1} / TSA_{t-1})
9	ΔTURN	OE	asset turnover ratio	(TSA_t / TAS_{t-1}) – (TSA_{t-1} / TAS_{t-2})

According to Piotroski (2000), each F-component is part of an overarching financial performance indicator, such as i) profitability (P), ii) leverage, liquidity and source of funds (LLSF) and iii) operating efficiency (OE). The third column of table 3.3 highlights the respective affiliation, and a tilt towards measures of profitability (P) should be noted.

Following this, a value of either 1 or 0 is assigned depending on the past year’s trend of each F-component. If, for instance, a company increased its share capital, the variable EQ_OFF would have received a value of 0 because this action is regarded negatively. Likewise, if the company reduced its financial gearing, ΔLEVER would have received a value of 1 because less debt reduces the interest payments and increases the return on equity, a positive signal for shareholders. After translating the results into a binary matrix for each firm and each year, the values of each F-component are added up, resulting in the final F-Score composite. The stocks with the highest (lowest) F-Score of 9 (0) are deemed to feature the highest (lowest) probability of generating abnormal one-year returns.

3.5.1.3 Benchmark definition and calculation of returns

After the construction of the composite score, a benchmark against which one-year stock returns are measured needs to be defined in the next step. In general, this goal can be achieved in various ways. To calculate the abnormal return, or α, one possibility is to use the annual return of an index such as the FTSE All-Share Index and deduct this value from each firm's stock return. This approach seems obvious at first because it represents a direct link to the investing practice. Fund managers pick certain stocks out of the available universe of securities and the portfolio returns are then compared with the overall market return. However, as not all companies supply the necessary data for the F-Score, fund managers are restricted in their choice of suitable stocks. Moreover, it is arguable which index is most suitable for use as a benchmark. The alternatives encompass the FTSE 100, FTSE 250 or FTSE 350, which are provided by the London Stock Exchange.[7] However, these indices mostly consist of companies with high market capitalisation and contain a relatively small amount of companies.

For this reason, the market-adjusted returns are computed using only those firms with available F-Score data (3,089), thus creating a bespoke F-Score Index as a benchmark. However, this approach poses another kind of problem. Some firms, especially for the early 1990s, provide accounting data only partially, so a maximum F-Score of 8 or less can be achieved. Stock returns would mostly be incomparable across firms and thus render any analysis futile. To allow for this variability, which is also caused by firms exiting the market or mergers and acquisitions, a flexible approach is implemented in that market-adjusted returns are computed based only on firms with all nine F-components. This implies that the number of firms representing the F-Score Index is allowed to vary from year to year.

However, because the UK data set is much smaller than the corresponding set from the US, it can be argued that outliers are expected to have a much bigger influence on the calculation of the F-Score Index and ultimately on the market-adjusted stock returns. This can be even more severe because no distinction is made with regard to the book-to-market value of firms compared with the original study. Although this problem is naturally mitigated for the F-components due to the binary nature of the F-Score, it is assumed to have a more serious impact on the respective market-adjusted returns, which are either in units of percentage or

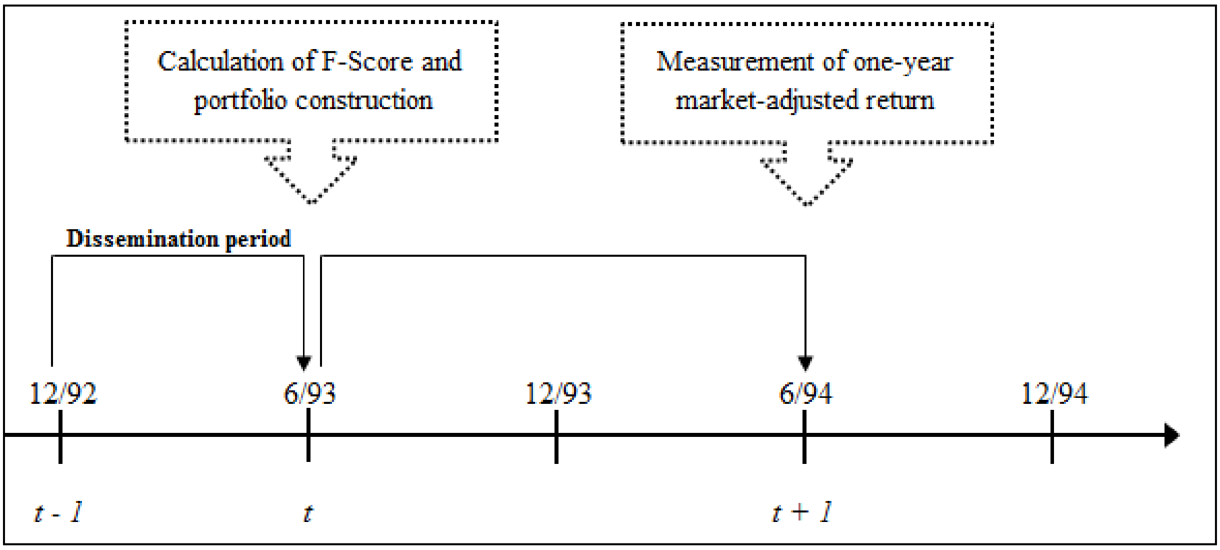
[7] Source: London Stock Exchange (http://www.londonstockexchange.com)

standardised ranks. For this reason, in addition to calculating the annual bespoke F-Score Index, the benchmark return is also winsorised at the 0.5% level using Stata before the market-adjusted returns for each stock are calculated.

Firm-specific returns are one-year buy-and-hold returns and are calculated on 30 June each year. The choice of this date is made to ensure that all the accounting information from the previous financial year is publicly available. The firms' individual buy-and-hold abnormal return (BHAR) or market-adjusted return equals the difference between the annual stock return and the benchmark return. It should be noted that the benchmark index is equally weighted rather than value weighted. This method is in line with Barber and Lyon (1997), who advocated the calculation of simple buy-and-hold returns on a sample firm less the simple buy-and-hold return on a reference portfolio or benchmark index, respectively. As the authors showed, the calculation of long-run abnormal returns can be subject to problems. However, this is mostly the case in studies in which either daily or monthly abnormal returns are summed with the result of a firm-specific one-year cumulative abnormal return (CAR). Problems can be caused, for instance, due to a rebalancing bias. In this case, the firm's individual CAR would be computed without rebalancing, whereas the reference portfolio would be subject to monthly rebalancing. The same applies if firms are newly listed and included in the reference portfolio. Since the reference portfolio is readjusted only once a year on exactly the same date as the stock returns are measured, these problems are mitigated in the present analysis. The BHAR in each of the fquintiles/terciles is equally weighted, similar to Zhang (2007) and Dechow et al. (2008). It should be noted that the term "fquintile" is introduced to distinguish between quintiles in the strict sense of the word (as used in referenced work) and the portfolios used in the present thesis. Due to an uneven distribution of firm years as presented in table 3.4, the fquintiles, i.e. portfolios, do not contain the same amount of firms except for H*/L* and H/L portfolios. This means that, for instance, the M and L portfolios are unequal. However, as the statistical analysis is only concerned with the H*/L* and H/L portfolios, this has no influence on the research outcomes. This terminology is maintained throughout the thesis and should not be confused with referenced work where quintiles are used according to their usual definition.

Equipped with these data, it is now possible to reconstruct Piotroski's (2000) original work as shown in the example of figure 3.1 below, in which the portfolios based on F-Scores from period *t - 1* are constructed in June of *t*. In the following year, at *t + 1*, the subsequent market-adjusted one-year returns are computed for each portfolio. This example replicates the situation of an investor in June 1994 and illustrates that ex post bias is not present.

Figure 3.1 Visualisation of F-Score construction



3.5.1.4 Portfolio construction with the original F-Score

To perform the statistical analyses, five portfolios are constructed each year. After calculating the F-Score for each company, stocks are assigned to one of these portfolio fquintiles depending on their F-Score. As mentioned earlier, portfolios are constructed using equal weighting to achieve consistency with the equally weighted BHAR described in the previous section, similarly to Hirshleifer et al. (2004). This approach deviates from Piotroski (2000) because, originally, portfolios were equally weighted and BHARs were value weighted. However, the main dispute exists with regard to the construction of equally weighted portfolios. According to Lewellen (2010), portfolios should be built using value weighting. However, three points justify the use of equal weighting in the present analysis. Firstly, since this thesis builds on the original F-Score strategy of Piotroski (2000), it appears useful to follow the original method to ensure comparability. Secondly, Baker and Wurgler (2006) formed equally weighted decile portfolios while researching the relationship of investor sentiment and stock returns in the cross-section. They argued that value weighting is likely to distort the relevant patterns because larger firms are generally less influenced by investor sentiment. Likewise, the underlying assumption of the thesis is that stock markets are not efficient at all times but rather subject to non-rational behaviour. Therefore, it is preferable to isolate behavioural biases to examine those using respective measures, such as information uncertainty or liquidity, which are described later. Thirdly and lastly, Benartzi and Thaler (2001) found that most investors use a diversification heuristic when faced with the task of asset allocation. This so-called 1/n rule is commonly applied in pension plans, in which contributors spread their funds evenly across asset classes (Huberman and Jiang, 2006). As cited by Benartzi and Thaler (2007), even Nobel laureate Harry Markowitz admitted: "I should have computed the historic covariances of the asset classes and drawn an efficient frontier. Instead, […] I split my contributions fifty-fifty between bonds and equities."

At this point, no distinction is made between stocks regarding their size or book-to-market ratio and therefore only five portfolio fquintiles exist. In other words, the portfolios are one-dimensional and referred to as one-way portfolios. The cut-off points of the fquintiles are derived from the distribution of firm years of the F-Score and remain constant throughout the entire analysis to make the results comparable. However, it is noteworthy that these cut-off points vary between the UK and Germany because the respective distributions are different. As presented in table 3.4, there are a total of 20,053 firm-year observations for the UK and 5,944 for the German stock markets, respectively. Firm years are defined as the cumulative amount of firms that provide a full set of relevant data, which is needed for the F-Score calculation for the period 1992 to 2010. According to the absolute percentages of the frequencies in the third column of table 3.4, the UK F-Scores appear to be relatively evenly distributed, with the lowest and highest fquintiles representing 5% of the data. The reason for that is most likely to be the higher amount of observations. More precisely, the lowest fquintile, defined as L*, contains the lowest 5% of F-Score stocks, ranging from 0 to 2. The lowest fquintile, in turn, is contained in the low (L) fquintile and therefore ranges from 0 to 3. Stocks with an F-Score between 4 and 6 are assigned to the middle fquintile, defined as M. Analogous to the bottom fquintile, the high fquintile (H), with F-Scores of 7 to 9, encompasses the highest (H*) scoring 5% of the data with F-Scores of 8 and 9, respectively. Therefore, the cumulative distribution is such that L = 30%, M = 40% and H = 30% of all firm years and both L* and H* are included in L and H, respectively. As mentioned, this segmentation is applied to all the fquintiles in the analyses. The general assumption is, of course, that the better the F-Score, the better the subsequent one-year market-adjusted return. The results for the German stock market are presented in chapter 6 and are based on different cut-off points. From table 3.4, it follows that L* contains F-Scores from 0 to 3, L from 0 to 4, M from 5 to 6, H from 7 to 9 and H* from 8 to 9, respectively. In general, the low fquintile contains about 30%, the middle fquintile 40% and the high fquintile 30% of the data, which is similar to the UK. However, the lowest and highest fquintiles contain approximately 10% of the data.

Table 3.4 Firm years of the F-Score

UK	Firm years			**DE**	Firm years		
F-Score	***Freq.***	***Percent.***	***Cum.***	***F-Score***	***Freq.***	***Percent.***	***Cum.***
0	13	0.06	0.06	0	7	0.12	0.12
1	203	1.01	1.08	1	43	0.72	0.84
2	875	4.36	5.44	2	182	3.06	3.90
3	2,177	10.86	16.30	3	529	8.90	12.80
4	4,091	20.40	36.70	4	892	15.01	27.81
5	4,875	24.31	61.01	5	1,317	22.16	49.97
6	4,109	20.49	81.50	6	1,297	21.82	71.79
7	2,550	12.72	94.22	7	1,036	17.43	89.22
8	999	4.98	99.20	8	527	8.87	98.08
9	161	0.80	100	9	114	1.92	100
Total	**20,053**	**100**		**Total**	**5,944**	**100**	

Key: Freq. = Frequency; Cum. = Cumulative

However, for reasons of practicability, the cut-off points are set exactly at 5% (10%) for the lowest and highest UK (German) fquintiles, with the remaining fquintiles being cut off at 30% (L), 40% (M) and 30% (H). To test for statistical significance and robustness, both the mean and the median returns are compared across the L/H and L*/H* portfolio fquintiles.

Having replicated Piotroski's (2000) original work using the F-Score measure, the next step is to test whether an alternative to the F-Score is able to deliver equally good results. The reasons for the extension of the original approach are manifold. Firstly, a point of criticism is that the original measure is based on a binary signal of either 0 or 1. Hence, it does not take into account the magnitude of the signal, which potentially could be very useful. Secondly, all nine signals are equally weighted. This is problematic because some signals could be more closely related to the return performance than others, thus omitting useful information once again. Allowing the weights to differ, for example related to past forecast accuracy, could potentially improve the trading gains. Thirdly, it is necessary to provide practitioners with better and/or additional selection tools to filter out the stocks with favourable return prospects. Finally, the alternatives can be used as a reciprocal robustness check for the initial F-Score measure. To distinguish Piotroski's F-Score from its alternative measure, the original version is referred to henceforth as the "F-Score".

3.5.1.5 Portfolio construction with an alternative to the F-Score

The alternative measure is the result of a ranking and standardisation process. This procedure is necessary to ensure commensurability between the nine F-components. The accounting information is ranked using the respective rank function in Excel, in which the same values are assigned the same ranks. The next step is to standardise the ranks so that each F-component takes on a value between 0 and 1 using the following equation:

$$F_{i,srank} = \frac{(F_{i,rank} - 1)}{(F_{max} - 1)} \tag{1}$$

The left side of the equation, $F_{i,srank}$, represents the standardised rank, which equals the individual rank of the F-component minus 1 divided by the highest rank of F-components in one particular year minus 1. To avoid confusion, it should be noted that “highest” in this regard does not mean the first but the highest rank in absolute value. Therefore, the highest rank for each single year is theoretically limited to 3,089 for the UK and 973 for the German data, respectively. Theoretically, because not all the firms report relevant data for the F-Score, the calculation and ties are retained. This means that the higher the rank, the worse the performance of the F-component relative to its peers. The only exemption applies to the variables capturing change in financial gearing and seasoned equity offering, which are listed in table 3.3 as numbers 5 and 7. In these cases, higher ranks are assumed to have a negative impact on future stock returns because more interest has to be paid for debt and less income is available due to an increased shareholder base, as mentioned before. To keep the analysis clear, the term “Fi-rank” henceforth refers to the standardised ranked F-component of the individual firm, while “F-rank” represents the aggregate of the nine Fi-ranks for each company. These terms are used throughout the remainder of the thesis.

Compared with the F-Score, the F-rank method takes into account the differences in values within each F-component and thus should lead to a more accurate representation of the financial strength of a company. For instance, while a company receives a 1 if its return on assets is positive, it does not matter how much greater than zero this value is due to the binary nature of the F-Score. Assuming that company A only marginally but consistently misses the threshold of obtaining a score of 1 for the respective F-component, it could end up receiving an F-Score of 0, while company B would be assigned a value of 5. In the case of the F-rank, the picture would look different, though. If company B narrowly exceeds the threshold in five out of the nine Fi-ranks but substantially misses them in the other ones, the overall F-rank would be low. The F-rank of company A by contrast would then be higher because it shows a constant performance in all nine F-ranks. Consequently, both the portfolios and the results would appear different in such a case.

After the ranks of each F-component are standardised (Fi-rank), the next step is to aggregate all the values into the F-rank. To avoid bias towards a specific method, both the mean and the median of the nine Fi-ranks is taken to construct the F-rank. Although the utilisation of ranks already ensures the moderation of outlier influence, this approach is another step in this direction compared with a simple addition of values. In comparison with the F-Score, with its absolute values, F-rank portfolio fquintiles necessitate flexible cut-off points that are recalculated on an annual basis. For this purpose, the percentage distributions summarised in column three of table 3.4 (UK) are used as a guideline. For instance, the lowest portfolio (L) will be generated by multiplying the average and median F-ranks by 30%, which will result in one cut-off point. Each firm with a lower F-rank will be assigned to this portfolio. The procedure is analogous for the middle, high, lowest and highest portfolio fquintiles.

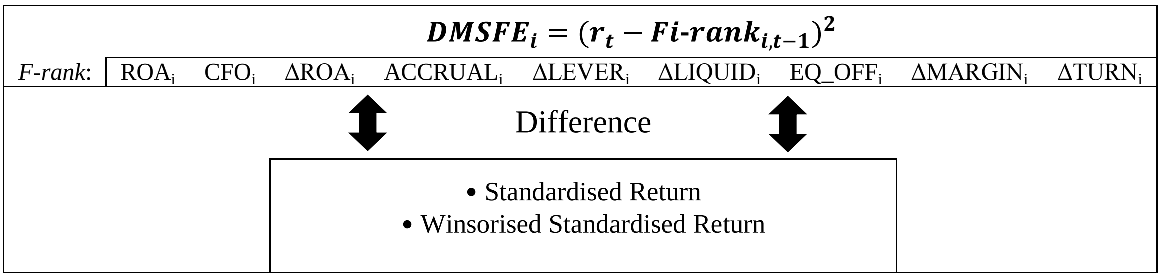
3.5.1.6 Alternative combination measures of financial strength: The weighted DMSFE

The original F-Score is a relatively crude measure of the financial health of a company because, apart from an all-or-nothing characteristic, it disregards historical information that might be useful for stock return forecasts. Therefore, the thesis extends the initial approach by using an ex ante means of forecasting, which weights the nine F-ranks using past data. The basic idea originates from a paper by Bates and Granger (1969) and makes the assumption that individual forecasts might contain independent "information" that can result in more accurate predictions if combined. As outlined earlier, the nine F-Score components cover a firm's i) profitability (P), ii) leverage, liquidity and source of funds (LLSF) and iii) operating efficiency (OE). Similar to the trends in the academic research focus, investors might pay particular attention to the cash flow (i.e. profitability) in the mistaken belief that this figure is hardly subject to manipulation. To moderate this effect, a combination of forecasts appears to be a fruitful strategy.

This idea was also picked up by Rapach et al. (2010) and the authors noted that forecast combination techniques have recently become popular again in economics, whereas the application in the finance literature is somewhat scanty. In contrast to the extensive use of time series applications in the economics area, this study implements forecasting methods in a cross-sectional environment. This means that to predict the *t + 1* stock return, the cross-section of firms in a specific year is used to establish the forecasting weights rather than using multiple years' data of the same firm. Because of the benefits of forecast combinations in time series settings, these should be also present in the cross-section.

To generate reliable forecasts, the forecast errors have to be computed first. This is achieved by squaring the differences between the nine Fi-ranks of a firm at time *t - 1* and its subsequent one-year (winsorised) stock return (r) at time *t*. The result is the discounted mean square forecast error of each Fi-rank, defined as $DMSFE_i$. To have consistent units of measurement, the percentage stock returns are ranked and standardised according to equation (1) in the previous section. Figure 3.2 illustrates the process.

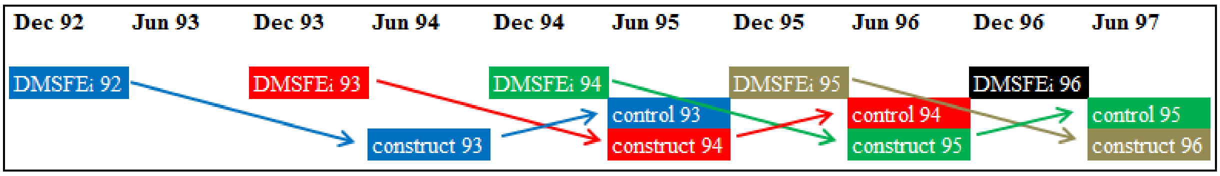
Figure 3.2 Illustration of $DMSFE_i$ computation



The discount factor is equal to 1 throughout the analysis because only a one-year rolling window is modelled. This approach is justified because portfolios are readjusted every year in the original F-Score strategy. It is also assumed that investors build their portfolios each year in June for two reasons. On the one hand, according to listing rule 12.42(e) of the Financial Conduct Authority (FCA), all public companies in the UK are obliged to publish audited annual accounts within six months of their period end.[8] As this thesis starts from the premise that markets are not fully efficient, this ensures that some time is given for the data to be processed by investors. On the other hand, the analysis aims to resemble the situation of an actual investor as closely as possible and thus to avoid look-ahead bias. The procedure is presented in figure 3.3.

[8] Source: Financial Conduct Authority (http://www.fca.org.uk)

Figure 3.3 Visualisation of DMSFE portfolio construction



From the perspective of an investor in June 1994, the financial statement information of 1992 is publicly available by June 1993 the latest. The investor is then able to calculate the one-year stock return from June 1993 to June 1994 and to construct the individual forecast errors or the $DMSFE_i$ using the 1992 financial statement information. Having constructed the portfolios in June 1994 based on the $DMSFE_i$, the investor will check the success of the investment strategy in June 1995. The underlying assumption is that the stock return should be higher for firms with a lower $DMSFE_i$.

However, the question is now how to assign stocks to the respective portfolio fquintiles in June 1994. Following Rapach et al. (2010), a more sophisticated weighting method is adopted in this thesis, which is derived from Stock and Watson (2004). According to this approach, weights are attached to each forecast contingent on the value of its $DMSFE_i$, whereby the lower (higher) the $DMSFE_i$, the higher (lower) the weight. All the weights are allowed to change every year during the period 1992 to 2010, similar to the original F-Score described earlier. The weights are computed using equation (2) and subsequently applied to next year's Fi-ranks, indicated by the blue path in figure 3.3. In other words, the average of each of the nine $DMSFE_i$s is computed first. After that, equation (3) yields the weightings for the forecasts of the forthcoming period.

$$\Phi_{j,t} = \frac{\sum_{i=1}^{N} \left(r_t - F\text{-}rank_{i,t-1}\right)^2}{N} \tag{2}$$

where

$$\omega_{i,t} = \Phi_{j,t}^{-1} \Big/ \sum_{j=1}^{9} \Phi_{j,t}^{-1} \tag{3}$$

Afterwards, the highest three, the next highest three and the lowest three weights are labelled as high, middle or low. Within each of these terciles, the respective means of the three weights are computed. Finally, this average weight is multiplied by next year's Fi-ranks that are members of the respective tercile. In other words, if, for instance, the forecast error is lowest for the first three profitability measures, that is, ROA, ΔROA and CFO, their weights are assigned to the highest tercile. After computing the average weight in this tercile, their individual Fi-ranks are multiplied by this value. Following the blue path in figure 3.3, an investor in June 1994 establishes the weights using the Fi-ranks of 1992 and the 1994 ranked returns and applies these weights to the 1993 Fi-ranks. The one-year-ahead returns are then measured in June 1995. This procedure is only undertaken in those cases in which all nine F-Score components of a firm can be computed. The portfolios and cut-off points are constructed as usual.

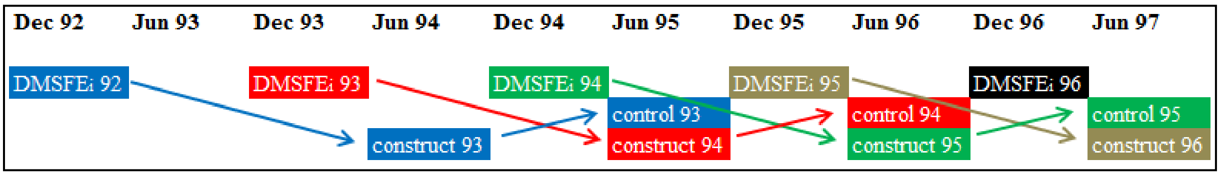
It should be noted that, due to a short holdout period at the start of 1992, the first portfolio is constructed in June 1994 only based on 1992 accounting data as described before. From this time onwards, the investment process and collection of accounting information are performed simultaneously. Furthermore, to check for consistency of the results, the analysis is extended in that it studies the performance of this strategy over a two-year period since portfolio construction.

3.5.1.7 Alternative combination measures of financial strength: Clusters

The next combination measure is motivated by a study conducted by Aiolfi and Timmermann (2006) and represents a variation of the previously introduced DMSFE method. The authors sorted forecasting models into clusters and found that these combination methods improve the quality of forecasts. Similar to the DMSFE procedure, clustering accounts for past financial statement data and could therefore improve an investment strategy that does not, such as the F-Score. The difference between the method of DMSFE and the method of clustering is that the latter filters out those forecasts ($DMSFE_i$) with the best past-year performance, that is, the lowest forecast error, and drops the remaining ones entirely. Besides, clusters are formed at time t but applied to accounting data at time $t + 1$. The communality is that the forecast clusters are allowed to change from year to year. In the present analysis, cluster sizes of $k = 3$ and $k = 2$ are considered. Because there are nine individual forecasts, the data set cannot be split evenly. To increase the depth of the analysis, five (three) forecasts are clustered, while the remaining four (six) are discarded in the case of $k = 2$ ($k = 3$) clusters. This is justifiable because not much difference between the performances of three ($k = 3$) and four forecasts (k

= *2*) is expected. Again, figure 3.4 depicts the process from the perspective of an investor in June 1994, highlighted in blue.

Figure 3.4 Visualisation of cluster portfolio construction



For each firm and year, the individual forecast errors are computed as the squared difference between the 1992 F-ranks and the standardised one-year return from June 1993 to June 1994 ($DMSFE_i$ 92). In the next step, each of the resulting nine forecast errors are averaged across the firms, as presented in equation (2). Following that, these nine averaged forecasts are assigned to either one of three or one of two clusters. This means that only the three (five) forecasts with the lowest average forecast error are considered in the case of $k = 3$ ($k = 2$) clusters. Finally, portfolios are constructed based on an average cluster score, or c-score, that is computed for each firm using the 1993 F-ranks, which are publicly available for June 1994. The cut-off points are generated as usual and the performance of each of the five portfolios is evaluated for June 1995.

3.5.1.8 Inclusion of a second dimension: The book-to-market ratio

Thus far, the analysis has only been concerned with firm characteristics that are company-dependent or internal. For these purposes, the F-Score and F-rank have been used to construct one-way portfolios ranging from L* to H*. This one-dimensional approach is now extended by including company-independent or external company characteristics. The first one is the book-to-market ratio or, in short, BM. This is also an extension of Piotroski (2000), as the original work on the F-Score is limited to high-BM firms only. In this thesis, two-dimensional or two-way portfolios are constructed by subdividing firms into BM terciles. Therefore, there are a total of fifteen portfolios per analysis compared with the five in the one-way setting. The advantage of having a finer distinction between portfolios is apparent for two reasons. On the one hand, an investor is interested in which BM tercile is the key driver of investment performance under the original F-Score and the alternative F-rank method. In general, the empirical results from the US provide evidence for the outperformance of high-BM or value firms over low-BM or growth firms (e.g. Rosenberg et al., 1985; Fama and French, 1992). More recently, Asness et al. (2013) confirmed these findings for the UK and continental Europe using portfolios based on BM terciles. On the other hand, this kind of segmentation is useful for the subsequent empirical chapters, which deal with various other characteristics. This two-way structure will be retained throughout the thesis.

3.5.2 Driving forces behind stock returns

This section describes the methods used for the second empirical chapter. The purpose of this chapter is to determine whether the relationship between size (instead of BM) and F-Score/F-rank performance is due to size proxying for information uncertainty or liquidity. More generally, chapter 5 addresses the question of which driving forces are behind stock returns under the F-Score and F-rank strategies. For this purpose, BM as the second dimension is replaced by size and two other measures of information uncertainty and two measures of liquidity. The process of constructing portfolios and calculating cut-off points is as described in section 3.5.1 and remains exactly the same.

3.5.2.1 Information uncertainty proxy: Size

The second empirical chapter is motivated by a study by Zhang (2006), who found that greater information uncertainty leads to higher (lower) stock returns following good (bad) news. By definition, information uncertainty, or IU, is a function of either the volatility of firm fundamentals or poor information. He made the point that investors usually underreact to publicly available information and thus should underreact even more the higher the degree of information uncertainty is. An investment strategy based on IU would generate the highest returns for high-uncertainty stocks. However, information appeared to be reflected reasonably well in the prices of low-uncertainty stocks and therefore the abnormal returns should be limited. This goes hand in hand with the assumption made in the present thesis that stock markets are not semi-strong-form efficient. To capture information uncertainty, the company size is used as the first proxy. This measure is also widely applied in the empirical literature on this topic. The underlying rationale is manifold. Firstly, small firms have fewer stakeholders and are less diversified (Zhang, 2006). Therefore, the assumptions about their future prospects should be less reliable. Secondly, investors are likely to have fixed costs for acquiring information on larger companies, in which they can also accumulate large stock positions. This, in turn, makes smaller companies less attractive in aggregate (Hong et al., 2000). Finally, small firms are likely to choose a listing in stock market segments that demand less timely reporting requirements, thus reducing the availability and increasing the uncertainty of information (cp. sections 3.3 and 3.4).

To perform the statistical analysis, the same two-way portfolios as before are constructed. However, this time, the construction of the original F-Score and the alternative F-rank approach is based on firm size rather than BM terciles. Likewise, another set of two-way portfolios is built comparing the stock return performances of the DMSFE and cluster methods in relation to size. As mentioned before, the stock returns of the respective portfolio fquintiles are calculated using both the mean and the median.

3.5.2.2 Information uncertainty proxy: The SDF and MADF

Amongst others, Zhang (2006) used cash flow volatility as an alternative proxy for information uncertainty. This procedure is justified because it is highly likely that each of those proxies, for example size, is influenced by other things that are not of interest in this analysis. However, it could be argued that reliance on just one accounting number is risky. Cash flow from operations (CFO) is reported by the company and thus there is a danger of manipulation. Firm size, by contrast, is largely determined by the aggregated stock market because the market capitalisation is dependent on the stock price. Lee (2012) found that the propensity to manipulate the CFO is based on four firm characteristics, such as financial distress and a higher correlation between stock returns and CFO. The results indicated that the likelihood of manipulations is highest if the incentives for the management are high. As the purpose of the F-Score and F-rank measures is to separate winning stocks from losing ones, it could be argued that the firms with the lowest scores are likely to be in some kind of distress. Consequently, the CFO values would be less reliable and would have undue bias in the analysis. Of course, this would be the case only if they were actually manipulated. Since the effort to determine which of the companies whitewashed their CFO figures is not practicable, additional measures should be applied.

For this purpose, it is helpful to involve the remaining eight Fi-ranks – not least because they are already available from previous analyses. In the present thesis, the second IU measure is the SDF, which is defined as the standard deviation of the F-rank of each firm and year. Equation (4) illustrates its computation more formally:

$$SDF_{i,t} = \sqrt{\frac{\sum(F\text{-}rank_{i,t} - \overline{F\text{-}rank}_t)^2}{(n-1)}} \tag{4}$$

Since the SDF measures the aggregated dispersion within the Fi-ranks, it should serve as a useful tool to give some indication of a firm's information uncertainty. Therefore, the higher (lower) the SDF, the higher (lower) the information uncertainty. As previously described, the two-way portfolios are constructed using the same fquintiles and cut-off points for the mean and median F-ranks in the first dimension. The respective SDFs are subdivided into terciles in the second dimension, which results in a total of 15 portfolios. Because the original F-Score is a discrete measure that takes values of either 1 or 0, the computation of F-Score volatility would not be meaningful. However, F-Score fquintiles are still constructed, but the subdivision into terciles is based on F-ranks. The mean and median one-year stock returns are computed as before.

The third and last IU measure is the mean absolute deviation, or MADF, of the F-ranks and is based on the same underlying assumptions as described for the SDF. It is an alternative summary of dispersion and is defined as follows:

$$MADF_{i,t} = \frac{\sum |F\text{-}rank_{i,t} - \overline{F\text{-}rank}_t|}{n} \tag{5}$$

As can be seen from equation (5), the MADF is the sum of the absolute differences between each of the nine F-ranks and their overall mean divided by the factor nine. The main reason for the use of the MADF is to check the robustness of the SDF. However, there are also more subtle reasons for that. Gorard (2005) summarised some advantages of the mean deviation method and questioned the usefulness of the standard deviation (SD) as the preferred measure of dispersion at the same time. In general, he made the point that the mean deviation (MD) is superior to the standard deviation in terms of efficiency, because the latter assumes no measurement errors and a normal distribution. The process of squaring increases each of the nine distances exponentially rather than additively, as can be seen from equation (4). Calculating the square root in the next step does not remove this bias in the SD, though.

Further, even if the observations are normally distributed, small measurement errors do arise in reality. In this case, the advantage of the SD over the MD will fade and actually reverse, as referring to by Huber (1981). Moreover, the normal distribution is an idealised assumption. Gorard (2005) showed an example of how the SD is inferior to the MD if the distributions are not Gaussian. He referred directly to Fisher (1920), who proved the greater efficiency of the SD for normal distributions and highlighted its inferiority in all other cases. Finally, Gorard (2005) noted that a simpler measure, such as the MD, should be sufficient in social science research. In his view, the reason for the preferred use of the SD is of a practical nature, because this measure is more convenient for mathematical manipulation. Applied to finance, the reason for the use of the SD may be "physics envy", as mentioned earlier. As Dempsey (2013a) argued, the prestige of natural sciences in the 1950s sparked the idea that financial markets are similar to physics and therefore could be described with the help of applied mathematics. It seems likely that this academic dispute will continue and probably never be settled. For this reason, both measures of dispersion are used in this thesis. Again, portfolio construction under the MADF approach follows exactly the same procedure as described for the SDF.

3.5.2.3 Causality of investment performance: The Amihud liquidity measure

While both the SDF and the MADF are pure proxies for information uncertainty, size may also proxy for liquidity (Amihud, 2002). Therefore, size as a measure of information uncertainty may be regarded as an indistinct term. Even when relying on either the SDF or the MADF, a positive correlation between information uncertainty and stock returns might not exist. To address these issues, another possible causality of investment performance is analysed in this thesis for the UK, but not for Germany. This is due to the problem of scant data availability for the German stock market, particularly at the beginning of the 1990s. In contrast to measures of uncertainty, liquidity is a more precise measure, because this kind of data is stored by the stock exchange and can easily be accessed via Datastream. As Zhang (2006) noted, financial statement data, such as the nine F-components, are influenced by a company's information system and thus may be subject to bias. Consequently, the measures of liquidity should be at least as useful as the IU measures, because they are not generated internally.

The first liquidity measure is taken from Amihud (2002), who found a positive relationship between market illiquidity and ex ante stock excess returns in the cross-section. This result implies the existence of an illiquidity premium. It is therefore of interest to investors to know whether this premium is also related to the performance of their F-Score and F-rank portfolios, respectively. The findings also challenge the efficient market view that higher returns are associated with higher risk, which is in line with the rationale of the F-Score/F-rank approach. Equation (6) shows how the liquidity measure is calculated.

$$LIQ_{i,t} = 1 \;/\; D_{i,t} \sum_{\mathrm{t}}^{\mathrm{D_{i,t}}} |\mathrm{R_{i,t,d}}| \;/\; \mathrm{VOLP_{i,t,d}} \tag{6}$$

For each firm and each year, LIQ is the result of the absolute daily stock return in one year divided by its respective volatility expressed in British pounds. In this context, $D_{i,t}$ represents the number of days in one specific year for each firm. As a result, greater values of LIQ are associated with less liquidity and vice versa. All the necessary information was downloaded and computed manually, because the Amihud liquidity measure is not readily available from Datastream. It should be noted that, in contrast to Amihud (2002), this thesis defines the Amihud measure as liquidity rather than illiquidity. The change is in the interest of convenience and similar to the approach of Asness et al. (2013). It has no influence on the subsequent analysis, though.

3.5.2.4 Causality of investment performance: The ToR liquidity measure

To test for robustness, it is appropriate to employ a second measure of liquidity. Although used in a different context, Kuo et al. (2013) used the turnover ratio, or ToR, as a proxy for liquidity. As summarised by Banerjee et al. (2007), they justified the use of this measure on the grounds that it is widely applied in the empirical and theoretical literature, which also motivates the use in the present thesis. By contrast, Kuo et al. (2013) based their choice on the lack of a different dimension in the Amihud measure. It is defined as the daily volume of shares traded divided by the number of outstanding shares or, more formally, as:

$$ToR_{i,t} = 1 \;/\; D_{i,t} \sum_{\mathrm{t}}^{\mathrm{D_{i,t}}} \mathrm{VOL_{i,t,d}} \;/\; \mathrm{NOSH_{i,t,d}} \tag{7}$$

Equation (7) states that the ToR for the individual firm *i* in year *t* equals the daily trading volume of its stock scaled by the number of shares outstanding on this day. Because either no trading takes place or no data are available for several days, the previous result is divided by the amount of days for which data are available. While the Amihud measure captures the price impact dimension, the turnover ratio captures the trading quantity dimension. The authors argue that in theory there is a negative correlation between stock liquidity and dividend payments. This implies that rational investors prefer highly liquid firms to illiquid firms. If this argument holds true, investors are assumed to apply higher discount rates in their evaluation process, which in turn will result in lower valuations for illiquid stocks. To increase their valuations, firms are more likely to pay dividends. Although this thesis is not concerned with the relationship between stock liquidity and a firm's decision to pay dividends, its logic can be applied to the F-Score/F-rank strategies. Assuming two exactly similar firms, A and B, of which B pays a dividend and A does not, the stock price of A should be lower due to the higher discount rate applied by investors. According to Bulan et al. (2007), dividend-paying firms in the US feature characteristics that would categorise them as high-BM rather than growth firms. More precisely, these are generally large, relatively profitable firms with cash balances and restricted growth opportunities. These results were substantiated by Denis and Osobov (2008) for the US and, amongst other countries, for the UK. This in turn raises the interesting question of whether the stocks of the more illiquid BM firms show a better one-year performance than their growth counterpart. The study by Piotroski (2000) is based on this assumption and his decision to focus entirely on high-BM firms is justified by previous findings that suggest a positive relationship between the BM ratio and the subsequent stock returns (e.g. Fama and French, 1995; Chen and Zhang, 1998). The reason for this is the fact that high-BM firms are financially distressed and therefore investors require a risk premium. However, this is directly contrary to the finding that dividend-paying firms are generally financially healthy and belong to the high-BM group of firms at the same time. This so-called distress risk puzzle dates back to earlier work, such as that of Dichev (1998). He concluded that investors who take on bankruptcy risk are not compensated by higher stock returns. Other studies have reached similar conclusions, most recently Kim (2013), who summarised the work carried out until then.

In summary, the ToR measure serves primarily as a liquidity measure, as described by Kuo et al. (2013). As an interesting side effect, it can also be used to categorise the data as either low-BM (growth) or high-BM (value) firms. Consequently, the analysis i) shows how liquidity is related to stock returns, ii) determines whether those returns are higher for high-BM firms, as assumed by Piotroski (2000), and iii) tests for robustness. The last point is based on the fact that Kuo et al. (2013) found that the Amihud measure was not suitable for predicting the propensity to pay dividends. However, the ToR was reported to be significant for the UK and all other markets subject to their research. This should increase the reliability of the results once applied to the F-Score and F-rank strategies.

3.5.3 F-Score and F-rank: The case of Germany

The final empirical chapter explores how the F-Score and F-rank investment strategies perform in Germany. This is of interest for three reasons. Firstly, as mentioned in the literature review, research on the German stock market is limited (Amel-Zadeh, 2011). This is rather surprising as the German economy is the largest within the EU and plays a vital part in the global context as well.

Secondly, the legal system is based on code law, in contrast to the UK, which has a common law system. Originally, the F-Score measure was tested for the US only, where a common law system is customary. Because of the similarity of the legal systems between the US and the UK, investors are likely to be interested in the applicability of the strategy outside common law countries. This desire may be rooted in the different regulations with regard to investor protection. In an influential paper by La Porta et al. (1997), the authors showed that countries with poorer investor protection usually have smaller capital markets, such as Germany. Another possible way of measuring investor protection is to assess the quality of accounting information. Bartov et al. (2005) found that earnings prepared under both the US GAAP and the IAS (now IFRS) show greater value relevance to subsequent stock returns than under the German GAAP (HGB). As they argued, the reason is that the focus of the US GAAP and IFRS rests on the shareholder rather than the stakeholder. This goes hand in hand with the results of Barth et al. (2008), who confirmed the better information quality of firms reporting under the IFRS in contrast to domestic accounting principles. However, differences in stock returns due to differing accounting standards are likely to fade in the future. This is because the EU decided in 2002 to harmonise the accounting standards of the member countries and make the IFRS a binding principle for listed companies with effect from 1 January 2005.[9]

[9] Source: Regulation EC No. 1606/2002: (http:eur-lex.europa.eu)

Nevertheless, as the majority of the data in this thesis are analysed before 2005, the cause of the differences between the UK and the German results may be differing accounting standards.

Thirdly and lastly, the German stock market is structurally different from that of the UK, because firms rely less on equity capital and more on debt financing, which has implications for stock valuations. As documented by Bartov et al. (2005), the German GAAP allows greater leeway in determining the amount of provisions than the IFRS. In general, those changes in provisions directly affect the three sections presented in table 3.3, for example return on assets, financial gearing or leverage and the gross profit margin. This, in turn, has direct implications for the computation of the F-Score and F-rank, respectively. Consequently, the rationale for the variations in provisions may not be obvious to investors who rely on an automated F-Score/F-rank investment strategy, as presented here. If, for instance, a firm has had a stable F-Score over a period of time and then the management decides to increase the provisions, its F-Score might drop dramatically. The investor would not be able to clarify easily whether this decision is really justified or whether the directors intend to smooth earnings, as exemplified by Bartov et al. (2005).

In general, the application of the investment strategy to the German stock market appears to be even more justified against the background of market efficiency. Traditionally, a large number of German companies rely on debt capital, which is usually provided by banks. Because the bank decides whether to approve the loan, the incentive for firms to provide timely and publicly available accounting information is relatively low (Bartov et al., 2005). The consequence is that stock markets should be less efficient and therefore more opportunities are available to earn abnormal returns. However, Daske and Gebhardt (2006) found a statistically and economically significant increase in the disclosure quality for German companies that introduced the IFRS. The results were confirmed for both Austria and Switzerland, which have a common accounting tradition with Germany. The portfolios in this last empirical chapter are constructed in exactly the same way as described previously. However, compared with the UK, the analysis with regard to the two liquidity measures (Amihud and ToR) was not conducted on the German stocks due to the limited data availability.

3.6 Empirical tests

The following section provides the rationale for choosing the respective statistical tests and describes how they were applied to the data set. It summarises the portfolio construction process and the subdivision into terciles through visualisation. The methods are explained using the UK stock market data but are transferable without any modification to the German counterparts. However, as mentioned earlier, no tests are undertaken for the liquidity measures of German companies.

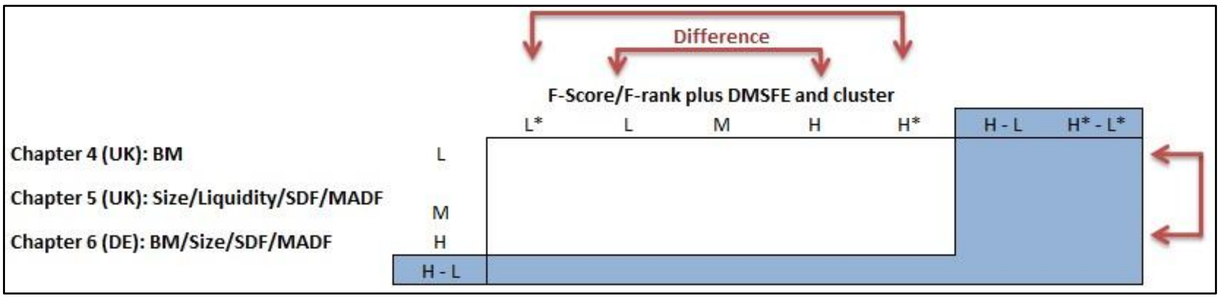
3.6.1 The rationale behind the choice of statistical methods

Referring to the aforementioned term “physics envy”, which was coined by Lo and Mueller (2010), the intention for the choice of statistical methods is to consider its practicability for the general investor. This is compliant with Piotroski’s (2000) original idea of introducing a simple accounting-based fundamentals analysis strategy that works for the US stock market. It follows from this consideration that if a highly sophisticated econometric analysis is needed to provide any statistically significant results, the investment strategy is probably not as efficient in a European setting. As Welch (2001) observed, academic journals are more likely to publish significant rather than insignificant results. In his view, the danger is not to scrutinise sufficiently whether the methods used are reasonable but to focus too much on mild statistical significance. This point ties in with Dempsey (2013), who progressed even further and stated that econometrics nowadays is an exercise in data mining in the finance discipline. Likewise, Moosa (2013) mentioned that econometrics should be used as a means to an end rather than an end in itself. This statement is corroborated by the invention of highly sophisticated models, such as autoregressive conditional heteroscedasticity (ARCH) and its successors. This trend has at least two consequences, which are considered in the subsequent analyses. On the one hand, some of the simpler econometric applications should provide results that are as good as the more complex techniques. This was shown, for instance, in multiple regressions in which the weighted combination of the exogenous variables predicts the outcome of the endogenous variable (Dawes, 1979). Equal weighting of the predictors can provide just as accurate forecasts if a score is used. To compare the scores, ranking and standardisation methods should be used. As described before, the preparatory work has already been conducted by establishing both the F-Score and the F-rank measure of financial health. On the other hand, Shiller (2010) noted student dissatisfaction with the teaching of economics. The reason is that students are under the impression that the lecture materials are hardly applicable in the real world. In contrast, this seems not to be the case in the physics profession. As the author argued, this perception is mainly rooted in the fact that some economic theories are disproved and therefore transitory, which in turn implies that they are rendered useless in the future. As this issue is innate in the economics profession and most likely in finance as well, Shiller (2010) advocated the influx of views from other social sciences, such as anthropology or psychology. This thesis intends to take a step in that direction.

3.6.2 Visualisation of the test methods

To illustrate the statistical analyses performed on the portfolios, figure 3.5 depicts the process of comparing the stock returns between the fquintiles.

Figure 3.5 Visualisation of the statistical analyses



As can be seen, the horizontal axis is divided into the F-Scores/F-ranks from their lowest (L*) to their highest (H*) values. The more sophisticated combination techniques of DMSFE and clusters are considered subsequently, as described above in sections 3.5.1.4 to 3.5.1.8. The vertical axis is subdivided into terciles and represents the second dimension, as detailed in sections 3.5.1.8 and 3.5.2. Of interest are the horizontal and vertical return differences between the high–low (H–L) and highest–lowest (H*–L*) portfolio fquintiles. Section 3.5.1.3 describes how the stock returns are computed. In general, there are four different types of those returns to avoid undue influence of extreme outliers. These are i) percentage returns, ii) standardised ranked returns, iii) winsorised percentage returns at the 0.5% level and iv) winsorised standardised ranked returns at the 0.5% level. This means that if, for instance, the SDF terciles (vertical) are combined with the F-rank (horizontal) in chapter 4, the results are four tables for which portfolios are compared both vertically and horizontally.

3.6.3 Parametric and non-parametric tests

The first statistical test[10] conforms to the methods used in the original work on the F-Score by Piotroski (2000). At the same time, it follows the analysis by Zhang (2006) regarding measures of uncertainty. As one part of the thesis intends to combine the original accounting-based investment strategy with IU measures, this procedure ensures consistency and comparability. In general, unpaired two-sample t-tests are conducted to test the equality of means of the percentage stock returns between fquintiles with high/low and highest/lowest F-Scores and F-ranks and their respective combination techniques, namely DMSFE and clustering. The same tests are then applied within fquintiles across terciles, which represent the second dimension, for example BM or size. All the tests are performed using Stata.

As noted by Kothari and Warner (1997), long-horizon parametric test results, such as those in the present thesis, might suggest abnormal performance when none actually exists. This is a general problem in the finance literature as it reduces the reliability and implementability of a trading strategy in the real world. However, the authors pointed out that there is a potential problem with their own results, because they may also be a result of model misspecification. If this is the case, it would still be unclear whether the existing results in the literature that speak for market inefficiencies are due to actual mispricing or misspecification. For this reason, the authors recommended the use of non-parametric tests. Consequently, the non-parametric version of the t-test, which was originally introduced by Wilcoxon (1945) and modified by Mann and Whitney (1947), is applied in this analysis. According to Neave and Worthington (1988), the Mann–Whitney–Wilcoxon (MWW) test is widely used and virtually as powerful as its parametric counterpart. Finally, Mood's (1950) median test for two independent samples is applied as a robustness check. All the test methods are presented in table 3.5.

Table 3.5 Parametric (p) and non-parametric (np) tests on portfolio returns

Portfolio return measure	***Test applied***
BHAR (in %) Winsorised BHAR (in %)	Unmatched two-sample t-test (p)
BHAR (ranked) Winsorised BHAR (ranked)	Unmatched MWW (np); median test (np)

[10] A summary of the statistical test methods is provided in the appendix.

3.7 Conclusions

This chapter provided a link between the literature review and the subsequent chapters, in which the empirical results are presented. It first classified the thesis within the respective philosophical setting, provided information about the UK and German stock markets that this thesis intends to analyse and finally detailed the research methods. The main focus was deliberately on the last part and provided answers to the paramount question of how the investigation is conducted to answer the research questions. The use of methods was substantiated by examples from the literature, thus ensuring common research standards.

Chapter 4 Empirical Analysis and Results – The UK

4.1 Introduction

This chapter presents and discusses the results of the F-Score and F-rank investment strategies for the UK. The results are based on the original work presented by Piotroski (2000) and extended in the further course of the analysis. Therefore, the chapter links back directly to section 3.5.1, in which the respective research methods were detailed. For the sake of conciseness, only those results that are deemed to be the most relevant are shown.

The key research issue of this chapter is twofold. First, it aims to test whether a simple value investing strategy is easily transferable to another stock market with a similar success rate. In addition, this strategy is extended to alternative combination techniques which represent the main contribution of this chapter. This is of interest from both a practical and a theoretical aspect. For fund managers, a trading strategy’s value is positively correlated with its successful implementation in as many different stock markets as possible. If this is the case, it provides managers with a much broader range of options to find suitable investments for their clients. Further, the present chapter links to the concept of generalisability in deductive research, which is used in the present thesis.

This means that, on a wider scale, this research is testing theory and therefore similar results are expected for the UK. In other words, both the efficient market hypothesis and the behavioural finance school of thought require research results to be consistent and therefore to confirm one of these theories. As this thesis assumes at least partially inefficient markets, the investment strategy should be readily applicable in the UK. If this is not the case, the research outcome would not support either school of thought. The reason lies in the variability of the factor that is common to both academic stances, human behaviour. To put it simply, Lo (2007) described human behaviour as either rational (EMH) or partly rational (BF). Irrespective of one’s own viewpoint, a strategy should provide consistent results, as mentioned earlier. However, testing a set of different strategies may lead to different outcomes amongst them. Because of these variations, the overarching concept of either rational or non-rational equity markets cannot ultimately be proven.

Second, this chapter intends to provide a range of different investment strategies that are aimed to improve the results of the original F-Score method. This is motivated by some of the drawbacks, such as the binary coding of financial statement signals as well as the equal weighting of the nine signals. To overcome them, a forecasting method that is used not in the finance but in the economics literature is borrowed and applied, as in Rapach et al. (2010). This is a novel approach because prior work focused at generating weights based on time-series forecasts while in this chapter weights are generated based on cross-sectional forecasting. Of course, investment strategies are only valuable if they remain profitable over time. However, there is disagreement in the literature about the length of a specific sample period. For instance, Nelson and Kim (1993) regressed real and excess returns on the dividend yield for the period from 1871 to 1987 and found no predictability for the pre-war period. Paye and Timmermann (2006) studied the instability of return prediction models in the UK and found that the relationship between certain state variables and stock returns may change substantially after a break. While these empirical results are valid, they should not pose too big a problem in practice as an investment period of over 100 years is unrealistic for most individual investors.

Finally, this chapter does not restrict its focus to firms with a high book-to-market ratio only. While the study by Piotroski (2000) demonstrated that yields of up to 23% p.a. within high-BM portfolios are possible in a long-short portfolio for a period of twenty years, there is no grounded reason why this should not also be possible for low-BM firms. This is of interest for two reasons. Firstly, it provides investors with another set of options to choose from and potentially to increase their stock returns. Secondly, building a long-short portfolio of high-BM (long) and low-BM (short) stocks may also lead to higher returns than choosing stocks that feature similar BM characteristics.

To concretise these research issues, the next section provides an overview of the key research questions examined and their underlying rationale. With regards to the methodology used, the reader is referred to chapter 3 and the appendix of this thesis.

4.1.1 Hypothesis development

Research on the strategy in the UK stock market, that is, the London Stock Exchange (LSE), is of interest not only from a European investor's perspective but also because it represents the second-largest market in global terms. In a study by Doidge et al. (2009), for instance, the authors quantified the foreign listings on the New York Stock Exchange (NYSE) as up to 30% compared with 16% on the LSE, with no other exchange attracting more than 7% of foreign investors. Therefore, it can reasonably be assumed that the two markets have similar characteristics and that the reasons for a firm's decision to cross-list its stock either in the US or in the UK are aligned. According to Roosenboom and van Dijk (2009), the most important determinants in this regard are generally the high investor protection (bonding) offered in both countries and the benefits of reducing the cost of capital (market segmentation), which is especially the case in the UK. Because those markets have similar features, their investor structure should be similar as well.

The benefits of investor protection as a determinant of choosing either the US or the UK for a listing relate directly to the legal environment. Both countries have a long tradition of the common law system and most academics agree that a common law environment reduces earnings management (e.g. Leuz et al.., 2003) and insider trading (e.g. Bhattacharya and Daouk, 2002) and increases information disclosure (e.g. La Porta et al., 2006), as well as that of shareholders (e.g. Djankov et al., 2008). Because of those similarities, the evaluation of the success of an investment strategy, which was originally tested in the US markets and which is now evaluated in the UK, is tantamount to a robustness check similar to that of Duong et al. (2014). To meet this requirement, the first hypothesis is as follows.

Hypothesis 4.1: The F-Score investment strategy and its alternatives can successfully separate future winning from losing stocks in the UK.

Having tested for robustness, the analysis then focuses on the efficiency of the investment strategy originally presented by Piotroski (2000). As presented in the previous chapter, the nine measures of the composite F-Score are binary and are therefore likely to disregard useful financial statement information. For instance, if a company has issued additional equity during its past business period, this is assessed negatively and a value of zero is assigned to this particular measure. Although this might be justified on the grounds that, ceteris paribus, additional equity dilutes net income, virtually no grading is applied. In other words, irrespective of whether, for instance, 5% or 25% of further equity was raised, both companies receive a value of zero. In the worst case, this scenario might apply to the remaining variables and ultimately categorise the company as non-investable. Due to a broad range of alternative, high-ranking F-Score firms, this should not be problematic in a long-only portfolio. However, if such an unduly negatively rated stock is shorted at the same time, this could adversely affect the overall performance of the portfolio. Of course, the same concept applies vice versa.

Despite the appeal of this simple all-or-nothing strategy, it might still be suboptimal and thus leave room for improvement. However, the benefits of potentially higher returns should be at least in proportion to the additional time and effort spent on the implementation of alternatives. Hence, the further analyses cover three aspects that explore whether the original method can actually be improved. The first one ensures a finer grading of the financial statement information and aims to ameliorate the simpler binary perspective of a company by implementing a ranking approach. Conceptionally, this follows the method used by Jegadeesh and Titman (1993), who found a trend continuation of formerly well and poorly performing stocks in the subsequent three to twelve months. Compared to the original F-Score, the alternative measures are not absolute but rather relative. In other words, in the F-Score environment, firms receive one defined score that ranges from 0 to 9 irrespective of the number of peer firms in one specific year. Because a grading approach is used in the alternative strategies, the alternative score is dependent on the performance of other firms in the sample. However, this is not a drawback as it follows the same argument as pointed out earlier regarding the benchmark definition and calculation of returns (section 3.5.1.3). Another reason is that all firms are exposed to systematic risk and therefore correlations between firms are likely to remain stable.

The second method measures the accuracy of the forecasting ability of each of the nine individual (Fi-rank) measures and therefore assumes that past accuracy holds true as well future accuracy. Finally, measures with higher accuracy are weighted more heavily to reduce the noise of inaccurate and therefore less useful forecasting variables. Both concepts have been widely used in economic time series applications but have attracted very limited attention in the cross-section as well as the finance literature in general (Rapach et al., 2010). The second hypothesis is therefore summarised as follows:

Hypothesis 4.2: Alternative combinations of F-Score components outperform the original investment strategy in the UK.

A fundamental difference from Piotroski's (2000) approach is that not only high book-to-market firms are analysed but the entire set of UK listed companies. The rationale for focusing on high-BM firms only is that less than half of the sample of those firms provided positive returns in the original US study. However, as discussed earlier, this outcome could be attributed to the binary or all-or-nothing characteristic of the F-Score strategy, which is sought to be improved as part of the previous hypothesis. The reasons for the inclusion of all listed firms are twofold. On the one hand, it has practical implications in that the overall amount of UK listed firms is smaller than that of the US. This is also true for other stock exchanges globally. Therefore, further reducing the amount of data by disregarding all low-BM firms could lead to the acceptance of the law of small numbers and therefore inconclusive results. Tversky and Kahneman (1971) summarised this law as the "exaggerated confidence in the validity of conclusions based on small samples". On the other hand, the focus on the entire data set is a means to prove the robustness of the original strategy. If the strategy is found to work on all stocks, this would save the practitioner the additional step of categorising stocks according to their BM ratios before applying the actual investment strategy. Although some evidence was provided by Duong et al. (2014) for the UK, the authors did not analyse the performance of alternatives to the original F-Score approach. Therefore, the third hypothesis is as follows:

Hypothesis 4.3: The F-Score investment strategy and its alternatives can successfully separate winners from losers irrespective of their affiliation to a specific book-to-market ratio group.

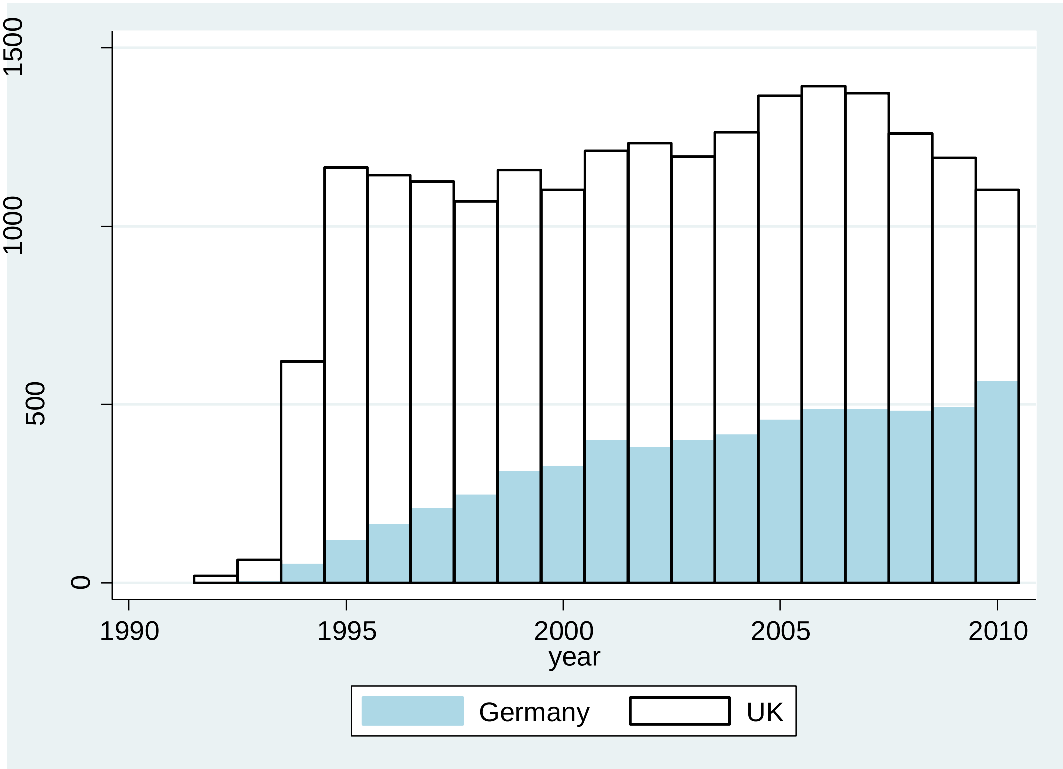
4.2 An overview of the UK data set

To summarise the data sets of the two stock markets analysed in the present thesis at a single glance, the UK data are contrasted with their German equivalents. To begin with, table 4.1 shows the amount of UK and German firms that provide complete accounting data, which is necessary for the functioning of the investment strategy. The results from the German data are included for comparative purposes only and are dealt with separately in chapter 6. Figure 4.1 shows the same data as a Stata histogram. The data correspond to table 3.4 in the previous chapter but are broken down by years rather than F-Scores.

Table 4.1 Firms with complete F-Scores

Year	*UK*	*DE*
1992	19	0
1993	63	1
Year	*UK*	*DE*
1994	620	49
1995	1,164	116
1996	1,144	160
1997	1,125	206
1998	1,069	244
1999	1,158	310
2000	1,102	324
2001	1,211	396
2002	1,233	377
2003	1,195	396
2004	1,264	412
2005	1,366	453
2006	1,393	485
2007	1,373	485
2008	1,260	479
2009	1,192	490
2010	1,102	561
Total	**20,053**	**5,944**

Figure 4.1 Histogram of firms with complete F-Scores



As can be seen, the amount of accounting data is fairly stable in the UK, with the exception of the first two years of the evaluation period. Regarding Germany, there is an increasing trend over time in the amount of data usable for the analysis. The reasons for this are likely to be found in the more sophisticated equity culture of the UK compared with Germany, which started to evolve with the IPO of Deutsche Telekom in 1996 (Börsch, 2004). More stringent disclosure requirements that ensure better protection for equity investors, especially since the revelation of the accounting frauds at the turn of the century, could be another reason for the greater data availability, especially in Germany (Tuschke and Sanders, 2003). This, in turn, can be attributed to both increased market capitalisation and a greater volume of shares traded during the period of 1990 to 1998 (Boutchkova and Megginson, 2000). Another possible explanation is the easier and most importantly cheaper access to stock markets for the individual investor (Colliard and Foucault, 2012). Because trading costs have dropped significantly since 1992, companies now have more options for raising capital and saving costs at the same time. This is because financial intermediaries, such as banks, are now competing with public capital sources, such as stock exchanges (Stulz, 2009).

4.2.1 Descriptive statistics of UK firms

Table 4.2 presents descriptive statistics for all of the nine F-Score components, with the exception of equity offerings. As can be seen, the average (median) UK firm has a BM ratio of 0.585 (0.502) and a market capitalisation of 835.65 (42.03) million GBP. Regardless of a firm's BM ratio, the average (median) return on assets amounts to -0.061 (0.043). The related change in ROA is -0.231 on average, with a median of 0.0004, respectively. The average (median) public listed company in the UK has total assets of 968.94 (53.13) million GBP. Further, it increased its financial gearing and decreased its liquidity compared with the previous year. It should be pointed out that the amount of observations for the BM ratio and market capitalisation figure are less than 20,053, because some figures were not available through Datastream even if all of the nine F-Score components were available.[11] However, these divergences are minor and are not assumed to have a significant influence on the results. The remaining columns show the respective standard deviations, the percentage of positive signals and the minimum/maximum values for each of the F-components. At the bottom of

[11] The book-to-market ratios are computed by dividing the book value by the market value at the financial year end. The book values were extracted from Datastream with the code WC05491. The market capitalisations were extracted with the code WC08002.

the table, the total assets are denominated in thousands and the market capitalisation in millions of GBP.

Table 4.2 Financial characteristics of UK firms

Variable	*Obs.*	*Mean*	*Median*	*Std dev.*	*% positive*	*Min.*	*Max.*
ROA	20,053	-0.061	0.043	1.793	69.05	-200.64	4.987
CFO	20,053	-0.015	0.072	4.226	74.29	-576.46	80.986
ΔROA	20,053	-0.231	0.0004	51.062	50.37	-7,010	1,666
ACCRUAL	20,053	-0.046	-0.043	4.350	73.49	-169.38	576.01
ΔLEVERAGE	20,053	0.139	0.062	0.741	66.44	0	71.33
ΔLIQUIDITY	20,053	-0.048	-0.010	15.84	47.36	-1,274	1,657
ΔGPM	20,053	9.115	-9.6E-06	666.57	34.13	-11,367	75,494
ΔTURN	20,053	-0.640	0.007	61.77	51.94	-8,500	122
Variable	*Obs.*	*Mean*	*Median*	*Std dev.*	*% positive*	*Min.*	*Max.*
BM	19,687	0.585	0.502	3.828	-	-387.67	59.65
TASSETS	20,053	968.94	53.13	6,246	-	1	203,324,559
Market Cap.	19,485	835.65	42.03	5,141	-	.03	158,543

The next table shows the one-year and two-year winsorised buy-and-hold returns of all the UK stocks included in the data set. The one-year returns consist of 20,053 firms, whereas their two-year counterparts are computed from 17,754 firms for the whole period. The difference arises because of the longer measurement period and because some of the firms ceased trading. The returns are compared with the original market-adjusted returns reported in Piotroski's (2000) study and marked as (US). It should be noted that the US results only contain high-BM firms, as opposed to the UK returns, which are based on low-BM firms as well. Despite this difference, there are noticeable similarities between the US and the UK findings, as table 4.3 shows. Exceptions to this are the one-year mean returns, which are only marginally positive for the UK and marginally negative in the two-year scenario. However, both markets indicate positive skewness for both investment horizons. This is underlined by the last column of table 4.3, which reports the percentage of positive stock returns over the two investment periods. Similar to the US results, the majority of stock returns are negative. However, this result confirms the assumption of Duong et al. (2014) that the F-Score approach could be used as an accounting-based investment strategy regardless of whether the stocks are part of the high-BM or the low-BM group. This is particularly interesting as the use of a separate stock valuation model, such as the G-Score presented by Mohanram (2005), would not be necessary, which increases the practicability in the day-to-day investment environment as a consequence.

Table 4.3 Comparison of buy-and-hold returns between UK and US firms

Returns	*Mean*	*10th percentile*	*25th percentile*	*Median*	*75th percentile*	*90th percentile*	*% positive*
1 Y (UK)	6.15E-17	-0.558	-0.306	-0.048	0.229	0.542	0.451
1 Y (US)	0.059	-0.560	-0.317	-0.061	0.255	0.708	0.437
2 Y (UK)	-2.36E-17	-0.802	-0.496	-0.118	0.304	0.822	0.424
2 Y (US)	0.127	-0.872	-0.517	-0.111	0.394	1.205	0.432

4.2.2 The correlation matrix and basic F-Score strategy

However, table 4.3 still gives only a crude picture of the situation in the UK stock market. To investigate the relationships between the nine individual F-components and the subsequent stock returns, table 4.4 represents the Spearman correlation matrix. In more specific terms, the matrix shows the correlations between the individual F-components, which can take values of either 1 or 0, the aggregate F-Score and the one- and two-year subsequent returns in the normal and winsorised versions. Most interesting is the relationship between the aggregate F-Score and the one- and two-year (winsorised) market returns, which is significantly positive at the 1% level with values greater than 0.129. In addition to the correlation matrix, a factor analysis was conducted. This technique was utilised to reveal different common dimensions, subsets or factors of the nine financial ratios which were assumed to exist in the dataset, compared to column 3 of table 3.3. However, results indicated that this was not a fruitful line of enquiry and therefore no further discussion on this topic is presented in the present thesis.

The matrix is subdivided into sections I–III of the firms' financial characteristics, as outlined in section 3.5.1.2 and presented in table 3.3. From this it follows that ROA, CFO and ΔROA appear to be the main contributors to the overall F-Score, because their correlations are greater than 0.4 throughout. This suggests that if, for instance, ROA receives a coded value of 1 then CFO will most likely also receive a value of 1 due to a correlation of 0.535 between those two variables. This observation is particularly of interest once weights are attached to forecasts (DMSFE) or forecasts are dropped entirely (clusters). The reason is that weak forecasts may be noisy and therefore might reduce the success of the original F-Score investment strategy. This might even go as far as to cause the statistical results to become insignificant without the knowledge of the researcher. Identifying and separating more reliable from the weaker forecasts should ameliorate these problems and also help to reduce the noise problem in general. As all forecasts are kept in the DMSFE scenario as compared to the clusters, it is of particular interest to analyse whether weak forecasts are still able to outperform their noise levels or whether it is more advantageous to drop these forecasts entirely.

Fairly high are the F-components that measure the change in liquidity (ΔLIQ) and equity offerings (EQO) in section II as well as the change in turnover (ΔTRN) with values above 0.3. The weakest contributors to the composite are accruals and change in the gross profit margin of sections I and III, respectively, with values only above 0.2. However, all nine F-components are significant at the 1% level and therefore can be considered as equally useful parts of the investment strategy.

The picture looks slightly different once the F-components are compared with the one- and two-year (winsorised) returns. Section I, profitability, is positively and significantly related to all of the four returns, although the correlation is highest for the ROA and CFO measures. Here, the values are consistently higher than 0.14 but much lower for the change in ROA and ACC. Regarding section II, which deals with the leverage, liquidity and source of funds, ΔLIQ and EQO are positively correlated with the returns except for the change in financial gearing measure (ΔLEV), which only explains the two-year (winsorised) returns. Interestingly, none of the F-components in section III, a firm's operating efficiency, are significant. Overall, the most highly correlated measures remain ROA and CFO.

Compared with the original study by Piotroski (2000), the results with regard to the overall F-Score are comparable with those of the UK. The correlations between the F-Score and the one- and two-year returns amount to 0.121 and 0.130, respectively. Further, ROA and CFO likewise appear to play dominant parts in explaining the subsequent stock returns for the two different investment periods. Generally, the variability in the US is smaller than that in the UK, which implies lower values for ROA and CFO but higher values for the measures of operating efficiency.

Table 4.4 Spearman correlation matrix between stock returns, F-components and F-Score[a]

Variable	*I) Profitability*				*II) Leverage, liquidity and source of funds*			*III) Operating efficiency*		*Returns*				*F-Score*
	ROA	CFO	ΔROA	ACC	ΔLEV	ΔLIQ	EQO	ΔGPM	ΔTRN	RET	W_RET	RET2	W_RET2	F-Score
ROA	1	-	-	-	-	-	-	-	-	-	-	-	-	-
CFO	0.535***	1	-	-	-	-	-	-	-	-	-	-	-	-
ΔROA	0.229***	0.073***	1	-	-	-	-	-	-	-	-	-	-	-
ACC	-0.146***	0.274***	-0.102***	1	-	-	-	-	-	-	-	-	-	-
ΔLEV	0.022***	0.010	0.024***	-0.007	1	-	-	-	-	-	-	-	-	-
ΔLIQ	0.117***	0.047***	0.062***	-0.100***	-0.052***	1	-	-	-	-	-	-	-	-
EQO	-0.065***	-0.020	0.010	0.008	0.065***	-0.059***	1	-	-	-	-	-	-	-
ΔGPM	-0.150***	-0.095***	0.050***	0.030***	0.044***	0.057***	0.104***	1	-	-	-	-	-	-
ΔTRN	0.031***	-0.007	0.303***	-0.016	-0.104***	-0.029***	-0.010	-0.113***	1	-	-	-	-	-
RET	0.145***	0.161***	0.025***	0.051***	0.018	0.020***	0.031***	-0.012	0.017	1	-	-	-	-
W_RET	0.144***	0.163***	0.023***	0.054***	0.018	0.020***	0.033***	-0.009	0.014	0.998***	1	-	-	-
RET2	0.185***	0.194***	0.029***	0.058***	0.019***	0.025***	0.028***	-0.011	-0.003	0.735***	0.737***	1	-	-
W_RET2	0.183***	0.195***	0.028***	0.061***	0.029***	0.025***	0.030***	-0.008	-0.002	0.735***	0.739***	0.999***	1	-
F-Score	0.463***	0.514***	0.527***	0.245***	0.289***	0.322***	0.310***	0.267***	0.336***	**0.129*****	**0.130*****	**0.149*****	**0.151*****	1

[a] The nine F-components are correlated between each other, one-year and two-year returns (RET and RET2), one-year and two-year winsorised returns (W_RET and W_RET2) and the F-Score. All the F-components are correlated using their binary score of either 1 or 0 representing good or bad financial performance.

In the next table, the stock returns are presented broken down by their respective F-Score affiliation. Panel A contains the one-year buy-and-hold returns and panel B the two-year equivalent. Apart from the mean return calculations for each F-Score portfolio, table 4.5 shows the subdivision of the same portfolio into fquintiles, the percentage of positive stock returns and the amount of observations in the last column. Similar to the firm years presented in table 3.4 of the previous chapter, the majority of observations are clustered around firms with F-Scores of 3 to 7. This is in accordance with Piotroski (2000), who interpreted this result as a sign that most of the firms have no clear-cut signs of financial strength or weakness. This leaves the lowest (L*: 1,091) and highest (H*: 1,160) F-Score firms to analyse any patterns with regard to their return performance. With reference to panel A, the average one-year stock returns are increasing steadily from the lowest to the highest F-Score portfolios. This is, with one exception, also the case for the winsorised returns, which appear in italics below the normal market-adjusted returns. Regarding the fquintiles, this trend continues, although it is not as clear-cut as for the mean. Parametric t-tests are conducted and the results are presented at the bottom of panels A and B. Two scenarios are used in this process. The first one tests for statistical significance between the highest F-Score portfolios and all the portfolios, whereas the other one compares the mean returns of the highest with the lowest F-Score portfolios.

The t-statistics for both the normal and the winsorised one-year returns are significant at the 1% level. Thereby, the t-statistics of 5.073 relate to the winsorised version of the one-year returns. The results for the H* and L* portfolios likewise indicate a significant difference in mean returns. This trend is repeated in the penultimate column, which shows the percentage of observations with positive stock returns in the related F-Score group. More than half of the observations in the highest F-Score are positive, whereas this is only the case for around one-third of the lowest portfolios. Panel B provides the results for the same analysis with regard to the two-year investment horizon. In general, the findings are similar to the one-year returns, although it might be stated that the improvement from the lowest to the highest portfolios occurs more in a straight line.

Table 4.5 One-year and two-year buy-and-hold returns of F-Score portfolios

Panel A: One-year buy-and-hold returns

	Mean	***10%***	***25%***	***Median***	***75%***	***90%***	***% positive***	***Obs.***
All firms	3.9E-17 *6.2E-17*	-0.578 *-0.558*	-0.323 *-0.306*	-0.060 *-0.048*	0.215 *0.229*	0.528 *0.542*	0.440	20,053
F-Score								
0	-0.341 *-0.312*	-0.762 *-0.761*	-0.673 *-0.670*	-0.251 *-0.224*	-0.041 *-0.042*	0.120 *0.120*	0.231	13
1	-0.114 *-0.106*	-0.849 *-0.833*	-0.602 *-0.596*	-0.262 *-0.232*	0.149 *0.175*	0.735 *0.751*	0.330	203
2	-0.072 *-0.148*	-0.748 *-0.729*	-0.529 *-0.509*	-0.239 *-0.218*	0.081 *0.085*	0.416 *0.437*	0.309	875
3	-0.067 *-0.085*	-0.715 *-0.687*	-0.463 *-0.448*	-0.164 *-0.151*	0.157 *0.170*	0.537 *0.549*	0.356	2,177
4	-0.011 *-0.010*	-0.612 *-0.598*	-0.362 *-0.345*	-0.082 *-0.068*	0.206 *0.216*	0.566 *0.587*	0.422	4,091
5	0.005 *0.008*	-0.550 *-0.536*	-0.313 *-0.294*	-0.048 *-0.034*	0.231 *0.238*	0.515 *0.529*	0.450	4,875
6	0.025 *0.035*	-0.482 *-0.451*	-0.261 *-0.232*	-0.033 *-0.021*	0.227 *0.234*	0.521 *0.537*	0.467	4,109
7	0.031 *0.043*	-0.469 *-0.443*	-0.240 *-0.216*	-0.002 *-0.010*	0.238 *0.242*	0.518 *0.526*	0.497	2,550
8	0.063 *0.073*	-0.432 *-0.418*	-0.229 *-0.210*	0.104 *0.018*	0.247 *0.262*	0.579 *0.583*	0.511	999
9	0.091 *0.108*	-0.428 *-0.413*	-0.172 *-0.143*	0.046 *0.049*	0.284 *0.294*	0.486 *0.504*	0.559	161
H*–All t-statistic *(p-value)*	2.950*** *(0.003)* 5.073*** *(0.000)*	-	-	-	-	-	-	-
H*–L* t-statistic *(p-value)*	2.679*** *(0.007)* 9.717*** *(0.000)*	-	-	-	-	-	-	-

Stars indicate statistical significance as follows: *** = 1%, ** = 5% and * = 10%.

Panel B: Two-year buy-and-hold returns								
	Mean	***10%***	***25%***	***Median***	***75%***	***90%***	***% positive***	***Obs.***
All firms	-5.9E-17 *-2.4E-17*	-0.823 *-0.802*	-0.514 *-0.496*	-0.137 *-0.118*	0.283 *0.304*	0.801 *0.822*	0.412	18,946
F-Score								
0	-0.433 *-0.404*	-1.153 *-1.131*	-0.915 *-0.818*	-0.579 *-0.572*	-0.157 *-0.094*	0.232 *0.236*	0.154	13
1	-0.126 *-0.179*	-1.010 *-1.073*	-0.840 *-0.801*	-0.517 *-0.475*	-0.001 *0.007*	1.062 *1.068*	0.249	197
2	-0.175 *-0.215*	-1.032 *-0.998*	-0.750 *-0.724*	-0.399 *-0.383*	0.030 *0.048*	0.687 *0.701*	0.263	839
3	-0.114 *-0.137*	-0.957 *-0.932*	-0.668 *-0.651*	-0.295 *-0.279*	0.151 *0.167*	0.714 *0.734*	0.312	2,102
4	-0.015 *-0.022*	-0.861 *-0.838*	-0.566 *-0.536*	-0.172 *-0.151*	0.242 *0.267*	0.785 *0.814*	0.396	3,903
5	0.001 *0.009*	-0.796 *-0.776*	-0.484 *-0.467*	-0.108 *-0.090*	0.297 *0.315*	0.784 *0.806*	0.427	4,647
6	0.048 *0.059*	-0.701 *-0.682*	-0.401 *-0.376*	-0.071 *-0.047*	0.322 *0.339*	0.809 *0.834*	0.444	3,842
7	0.069 *0.089*	-0.696 *-0.683*	-0.400 *-0.375*	-0.032 *-0.013*	0.381 *0.396*	0.889 *0.917*	0.482	2,363
8	0.105 *0.105*	-0.662 *-0.642*	-0.381 *-0.360*	-0.021 *-0.000*	0.384 *0.397*	0.906 *0.913*	0.482	894
9	0.178 *0.169*	-0.748 *-0.713*	-0.351 *-0.330*	-0.003 *0.023*	0.389 *0.406*	1.067 *1.071*	0.493	146
H*–All t-statistic *(p-value)*	3.456*** *(0.001)* 4.527*** *(0.000)*	-	-	-	-	-	-	-
H*–L* t-statistic *(p-value)*	4.553*** *(0.000)* 8.905*** *(0.000)*	-	-	-	-	-	-	-

[a] Values in the upper (*lower*) half of each cell represent non-winsorised (winsorised) one- and two-year returns.

The differences in the respective mean returns remain significant at the 1% level. The only two main differences can be found in the last two columns. On the one hand, fewer firms show positive returns over the two-year period and even firms with the highest three F-Scores stay below the 50% mark. On the other hand, there are fewer observations. The reason is that more data are needed for the calculation, which would exceed the defined period under research. In addition, relevant data are simply not available due to takeovers, delisting or withdrawal from business, as mentioned earlier.

In summary, the basic test results show that the F-Score strategy is indeed able to discriminate between high and low future stock returns, confirming Piotroski's (2000) results. However, there appears to be a tendency for the major part of the returns to be most likely to be earned within the first year of portfolio construction, as can be deduced from the penultimate columns of panels A and B, respectively. In addition, most of the overall profit of a long-short investment strategy, that is, H*–L*, comes from firms with a low F-Score. In other words, the shorted stocks generate higher returns than their long counterparts. This is potentially problematic for investors as beta hedging might turn out to be unattractive if the group of low-F-Score firms is characterised by high betas compared with the high-F-Score firms. However, the findings are generally in line with the original results from the US study, which indicates a promising application in the UK. This is especially noteworthy as not only high-BM firms are analysed but rather the overall UK stock market.

4.2.3 One-way portfolios with the original F-Score

The analysis now moves on to prepare the five one-way portfolios for deeper statistical analyses, as mentioned in section 3.5.1.4. Table 4.6 presents the test results of the differences in the one-year returns with regard to the high/low and highest/lowest F-Score portfolios. In the F-Score column, the middle fquintile (M) is omitted because it is not of interest in the analysis and for succinctness. The first column describes how those percentage returns are calculated, specifically either as average or as median. In the bottom half of the table, the respective winsorised (winsd.) percentage returns are displayed. This means, for instance, that the winsorised mean return of the low (L) F-Score portfolio was -10.4% per annum. Moving further across, the three remaining sections are concerned with the presentation of the test statistics. The absolute difference between the highest and the lowest portfolio for the non-winsorised returns amounts to about 15 percentage points and is significant at the 1% level with t-statistics of 2.679. In the adjacent column, MWW describes the Mann–Whitney–Wilcoxon test and indicates significance for both portfolio differences. Lastly, the median column reports the test results for differences in the median returns, which are likewise significant. It should be noted that both the MWW and the median represent the non-parametric tests, as described in section 3.6.3.

Table 4.6 One-way F-Score portfolios

	F-Score				*t-test*		*MWW*		*median*	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.083	-0.073	0.042	0.067	0.12*** *(4.875)*	0.15*** *(2.679)*	*(17.59)****	*(13.08)****	-	-
median	-0.242	-0.193	0.004	0.014	-	-	-	-	0.20*** *(238.42)*	0.26*** *(131.09)*
winsd.										
mean	-0.142	-0.104	0.054	0.078	0.16*** *(12.822)*	0.22*** *(9.717)*	*(17.65)****	*(13.12)****	-	-
median	-0.221	-0.183	0.013	0.026	-	-	-	-	0.20*** *(237.96)*	0.25*** *(131.09)*

Stars indicate statistical significance as follows: *** = 1%, ** = 5% and * = 10%.

Overall, the results in table 4.6 highlight that H (H*) portfolios outperform their L (L*) peers with regard to their mean and median one-year returns. In other words, portfolios consisting of stocks with high F-Scores perform relatively better, as originally documented by Piotroski (2000).

4.2.4 One-way portfolios with an alternative: The F-rank approach

In the next step, portfolios are built as detailed in section 3.5.1.5. As elucidated earlier, the ranking approach is applied to overcome the drawbacks of the binary nature of the F-Score. In general, the ranking of stock returns is widely used in time series regressions in the finance literature. For instance, Jegadeesh and Titman (1993) used ranked returns as evidence of market under-reaction to firm-related information and derived the success of certain momentum strategies from it. Further, Asness et al. (2013) used ranking to construct value and momentum factors to establish portfolio weights. In time series regressions, the portfolio returns are then regressed on the return of a market index. As they highlighted, their intention is to reduce the undue influence of outliers, whereby portfolios usually perform better using the raw values. Although this study adopts a more cross-sectional approach, results based on ranks should be more reliable still.

Table 4.7 presents the respective test results, which are then compared with the original F-Score. The difference between the upper and the lower half of the table is the method of calculating the F-rank, which, in turn, has a direct effect on the construction of the fquintiles. As outlined before, in comparison with the F-Score, the F-rank is the composite measure of all nine ranked and standardised F-components (i.e. the Fi-ranks). Because the standardisation process results in a continuous value for each Fi-rank, limited by 0 and 1, there is the possibility to calculate either the average or the median of the F-rank. The computation of the one-year returns is performed using the mean and median in the same way as described for the F-Score above. Both versions of the F-rank investment strategy are able to discriminate between the low/high-returning portfolios similarly to the original F-Score as the t-statistics are highly significant. However, this is not always the case for the highest/lowest portfolios observable in the non-winsorised sections under the parametric t-test. However, both non-parametric test statistics indicate that the mean and median portfolio differences are significant.

Table 4.7 One-way F-rank portfolios

A	***F-rank (mean)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.032	-0.044	0.022	0.035	0.07*** *(4.388)*	0.07 *(1.009)*	*(13.03)****	*(11.07)****	-	-
median	-0.249	-0.130	-0.032	-0.022	-	-	-	-	0.10*** *(100.69)*	0.23*** *(110.18)*
winsd.										
mean	-0.129	-0.058	0.031	0.048	0.09*** *(9.318)*	0.18*** *(6.71)*	*(13.05)****	*(11.07)****	-	-
median	-0.230	-0.120	-0.020	-0.010	-	-	-	-	0.10*** *(101.06)*	0.22*** *(112.06)*
B	***F-rank (median)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.085	-0.040	0.021	0.017	0.06*** *(3.770)*	0.10** *(2.188)*	*(13.72)****	*(8.93)****	-	-
median	-0.280	-0.148	-0.038	-0.097	-	-	-	-	0.11*** *(129.99)*	0.18*** *(48.43)*
winsd.										
mean	-0.147	-0.062	0.029	0.012	0.09*** *(9.190)*	0.16*** *(5.51)*	*(13.73)****	*(8.94)****	-	-
median	-0.267	-0.136	-0.026	-0.091	-	-	-	-	0.11*** *(129.16)*	0.18*** *(49.67)*

However, in general, it can be observed that, compared with the F-Score one-way strategy, stock selection and portfolio construction using the F-rank as an alternative tool are not as powerful. Despite the initial appeal of ranking the financial statement items, this method is not as statistically powerful as the binary approach, which is substantiated by lower t-statistics in all the tests. However, the basic idea of utilising accounting data to pick winning stocks and to avoid losing ones is confirmed by the F-rank strategy.

4.3 The first set of alternative combinations of F-ranks: An overview

The following subsection presents and evaluates the results of weighting and combining the F-ranks, which relates to section 3.5.1.6 of the previous chapter. As before, this weighted strategy is generally based on forecast errors, that is, DMSFEs. In general, portfolios are established by grouping firms according to their weighted forecast errors, which are defined as the squared differences between each Fi-rank, specifically the individual forecast, and the ranked one-year return. The lower the sum of the combined forecast errors, the higher the return should be at $t + 1$. This concept is derived from a study by Rapach et al. (2010), who combined 15 individual forecasts to improve both statistical and economic significance. Allowing the weights of individual forecasts to fluctuate, that is, to be non-equal, is an idea borrowed from Bates and Granger (1969) for the same reasons. The authors concluded that the pooling of forecasts reduces the mean-square error compared with any individual forecast. One possible explanation for this so-called forecast combination puzzle (Stock and Watson, 2004) is that a combination of idiosyncratic instabilities in the forecasts could itself be stable. The final part of this subsection discusses the results and contrasts them with the existing forecasting literature.

4.3.1 Alternative combinations of F-ranks: The weighted DMSFE

Table 4.8 shows an overview of the amount of companies within each of the five portfolio fquintiles. As described in the methodology section, the L* (H*) fquintiles are part of the L (H) fquintiles and the portfolios are built according to the aforementioned cut-off points, which are as follows. Each of the L* and H* star fquintiles contain 5% of the data, the respective L and H fquintiles contain 30% each and the remaining 40% are assigned to the M fquintile.

The strategy starts in 1992. The first set of complete data is available in June 1993 and evaluated according to the June 1994 results. Based on these, the first weighted portfolios are then constructed in June 1994 and evaluated according to their returns in June 1995. The strategy ends in 2012.

Table 4.8 An overview of DMSFE portfolios by year

DMSFE	L*	L	M	H	H*	Total (L, M, H)
1993	4	19	27	19	4	65
1994	31	187	248	186	31	621
1995	59	349	467	349	59	1,165
1996	58	343	460	342	58	1,145
1997	57	338	450	338	57	1,126
1998	54	321	428	321	54	1,070
1999	58	348	463	348	58	1,159
2000	55	331	441	331	56	1,103
2001	61	364	485	363	61	1,212
2002	62	370	494	370	62	1,234
2003	60	359	478	359	60	1,196
2004	63	379	507	379	64	1,265
2005	69	410	547	410	69	1,367
2006	70	418	558	418	70	1,394
2007	68	412	550	412	69	1,374
2008	63	378	505	378	63	1,261
2009	60	358	477	358	60	1,193
2010	56	331	441	331	56	1,103
						20,053

Regarding table 4.9 below, the difference between panel A and panel B is as follows. The construction of the F-rank (weighted) portfolios is based on the forecast errors between the nine individual F-ranks (Fi-rank) and the future non-winsorised one-year return. The winsorised weighted F-ranks, on the contrary, are computed similarly, with the exception that the winsorised one-year returns are used instead. However, the portfolio returns are calculated as usual. The respective test results are shown in table 4.9. Without any exception, an improvement of the overall returns is observable beginning with the portfolios with the highest (high) forecast errors to the ones with the lowest (low). The improvements follow a very similar pattern both in panel A and in panel B, which suggests that the use of either non- or winsorised one-year returns in calculating forecast errors is interchangeable. However, this is not the case once the fquintile returns are computed. Here, the improvement is particularly pronounced for the winsorised mean returns at the bottom of panels A and B, where the return difference amounts to nearly ten (6.6) percentage points for the high/low (highest/lowest) portfolios. Thus, the results suggest that winsorising at the 0.5% level is a small but effective means to reduce the undue influence of outliers in the data set, in particular for the H*/L*fquintiles, which now become highly significant.

Table 4.9 One-way weighted DMSFE portfolios

A	***F-rank (weighted)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	0.028	0.020	-0.049	-0.038	-.069*** *(-4.601)*	-.066 *(-0.992)*	*(-14.11)****	*(-11.22)****	-	-
median	-0.024	-0.031	-0.139	-0.269	-	-	-	-	-0.109*** *(129.71)*	-0.244*** *(109.88)*
winsd.										
mean	0.041	0.030	-0.064	-0.136	-.094*** *(-9.825)*	-.18*** *(-6.638)*	*(-14.12)****	*(-11.22)****	-	-
median	-0.010	-0.019	-0.128	-0.238	-	-	-	-	-0.109*** *(129.71)*	-0.228*** *(111.75)*
B	***F-rank (winsorised weighted)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	0.028	0.020	-0.049	-0.038	-.069*** *(-4.601)*	-.066 *(-0.994)*	*(-14.11)****	*(-11.22)****	-	-
median	-0.024	-0.031	-0.139	-0.269	-	-	-	-	-0.109*** *(129.71)*	-0.244*** *(109.88)*
winsd.										
mean	0.041	0.030	-0.064	-0.136	-.094*** *(-9.825)*	-.18*** *(-6.643)*	*(-14.12)****	*(-11.22)****	-	-
median	-0.010	-0.019	-0.128	-0.238	-	-	-	-	-0.109*** *(129.71)*	-0.228*** *(110.81)*

In general, the results of the weighting method improve the simple F-rank investment strategy shown in table 4.7 considering the winsorised version at the bottom of each panel. The results of the t-test in panel A are highly significant for the H–L portfolio differentials. Regarding the H*–L* differentials, significance is only observable for the winsorised version (showing only significance for the winsorised mean returns, which are computed from the weighted forecast error portfolios. At least, the mean return differences between the highest and the lowest portfolio returns are significant, albeit only at the 10% level. However, this is not the case for the winsorised mean returns shown in panel A of table 4.7. Additionally, the t-statistics for both portfolio differences are higher, which suggests that the weighting method is preferable to non-weighting.) However, slightly better results are identifiable as a result of the non-parametric tests. Whereas the return differences are similar amongst the H–L fquintiles, they are more pronounced with regard to the H*–L* fquintiles. All the mean and median portfolio returns are highly significant at the 1% level in this scenario, as presented in the last two columns.

Of the two weighting approaches, one of which uses the non-winsorised (panel A) and the other uses the winsorised forecast errors (panel B), the former method is preferable. Apart from nearly identical results, this investment strategy has the advantage that it is easier to implement. Although in both instances the forecast errors have to be computed, the first strategy does not necessitate additional winsorising before the portfolios are actually constructed. Even though this working step appears to be minor, it involves a thoughtful process about the level of winsorising, because choosing too high a percentage level might be regarded as inappropriate manipulation of the data at hand. This, in turn, might turn out to undermine the principal aim of building a portfolio with high future returns. However, the decision not to winsorise the forecast errors (DMSFEs) seems to be more obvious at second glance. The reason is that the DMSFEs are ranked and standardised, which naturally eradicates the unwanted influence of outliers. Interestingly, this observation serves as an example of how the excessive modification of the data turns out not to be targeted and positive at the same time. In other words, an investor is likely not to be able to arrive at meaningful results due to the over-modification of data, which is due to ranking, standardising and winsorising the DMSFEs.

4.3.2 Discussion and links to the literature: DMSFE

One insight from the results of Rapach et al. (2010) was that the highly sophisticated forecast combination methods did not outperform the simpler ones, such as the weighted DMSFE, which is generally in line with the evidence in the literature. Applying the combination forecasts to three different time horizons, the results clearly indicated that a combination of 15 economic variables significantly outperformed the individual forecasts. In that regard, the benefits of combining to forecast the equity premium in the US were most pronounced in the longest period under scrutiny, from 1965 to 2005. However, the combinations were at least significant at the 10% level in all three investment windows.

In terms of the weighted forecast errors, the picture in the original study looked very similar, although the R^2 was consistently lower for the weighted than for the simple mean F-rank combination method. This is not necessarily the case in the present analysis. Once the t-statistics and percentage return differentials are compared with each other, the preference for one specific method is inconclusive. The reason for the difference may lie in the nature of the variables used. Rapach et al. (2010) examined a mixture of economic and stock market indicators, whereas Piotroski (2000) focused entirely on firm-specific accounting variables. From this, it follows that a limit appears to exist to the extent to which macroeconomic variables are able to explain stock returns and thus the benefits of weighting seem to diminish. The causes are likely to be related to what is called parameter instability or time variation in regression coefficients and on which the literature agrees, according to Dangl and Hailing (2012). In their study, Stock and Watson (2004) pre-empted the outcomes of Rapach et al. (2010) by asserting that the forecast of output growth in an economy is more stable using simple weighting, that is, near-equal or mean weighting. The effect of parameter instability, as mentioned before, seems to be based, among other things, on changes in regulations or monetary policies and is mainly associated with the economy in general. Of course, this process has ramifications for firms that ultimately affect their financial statements. However, individual firms are more insulated against these effects as the F-rank method focuses on firm-related measures of financial health, which allow a finer evaluation. Therefore, the investment strategy based on Piotroski's (2000) accounting approach suggests that the weighting of individual forecast errors is a more dependable method of discriminating between winning and losing stocks than non-weighting. This is particularly the case for the winsorised mean returns, as shown in panel A of table 4.9, given the significant results in both parametric and non-parametric tests.

As opposed to the reference studies by both Rapach et al. (2010) and Stock and Watson (2004), the equally good results for the weighting approach may also arise from differences in the way in which the general analyses were conducted. The present study is based on a cross-sectional setting, whereas the two mentioned above use time series data. While combination methods have been used extensively in time series in the past, the application of forecast combination techniques in the cross-section is a novel way of distinguishing between future winning and future losing stocks for two reasons. On the one hand, establishing the respective weights and forecasting the next year's return appear to be practically more relevant as investors' main focus is on the annual accounts rather than on the results from longer reporting periods. On the other hand, they free investors from determining a time frame for the holdout period, which seems arbitrary at best.

4.4 The second set of alternative combinations of F-ranks: An overview

The following three sections have a similar structure to section 4.3. First, the results for the two clustering approaches are presented. As described in section 3.5.1.7, portfolios are constructed in a slightly different way from their DMSFE counterparts. However, the parametric and non-parametric tests are performed in exactly the same way and report t-statistics for the mean and median differences in the respective one-year portfolios. Finally, the findings are compared with the current status of research.

4.4.1 Alternative combinations of F-ranks: Clustering (*k = 2)*

The first application of the clustering method ($k = 2$) divides the nine individual forecast errors into two parts; those five with the lowest errors (c-score) are retained and the remaining four are dropped entirely before portfolio construction. Distribution-wise, the portfolios follow exactly the same pattern as presented for the DMSFE approach in table 4.9. Table 4.10 summarises how the F-rank strategy performs following this method. To uphold the logic of the F-Score, the c-score is adjusted so that higher values should be accompanied by higher portfolio returns.

Table 4.10 One-way clustered portfolios (*k* = 2)

	F-rank (c-score)				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.057	-0.031	0.024	0.022	.054*** *(3.407)*	.079 *(1.204)*	*(15.81)****	*(11.34)****	-	-
median	-0.280	-0.149	-0.017	-0.026	-	-	-	-	0.13*** *(205.73)*	0.25*** *(113.53)*
winsd.										
mean	-0.149	-0.054	0.034	0.036	.088*** *(9.124)*	.185*** *(7.011)*	*(15.80)****	*(11.33)****	-	-
median	-0.262	-0.136	-0.004	-0.009	-	-	-	-	0.13*** *(200.02)*	0.25*** *(113.53)*

The results show that the one-year returns improve substantially between the lowest (low) and the highest (high) portfolios. The difference is most pronounced for the winsorised mean returns, for which the high/low divide amounts to nearly 9 and the highest/lowest one to 18.5 percentage points, respectively. This is underlined by all three test results, which attest to statistical significance at the 1% level, with the only exception of the highest/lowest non-winsorised mean returns lacking any significance. Consequently, amongst the three alternative combination methods presented up to now, clustering appears to be the dominant choice because it distinguishes portfolio returns with statistical and economic significance.

4.4.2 Alternative combinations of F-ranks: Clustering (*k = 3*)

Finally, the results of the last combination approach are presented in table 4.11. Following this method, the c-score is computed as the average of those three individual F-ranks, that is, Fi-ranks, with the lowest forecast errors. All of the remaining six Fi-ranks are left unconsidered for the analysis. This approach follows directly from the results of the Spearman correlation matrix in table 4.4. The table provides evidence that the correlation between the F-components and the one-year returns is the highest amongst the members of the profitability measure. Building clusters of only three F-ranks ensures that the portfolio construction is based only on the most promising forecasts, meaning those with the lowest errors.

Table 4.11 One-way clustered portfolios (*k = 3*)

	F-rank (c-score)				*t-test*		*MWW*		*median*	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.033	-0.029	0.025	0.027	.054*** *(3.363)*	.061 *(0.939)*	*(16.10)****	*(12.33)****	-	-
median	-0.281	-0.153	-0.018	-0.011	-	-	-	-	0.14*** *(206.25)*	0.27*** *(134.76)*
winsd.										
mean	-0.127	-0.054	0.037	0.043	.091*** *(9.321)*	.170*** *(6.427)*	*(16.12)****	*(12.34)****	-	-
median	-0.267	-0.140	-0.006	0.006	-	-	-	-	0.13*** *(201.05)*	0.27*** *(134.76)*

As can be seen from the above table, the overall returns improve monotonically from the lowest c-score portfolio to the highest one. The results of the previous combination methods only show better returns between the highest/lowest and the high/low portfolios. However, with the establishment of three clusters, that is, *k = 3*, the returns are now improving even between the lowest/low and the high/highest ranking c-score portfolios, and this applies to both the non-winsorised and the winsorised mean and median returns. Similar to the previous method with *k = 2* clusters, all but one mean and all median differences in portfolio returns remain highly significant. The respective t-statistics of the parametric tests are comparable with the previous *k = 2* clustering but are persistently higher with regard to the non-parametric tests. However, the findings are mixed considering the economic benefits. With the *k = 3* procedure, the one-year winsorised mean return difference amounts to 9.1% (17%) for the high/low (highest/lowest) portfolios. By contrast, the *k = 2* portfolios yield 8.8% and

18.5%, respectively. The differences are more definite regarding the median return difference for the highest/lowest portfolio, in which the $k = 3$ method earns above 27% and the $k = 2$ slightly over 25%.

In summary, the $k = 3$ clustering approach ensures a finer division of stocks, which results in monotonically increasing portfolio returns. Additionally, the return differences mostly remain highly significant and are economically relevant. Therefore, investors are well advised to cluster stocks according to the $k = 3$ method.

4.4.3 Discussion and links to the literature: Clustering

The idea of sorting forecasts into clusters based on their past performance stems from Aiolfi and Timmermann (2006). Similar to their findings, the above results demonstrate that this method leads to better results regarding the separation of future winning and losing portfolios. This holds particularly true for the $k = 3$ cluster method. A new element is that this approach is applied to the F-Score strategy rather than using a mixture of macro-economic variables to forecast a country's output growth. Nevertheless, we find that clustering forecasts appears to work as well as in a firm-specific setting. This is possibly related to the performance stability of individual forecasts, as noted by Stock and Watson (2004). In the macro-economic environment, this means that the reliability of individual predictors depends on the sensibility towards unforeseen events, such as economic shocks. Usually, the higher this sensibility is, the worse the individual forecast performs. Although the study by Stock and Watson (2004) is related more to the DMSFE method than clustering, it has implications for the latter. To their surprise, the authors found that the simplest forecast combinations, such as computing the mean of the individual forecasts, performed best. However, this approach does not consider the sensibility of the nine individual errors, because it averages them in such a way that the weights are equal for all the constituents. The weighted DMSFE method ameliorates this problem by weighting the more stable forecasts more heavily. The problem is that some weight is still applied to the more unstable individual forecasts irrespective of their sensitivity towards any kind of firm-related shock. The clustering approach, in contrast, picks only the three most accurate forecasts from the previous period (in the case of $k = 3$), namely those with the lowest forecast errors, and averages them. Consequently, it combines the benefits of both the forecast stability and the simplified combining procedure due to the use of the mean, which ultimately leads to superior results.

4.5 Empirical results in the two-way environment

The second half of this chapter presents the results of both the original F-Score and the F-rank investment strategy with regard to a second dimension, the book-to-market (BM) ratio. This enables a more precise demarcation between portfolios and allows the testing of one of the fundamental assumptions of value investing, that is, high-BM firms should feature higher future returns than low-BM firms. This procedure also follows the structure of Piotroski's (2000) study and questions whether the newly introduced rank strategies work as well for different BM groupings. From an investor's point of view, it is initially not necessarily important to know why this so-called value-glamour anomaly exists. This thought is directly related to the logic of Friedman's (1953) anecdote about billiard players who are not familiar with the complicated mathematics with which their skill is associated. The question of whether the anomaly is due to investor irrationality, the stance of behavioural finance or the higher risk–return relationship, as EMH advocates propose, only arises once the basic observation is no longer given.

4.5.1 The F-Score and the book-to-market ratio

The results of the F-Score two-way portfolios are presented first in table 4.12. Compared with their one-way equivalents, some changes were made to keep the tables as clear and concise as possible. Again, the presentation of the middle portfolios (M) is omitted for succinctness. In addition to the retained division of the horizontal axis of the F-Score portfolios, the stocks are now further subdivided into either low-BM or high-BM portfolios, which are represented by the vertical axis. It should be noted that the subdivision of the BM portfolios was originally made into terciles. For the same reasons as mentioned before, the middle tercile is omitted. As usual, the returns are in percentages, panel A containing the (winsorised) mean returns and panel B the (winsorised) median counterpart. For this reason, the results of the t-test and the Mann–Whitney–Wilcoxon (MWW) test are documented in panel A, whereas panel B shows the median test results only. An example helps to illustrate how to read the table. In panel A, the portfolio with the highest (H*) F-Score and the lowest (L) BM ratio returns 2%. A portfolio with an equally high F-Score but a higher (H) BM ratio earns an 8.5% annual return. The difference between the two portfolios amounts to 6.5 percentage points (H–L), which is significant at the 10% level according to both the t-test and the MWW test, the t-statistics of which are presented in italics. The structure is analogous for the (winsorised) median returns and median test in panel B.

Table 4.12 Two-way F-Score portfolios

A	*F-Score (mean returns)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.053	-0.116	-0.004	0.020	.112* *(1.83)*	.073 *(0.45)*	*(12.38)****	*(8.08)****	-	-
H	-0.091	-0.045	0.064	0.085	.109*** *(4.92)*	.176*** *(4.86)*	*(7.36)****	*(5.36)****	-	-
H–L	-.038	.071	.068***	.065*						
t-test	*(-.25)*	*(1.09)*	*(3.65)*	*(1.85)*						
MWW	*(4.90)****	*(8.46)****	*(3.55)****	*(1.77)**						
winsd.										
L	-0.195	-0.186	0.010	0.034	.196*** *(8.88)*	.229*** *(5.37)*	*(12.41)****	*(8.09)****	-	-
H	-0.075	-0.039	0.076	0.095	.115*** *(5.63)*	.170*** *(4.85)*	*(7.39)****	*(5.39)****	-	-
H–L	.120***	.147***	.066***	.061*						
t-test	*(2.76)*	*(6.04)*	*(3.60)*	*(1.80)*						
MWW	*(4.87)****	*(8.44)****	*(3.55)****	*(1.76)**						
B	*F-Score (median returns)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.285	-0.274	-0.013	0.026	-	-	-	-	.261*** *(140.13)*	.311*** *(58.51)*
H	-0.151	-0.122	0.015	0.026	-	-	-	-	.137*** *(37.07)*	.177*** *(14.53)*
H–L	.134*** *(20.94)*	.152*** *(65.86)*	.028 *(2.33)*	.000 *(-0.72)*						
winsd.										
L	-0.270	-0.261	-0.005	0.032	-	-	-	-	.256*** *(138.76)*	.302*** *(58.51)*
H	-0.140	-0.110	0.026	0.032	-	-	-	-	.136*** *(35.07)*	.172*** *(13.45)*
H–L	.130*** *(22.30)*	.151*** *(67.26)*	.031 *(2.58)*	.000 *(-0.01)*						

It is evident from panels A and B that the portfolio returns are consistently positive for stocks with high or the highest F-Scores paralleled by a high BM ratio. However, the vertical differences in returns between high F-Score/high-BM and high F-Score/low-BM portfolios are only significant for the (winsorised) mean returns in panel A and not for the median as a result of the median test in panel B. This is not the case once the focus is shifted to BM portfolios in the low and lowest F-Score fquintiles. Here, the vertical differences in the mean and median returns are mostly significant and to a larger extent. With one exception, all the vertical differences between the high-BM and the low-BM firms have the correct sign, which agrees with the literature in that high-BM stocks feature higher future returns (e.g. Piotroski, 2000).

Compared with the one-way F-Score portfolios in table 4.6, the horizontal mean return differences are mostly lower. For instance, the one-way difference between high and low portfolios is 12 percentage points, whereas the same difference in the high-BM environment amounts to only 10.9 points. The same can be observed for the winsorised one-way portfolios in comparison with the two-way high-BM portfolios. Because of the lower return differences, subdividing portfolios according to their BM ratio could thus be viewed as a drawback, but only at first sight. If the advocates of the EMH are correct and stock returns are indeed associated with greater risk taking, then the further subdivision into BM portfolios can help the investor to choose whether a high-risk or a low-risk investment strategy is preferable. Consequently, investors would abstain from investing in high-BM portfolios, which supposedly feature a higher level of risk as a result of financial distress, and invest in the low-BM equivalent instead. The EMH argument would then be unjustified, but this is contingent on one condition. Investors are required to hold a long-short portfolio. As the winsorised portfolio returns in panel A show, a long-only portfolio consisting of a high F-Score and low-BM (high-BM) firms would yield only 1% (7.6%) annually. However, if investors short the respective low F-Score portfolio within the same BM tercile at the same time, the return difference amounts to 19.6 (11.5) percentage points with a high level of significance in the t-tests and MWW tests, respectively. Due to the overall significant horizontal differences and, with some exceptions, mostly highly significant vertical return differences, the two-way F-Score portfolios can be regarded as a reliable method to filter out stocks with high return prospects.

4.5.2 The F-rank and the book-to-market ratio: The mean F-rank

The next section presents the results for the alternative investment strategy of the F-rank with regard to a second dimension, the BM ratio. Table 4.13 is structured in exactly the same way as for the original F-Score two-way portfolios. It should be pointed out that, similar to the one-way tables, the F-ranks can be calculated in two different ways, as described in section 4.2.4. As a result, table 4.13 first shows the results following the mean F-rank approach.

Table 4.13 Two-way mean F-rank portfolios

A	F-rank (mean)				t-test		MWW		median	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.029	-0.082	-0.014	-0.012	.068** *(1.98)*	.017 *(0.12)*	*(6.43)****	*(7.39)****	-	-
H	-0.055	-0.018	0.040	0.064	.058*** *(3.31)*	.119** *(2.19)*	*(8.52)****	*(3.77)****	-	-
H–L	-.026	.064*	.054***	.076*						
t-test	*(-.14)*	*(1.77)*	*(3.26)*	*(1.96)*						
MWW	*(4.14)****	*(9.01)****	*(4.64)****	*(2.60)****						
winsd.										
L	-0.183	-0.122	-0.005	0.002	.117*** *(6.82)*	.185*** *(4.47)*	*(6.43)****	*(7.39)****	-	-
H	-0.055	-0.012	0.053	0.076	.065*** *(3.99)*	.131*** *(2.74)*	*(8.52)****	*(3.77)****	-	-
H–L	.128**	.110***	.058***	.074**						
t-test	*(2.39)*	*(6.15)*	*(3.74)*	*(1.97)*						
MWW	*(4.11)****	*(9.01)****	*(4.63)****	*(2.59)****						
B	***F-rank (mean)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.309	-0.215	-0.062	-0.073	-	-	-	-	.153*** *(75.66)*	.236*** *(34.48)*
H	-0.183	-0.083	-0.011	0.013	-	-	-	-	.072*** *(19.26)*	.196*** *(18.09)*
H–L	.126*** *(14.68)*	.132*** *(67.59)*	.051*** *(11.88)*	.086** *(5.38)*						
winsd.										
L	-0.290	-0.201	-0.048	-0.067	-	-	-	-	.153*** *(75.50)*	.223*** *(34.48)*
H	-0.161	-0.072	0.006	0.025	-	-	-	-	.078*** *(19.26)*	.186*** *(18.09)*
H–L	.129*** *(14.68)*	.129*** *(66.56)*	.054*** *(11.03)*	.092** *(6.11)*						

Panel A (B) contains the (winsorised) mean (median) portfolio returns. Based on the horizontal return differences presented in panel A, compared with the F-Score two-way portfolios, the mean F-rank strategy produces poorer results, especially regarding economic significance, but also lower t-statistics. However, this observation is reversed once the vertical return differences are analysed in the non-parametric test setting, with the exception of the lowest F-rank portfolios. Although the returns remain significant and comparable with those of the F-Score method, the higher MWW t-statistics indicate that the F-rank can improve the investment results by subdividing the portfolios into BM terciles. If BM is regarded as a proxy for the financial distress of a company, this method could serve as a means to adjust better to the investor's risk requirements, as explained before with regard to the F-Score two-way portfolios.

Essentially the same applies to the median portfolios, as illustrated in panel B. In addition, the vertical return differences now also turn out to be significant for the high and highest F-rank portfolios at the 10% and 5% level, respectively, compared with the F-Score portfolios. In view of the overall performance of the F-Score and F-rank methods in the two-way setting, the two strategies could be considered equivalent. Using the F-Score method results in larger horizontal portfolio differences in all the cases of the mean returns (panel A) and most of the time for the median returns in panel B compared with the F-ranks. This is particularly interesting for investors who are familiar and allowed to take short positions, such as hedge funds. The disadvantage is that the results are not continuously significant and thus partly lack reliability. This is true in particular for the high F-Score median portfolios in panel B. Therefore, the F-rank method seems to be the better choice overall.

4.5.3 The F-rank and the book-to-market ratio: The median F-rank

To complete the analysis of the basic F-rank strategy, the F-rank portfolios are now constructed by computing the median of the Fi-ranks. Table 4.14 presents the findings, which, with regard to the vertical differences, show no significance across all the portfolios for the highest (H*) F-Score fquintile. This applies to all three statistical test methods. Compared with the original F-Score and especially the mean F-rank portfolios, this is a clear drawback. In terms of the vertical percentage portfolio differences between the high-BM and the low-BM firms, the results are also weaker than following the F-Score and mean F-rank investment strategy. The reason is that, although the median F-rank approach can generate strong economically and statistically significant results, this is limited to the lowest fquintile. For all the remaining fquintiles, the other two methods are clearly superior.

Table 4.14 Two-way median F-rank portfolios

A	*F-rank (median)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.139	-0.066	-0.006	-0.002	.060* *(1.69)*	.137** *(2.05)*	*(10.96)****	*(7.36)****	-	-
H	-0.012	-0.019	0.039	-0.018	.058*** *(3.09)*	-.006 *(-0.07)*	*(5.09)****	*(0.75)*	-	-
H–L	.127	.047	.045**	-.016						
t-test	*(1.16)*	*(1.24)*	*(2.62)*	*(-.28)*						
MWW	*(4.60)****	*(9.17)****	*(3.87)****	*(0.24)*						
winsd.										
L	-0.210	-0.118	0.001	-0.008	.119*** *(7.03)*	.202*** *(5.01)*	*(11.00)****	*(7.38)****	-	-
H	-0.054	-0.013	0.052	-0.013	.065*** *(3.80)*	.041 *(0.66)*	*(5.07)****	*(0.74)*	-	-
H–L	.156***	.105***	.051***	-.005						
t-test	*(2.60)*	*(5.79)*	*(3.16)*	*(-.10)*						
MWW	*(4.58)****	*(9.18)****	*(3.85)****	*(0.20)*						
B	*F-rank (median)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.375	-0.233	-0.063	-0.134	-	-	-	-	.170*** *(99.34)*	.241*** *(40.76)*
H	-0.166	-0.094	-0.016	-0.116	-	-	-	-	.078*** *(20.16)*	.050 *(0.25)*
H–L	.209*** *(24.34)*	.139*** *(79.05)*	.047*** *(7.53)*	.018 *(0.48)*						
winsd.										
L	-0.355	-0.207	-0.048	-0.124	-	-	-	-	.159*** *(103.55)*	.231*** *(39.23)*
H	-0.154	-0.085	-0.006	-0.096	-	-	-	-	.079*** *(18.99)*	.058 *(0.91)*
H–L	.201*** *(22.66)*	.122*** *(79.05)*	.042*** *(7.18)*	.028 *(0.48)*						

In terms of the horizontal differences, the results are slightly inferior to the mean F-rank but to a much greater extent to the F-Score method regarding both statistical and economic significance. In fact, the F-Score clearly outperforms both of the F-rank approaches with only very few exceptions. As an overall result, therefore, one can say that, due to less statistical power and lower return differentials both vertically and horizontally, the median version of the F-rank method appears not to be sufficiently qualified to be considered by investors for their investment needs. However, as a means for checking robustness, this method confirms both the underlying logic of the original F-Score approach and the better performance of high-BM firms.

4.6 Two-way alternative F-rank combinations: An overview

In the next two sections, the analysis turns its attention towards the different alternative combination measures of the F-rank method, namely DMSFE and clusters, which are similar to the ones presented for the one-way portfolios earlier. So far, the empirical analysis has shown that the most reliable results are associated with the winsorised version of the percentage portfolio returns. This is true for all the one-way portfolios, which were structured according to the original F-Score, F-rank, weighted DMSFE and clusters (both $k = 2$ and $k = 3$). To test for this persistence in the two-way environment, that is, with the inclusion of the BM ratio as a second dimension, the results are shown taking into account the non-winsorised returns as well. Those results are displayed in tables 4.12 to 4.14. Because the non-winsorised portfolio returns generally follow the trend of the winsorised counterpart, the presentation of the non-winsorised returns is waived in the further course of the analysis. The reason for the omission is twofold. On the one hand, winsorising the returns at the 0.5% level constitutes only a slight modification of the data set, which, however, improves the test results noticeably due to the reduced influence of outliers. On the other hand, the result tables are more focused and can be read more clearly.

4.6.1 Two-way alternative F-rank combinations: The weighted DMSFE

Similar to section 4.3.1, table 4.15 highlights the results for the two-way weighted forecast errors including the BM ratio as a second dimension. In short, the weighting procedure improves the horizontal results of the mean F-rank with respect to the mean returns but underperforms in the case of the median returns. This indicates that the additional effort of constructing the weights does not create sufficient added value and is therefore not justified. However, the general trend of improving results from the L* (L) to the H* (H) fquintiles is upheld and all of the signs of the horizontal portfolio return differences are as expected under the t-test. It should be noted that a trend is only observable for H/L and H*/L* portfolios rather than across L*/L/H/H* portfolios. The reason is that both L*/H* fquintiles are also part of the L/H portfolios. Further, all of the differences between the high and the low portfolios are significant at least at the 5% level in this scenario, as documented in panels A.1 and B.1. Overall, it can be seen that, horizontally, the weighted and the winsorised weighted approach are very similar statistically and economically. This observation is mostly confirmed with a view to the vertical differences. Both approaches yield significant differentials under the t-test for all but the L* fquintiles with the correct, specifically positive, signs. Compared with the mean F-rank approach (table 4.13), the vertical differences are more pronounced in the H and H* fquintiles with strong significance. Although the L fquintiles are significant, their percentage returns are lower, while the L* fquintiles are insignificant. The picture looks very similar to the F-Score results in table 4.12 as the weighted DMSFE is only able to outperform vertically in the H and H* fquintiles.

Table 4.15 Two-way weighted DMSFE portfolios

A.1	***F-rank (weighted)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	0.007	-0.010	-0.132	-0.187	-.122*** *(-7.29)*	-.194*** *(-4.78)*	*(-11.31)****	*(-7.92)****	-	-
H	0.059	0.052	-0.014	-0.043	-.066*** *(-4.04)*	-.102** *(-2.01)*	*(-5.40)****	*(-3.96)****	-	-
H–L	.052	.062***	.118***	.144***						
t-test	*(1.34)*	*(4.06)*	*(6.62)*	*(2.61)*						
MWW	*(1.90)**	*(4.65)****	*(10.05)****	*(4.20)****						
A.2	***F-rank (weighted)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.044	-0.010	-0.154	-0.181	-	-	-	-	-.144*** *(109.23)*	-.137*** *(41.15)*
H	-0.000	0.009	-0.054	-0.165	-	-	-	-	-.063*** *(21.78)*	-.165*** *(15.61)*
H–L	.044 *(1.92)*	.019*** *(9.73)*	.100*** *(94.56)*	0.016*** *(16.00)*						
B.1	***F-rank (winsorised weighted)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	0.009	-0.010	-0.131	-0.185	-.121*** *(-7.24)*	-.194*** *(-4.77)*	*(-11.25)****	*(-7.89)****	-	-
H	0.062	0.054	-0.014	-0.041	-.068*** *(-4.21)*	-.103** *(-1.99)*	*(-5.58)****	*(-3.89)****	-	-
H–L	.053	.064***	.117***	.144***						
t-test	*(1.35)*	*(4.24)*	*(6.58)*	*(2.60)*						
MWW	*(1.86)*	*(4.81)*	*(9.98)****	*(4.18)*						
B.2	***F-rank (winsorised weighted)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.044	-0.010	-0.154	-0.174	-	-	-	-	-.144*** *(108.62)*	-.130*** *(39.55)*
H	-0.000	0.010	-0.053	-0.164	-	-	-	-	-.063*** *(22.43)*	-.164*** *(15.51)*
H–L	.044 *(2.07)*	.020*** *(10.91)*	.101*** *(93.39)*	.010*** *(16.13)*						

Panels A.2 and B.2 report the median portfolio returns and show solid return differences horizontally, underlined by high t-statistics. Still, this performance succumbs both statistically and economically to the F-rank and F-Score methods. In terms of the vertical differentials, the two-way weighted DMSFE procedure appears to be particularly strong in the H and H* fquintiles, the test statistics of which outperform those of the F-rank and F-Score. However, this observation is reversed entirely once the focus shifts to the L and L* fquintiles.

Nevertheless, all the signs are as expected, thus confirming the fundamental tenet of value investing. Considering the additional time, work and potential costs involved in obtaining the weighted DMSFE portfolios, the findings do not encourage investors to prefer this method to the simpler F-Score approach.

4.6.2 Two-way alternative combination measures: Clustering

For the last combination method, the $k = 2$ and $k = 3$ clusters are combined in a single table. Similar to the presentation of the two-way DMSFE results, table 4.16 contains the mean (median) portfolio returns in panels A.1 (A.2) and B.1 (B.2). A comparison between panel A.1 and panel B.1 shows that the $k = 2$ clusters provide better results horizontally with regard to the percentage differentials and statistical significance, although the difference is marginal. Vertically, both approaches result in reasonably high t-statistics, whereby the $k = 3$ approach outperforms (underperforms) in the lowest and low (highest and high) F-rank portfolios. Overall, the signs are as expected, which is indicated by positive return differentials and t-statistics. In other words, higher c-score portfolios and high-BM firms should perform relatively better.

Principally, the same trend is observable for the median portfolio returns shown in panels A.2 and B.2. All of the horizontal median differences are significant, which is not always true in the vertical setting. Nevertheless, the $k = 2$ clustering yields better results as even the high F-rank portfolio differences have significant t-statistics of 10.02. Within the two-way setting, it can therefore be concluded that clustering with the $k = 2$ method clearly outperforms its $k = 3$ counterpart.

Table 4.16 Two-way clustered portfolios

A.1	***F-rank (c-score, k = 2)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.196	-0.128	0.000	0.006	.128*** *(7.91)*	.202*** *(5.51)*	*(13.77)****	*(8.91)****	-	-
H	-0.087	-0.003	0.069	0.028	.072*** *(4.07)*	.115** *(2.02)*	*(6.30)****	*(3.05)****	-	-
H–L	.109**	.125***	.069***	.022						
t-test	*(2.19)*	*(7.01)*	*(4.49)*	*(0.48)*						
MWW	*(4.26)****	*(11.42)****	*(4.85)****	*(0.73)*						
A.2	***F-rank (c-score, k = 2)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.350	-0.230	-0.030	-0.038	-	-	-	-	.200*** *(180.12)*	.312*** *(71.52)*
H	-0.183	-0.071	0.024	-0.004	-	-	-	-	.095*** *(26.44)*	.179** *(10.34)*
H–L	.167*** *(20.51)*	.159*** *(127.16)*	.054*** *(10.02)*	0.034 *(0.03)*						
B.1	***F-rank (c-score, k = 3)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.177	-0.126	0.009	0.020	.135*** *(8.49)*	.197*** *(5.65)*	*(15.08)****	*(10.17)****	-	-
H	-0.000	0.005	0.068	0.031	.063*** *(3.44)*	.031 *(0.43)*	*(4.88)****	*(2.25)***	-	-
H–L	.177***	.131***	.059***	.011						
t-test	*(3.05)*	*(7.25)*	*(3.70)*	*(0.25)*						
MWW	*(4.49)****	*(12.30)****	*(3.45)****	*(0.21)*						
B.2	***F-rank (c-score, k = 3)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.341	-0.231	-0.016	-0.020	-	-	-	-	.215*** *(210.82)*	.321*** *(97.20)*
H	-0.135	-0.060	0.017	0.015	-	-	-	-	.077*** *(15.36)*	.150** *(6.45)*
H–L	.206*** *(23.94)*	.171*** *(146.65)*	.033 *(2.67)*	0.035 *(0.17)*						

As mentioned earlier, with regard to the one-way portfolios, the *k = 3* method was recommended, although the differences between the two were not as clear-cut. This has now changed with the inclusion of the BM ratio. For instance, the horizontal H*–L* difference in panel B.1 is insignificant, while it is significant at the 5% level in panel A.1 (t-statistics of 2.02). Additionally, the vertical differences are significant in three out of four cases. However, in general, both cluster methods provide better results than the weighted DMSFE approach. For this reason, clustering is the best alternative choice for investors, as this technique provides a tool for practitioners to construct portfolios with reliable return differences both horizontally and vertically. This insight is based on significant parametric and non-parametric test results and is highly relevant to long-only and long-short investors.

4.7 Relation and relevance to the empirical finance literature

On the one hand, the purpose of the present chapter was to replicate a simple accounting-based investment strategy that was originally proposed by Piotroski (2000) in the UK stock market. On the other hand, this work was extended by examining alternative ways of separating winning from losing stocks. For this purpose, different methods of forecasting were employed, which commonly used the original nine F-Score components as individual stock return forecasts. Ultimately, the latter part borrowed concepts from two strands in the literature, namely fundamental analysis and forecasting. Although investors engage in fundamental analysis to predict stock returns, the methods that they use differ. In the present research, the main difference between the F-Score/F-rank and the alternative combination measures of the F-rank is that the former investment strategy does not require any adjustments and thus can be put into practice much more easily.

As noted by Richardson et al. (2010), the results from accounting-based investment strategies are subject to criticism for two reasons. Firstly, the best working strategies might just be the outcome of in-sample fitting without any external or out-of-sample validity. Secondly, researchers use a so-called "kitchen sink" approach in which stock returns are regressed on varying amounts of financial statement data in arbitrary combinations without any underlying theory. This research is not affected by the first point of criticism. As stated in the objectives of the thesis in chapter two, this research intends to test rather than develop theory. It therefore automatically avoids in-sample fitting and satisfies the requirement of out-of-sample validity because it uses UK stock market data. The second point, however, is much harder to tackle. Although Piotroski (2000) reduced the amount of financial statement variables substantially compared with other studies (e.g. Ou and Penman, 1989; Lev and Thiagarajan, 1993; Abarbanell and Bushee, 1997), he stands accused of choosing the nine F-Score components at will. Although, for instance, the study by Lev and Thiagarajan (1993) was guided by a qualitative approach to finding key variables that drive value, only one of those is considered in the composite F-Score. However, this problem is partly reduced by including a second dimension, which in this chapter is represented by the BM ratio. Hence, the further discussion focuses on the two-way portfolios.

In accordance with the findings of Piotroski (2000), the results from the UK stock market confirm that the F-Score strategy works reliably outside the US. As stated earlier, the alternative F-rank strategy serves as a robustness check as well as determining whether some relevant information is lost during the transformation of accounting data into the binary F-components. Considering the winsorised portfolio returns, both the mean and the median F-rank underperform the F-Score strategy, albeit on a small scale. Likewise, Piotroski (2000) used ranking to test for the robustness of the F-Score and the results showed relative underperformance. He argued that this finding can most likely be explained by the loss of each signal's sign during the ranking process. However, this explanation is partly challenged by some of the results of the vertical differentials, particularly following the clustering method in table 4.16. In this case, this version of the F-rank strategy provides better statistical and economic overall performance. However, compared with this study, Piotroski (2000) does not provide evidence on the performance of portfolios with returns that are calculated using the median. Instead, he mentions that alternative specifications yield better results without providing any further details. For this reason, it is fair to assume that a binary strategy, such as the F-Score, is equivalent to a continuous strategy, such as the F-rank, at least in the case of the UK stock market. Equivalent in this context means that both strategies are clearly useful in discriminating between high-returning and low-returning portfolios within the low-BM and high-BM terciles and across those terciles.

The sequential sorting of firms into F-Score fquintiles with a subsequent division into BM terciles within each of the fquintiles controls for the effect of a company's financial health according to the F-Score when analysing the effect of the BM ratio. As it turns out, high-BM firms outperform their lower counterparts, as indicated by the mostly significant vertical return differences. These observations are consistent with the empirical literature (e.g. Rosenberg et al., 1984; Fama and French, 1992; Lakonishok et al., 1994) and apply to the broad UK stock market as well. Based on evidence from Piotroski (2000) and Asgharian and Hansson (2010), however, risk can be ruled out as a driving force behind the outperformance of high-BM firms. Naturally, the question arises of which other factor is responsible for this conundrum of better-performing high-BM stocks. This question is dealt with in the next chapter.

The second part of this chapter extended the original idea of the F-Score and analysed the performance of different forecasting models with the aim of improving the results of the simple accounting-based investment strategy. It comes as no surprise that these methods are likewise subject to the criticisms mentioned, for instance, by Richardson et al. (2010), and although they have their justification, forecasting future returns with accounting data is a step in the right direction. Another problem is caused by the persistent past attempts of researchers to forecast stock returns. In this regard, Welch and Goyal (2008) noted that it proves difficult to grasp the literature entirely because different methods and time periods are used by researchers. In general, however, although some agreement exists about stock return predictability, the authors re-evaluated the plethora of potential predictor variables and concluded that most of them perform poorly. However, this applies in particular to macro-economic variables, suggesting that a top-down investment approach falls short of superior return expectations. By contrast, the literature on forecasting stock returns according to the bottom-up method is scarce but evolving, as mentioned by Rapach et al. (2010). This is in particular true of the combination of the F-Score with the forecasting models applied in this research.

As Piotroski (2000) did not provide insights into the performance of those models in the US stock market, unfortunately no direct comparison can be made between outcomes of the same model, only between outcomes of different models. With this in mind, it can be said that the two main forecasting models employed in this thesis, DMSFE and clustering, deliver acceptable results but generally underperform both the original F-Score and the F-rank method. Amongst the two inferior models, the cluster approach works best, particularly once the subdivision into $k = 2$ clusters is made. To establish some comparability, the performance of these forecast methods with related applications in the literature is likely to shed light on their overall usefulness.

Similarly to this research, Jordan et al. (2014) analysed whether stock returns can be forecast out of sample (OOS) for 14 European markets. Although the authors used different fundamental variables and included macro-economic and technical predictors, they confirmed the earlier findings that combinations of single forecasts outperform single-variate models (e.g. Bates and Granger, 1969; Stock and Watson, 2004; Rapach et al., 2010). However, apart from analysing the correlations between the F-components and the one-year stock returns, this thesis does not specifically test the performance of those individual forecasts. In fact, though, this is not necessary on logical grounds. Assuming the validity of the previous empirical findings in the literature, if the combined methods, in this case DMSFE and clusters, underperform the F-Score/F-rank approaches, then the nine individual forecasts must perform even worse. Consequently, the question to be answered is not whether combined forecasts outperform individual ones but whether combining them can improve the results to such an extent that it outperforms the original investment strategy, that is, the F-Score introduced by Piotroski (2000).

Due to the inferior results of the alternative combination techniques described earlier, it is justifiable to raise at least two follow-up questions. On the one hand, it should be investigated whether a nonlinear forecasting model would lead to better results. On the other hand, reasons for the superior performance of the F-Score/F-rank should be given. Regarding the first question, Deutsch et al. (1994) documented that nonlinear combination methods improved the accuracy of forecasting the UK inflation rate. Although their model is easy to implement, it is not practical because the data sample in this research is too large. Similar problems were observed by Stock and Watson (2004), who concluded that this requirement made nonlinear combination methods ineffective. While nonlinear models are developed because researchers reject the assumption of an innate linear relationship between variables, the same logic applies vice versa. Another explanation is the unsatisfactory performance of linear models in the finance literature, as remarked by Clements et al. (2004). However, the performance of nonlinear models indicates that they are unable to simulate reality any better than their linear counterparts. Generally, nonlinear models feature a large number of variables, which even exacerbates the criticism of the kitchen sink approach put forward by Richardson et al. (2010).

With regard to the second question, no clear answer was given by Piotroski (2000). The question of whether the success of his investment strategy is based on market inefficiencies or rational finance will be addressed in the following chapter. An alternative explanation for the superiority of the F-Score/F-rank methods might be their better ability to reduce noise. This was picked up on by Jordan et al. (2014). Although the UK has a developed stock market, its size is smaller than the US market, which implies greater noise and therefore noticeable differences in performance.

4.8 Chapter summary

This chapter presented the results of the original F-Score investment strategy from the UK stock market. Firstly, the main parts of this strategy were replicated to evaluate its usefulness in a different investment environment. Secondly, the individual components of the F-Score were ranked, standardised and tested. Thirdly, the differences between the resulting nine Fi-ranks and the ranked one-year future returns were then used as single forecasts and subsequently combined in alternative ways with the aim of improving the original F-Score method. The two main approaches used in this process were the discounted mean square forecast error (DMSFE) and cluster models. Fourthly, the results of those models were compared with each other and evaluated. This process was repeated by introducing a second dimension, the book-to-market ratio. In accordance with the original work, the overall findings showed that the F-Score method generally dominates the two combination strategies. Finally, the results were discussed and linked to the present state of the literature.

With regard to the research questions posed at the beginning of this chapter, the analyses yielded the following results. In general, the first hypothesis is confirmed. Both the original F-Score and its alternatives are suitable means for differentiating promising from underperforming stocks in the UK. Many explanations seem valid, but the argument of huge overlaps between the characteristics of the US stock market and those of the UK stock market is most likely. As suggested by Roosenboom and van Dijk (2009), similarities are determined by either bonding, meaning the voluntary adoption of stricter regulatory requirements by a company through a stock listing abroad, or market segmentation, that is, increasing the general access to capital and/or reducing the cost of capital. With regard to the first point, Piotroski and Srinivasan (2008) analysed the listing preferences of foreign firms in the US (NYSE and NASDAQ) and the UK (Main Market) after the introduction of the Sarbanes–Oxley Act (SOX). As opposed to small firms, which have to bear increased listing costs due to the stricter requirements of the SOX, the listing preferences of larger firms did not change. With respect to the second point, Baker et al. (2002) showed that international firms can substantially increase their visibility by listing their stock on either the NYSE or the LSE, which is accompanied by a reduction of the cost of equity capital. Because either of those markets fulfils the requirements of bonding and market segmentation, a high degree of homogeneity can be inferred, which makes the investment strategy transferable.

The main contribution of this chapter concerns the second hypothesis, which states that alternatives to the original F-Score investment strategy should outperform the original F-Score strategy. However, the actual results indicate otherwise. Although in general these strategies can also separate winning from losing stocks, they are less efficient in doing so and investors are not rewarded for the additional effort that they exert on constructing them. In this regard, therefore, the simplest method, that is, the F-Score, turns out to be the best, which agrees with the findings of Rapach et al. (2010), because introducing complexity into forecasts does not improve the results. However, the assumption of isolating only those Fi-ranks that most accurately predict future returns proves to be successful insofar as the return results beat the no predictability benchmark. In particular, the cluster approaches usually deliver the best results within the group of alternative investment strategies. While complexity is introduced into both the weighted DMSFE and the cluster approach due to the weighting process, this differs with regard to the F-rank approach. Here, the intention was to counter the seemingly negative effects of the binary nature of the F-Score. Because of the inferior results, it can therefore be concluded that the expected benefits of forecast combinations and a more detailed measurement of financial statement data are outweighed by the drawbacks of the improvement process itself. Despite these results, however, it needs to be highlighted that those strategies can be regarded as a robustness check for the overarching concept of successful stock picking according to a firm's underlying financial signals.

As mentioned earlier, Piotroski (2000) tested the F-Score approach only amongst high-BM firms, as opposed to this study, which contains all firms. For this reason, the results are not directly comparable. Nevertheless, the findings indicate that both the original and the alternative strategy are indeed useful tools to be applied successfully to low-BM firms. Interestingly, the horizontal return differentials of low-BM firms almost always exceed the ones of high-BM firms, which makes this strategy especially interesting for long/short traders. Therefore, the third hypothesis is confirmed.

The findings of this chapter have advanced the knowledge in three ways. Firstly, it was shown that the original F-Score investment strategy is transferable and works reliably in the UK. Secondly, the simplicity of this approach can generally be neither matched nor improved by more sophisticated alternatives as presented here. Lastly, the strategy is not limited to high BM firms but is also applicable to stocks with a wider range of characteristics than initially claimed.

However, there is further need for research on the question of whether these investment strategies could be extended by a measure that captures intangible information. For instance, Jiang (2010) documented that institutions trade in the direction of intangible information and thus fuel the book-to-market effect, that is, the value premium. This outcome was further substantiated by Edmans (2011), who equated intangibles with employee satisfaction and reported an annual alpha of 2.1% above the industry benchmarks. Hence, a combination of quantitative and qualitative measures could improve the strategy without following a kitchen sink approach.

Chapter 5 Information Uncertainty and Liquidity – The UK

5.1 Introduction

This chapter builds further on the findings made previously regarding the book-to-market ratio by shifting the focus to the potential causes of the success of the various investment strategies. In other words, it is particularly concerned with the questions of what the key drivers of the F-Score-based investment strategy are and whether these factors are common amongst the alternatives. It first explores how firm size, measured as a company's market capitalisation, is relevant to the success of the original F-Score and variant alternative strategies. This question is of interest because, unlike a generally positive correlation between the book-to-market ratio and the subsequent returns, which was analysed in conjunction with investment strategies in the previous chapter, various anomalies are concentrated particularly in small firms (Fama and French, 2008; Fama and French, 2012). For instance, both value premiums and return momentum are usually higher for smaller companies.

Previous work also suggests that anomalies are linked to information uncertainty and liquidity risk. However, this has not been extensively previously examined in the case of the financial health anomaly (Piotroski, 2000). As size proxies for either information uncertainty (IU) or liquidity, the analysis in the next step evaluates how dedicated measures of IU and liquidity are correlated with the variant strategies. Regarding information uncertainty, Jiang et al. (2005) studied the role of IU in forecasting cross-sectional stock returns. According to their definition, IU does not relate to information asymmetry between insiders and others but rather to how dependably their cash flows are predictable. The authors noted that in the classic asset pricing models, such as the CAPM, non-systematic risk is not priced. However, the empirical results show that it is indeed priced but in the wrong direction. Stocks with a higher degree of IU should feature higher, not lower, expected returns. In their study, they used four variables,[12] two of which are of interest in the present research, namely return volatility and daily turnover. Firm age, for instance, is an absolute variable that changes by the same degree amongst the entire sample. Return volatility and daily turnover, however, change in relation to the other firms in the sample. Due to their more pronounced short-term nature, they are considered more suitable for predicting one-year subsequent returns. Nevertheless, Jiang et al.

[12] Firm age, return volatility, average daily turnover and duration of future cash flows

(2005) found a negative relationship between all four variables and future returns. In a related study by Zhang (2006), a different set of IU variables was used but with similar results. One disadvantage was pointed out by Jiang et al. (2004), who were aware of Zhang's (2006) study during its working paper status. The sample is confined because it only contains stocks with analyst coverage. As opposed to the US, this may have undue influence on the research results with regard to the UK stock market. For this reason, firm size and two novel volatility measures that are derived from the F-Score strategy are used in this study.

The remainder of the chapter is concerned with analysing the interplay between liquidity as a measure of information uncertainty and stock returns. The link between accounting information, information uncertainty and liquidity might not be apparent at first but was studied by Lang and Maffett (2011). Their results suggest a negative relation between accounting transparency, that is, low information uncertainty, and a stock's liquidity. This effect was found to be even stronger during crisis periods because of "flights to quality". While their study concerned the liquidity on the firm level, Ng (2011) investigated the correlation between a stock's return and market liquidity. In this context, the author highlighted the difference between liquidity and liquidity risk. The latter term refers to the sensitivity of a stock's (i) return (r) to unanticipated changes in market (m) liquidity (l), namely $\mathrm{cov}(r_i, l_m)$, whereas the former deals with the ability to trade large quantities of a share quickly, cheaply and without considerably affecting its stock price, that is, $\mathrm{cov}(l_i, l_m)$ and $\mathrm{cov}(l_i, r_m)$. Although liquidity is most likely to play a part in measuring information uncertainty, this study focuses on liquidity risk. Similar to Ng (2011), it is anticipated that lower information uncertainty is negatively connected with liquidity risk.

To maintain the focus on the analysis at hand, the development of hypotheses is kept separate and appears where appropriate. Two-way results are presented in tabular form, each having a similar structure to the tables in the previous chapter. Similar to chapter 4, the reader is referred to chapter 3 and the appendix of this thesis with regards to the methodology used. Finally, a discussion of the results obtained is provided, which concludes this chapter.

5.2 Hypothesis development: Firm size as an indirect information uncertainty proxy

Firm size appears to be a natural variable for measuring information uncertainty (Zhang, 2006). The reason is less diversification and less information availability due to fewer stakeholders and fewer disclosure requirements by the stock exchange for smaller firms. Generally, information on large firms is available both at lower costs and more publicly (Bhattacharya, 2001). As most investors have predefined information acquisition costs, some smaller firms are unattractive to them as investment targets. However, firm size is most likely to be only an approximation of information uncertainty according to Zhang (2006), because it might also capture other things, such as earnings management. This activity appears to be less pronounced amongst larger firms because institutional investors, who are regarded as sophisticated, are usually keen to monitor the likelihood of earnings management (Balsam et al., 2002).

Therefore, Zhang (2006) introduced the concept of information uncertainty, which is a measure of investor reaction to new information. In his study, the horizontal subdivision into fquintiles was based on average monthly portfolio returns that proxy for momentum. The secondary or vertical subdivision is according to fquintiles of six information uncertainty proxies, of which firm size is just one. He found that in general smaller firms with supposedly greater information uncertainty provide better (worse) returns following good (bad) news. By definition, therefore, large (small) firms should feature lower (higher) information uncertainty, because, for instance, larger firms are required to meet higher disclosure standards and thus leave less room for interpretation. As defined by Zhang (2006), information uncertainty is a function of a firm's fundamental value and a noise term. Therefore, the variation of the information uncertainty signal is, in turn, a function of the volatility of the fundamental value and the volatility of the noise term. More formally:

$$IU = v + e \tag{1}$$

and

$$var(IU) = var(v) + var(e) \tag{2}$$

The fundamental value is represented by *v* and the noise term or measure of information quality by *e*. Zhang (2006) did not distinguish between *v* and *e* and justified this choice by stating that both terms contribute to the measure of information uncertainty and separating the two variables is problematic as they are intertwined. Applied to firm size, therefore, smaller firms are characterised by lower information quality which results in more noise (*e*). If firm size is equated with market capitalisation then market capitalisation is representative of fundamental value (*v*). However, this argument is only valid on the assumption that markets are efficient under the efficient market hypothesis (EMH). The underlying idea of both equations is shared in the present research and therefore the further analysis is based on Zhang's (2006) approach. In general, the first hypothesis assumes a gradual investors' response to information which is in line with the behavioural approach taken in the present thesis. Further, according to Hirshleifer (2001), investor biases are augmented when information uncertainty is high. In response to that, Zhang (2006) directly builds on these findings and conjectures that because of larger biases, information will be reflected slower in stock prices when there is more ambiguity about its impact on firm value.

Referring back to the concept of noise mentioned earlier, a clarification is important because noise is likely to have a different level of influence on the research results depending on the time horizon. For instance, Jordan et al. (2014) mentioned that more (less) noise is expected in the short run (long run). However, this argument only applies if the term "noise" is used in a certain way. Black (1986) used this term in many different senses while describing the effects of noise on the world. He observed that a difference should be made between investors who trade on actual information and investors who trade solely on noise that they believe is information. Consequently, if size is regarded as a proxy of information uncertainty, noise should be equal both in the short and in the long run. The reason is that firm size measured as market capitalisation is an external factor, determined by the stock market every trading day, and thus there should not be any ambiguity about this measure. In other words, firm size is believed to be actual information. This reasoning should therefore ensure the comparability of Zhang's (2006) findings, which are based on monthly portfolio returns, and the results of the annual return portfolios used in this thesis.

Of primary interest are studies on the interplay between a firm's stock return, on the one side, and its size and financial strength indicators, particularly those that are based on Piotroski's (2000) idea of the F-Score as a summary measure, on the other side. One of the interesting studies was conducted by Fama and French (2006b), who used the original F-Score method as a means to evaluate the forecasting ability of a stock's return. In general, the authors found that in monthly cross-section return regressions, this simple aggregation of a firm's accounting information was indeed able to provide reliable return forecasts. As presented in the previous chapter, this is also the case in the UK. Firm size, on the contrary, was found to have a negative relationship with subsequent stock returns. Generally, this confirms the so-called size effect, which is well documented in the literature (e.g. Banz, 1981; Fama and French, 1992; Van Dijk, 2011) and agrees with Zhang's (2006) intuition. The difference, however, is that in Zhang's (2006) scenario, small firms do not outperform larger ones per se, because investors require a premium due to less stringent disclosure requirements, but only under the condition of good news. In the present study, the quality of news is determined by the F-Score; therefore, this scenario can be tested in the UK through the first hypothesis.

Hypothesis 5.1: Firms with higher (lower) information uncertainty, proxied by firm size, provide lower (higher) subsequent returns following bad news and higher (lower) returns following good news.

5.2.1 Discussion of the results and links to the literature

The first step in answering questions about the driving forces behind the F-Score strategy is to analyse the relationship between company size terciles and F-Score return fquintiles. Table 5.1 documents the respective findings. To keep the amount of tables to a minimum, only the winsorised mean (panels A.1, B.1 and C.1) and winsorised median portfolio returns (panels A.2, B.2 and C.2) are reported. This structure is maintained throughout the present chapter. Further, table 5.1 contains the results of all three of the investment strategies: the A panels represent the original F-Score, the B panels the averaged F-ranks and the C panels the median F-rank method. Beginning with panel A.1, the horizontal findings show that the F-Score strategy is particularly fruitful for smaller-sized companies and that those results are highly significant. For instance, a long-only strategy based on the highest F-Score fquintile and the smallest firm size thereby yields an annual return of 10.4%. With the additional option to short the lowest F-Score portfolio in the same size tercile, the horizontal mean return difference extends to 23.6 percentage points. The analysis of the median portfolio returns and return differences in panel A.2 reveals a very similar trend, with highly significant median test results. This performance of small firms is unmatched by the larger firms with one exception in panel C.2.

Table 5.1 Two-way F-Score/F-rank-size portfolios

A.1	***F-Score, mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.132	-0.110	0.072	0.104	.182*** *(7.86)*	.236*** *(6.13)*	*(11.14)****	*(8.46)****	-	-
B	-0.101	-0.070	0.046	0.049	.116*** *(6.15)*	.150*** *(3.62)*	*(7.24)****	*(4.76)****	-	-
B–S	.031	.040	-.026	-.055						
t-test	*(0.58)*	*(1.49)*	*(-1.40)*	*(-1.47)*						
MWW	*(1.84)**	*(4.59)****	*(2.35)***	*(0.67)*						
A.2	***F-Score, median returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.230	-0.210	-0.016	0.003	-	-	-	-	.194*** *(100.11)*	.233*** *(55.69)*
B	-0.137	-0.079	0.037	0.051	-	-	-	-	.116*** *(29.40)*	.188*** *(11.10)*
B–S	.093** *(4.48)*	.131*** *(20.42)*	.053*** *(9.89)*	0.048 *(1.72)*						
B.1	***F-rank (mean), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.127	-0.075	0.042	0.083	.117*** *(6.30)*	.210*** *(4.82)*	*(10.21)****	*(7.59)****	-	-
B	-0.048	-0.006	0.032	0.051	.038*** *(2.73)*	.099* *(1.88)*	*(3.04)****	*(3.17)****	-	-
B–S	.079	.069***	-.010	-.032						
t-test	*(1.34)*	*(3.91)*	*(-0.62)*	*(-0.70)*						
MWW	*(2.76)****	*(10.28)****	*(3.02)****	*(0.46)*						
B.2	***F-rank (mean), median returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.231	-0.201	-0.037	-0.008	-	-	-	-	.164*** *(84.23)*	.223*** *(42.94)*
B	-0.116	-0.023	0.003	0.019	-	-	-	-	.026* *(2.75)*	.135** *(4.78)*
B–S	.115*** *(7.90)*	.178*** *(129.06)*	.040** *(6.44)*	0.027 *(1.01)*						
C.1	***F-rank (median), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.129	-0.066	0.047	0.042	.113*** *(6.10)*	.171*** *(3.81)*	*(9.36)****	*(5.59)****	-	-
B	-0.153	-0.012	0.034	0.011	.046*** *(3.04)*	.164** *(2.48)*	*(3.93)****	*(4.18)****	-	-
B–S	-.024	.054***	-.013	-.031						
t-test	*(-0.33)*	*(2.75)*	*(-0.77)*	*(-0.54)*						
MWW	*(-0.41)*	*(7.63)****	*(3.46)****	*(1.24)*						
C.2	***F-rank (median), median returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.289	-0.192	-0.048	-0.105	-	-	-	-	.144*** *(63.35)*	.184*** *(19.37)*
B	-0.261	-0.037	0.005	-0.033	-	-	-	-	.042*** *(8.67)*	.228*** *(11.11)*
B–S	.028 *(-0.10)*	.155*** *(69.65)*	.053*** *(9.80)*	0.072 *(1.17)*						

With regard to the two alternative combination methods of the F-rank, the findings confirm the benefits of the original F-Score strategy, which is to focus on companies with small market capitalisation. Overall, though, both the absolute returns and the return differences between the high/low and the highest/lowest portfolios are less pronounced than with the F-Score method. Although all the t-statistics of the small firms' portfolios are significant at the 1% level, the return differences between the highest/lowest portfolios within the large-firm tercile turn out to be differences between the highest/lowest portfolios within the large-firm tercile turn out to be only marginally significant at the 10% (panel B.1) or 5% level (panel C.2), respectively. The evaluation of the corresponding median portfolio returns provides a very similar picture again. However, both F-rank methods underperform the F-Score strategy in absolute percentage terms with only one exception in panel C.2. Comparing the two F-rank methods with each other, no clear answer can be given regarding which of the two is preferable. The reason is that the F-rank mean return portfolio differences in panel B.1 perform better than the median F-rank counterpart for small firms. For instance, the highest/lowest horizontal difference amounts to 21% for the mean F-rank, whereas the median F-rank method yields only 17.1%. However, this observation is completely reversed for large firms and likewise applies to the median portfolio returns. Overall, therefore, the greater benefits of one strategy over the other cancel each other out depending on the area of scrutiny.

The more interesting part is the analysis of the vertical return differences between portfolios consisting of the largest and smallest firms (B–S or big minus small) as it provides answers to the first hypothesis. It aims to determine whether firm size in combination with the three strategies as an indicator of information quality behaves as anticipated by Zhang (2006). If this is the case, small firms should provide better (worse) results following good (bad) news than large firms. The quality of news is thereby determined by the respective fquintile and follows the logic of Piotroski (2000) insofar as higher (lower) fquintiles are associated with good (bad) news. Focusing on the F-Score strategy in panel A.1 first, this trend is generally confirmed. Small firms in the lowest and low fquintiles are outperformed by their larger competitors, as predicted by Zhang (2006). The opposite is the case for the high and highest F-rank fquintiles. However, the results of the parametric t-tests are insignificant and suggest that the portfolio returns within the F-Score fquintiles are not driven by firm size.

In view of the non-parametric tests, the MWW results document some significance for all the vertical return differentials except for the highest F-Score portfolio. On the one side, small firms in the lowest F-Score portfolio return -13.2% over the year, whereas large firms return -10.1%. The difference of 3.1 percentage points is significant at the 10% level (t = 1.84). This seems to imply that size proxying for information uncertainty is positively correlated with one-year returns for companies with the lowest/low F-Scores, as initially observed by Zhang (2006) for the US stock market. On the other side, the high and highest F-Score portfolios within the small-size tercile have higher one-year returns of 7.2% and 10.4%, respectively. However, although the return difference between the big minus small (B–S) tercile is only significant for the high but not the highest F-Score portfolios, the signs are not as expected. In other words, large firms continue to outperform smaller ones based on positive MWW t-statistics throughout.

These findings are further supported by the mean portfolio returns in panels B.1 and C.1. While large firms significantly outperform smaller ones in the low F-Score fquintile, the remaining parametric test results are not significant. All the significant differentials in those two panels feature a positive sign in the case of the non-parametric t-statistics. Consequently, larger firms are more profitable within their respective fquintile. In addition, these findings are consistent with the results of the median return portfolios reported in panels A.2, B.2 and C.2 and therefore point to the same conclusion. As can be seen from there, larger firms consistently outperform their smaller peers, with the majority of the return differences being highly significant. In summary, in relation to the mixed parametric test results, both the MWW and the median test statistics appear to be strong and thus should be regarded as determinative in this case.

With regard to the UK stock market analysed in this chapter, the results of the F-Score and both of the F-rank methods as presented in table 5.1 confirm Fama and French’s (2006b) findings. However, this depends on the scenario under scrutiny. In the case of the horizontal return differentials, both the parametric and the non-parametric test results strongly underline the strength of the F-Score/F-rank as a forecasting tool. In the vast majority of cases, those results are significant at the 1% level in both of the test settings. Taking into account firm size, that is, market capitalisation next, the vertical big minus small (B–S) firm size differentials are mostly positive, meaning that larger firms outperform smaller ones, at least in those cases in which the results are significant. This holds true for all the panels reported in table 5.1.

Given the weak results of the vertical t-tests, which show no significance in the majority of cases, the question could be posed of whether the size effect is still a determinant in today’s stock market of the UK. In general, the debate about the death of the size premium is still ongoing amongst academics and the discussion about the reasons for its existence appears to be far from settled (van Dijk, 2011). Earlier evidence for a reversal of the vanishing small-size premium in the UK was provided by Dimson and Marsh (1999). In the periods of 1983 to 1997 and 1988 to 1997, the premium reverted to minus 2.4% and 5.6%, respectively. More recent evidence for the UK in particular was presented by Andrikopoulos et al. (2008), who found a continuation of the small-size premium. The authors thereby focused on the period from January 1988 to December 2004, which widely overlaps with the time period used in this study. However, they noted that the presence of the size premium was influenced by i) how the term “small firm size” is defined, ii) what kind of stocks are considered and iii) whether equal or value weighting was used in constructing the portfolios. Overall, they conceded that it was also time-varying and generally unreliable and concluded that overall their results corresponded to the study by Dimson and Marsh (1999). Further international evidence was brought forward by Fama and French (2006) in an international setting. Between January 1974 and December 2004, the authors tested a merged data set of 14 industrialised stock markets, including the UK and Germany, sorted on both BM and EP (earnings to price) ratios. Although a value premium was documented in small and large firm portfolios, they were nearly identical, albeit with slightly higher premiums for large firms.

With regard to point i) of the previous paragraph, one striking difference between Zhang’s study (2006) and this research is the sub-division of firm size into quintiles rather than terciles. Which one of these approaches is the right choice is arguable. On the one hand, the larger US data set might facilitate a finer separation between portfolios to observe any significant effect. Although this method may raise criticisms about data mining, Zhang (2006) also reported similar but untabulated results once subdivision was made into terciles and even deciles. For this reason, table 5.1 presents size terciles only for the UK, assuming that the effect is pervasive irrespective of the subdivision method with regard to the mean portfolio returns. Further subdivision of portfolios according to firm size bears another drawback, which is rooted in the asymmetric effect of delistings on portfolios depending on their size. In other words, portfolios of large stocks are virtually unaffected by dropouts due to delisting, whereas small-size portfolios are not. The reason is that the probability of delisting is about ten times higher for small firms in the UK (Andrikopoulos et al., 2008). Overall, the authors confirmed the findings of earlier studies insofar as the small-size premium is time-dependent and unreliable, which concurs with Fama and French’s (2006) findings. Most importantly, the size effect also depends on the choice of the investment strategy to exploit it. With this in mind, the results of the original F-Score and the two F-rank alternatives indicate that the absorption of new information into stock prices is incomplete due to significant horizontal return differentials. However, this effect is not dependent on a firm's market capitalisation, which is confirmed by spurious parametric evidence for and strong non-parametric evidence against Zhang's (2006) findings.

Finally, table 5.1 shows not only the results for the mean portfolio returns but also the respective median returns. One reason why the findings from the US study by Zhang (2006) are not entirely confirmed for the UK could be attributed to a different degree of positive skewness in the data. As presented in table 4.3, both the UK's and Piotroski's (2000) data sets are skewed to the right because the mean buy-and-hold returns are higher than their median counterparts. Zhang (2006) also reported slight positive skewness but failed to provide median portfolio returns. If the median test results are akin to those in the present thesis, this would challenge the validity of his initial hypothesis.

5.3 Size and alternative combinations of the F-rank: An overview

The next section presents the results from the analysis of the alternative DMSFE combination techniques. The aim is to keep the discussion of the test results with regard to the recent findings in the literature as concise as possible. This procedure is opposite to the previous section and was chosen to keep the focus on the research findings first. However, before moving on to the second part of this chapter, which deals with measures of liquidity, the link between the research findings of this thesis and the literature will be provided.

5.3.1 The weighted DMSFE in the UK

In the next step, the analysis focuses on the relationship between the firm size and the alternative combination methods of the F-rank, that is, the weighted DMSFE. Table 5.2 below presents the results for the weighted DMSFE portfolios including size as the second dimension. Here, too, only the winsorised mean portfolio returns are included in panels A and B because the median results are indicative that large (B) firms consistently yield higher annual returns than smaller (S) ones. The horizontal return differentials are almost identical to those of the F-rank approach shown in table 5.1, indicating a high level of significance and the correct sign. This applies to both the parametric and the non-parametric test statistics, with the majority of return differences being significant at the 1% level. Although the results are generally solid, they lag behind those of the F-Score strategy.

Table 5.2 Two-way weighted DMSFE–size portfolios

A	***F-rank (weighted), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	0.068	0.039	-0.078	-0.126	-.117*** *(-6.35)*	-.194*** *(-4.45)*	*(-10.10)****	*(-7.49)****	-	-
B	0.037	0.034	-0.018	-0.064	-.052*** *(-3.73)*	-.101* *(-1.78)*	*(-4.13)****	*(-3.30)****	-	-
B–S	-.031	-.005	.060***	.062						
t-test	*(-0.71)*	*(-0.30)*	*(3.35)*	*(0.96)*						
MWW	*(0.25)*	*(3.30)****	*(9.13)****	*(2.06)***						
B	***F-rank (winsd weighted), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	0.068	0.039	-0.079	-0.126	-.118*** *(-6.34)*	-.194*** *(-4.45)*	*(-10.09)****	*(-7.49)****	-	-
B	0.037	0.034	-0.018	-0.062	-.052*** *(-3.74)*	-.099* *(-1.76)*	*(-4.16)****	*(-3.24)****	-	-
B–S	-.031	-.005	.061***	.064						
t-test	*(-0.71)*	*(-0.27)*	*(3.35)*	*(0.99)*						
MWW	*(0.25)*	*(3.35)****	*(9.13)****	*(2.13)**						

Regarding the vertical return differences, the t-test statistics of 3.35 in panels A and B only indicate significance between small and big firms in the H-fquintile, while the other fquintiles remain insignificant. Therefore, the performance of the weighted DMSFE has not improved compared with the F-rank, for which only the L-fquintile is highly significant. However, both strategies outperform the F-Score in table 5.1, in which none of the parametric test statistics are significant. Taking into account the MWW results, the differences are significant at least at the 5% level, except for the L-fquintile. This is opposed to the F-rank and F-Score methods, in which all but the H* differences are significant.

In summary, the research findings of the alternative combinations of F-rank, that is, the weighted DMSFE, do not support Zhang's (2006) hypothesis in its entirety. This is because smaller firms perform worse than their larger counterparts following bad news. However, in contrast to Zhang (2006), this is also the case regarding good news. Despite this outcome, the weighted DMSFE methods are still useful in implementing an investment strategy that is based on firm size. This is due to the relatively strong test statistics with respect to the horizontal and vertical portfolio return differences.

5.3.2 Clusters

As previously reported, the results of the F-Score/F-rank methods and their alternative combinations uphold the basic idea that firms with better fundamental results or alternatively with lower forecast errors usually perform better (Piotroski, 2000).

Table 5.3 Two-way clustered-size portfolios

A.1	***F-rank (c-score, k = 2), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.151	-0.041	0.046	0.055	.087*** *(4.25)*	.206*** *(4.28)*	*(7.78)****	*(6.41)****	-	-
B	-0.161	-0.033	0.040	0.054	.073*** *(4.89)*	.215*** *(4.00)*	*(6.49)****	*(5.42)****	-	-
B–S	-.010	.008	-.006	-.001						
t-test	*(-0.14)*	*(0.34)*	*(-0.40)*	*(-0.01)*						
MWW	*(-0.16)*	*(4.09)****	*(2.98)****	*(0.79)*						
A.2	***F-rank (c-score, k = 2), median returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.268	-0.162	-0.033	-0.004	-	-	-	-	.129*** *(49.34)*	.264*** *(27.71)*
B	-0.230	-0.061	0.019	0.010	-	-	-	-	.080*** *(29.41)*	.240*** *(20.81)*
B–S	.038 *(0.04)*	.101*** *(19.75)*	.052*** *(12.15)*	0.014 *(0.32)*						
B.1	***F-rank (c-score, k = 3), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.108	-0.042	0.049	0.038	.091*** *(4.32)*	.148*** *(2.69)*	*(7.81)****	*(6.29)****	-	-
B	-0.080	-0.028	0.034	0.032	.062*** *(4.08)*	.112** *(2.11)*	*(5.33)****	*(3.76)****	-	-
B–S	.028	.014	-.015	-.006						
t-test	*(0.33)*	*(0.63)*	*(-0.92)*	*(-0.17)*						
MWW	*(1.05)*	*(4.36)****	*(2.44)***	*(0.33)*						
B.2	***F-rank (c-score, k = 3), median returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.273	-0.163	-0.036	0.002	-	-	-	-	.127*** *(45.94)*	.275*** *(34.87)*
B	-0.189	-0.046	0.009	0.006	-	-	-	-	.056*** *(12.65)*	.195*** *(10.07)*
B–S	.084* *(2.84)*	.117*** *(30.09)*	.045** *(5.48)*	0.004 *(-0.02)*						

However, this does not always hold true once a company's size is included in the analysis. Following the structure of the previous chapters, the remaining two combination methods with $k = 3$ and $k = 2$ clusters are presented now.

Table 5.3 presents four panels that report the portfolio return results according to the $k = 2$ and $k = 3$ clustering. Panels A.1 (A.2) and B.1 (B.2), respectively, contain the mean (median) winsorised one-year returns. Again, to replicate the logic of the F-Score, firms with a high or alternatively the highest c-score are supposed to perform best. Therefore, the t-statistics should be positive. With regard to the hypothesis of Zhang (2006), this implies that high (low) c-scores are tantamount to good (bad) news. Small firms in the high and highest fquintiles in panels A.1 and B.1 outperform their larger peers as hypothesised, but the results are not significant under the t-test. The opposite is mostly the case in the low and lowest fquintiles but again without significance. The portfolios in the high c-score fquintile of panel A.1 are used as an example. There it can be observed that the small-size portfolio earns 4.6% per annum, whereas the large one only returns 4%. Apart from the t-statistics of -0.40, it can also be observed that all the other vertical test statistics are insignificant as well. This is also true for panel B.1, and both panels show insignificant MWW results for the lowest and highest portfolios. Besides, the t-statistics of 2.98 (2.44) for the high and 0.79 (0.33) for the highest fquintile in panel A.1 (B.1) have the wrong sign and thus do not confirm the information uncertainty hypothesis by Zhang (2006).

Before moving on to the horizontal differences, the vertical median portfolio returns are briefly evaluated next. Panels A.2 and B.2 principally do not deliver any new insights. Of statistical significance are only the low and high portfolio differentials, with the latter having the wrong sign once more. In general, this confirms the initial observations of the $k = 2$ and $k = 3$ cluster approaches in which the winsorised mean annual returns were used.

The results of the horizontal return differentials are very strong throughout. Both the parametric and the non-parametric test statistics indicate significance at least at the 5% level but overwhelmingly at the 1% level. In addition, all the signs are positive, which means that the c-score strategy, that is, clustering, works as reliably as the original F-Score. With regard to firm size, this represents a strong benefit for investors compared with the variants of the DMSFE approach presented earlier.

In conclusion, one can say that the findings of the two cluster methods are in accordance with the previous research findings in this chapter with regard to firm size. This means that for the UK stock market the results of neither the DMSFE nor the cluster method indicate that greater information uncertainty, that is, small firm size, is correlated with higher future returns following good news. The non-parametric test results rather indicate that large firms are generally associated with higher stock returns irrespective of the information quality. However, the underlying logic of the F-Score is not affected, which means that in general firms with stronger c-scores always perform better.

5.4 Measures of fundamental value volatility: An overview

The next part of the analysis deals with other measures of information uncertainty. The reason is that, according to Zhang (2006), firm size most likely does not solely proxy for information uncertainty. As described earlier, variable *v* equals the fundamental value measure, which is also an indicator of good or bad news about a firm and is subject to volatility. The question now is which combination of explanatory accounting variables is ultimately best suited to predicting stock returns. Consulting the literature on this topic, Richardson et al. (2010) advised the researcher to let this exercise be guided by theory. Due to the use of a wide range of accounting ratios that cover different aspects of a firm's financial health, the F-Score appears to be a good starting point. Besides, the good performance originally reported by Piotroski (2000) for the US is confirmed by this study with regard to the UK stock market.

5.4.1 Hypothesis development: Direct information uncertainty measures

As noted by Zhang (2006), firm size can be regarded as a more indirect measure of information uncertainty because it captures other things in addition and therefore is likely to distort the research findings. Hence, a more direct indicator should alleviate this potential drawback. In this regard, the F-Score is utilised not only to distinguish between high-returning and low-returning firms but also to generate breakpoints for information uncertainty terciles. One way to achieve that is to quantify the volatilities of each of the nine components and ultimately the composite itself. The underlying rationale behind both the standard deviation (SDF) and the mean absolute deviation (MADF) of the F-Score is that greater volatility of the composite is associated with a higher degree of information uncertainty. In the further course of this overview, the attention turns to the SDF, but the reasoning applies equally to the MADF. There are two main advantages of using the SDF as a proxy for information uncertainty compared with firm size. Firstly, as noted earlier, the SDF is a more direct measure because it is not influenced by the local market conditions. Every exchange supervisory authority might have different disclosure requirements depending on a firm's size. These, in turn, are the direct realisation of a country's legal requirements, which are in place to ensure investor protection. Leuz et al. (2003), for example, found higher earnings management for firms in Germany than in the UK, which suggests different levels of protection between the two markets. Better investor protection in the UK was also confirmed more recently by McLean et al. (2012), who stressed that both the strength of the law itself and its enforcement are positively correlated with the level of investor protection. Consequently, information uncertainty is likely to vary between jurisdictions. These findings could explain why firm size in the UK is not a relevant factor regarding information uncertainty, as documented in this research. The reason is that if the disclosure requirements are very similar between small and large firms, meaning that they apply to all listed firms irrespective of their size, information uncertainty based on size alone cannot be measured as a consequence. This interpretation agrees with the studies by Leuz et al. (2003) and McLean et al. (2012), who remarked that firm size did not influence their findings.

Secondly, the SDF measure captures the volatility of nine different accounting ratios and should therefore give a more faithful representation of the financial health of a company. If the disclosure requirements apply equally to all firms, as mentioned in the previous paragraph, the only remaining possibility to measure the level of information uncertainty is the volatility of key accounting ratios. The idea is that firms with higher volatility of the F-Score/F-rank, that is, information uncertainty, should generate lower subsequent stock returns. This follows on from research by Dichev and Tang (2009) for the US and Clubb and Wu (2014) for the UK. Both studies found a negative correlation between earnings volatility and earnings predictability. Of course, the question could be asked of why earnings in particular should be the determining force in equity valuation. Haugen and Baker (1996) found evidence on the UK and German stock markets that a range of profitability measures can indeed forecast future stock returns. Of course, this might be due to commonalities in investor behaviour. In this regard, Barton et al. (2010) analysed which performance measures are used the most by investors globally. With the highest values for the UK, earnings before interest, taxes, depreciation and amortisation (EBITDA) are the most commonly used. Because both the profitability factors and the fundamentals-based investment strategy are suitable for forecasting stock returns, higher variations in these fundamentals should also be negatively correlated with subsequent stock returns. For this reason, the second hypothesis is stated in the following way.

Hypothesis 5.2: Firms with higher (lower) information uncertainty, proxied by the volatility of accounting fundamentals, provide lower (higher) subsequent returns following bad news and higher (lower) returns following good news.

5.4.2 Direct measures of fundamental value volatility: The SDF

Following the theoretical foundations of the SDF measure, this section deals with the evaluation of the practical results. The method of calculating the standard deviation (SDF) of the F-Score/F-rank is detailed in section 3.5.2.2.

Table 5.4 Two-way SDF portfolios

A.1	***F-Score, mean returns***				***t-test***		***MWW***		***median***	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.134	-0.069	0.054	0.059	.123*** *(6.53)*	.193*** *(5.87)*	*(8.40)****	*(7.32)****	-	-
H	-0.148	-0.134	0.034	0.107	.168*** *(6.18)*	.255*** *(4.17)*	*(8.42)****	*(5.71)****	-	-
H–L	-.014	-.065***	-.020	.048						
t-test	*(-0.31)*	*(-2.63)*	*(-0.99)*	*(1.11)*						
MWW	*(-1.61)*	*(-5.00)****	*(-3.01)****	*(0.54)*						
A.2	***F-Score, median returns***				***t-test***		***MWW***		***median***	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.202	-0.129	0.020	0.013	-	-	-	-	.149*** *(38.67)*	.215*** *(33.10)*
H	-0.262	-0.230	-0.044	0.043	-	-	-	-	.186*** *(46.09)*	.305*** *(24.69)*
H–L	-.060** *(-4.03)*	-.101*** *(-24.87)*	-.064*** *(-8.74)*	0.030 *(0.70)*						
B.1	***F-rank (mean), mean returns***				***t-test***		***MWW***		***median***	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.125	-0.020	0.045	0.077	.065*** *(4.65)*	.202*** *(5.00)*	*(5.09)****	*(6.75)****	-	-
H	-0.099	-0.091	0.005	0.007	.096*** *(5.06)*	.106* *(1.92)*	*(8.45)****	*(4.01)****	-	-
H–L	.026	-.071***	-.040**	-.070						
t-test	*(0.52)*	*(-4.10)*	*(-2.53)*	*(-1.61)*						
MWW	*(-1.16)*	*(-9.66)****	*(-5.51)****	*(-2.78)****						
B.2	***F-rank (mean), median returns***				***t-test***		***MWW***		***median***	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.200	-0.036	0.011	0.008	-	-	-	-	.047*** *(8.57)*	.208*** *(39.62)*
H	-0.240	-0.207	-0.073	-0.092	-	-	-	-	.134*** *(55.21)*	.148*** *(7.39)*
H–L	-.040 *(-1.34)*	-.171*** *(-112.95)*	-.084*** *(-32.85)*	-0.100*** *(-7.38)*						
C.1	***F-rank (median), mean returns***				***t-test***		***MWW***		***median***	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.221	-0.041	0.044	0.115	.085*** *(5.19)*	.336*** *(4.33)*	*(6.52)****	*(5.32)****	-	-
H	-0.134	-0.075	-0.001	-0.007	.074*** *(4.36)*	.127*** *(3.52)*	*(6.92)****	*(5.86)****	-	-
H–L	.087	-.034*	-.045***	-.122**						
t-test	*(1.33)*	*(-1.75)*	*(-2.70)*	*(-1.91)*						
MWW	*(1.06)*	*(-5.53)****	*(-6.66)****	*(-2.86)****						
C.2	***F-rank (median), median returns***				***t-test***		***MWW***		***median***	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.332	-0.071	0.013	0.014	-	-	-	-	.084*** *(20.72)*	.346*** *(28.86)*
H	-0.269	-0.191	-0.086	-0.123	-	-	-	-	.105*** *(34.96)*	.146*** *(17.85)*
H–L	.063 *(0.09)*	-.120*** *(-44.88)*	-.099*** *(-44.11)*	-.137*** *(-9.46)*						

Table 5.4 contains the winsorised mean portfolio returns in panels A.1, B.1 and C.1 and the respective median equivalents in panels A.2, B.2 and C.2. As usual, the first two panels display the F-Score fquintiles, whereas the remaining four represent the F-ranks based on their mean and median combination methods. As a new second dimension, the terciles of the SDF are now shown on the vertical axis.

As can be observed from panel A.1, the F-Score once again is able to separate future winning stocks reliably from losing ones. This is corroborated by both the parametric and the non-parametric test results, which are highly significant at the 1% level throughout. For instance, the highest F-Score portfolio in the highest SDF tercile returns a positive 10.7% annually compared with a negative 14.8% for the lowest portfolio. Hence, the difference between the returns of those portfolios amounts to 25.5 percentage points, with parametric t-statistics of 4.17.

In general, this trend is strongly confirmed for the two alternative measures in panels B.1 and C.1. The only exception is the return difference between the highest and the lowest F-rank portfolio in the high-SDF tercile, which is only significant at the 10% level according to the t-test result.

Similarly, the horizontal winsorised median portfolio differentials are highly significant for all three investment strategies. Thereby, the most pronounced difference of 34.6 percentage points can be found in the low information uncertainty tercile between the highest and the lowest portfolio, shown in panel C.2. For the high-SDF tercile, the greatest difference is observable for the F-Score strategy in panel A.2, which amounts to 30.5 percentage points. Overall, these two strategies perform best with regard to a long-short investment strategy. Further, the absolute returns are highest for those two as well considering investors who prefer to pursue a long-only investment style. For instance, the H*/L F-Score portfolio in panel A.2 returns 1.3% per annum, the median F-rank portfolio 1.4% (panel C.2) and the mean F-rank equivalent 0.8% (panel B.2) only.

In view of the new information uncertainty measure, the vertical differences in the winsorised mean portfolio returns are significant for some fquintiles under parametric test conditions. In panel A.1, the low F-Score fquintile yields highly significant t-statistics of -2.63 and a return differential of -6.5 percentage points. This implies that firms with lower (higher) uncertainty perform better (worse). However, regarding statistical significance, this is not the case for the remaining vertical differences shown in this panel. Nonetheless, apart from the highest F-Score fquintile, the negative signs of the differentials suggest that the trend is towards better performance of lower-uncertainty portfolios. This is confirmed by the parametric test results documented in panels B.1 and C.1. Apart from the L* and H* portfolios of the former, the return differences are significant at least at the 5% level, with t-statistics of -4.10 (-2.53) for the low (high) mean F-rank fquintiles. Another improvement of these results is displayed in panel C.1 of the median F-rank approach. In addition to the low and high fquintiles, the highest fquintile now also features significance at the 5% level, with t-statistics of -1.91.

The analysis of the non-parametric results provides a similar picture. The respective MWW t-statistics are highly significant at the 1% level for all of the low and high fquintiles in panels A.1, B.1 and C.1. The negative signs of all the return differences indicate better performance of firms with low information uncertainty for all of those portfolios. In addition, the highest fquintiles of both the mean and the median F-rank combination method are reported to be highly significant. In other words, the results of the MWW test in panel B.1 (C.1) report t-statistics of -2.78 (-2.86) at the 1% level of significance. Overall, however, no significant return difference is observable between the two portfolios in the lowest fquintile of all the aforementioned panels.

Finally, the evaluation of the median portfolio returns in panels A.2, B.2 and C.2 confirm the observations of the mean portfolio returns. The return differentials in the lowest F-Score fquintile are now found to be significant at the 5% level in panel A.2. With regard to the two F-rank methods, the median results are exactly as significant as the respective mean results, as already described in the previous paragraph.

The conclusions from the analyses of the two-way SDF portfolios are threefold. Firstly, even under the new information uncertainty measure, investors are able to construct high-returning portfolios with the help of all three of the investment strategies. In this context, the F-Score, mean F-rank and median F-rank serve as a proxy for good or bad company news. This is relevant in particular to market participants who are either willing or allowed to engage in a long-short portfolio strategy. Proof for this is given by the high level of significance of the parametric and non-parametric horizontal test results in the overwhelming majority of cases. Therefore, there is clear evidence that the market participants do not fully react to newly issued public information, which agrees with Zhang (2006). Secondly, better performance of stock portfolios with low information uncertainty as opposed to high-uncertainty portfolios is detectable in most cases. This observation is especially distinct for the median F-rank approach but also for the remaining two, albeit to a lesser extent. Thirdly and lastly, as a consequence of the second point, these results only partly confirm the findings of Zhang (2006). Partly, in this regard, this means that it is confirmed that firms with high information uncertainty relatively underperform after bad news is published, that is, those firms in either the low or the lowest fquintile. However, higher information uncertainty in the high and highest fquintiles likewise leads to worse results following good news, which is opposed to the prediction by Zhang (2006). This speaks against the assumption that investors react differently to good and bad news about a company, rather suggesting that information is slowly reflected in the stock prices. The conclusion from this observation is that investors generally prefer stocks with low information uncertainty because the one-year portfolio returns tend to be higher.

5.4.3 Direct measures of fundamental value volatility: MADF

The final information uncertainty measure is the mean absolute deviation, or the MADF, of either the F-Score or the F-rank. It represents an alternative route to the SDF and thus can be regarded as an overall robustness check of the aforementioned SDF result. The calculation of the MADF can be referred to in section 3.5.2.2. Table 5.5 summarises the test results of the vertical and horizontal portfolio return differences. As usual, panels A.1 (A.2), B.1 (B.2) and C.1 (C.2) contain the mean (median) winsorised portfolio returns. Beginning with the former, all three of the investment methods show a strong ability to discriminate between future losing and future winning stocks. This outcome is underlined by highly significant horizontal return differences in the vast majority of cases as a result of both the parametric and the non-parametric tests. In panel A.1, for example, the H*/H portfolio delivers an absolute return of 16.2% and the respective difference from the L*/H portfolio amounts to a total of 32.7 percentage points, with t-statistics of 4.98. With only one exception, the F-Score method works best insofar as it is able to generate the deepest gulf between the high/low and the highest/lowest fquintiles. Additional evidence for the validity of those results is provided by the median tests. As can be seen in panels A.2 and C.2, the test statistics are significant at the 1% level throughout, whereas the differences in panel B.2 are at least significant at the 5% level. Again with only one exception, the F-Score method is the preferable choice for investors regarding the generation of return differences. In summary, the MADF portfolios mirror exactly the results of the two-way SDF portfolios of table 5.4. Hence, both strategies are robust to changes in the two different ways of calculating information uncertainty with regard to the horizontal return differentials.

Table 5.5 Two-way MADF portfolios

A.1	***F-Score, mean returns***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.159	-0.081	0.063	0.069	.144*** *(7.66)*	.228*** *(7.26)*	*(9.78)****	*(8.51)****	-	-
H	-0.165	-0.135	0.045	0.162	.180*** *(6.57)*	.327*** *(4.98)*	*(8.41)****	*(5.65)****	-	-
H–L	-.006	-.054**	-.018	.093*						
t-test	*(-0.13)*	*(-2.24)*	*(-0.85)*	*(1.96)*						
MWW	*(-1.11)*	*(-4.10)****	*(-3.21)****	*(0.74)*						
A.2	***F-Score, median returns***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.207	-0.147	0.026	0.013	-	-	-	-	.173*** *(57.24)*	.220*** *(44.37)*
H	-0.262	-0.228	-0.044	0.069	-	-	-	-	.184*** *(40.24)*	.331*** *(17.85)*
H–L	-.055* *(-3.45)*	-.081*** *(-20.29)*	-.070*** *(-8.57)*	0.056 *(0.95)*						
B.1	***F-rank (mean), mean returns***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.136	-0.034	0.048	0.066	.082*** *(5.70)*	.202*** *(5.19)*	*(6.43)****	*(7.39)****	-	-
H	-0.117	-0.092	0.009	0.009	.101*** *(5.33)*	.126* *(1.93)*	*(8.52)****	*(3.77)****	-	-
H–L	.019	-.058***	-.039**	-.057						
t-test	*(0.36)*	*(-3.30)*	*(-2.41)*	*(-1.26)*						
MWW	*(-1.06)*	*(-8.24)****	*(-5.32)****	*(-2.50)***						
B.2	***F-rank (mean), median returns***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.210	-0.052	0.013	0.006	-	-	-	-	.065*** *(19.07)*	.216*** *(43.30)*
H	-0.252	-0.207	-0.071	-0.095	-	-	-	-	.134*** *(50.75)*	.157** *(5.22)*
H–L	-.041 *(-1.27)*	-.155*** *(-86.70)*	-.084*** *(-30.87)*	-0.101** *(-4.30)*						
C.1	***F-rank (median), mean returns***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.200	-0.052	0.051	0.098	.103*** *(5.98)*	.298*** *(4.13)*	*(7.98)****	*(5.50)****	-	-
H	-0.150	-0.073	0.007	0.003	.080*** *(4.69)*	.153*** *(4.13)*	*(6.96)****	*(6.36)****	-	-
H–L	.050	-.021	-.044***	-.095						
t-test	*(0.85)*	*(-1.11)*	*(-2.59)*	*(-1.54)*						
MWW	*(0.67)*	*(-3.99)****	*(-6.55)****	*(-2.62)****						
C.2	***F-rank (median), median returns***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.298	-0.098	0.015	0.013	-	-	-	-	.113*** *(38.44)*	.311*** *(29.69)*
H	-0.288	-0.184	-0.082	-0.112	-	-	-	-	.102*** *(30.46)*	.176*** *(23.11)*
H–L	.010 *(0.00)*	-.086*** *(-31.41)*	-.097*** *(-46.74)*	-.125*** *(-6.80)*						

In view of the vertical differences, the parametric test results are mostly disappointing. For the F-Score method in panel A.1, the only significant t-statistic of the H–L portfolios is located in the low fquintile, with a value of -2.24. Further, the highest fquintile indicates that high-uncertainty (low-uncertainty) firms return 16.2% (6.9%), as opposed to what is expected following both the SDF method and Zhang's (2006) findings. Panel B.1, which contains the mean F-rank results, produces somewhat better results. Both the low and the high fquintiles are significant at least at the 5% level. Lastly, the median F-rank performs poorly except for the high return differentials, with strong t-statistics of -2.59.

However, the parametric test results are relativised once the non-parametric alternatives are employed. For instance, all but one of the H–L outcomes in panel C.1 are highly significant for the MWW test. The highest return differential can be found in the highest median F-rank fquintile, which amounts to 9.5 percentage points, with MWW t-statistics of 2.62. To a lesser extent, this is also true for panels B.1 and A.1. The test results for the median winsorised returns in panels A.2, B.2 and C.2 are confirmative of the results that have just been presented. The F-Score method produces significant results for the high and low fquintiles and the two F-rank methods for the highest fquintile in addition.

Here, too, the MADF results are generally in accordance with the SDF findings presented in section 5.4.2. Particularly the non-parametric test results are analogous, whereas the parametric equivalents are slightly weaker under the MADF method. All in all, however, all three methods provide results that reasonably indicate better performance for the low information uncertainty stocks and vice versa.

5.5 Size, SDF and MADF: A discussion and links to the literature

This part of the chapter discusses the findings of sections 5.3 and 5.4 against the background of the existing literature on this topic. It should be regarded as a continuation of section 5.2.1, in which firm size was used as information uncertainty measure within the original F-Score and F-rank investment strategies. Sections 5.3 and 5.4 dealt with the alternative combinations of the F-rank, specifically the DMSFE and clusters, and firm size and also introduced two alternative measures of information uncertainty, namely the SDF and MADF. The segmentation was deliberately chosen to keep the discussion coherent and clear.

In general, the previous discussion about the vertical return differentials applies analogously to the DMSFE and cluster approaches. This is because the inclusion of firm size as a second dimension to the DMSFE and clusters yields very similar results to those obtained when the original F-Score and F-rank methods were used. However, the horizontal differences constitute a difference from Piotroski (2000). Both the $k = 2$ and the $k = 3$ cluster method used in this study are highly significant under parametric and non-parametric circumstances. As his analysis showed, the p-values of differentiation within size terciles are monotonically decreasing from the small to the large size terciles. He noted that some of the return differences in the highest size terciles were insignificant and concluded that investors would benefit most from investing in small or medium-sized companies. This trend cannot be confirmed for the UK stock market as the significance remains high for the cluster method irrespective of the affiliation to a certain size tercile. Therefore, F-rank clustering in combination with firm size yields fruitful results for investors but is mostly limited to horizontal return differences. This insight changes, however, for the three different F-rank combination approaches, in which most of the parametric (non-parametric) test results are insignificant (significant). Consequently, the combination of simple and more elaborate forecasting techniques in association with firm size leads to inconclusive results. Hence, investors should use these methods as a second-best solution only.

This study does not intend to test the size effect explicitly but rather the underlying cause of it, as in Zhang (2006). However, if the size effect is not identifiable, there is also no cause, which implies that size as a measure of information uncertainty is futile. Due to the findings in all the two-way portfolios that include firm size as a second dimension, there is reasonable evidence that this is the case for the UK stock market. Of course, the absence of the size effect might be due to the method employed, in this case the F-Score, the F-rank and its combination variants. This would confirm the mixed UK findings on the size effect by Andrikopoulos et al. (2008), who conceded that the detection of the effect is model-dependent. Although the existence of a size effect is conceivable in efficient markets, it should persist only temporarily. To be measurable and even statistically significant, the effect has to be more pronounced. Hence, testing the size effect is an indirect way of testing the market efficiency itself. Then, though, Fama's (1970) argument of the joint hypothesis problem would come into play again. With reference to the size effect, this would mean that the model used to test for the size effect is defective.

Although the initial hypothesis stated by Zhang (2006), that stock portfolios with high information uncertainty perform better following good news and underperform following bad news, was generally confirmed in his study, the results of the size proxy in particular are mixed. The author reported both Pearson and Spearman correlation matrices that are contradictory. In other words, firm size as one of six information uncertainty proxies is correlated positively with the subsequent return in the first case but negatively in the second case. Consequently, the US evidence likewise suggests that the use of firm size as a tool for measuring information uncertainty is not advisable. In another study, Duong et al. (2012) analysed investors' confirmation bias in the UK stock market for the period from 1991 to 2007. As part of their study, cross-sectional stock returns in a value, neutral and glamour context were regressed on book-to-market ratio, size, momentum and F-Score variables. A second regression was performed with the same four control variables with the distinction that the F-Score was split into its high and low components. None of the regression results indicated statistical significance of the size coefficient. The study by Duong et al. (2012) most likely resembles the previous part of the present chapter, because it used the F-Score across the value and glamour stocks and did not apply it exclusively to the value stocks, such as in Piotroski (2000). Hence, it can be used as a means to compare the current findings, which are compliant regarding the vertical parametric test results, which show no statistical significance. In terms of the horizontal differentials, the regression coefficients of the F-Score variable were significant at least at the 5% level in the value, neutral and glamour contexts. However, the coefficients for both the high and the low F-Score in the second regression were not always significant, especially in the value context. This confirms some of the weak parametric test results, which are predominant in the weighted DMSFE method presented in table 5.2. Finally, Choi and Sias (2012) employed the F-Score as a measure of financial strength in conjunction with firm size as a measure of information uncertainty in the US. Although a return differential was observable, it was not statistically significant.

Despite the presence of evidence that the size effect is existent in an international context, this does not imply that studies in this area of finance are free from criticism. Van Dijk (2011) provided a summary of the various academic discussions in the literature, which range from the methods employed in empirical studies to the various explanations of the size effect. One severe point of critique is data mining, in which researchers repeatedly analyse the same data set to find statistically significant results. As MacKinlay (1995) noted, because of the characteristics of the economics discipline in general and the area of finance in particular, deviations from efficient markets are always found ex post. The cause of data mining is related to the publishability of research in general. Welch (2001) remarked that only significant results are likely to be released in academic journals. In this regard, it is interesting that efficient market advocates receive support from some of the behaviouralists, such as Hirshleifer (2001), who stated that the size effect has disappeared altogether. As this appears to be the case in the UK stock market as well, other measures of information uncertainty were tested following the approach of Zhang (2006).

In this research, the standard deviation (SDF) and mean absolute deviation (MADF) of the F-Score/F-rank were used as alternative proxies for information uncertainty. This is in contrast to Zhang (2006), who analysed six other proxies with regard to momentum quintiles.[13] However, three out of the six are also measures of some sort of volatility, which, in turn, are related to financial statement information. According to the author, higher volatility is associated with higher information uncertainty. However, there is a distinction to be made between the types of volatility, which are not further specified by Zhang (2006). On the one hand, cash flow volatility could be interpreted as a direct measure of a certain financial statement item. On the other hand, forecast dispersion could be regarded as the outcome of financial statement interpretation. Consequently, the last measure is of an indirect nature because the amplitude of the volatility depends on analysts' information processing ability. This leads straight to the root of behavioural finance and is the initial step before any decision is made under uncertainty (Tversky and Kahneman, 1974). Because investors apply certain heuristics to reduce the complexity of the information at hand, volatility is highest if there is deep disagreement on the interpretation of information. The third and final measure, stock volatility, could therefore be interpreted as a function of the previous two. Zhang (2006) reported higher negative correlations between cash flow volatility and stock returns than between analyst forecast dispersion and returns. Both the SDF and the MADF were used to extend the underlying logic of the cash flow volatility measure and to circumvent the problem of information processing. This is due to the inclusion of eight additional financial statement items in the measure, that is, the F-Score/F-rank, and because this measure is based on theory, as required by Richardson et al. (2010). Further, it strictly defines whether a financial statement item should be regarded as positive or negative. Therefore, it is as simple and

[13] The six information uncertainty proxies in Zhang (2006) are: firm size, firm age, analyst coverage, forecast dispersion, stock volatility and cash flow volatility.

effective a measure as the Apgar score, which is widely used in medicine to evaluate the health of newborns (Finster and Wood, 2005).

As described earlier, the SDF and MADF do not fully confirm the findings of Zhang (2006). On the one side, he found that the market reaction to newly published information is incomplete, as confirmed by the horizontal differentials in this study. On the other side, investors' reaction to information uncertainty is likely to be time-dependent. Zhang (2006) used one-month-ahead stock returns and found that greater information uncertainty generates lower (higher) returns following bad (good) news. As the analysis of the previous results in this chapter has shown, this is not always the case for the UK, because investors have a preference for portfolios with low information uncertainty regarding the one-year investment horizon.

There are two points that need to be noted. Firstly, the UK stock market appears to be more efficient in the long than in the short run, although investors are still able to generate positive annual returns following the F-Score/F-rank strategy. One interpretation can be made with the argument of limits to arbitrage proposed by behavioural finance proponents, such as Barberis and Thaler (2005), and the disposition effect by Barberis and Xiong (2009). Assuming market efficiency, extreme month-long abnormal returns should be arbitraged away in the long run. This is mostly the case for the L* portfolios because of non-significant parametric and non-parametric results. Regarding the H* portfolios, portfolios with low information uncertainty perform better, which indicates that investors start selling their stock that they bought during the one-month investment horizon when arbitrage opportunities still existed. The different behaviours of the lowest and highest portfolios could then be explained by the disposition effect. Because market participants want to lock in a profit, the trading volume should be much higher than for the L* portfolios, for which losses are not realised. If investors subsequently reallocate their money to the low-uncertainty stocks with the brightest prospects, the returns should be relatively higher one year ahead because the demand for them increases. This is confirmed by the non-parametric results. However, due to the relatively better performance of stock portfolios with low uncertainty, this implies that market inefficiency is still present to some extent. In the case of inefficient markets, as suggested by BF advocates, stock prices mean revert and former losers become future winners (e.g. De Bondt and Thaler, 1985; Fama and French, 1988; Poterba and Summers, 1988). Because this kind of overshooting is evidenced by significant test results, this would support the hypothesis of inefficient markets, because prices are continuously in disequilibrium (Kothari et al., 2010). This overshooting may even be fuelled by usually rational investors. As a consequence of rational finance theory, investors would possess the ability to recognise this disequilibrium fully. However, if they thought they would be able to profit from it, there would be no incentive to counter the trend, leading to a larger disequilibrium (O'Hara, 2008).

Of course, one major drawback of this interpretation is that it plays straight into the hands of EMH supporters. This is because, in retrospect, anomalies that could not be spotted by most of the market participants in real time always seem to become discernible to the researcher. However, this logic could also be applied to refute the efficient market hypothesis. If markets are believed to be efficient, investors should indeed be able to perceive mispricing and therefore no disequilibrium would show up in hindsight. In summary, the results of the F-Score/F-rank approaches under various information uncertainty measures show strong evidence in favour of an incomplete market reaction to newly published information. This is underlined by strong horizontal return differences within a specific information uncertainty tercile and across the lowest/low and high/highest fquintiles. Firm size, in turn, is positively correlated with subsequent one-year stock returns, although the parametric test results are weak. Finally, the volatility of financial statement items, as measured by the SDF and MADF, indicates investors' preference for portfolios with low information uncertainty in the UK stock market.

5.6 Liquidity and stock returns

The remaining part of this chapter develops the hypotheses and presents and discusses the results of the interplay between the F-Score/F-rank methods on the one hand and liquidity as the second dimension on the other hand. It thereby uses both the Amihud measure and the turnover ratio (ToR) as liquidity proxies. Following the discussion of the results, the chapter concludes with an overall summary of the findings of both measures and links them to the present state of the literature.

5.6.1 Hypothesis development: Liquidity as a key driver of investment performance

In a seminal paper, Amihud and Mendelsohn (1986) were amongst the first to analyse the effect of liquidity on stock returns. In their model, which represents a market consisting of rational investors with differing investment horizons and different asset spreads, they found a positive correlation between the average returns and the bid–ask spread in the cross-section. As they argued, this premium for illiquidity should not be regarded as a sign of a market anomaly but rather points to rational investors who are faced with trading frictions. Therefore, investors require higher future returns to compensate them for higher spreads. Using an alternative to the Amihud measure, the turnover rate (ToR), as another liquidity proxy, Datar et al. (1998) confirmed these previous results for cross-sectional stock returns. More recently, further evidence was presented by Amihud (2002) and Hasbrouck (2009). However, other research results suggest that there has been no consensus amongst researchers up to now. Two examples of contradictory findings are those of Brennan and Subrahmanyam (1996), who showed that higher spreads lead to lower returns, and Spiegel and Wang (2005), who found no correlation whatsoever.

One reason for that is most likely to be the concept of commonality. As the above-mentioned research is only concerned with a stock’s idiosyncratic liquidity–return relationship, the studies by Chordia et al. (2000) and Huberman and Halka (2001) initiated a stream of research, which is considered with commonality in liquidity. According to this idea, an additional common or market-wide liquidity parameter is assumed to have a direct impact on a firm’s spread and therefore its subsequent return. The findings by Pástor and Stambaugh (2003), for instance, documented a positive relationship between liquidity risk and expected returns. Generally, however, the literature on tests of the direct relationship between a firm’s spread and its stock return in the UK is limited. For this reason, the aim of the first part of the next hypothesis is to analyse whether liquidity is a determinant of the success of the F-Score strategy.

Hypothesis 5.3a: Firms with lower liquidity, as measured by the Amihud and ToR measures, contribute a larger share to the success of the F-Score and alternative investment strategies.

A direct continuation of the above hypothesis is the second finding in the study by Amihud and Mendelsohn (1986), known as the liquidity hypothesis. According to that, higher spreads may theoretically incentivise firms to increase the liquidity levels in their stock to reduce their cost of capital. This process is accompanied by a decrease in the investors’ required rate of return, because the formerly illiquid stocks are now discounted less heavily. Ultimately, the firm value increases. Several studies have supported this concept with various pieces of practical evidence. Muscarella and Vetsuypens (1996), for instance, documented wealth gains for investors after stock split announcements. As they conjectured, it appears likely that a certain trading range that minimises transaction costs improves liquidity, which, in turn, is causal for stock price appreciations.

Erwin and Miller (1998) studied the changes in stock liquidity, proxied by the bid–ask spread, after the inclusion of a stock in the S&P500 index and found a significant and permanent stock price increase, albeit only for those without tradable options. More recent evidence for the liquidity hypothesis and spread increases/decreases following index reconstitutions was provided by Chen (2006) for Russell 1000 additions and Russell 2000 deletions. Finally, Biktimirov and Li (2014) reported strong evidence for higher (lower) stock liquidity and permanent stock price increases (decreases) for added (demoted) stocks in different FTSE indexes. However, this was only observable for stocks that were part of an index already.

Further research on the UK stock market was conducted by Bortolotti et al. (2007). In their study, they used related measures of privatisation to stock market development, proxied by market liquidity, and in this way indirectly confirmed the previous results. The research period deliberately coincided with the privatisation process of the Thatcher Government in the 1980s (Aghion et al., 2009). Their underlying rationale was that more developed markets with their inherently higher liquidity reduce the cost of capital for firms, which enhances liquidity and might propel firm values, as mentioned before. Indeed, the authors found an improvement in liquidity since 1994, not just for the UK but also in 18 other developed stock markets.

It should be noted that the above-referenced papers are concerned with liquidity and price changes of the same stock after a certain event has taken place. In contrast, there is additional research on the pricing implications of two assets that should normally have the same values because they represent claims to identical cash flows. In this regard, Datar (2001) reported premia for more liquid shares of closed-end funds and vice versa. This correlation was detected for both equity and bond funds. According to their argument, a premium materialises once trading costs, that is, bid–ask spreads, are lower by holding an indirect claim to certain cash flows through a fund than by directly holding the stock. One reason for the reduced transaction costs might be through economies of scale or the possibility of the fund to trade in round lots from which especially small investors may profit (Muscarella and Vetsuypens, 1996). However, it could be argued that transaction costs in this scenario do not present a structural hindrance of trading the stock and therefore should not be priced. In other words, investors have the opportunity to set up a fund on their own or stop trading in odd lots. This assumption was addressed by Chan et al. (2008) in the market for American Depository Receipts (ADRs). Although those claims entitle investors to exactly the same cash flows, they might be barred from trading in the home market of the ADR for various reasons. In this case, therefore, there would indeed be a structural problem for investors that cannot be circumvented. However, this problem would only exist theoretically because the asset should trade at the same price in both the ADR and the home market. If this is not the case, it is of interest whether liquidity is a determinant of the pricing process of this asset. The results obtained by Chan et al. (2008) indicated that premia in the ADR market were associated with a liquidity increase even after controlling for various measures of country characteristics, such as transparency and openness of the home market.

This basic principle is directly applicable to the relationship between liquidity measures and the F-Score strategy of this study. Each fquintile, meaning vertically, represents a portfolio in which all the constituents are deemed to have similar return prospects. For this reason, no significant return differentials should be measurable. However, in light of the strong support of a liquidity–return relationship, the second part of the hypothesis is stated as follows.

Hypothesis 5.3b: Firms with higher (lower) liquidity, as measured by the Amihud and ToR measures, generate higher (lower) subsequent returns.

5.6.2 The first liquidity proxy: The Amihud measure

As Amihud (2002) hypothesised, illiquid stocks should feature higher future returns to compensate for liquidity risk. However, it is of interest whether this is also the case in conjunction with the F-Score/F-rank strategies. The rationale is straightforward. On the one hand, it should be tested first whether the investment strategy is working reliably irrespective of the level of liquidity in a certain stock. If this requirement is met, then the horizontal portfolio return differentials should be positive, indicating that the higher (lower) the F-Score/F-rank, the better (worse) the subsequent performance. It is only then that it can be evaluated whether these differentials are driven by illiquid or liquid stocks (hypothesis 5.3a). On the other hand, the logic of an illiquidity premium could be questioned. For instance, stocks that are part of a high F-Score portfolio have favourable future prospects in terms of stock returns. If this portfolio is now split into two separate portfolios according to the levels of liquidity of their contained stocks, the following question arises: why should the portfolio with lower liquidity yield higher returns than the one with higher liquidity? If rational investors regard both portfolios as equally likely to deliver satisfying yields, they should normally prefer the one with higher liquidity (hypothesis 5.3b). In this case, the stock returns should turn out to be relatively higher because of increased investors' demand for more liquid stocks, indicated by the Amihud measure. The process of calculating the Amihud measure is described in section 3.5.2.3. Table 5.6 presents the results of the winsorised percentage portfolio returns, similar to the procedure employed earlier.

Table 5.6 Two-way Amihud portfolios

A.1	***F-Score, mean returns***				***t-test***		***MWW***		***median***	
Amihud	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.125	-0.117	0.054	0.098	.171*** *(8.70)*	.223*** *(6.39)*	*(12.53)****	*(9.15)****	-	-
H	-0.117	-0.072	0.052	0.039	.124*** *(6.01)*	.156*** *(3.72)*	*(7.30)****	*(4.72)****	-	-
H–L	.008	.045*	-.002	-.059						
t-test	*(0.16)*	*(1.70)*	*(-0.12)*	*(-1.59)*						
MWW	*(1.54)*	*(4.44)****	*(2.94)****	*(0.24)*						
A.2	***F-Score, median returns***				***t-test***		***MWW***		***median***	
Amihud	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.230	-0.213	-0.017	0.009	-	-	-	-	.196*** *(137.57)*	.239*** *(73.75)*
H	-0.170	-0.096	0.044	0.036	-	-	-	-	.140*** *(31.54)*	.206*** *(13.96)*
H–L	.006 *(1.76)*	.117*** *(23.88)*	.061*** *(9.79)*	.027 *(0.39)*						
B.1	***F-rank (mean), mean returns***				***t-test***		***MWW***		***median***	
Amihud	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.112	-0.075	0.036	0.064	.111*** *(6.98)*	.176*** *(4.52)*	*(11.21)****	*(7.83)****	-	-
H	-0.134	-0.015	0.031	0.039	.046*** *(3.00)*	.173*** *(3.11)*	*(3.07)****	*(3.99)****	-	-
H–L	-.022	.060***	-.005	-.025						
t-test	*(-0.37)*	*(3.48)*	*(-0.29)*	*(-0.54)*						
MWW	*(0.79)*	*(9.29)****	*(2.57)***	*(0.06)*						
B.2	***F-rank (mean), median returns***				***t-test***		***MWW***		***median***	
Amihud	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.230	-0.183	-0.034	-0.010	-	-	-	-	.149*** *(105.35)*	.220*** *(49.85)*
H	-0.199	-0.026	-0.004	-0.010	-	-	-	-	.022 *(1.35)*	.189*** *(11.29)*
H–L	.031 *(0.33)*	.157*** *(109.23)*	.030*** *(6.71)*	.000 *(0.24)*						
C.1	***F-rank (median), mean returns***				***t-test***		***MWW***		***median***	
Amihud	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.123	-0.069	0.034	0.035	.103*** *(6.51)*	.158*** *(3.91)*	*(9.84)****	*(5.95)****	-	-
H	-0.231	-0.022	0.039	0.045	.061*** *(3.62)*	.276*** *(4.07)*	*(4.63)****	*(5.40)****	-	-
H–L	-.108	.047**	.005	.010						
t-test	*(-1.55)*	*(2.40)*	*(0.29)*	*(0.15)*						
MWW	*(-1.68)**	*(6.40)****	*(4.18)****	*(1.61)*						

C.2	F-rank (median), median returns				t-test		MWW		median	
Amihud	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.240	-0.180	-0.048	-0.081	-	-	-	-	.132*** *(82.75)*	.159*** *(20.43)*
H	-0.375	-0.046	0.009	-0.005	-	-	-	-	.055*** *(10.59)*	.370*** *(21.77)*
H–L	-.135* *(-2.85)*	.134*** *(42.58)*	.057*** *(16.92)*	.076* *(2.94)*						

The structure of the table follows exactly the presentation of the previous results regarding the F-Score/F-rank and the information uncertainty proxies. Panels A.1 (A.2), B.1 (B.2) and C.1 (C.2) contain the winsorised mean (median) portfolio returns and the respective parametric and non-parametric test results. The analysis of the horizontal mean return differentials indicates strong support for the ability of all three investment strategies to distinguish between future outperformers and underperformers. The highest percentage differences can thereby be found in panel A.1 within the low-liquidity tercile. For instance, the percentage differences between the high/low and the highest/lowest portfolios amount to 17.1 and 22.3 points, respectively, which are significant at the 1% level. Although the median F-rank strategy shown in panel C.1 earns a 27.6 percentage point highest/lowest differential within the high-liquidity portfolios, the difference for the high/low portfolio is only 6.1 points. In comparison, the F-Score strategy returns 12.4 and 15.6 percentage points in the same setting. This speaks for a smoother return pattern than in the case of both the mean and the median version of the F-rank. However, without any exception, all the parametric tests are highly significant; the F-Score t-statistics are the highest with only one exception. Exactly the same results are observable for the MWW tests. Overall, therefore, it can be concluded that the analysed strategies perform as expected within the same liquidity tercile.

Regarding the vertical return differences, the test statistics are weaker, especially for both of the extreme portfolio fquintiles, for which no significance was found. Panels A.1, B.1 and C.1 document some significance but only for the low fquintile differentials without any exception. In this regard, the high-liquidity portfolio of the mean F-rank approach in panel B.1 outperforms its low counterpart within the low fquintile by 6.0 percentage points and with significant t-statistics of 3.48. Higher levels of significance are reported in the non-parametric test environment. This is the case across all three investment strategies but is again limited to the low and high fquintiles only.

Regarding the median portfolios in panels A.2, B.2 and C.2, the test results generally resemble those of the previous portfolios. The horizontal differentials are with one exception significant at the 1% level throughout, while the vertical counterparts are mostly relevant only within the low and high fquintiles. This means that high-liquidity portfolios perform relatively better, which is in favour of the liquidity hypothesis.

Hence, the observations in table 5.6 lead to the following conclusions. Firstly, all three of the proposed investment strategies perform as expected in the two-way environment, in which liquidity represents the second dimension. Investors with a long-short strategy profit in particular due to the large spreads between the lowest/low and the high/highest portfolios. Hence, there is strong evidence that the lion’s share of the success of the investment strategies is attributable to illiquid stocks. Secondly, the assumption of investors’ liquidity preference is confirmed. This is due to the consistent outperformance of the high-liquidity terciles as opposed to the low-liquidity terciles. In other words, this means that investors are more willing to buy (sell) high-liquidity (low-liquidity) stocks because all of these stocks share common firm characteristics, as indicated by their respective fquintile. Because the probability is very high that firms in one specific fquintile will have similar returns over the next year, investors are better off choosing more liquid stocks. Thirdly and lastly, long-short investors especially should be interested in the higher return differentials predominantly in the low-liquidity terciles. However, this does not automatically imply that low-liquidity portfolios come with a premium, which would be the exact opposite of the previous point. Consequently, the question about the cause arises. The answer is likely to be found in the low fquintiles across all the panels in table 5.6, in which both the parametric and the non-parametric t-statistics are at least significant at the 10% level. This leads to the conclusion that investors heavily sell low-liquidity portfolios and buy high-liquidity portfolios instead. Due to this aggressive selling process, the high/low horizontal return differential of the low-liquidity tercile increases as a consequence. This can be seen in panel B.1, for example. Vertically, the percentage difference in the low (L) F-rank fquintile amounts to 6.0 points, with relatively high t-statistics of 3.48. The selling of the low-liquidity portfolio leads to a negative one-year return of 7.5%. This, in turn, is partly responsible for the horizontal 11.1 percentage point difference between the high and the low F-rank fquintiles.

In all of those cases in which the vertical non-parametric test results are consulted, the same behaviour is identifiable for the high fquintiles, the differentials of which now become significant. Interestingly, the t-statistics are also consistently lower for these high fquintiles. For instance, the high mean F-rank portfolio in panel B.1 exhibits vertical MWW results of

2.57 compared with the low fquintile with 9.29. As mentioned before, the horizontal high/low differences for the low-liquidity terciles outperform their more liquid counterparts. Therefore, the trading behaviour between the high and the low fquintiles must be asymmetric. In other words, the negative (positive) price move of the low-liquidity (high-liquidity) tercile and low fquintile is more pronounced than for the low-liquidity (high-liquidity) tercile and high F-fquintile. With regard to the Amihud measure, it can therefore be concluded that illiquidity is a key driver of the F-Score/F-rank investment strategies, as stated in hypothesis 5.3a. However, investors prefer more liquid stocks once the portfolio constituents feature the same attributes determined by the same fquintiles. Hence, hypothesis 5.3b is confirmed.

5.6.2 The second liquidity proxy: The ToR measure

The last part of this chapter deals with the turnover ratio, or ToR measure, which represents an alternative to the Amihud liquidity measure. It is an alternative in the way that it captures the trading quantity dimension rather than the price impact dimension. Hence, its introduction serves as a robustness check for the latter and therefore the results are expected to be analogous. Finally, it is also used as a tool to categorise stocks as either low or high book-to-market firms. The method of calculating the ToR ratio itself can be referred to in section 3.5.2.4. The results of the two-way portfolios with ToR as the second dimension are summarised in table 5.7.

Table 5.7 Two-way ToR portfolios

A.1	***F-Score, mean returns***				***t-test***		***MWW***		***median***	
ToR	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.214	-0.165	0.050	0.074	.215*** *(9.60)*	.288*** *(7.63)*	*(12.62)****	*(9.69)****	-	-
H	-0.096	-0.068	0.048	0.070	.116*** *(6.05)*	.166*** *(4.48)*	*(9.10)****	*(6.16)****	-	-
H–L	.118***	.097***	-.002	-.004						
t-test	*(2.77)*	*(4.03)*	*(-0.15)*	*(-0.12)*						
MWW	*(3.46)****	*(5.23)****	*(0.99)*	*(0.56)*						
A.2	***F-Score, median returns***				***t-test***		***MWW***		***Median***	
ToR	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.267	-0.230	-0.019	-0.010	-	-	-	-	.211*** *(128.48)*	.257*** *(71.62)*
H	-0.176	-0.132	0.013	0.027	-	-	-	-	.145*** *(60.23)*	.203*** *(26.66)*
H–L	.091*** *(7.34)*	.098*** *(22.09)*	.032* *(2.85)*	.037 *(2.47)*						
B.1	***F-rank (mean), mean returns***				***t-test***		***MWW***		***Median***	
ToR	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.216	-0.131	0.014	0.043	.145*** *(8.17)*	.259*** *(5.89)*	*(11.77)****	*(9.58)****	-	-
H	-0.085	-0.012	0.043	0.036	.055*** *(3.67)*	.121*** *(2.92)*	*(5.09)****	*(4.84)****	-	-
H–L	.131***	.119***	.029*	-.007						
t-test	*(2.76)*	*(6.94)*	*(1.86)*	*(-0.17)*						
MWW	*(4.37)****	*(9.73)****	*(2.30)***	*(-1.08)*						
B.2	***F-rank (mean), median returns***				***t-test***		***MWW***		***Median***	
ToR	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.343	-0.208	-0.044	0.008	-	-	-	-	.164*** *(102.00)*	.351*** *(88.42)*
H	-0.191	-0.061	-0.010	-0.020	-	-	-	-	.051*** *(16.56)*	.171*** *(16.21)*
H–L	.152*** *(12.14)*	.147*** *(86.91)*	.034** *(7.45)*	-.028 *(-0.45)*						
C.1	***F-rank (median), mean returns***				***t-test***		***MWW***		***Median***	
ToR	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.196	-0.116	0.013	0.031	.129*** *(7.06)*	.227*** *(4.55)*	*(10.49)****	*(6.87)****	-	-
H	-0.065	-0.020	0.038	0.009	.058*** *(3.77)*	.074 *(1.52)*	*(6.68)****	*(3.22)****	-	-
H–L	.131***	.096***	.025	-.022						
t-test	*(2.59)*	*(5.32)*	*(1.63)*	*(-0.45)*						
MWW	*(3.30)****	*(7.24)****	*(2.91)****	*(-0.67)*						
C.2	***F-rank (median), median returns***				***t-test***		***MWW***		***Median***	
ToR	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.348	-0.206	-0.055	-0.044	-	-	-	-	.151*** *(87.84)*	.304*** *(39.17)*
H	-0.212	-0.101	-0.010	-0.095	-	-	-	-	.091*** *(37.64)*	.117*** *(5.82)*
H–L	.136*** *(10.48)*	.105*** *(42.93)*	.045*** *(10.82)*	-.051 *(0.92)*						

Of interest first are the results of the horizontal differentials. An analysis of all the panels from A.1 to C.2 reveals that, with one exception only, both the parametric and the non-parametric test outcomes are significant at the 1% level. This corresponds to and therefore confirms the previous findings of the Amihud two-way portfolios. However, the difference is that those horizontal differences are more pronounced for the ToR measure. This means that long/short investors, for example, are able to increase the gap between the high (highest) and the low (lowest) fquintiles in ten (seven) out of twelve cases and overall in seventeen out of twenty-four cases. It is also notable that amongst the three investment strategies, the original F-Score persistently generates the largest differentials with regard to the winsorised mean portfolio returns. For instance, the highest/lowest differential in panel A.1 amounts to 28.8 percentage points, whereas the respective value in panel B.1 (C.1) is only 25.9 (22.7) points. This is not always true in the two-way Amihud portfolios. Regarding the median portfolios, this tendency is almost always observable for the F-Score strategy in panel A.2 except for the highest/lowest differential in the low-liquidity tercile, which underperforms the other two strategies.

The next section deals with the vertical portfolio return differences with a focus on the mean returns of panels A.1, B1 and C.1. It is conspicuous that the lowest and low fquintiles are highly significant in the parametric test results and, in the case of the mean F-rank, this is also the case for the high fquintile, albeit with weaker significance. For instance, the lowest and low F-Score fquintiles in panel A.1 show t-statistics of 2.77 and 4.03, respectively. Throughout the three panels, higher returns are tilted towards the high-liquidity terciles. Evaluating the non-parametric test results yields a similar picture. Although the high F-Score fquintile remains insignificant in the MWW test, both high F-rank fquintiles are significant, with t-statistics of 2.30 (mean F-rank) and 2.91 (median F-rank). The vertical return differentials of the winsorised median return portfolios presented in panels A.2, B.2 and C.2 further substantiate these findings in this paragraph, which means that except for the highest fquintiles investors prefer stocks with higher liquidity. Based on those findings, it should be remarked, however, that this trend manifests itself particularly in the low and lowest F-fquintiles. However, a distinction has to be drawn between the return and return differentials, respectively, and the test statistics. For instance, the vertical percentage difference within the low mean F-rank fquintile in panel B.2 amounts to 14.7 points, while the lowest one is higher with 15.2 points. However, the respective t-statistics are 86.91 and 12.14, which indicates that the preference for liquidity is more pronounced in the low fquintile.

In comparison with the Amihud two-way portfolios, the ToR measure appears to be even better suited to uncovering the interplay between a portfolio's liquidity and its F-Score/F-rank. This can be derived from the lowest fquintiles, which become highly significant following this approach as a result of the vertical parametric and non-parametric tests. In addition, the generally positive sign thereby suggests that investors prefer highly liquid stocks in the low, lowest and high fquintiles, whereas they seem to be indifferent in the highest fquintile. This implies that the arrival of good news does not trigger any buying and/or selling behaviour in investors who already hold a portfolio composed of promising stocks, as indicated by any of the F-strategy. This is consistent with the earlier findings in the Amihud environment of section 5.6.2. Opposed to that, most action is taken by investors who hold low-quality portfolios. Apparently, bad news is perceived as a signal that stocks that already underperform are deemed to continue this trend in the future. In the two-way ToR tables only, investors seem less surprised by the publication of bad news and the trading activity abates within the lowest fquintiles as a consequence. Once the ToR measure is equated with the BM ratio, the results are also in line with the research on the value premium, as higher liquidity is associated with a higher BM ratio and therefore higher returns for these stocks. Regarding both liquidity hypotheses, the results are generally as expected.

5.7 Discussion of the research findings and links to the literature

The final section of this chapter discusses the previous research outcomes of the two liquidity proxies with regard to the extant literature. Because those findings, on the whole, coincide, the following discussion distinguishes between the Amihud and the ToR measure only if explicitly stated.

As summarised by Bekaert et al. (2007), liquidity is a decisive factor in asset pricing, although the academic dispute on how exactly prices are affected appears to be far from settled. Nevertheless, Amihud's (2002) findings of an illiquidity premium were generally in support of hypothesis 5.3a during the analysis of the UK stock market. This means that less liquid stocks are the decisive factor for the success of both the F-Score and the alternative investment strategies. Generally, this is not surprising in the light of some of the established models that were developed by Amihud and Mendelson (1986), Acharya and Pedersen (2005) and Liu (2006) and that have been used by researchers to document this illiquidity premium. Liquidity itself is a term that features multiple facets. Based on the general description of the term, the presence of liquidity allows investors to trade large quantities in a timely manner, at low costs and with little impact on an asset's price (Liu, 2006). As a consequence, there are four dimensions to liquidity, which are the trading quantity, trading speed, trading cost and price impact. The present study employs two of these, namely the price impact dimension (Amihud) and the trading quantity dimension (ToR). However, as they are used singly, that is, not as a composite that captures both of those dimensions, it is likely that some features of liquidity are omitted, even the ones that they are supposed to capture (Liu, 2006). This disadvantage was noted earlier. Amihud (2002) himself described his liquidity measure as crude but pointed out its advantages, such as data availability. Likewise, Pástor and Stambaugh (2003) described liquidity proxies in general as arbitrary.

It is most likely for this reason that other researchers will call the findings on the illiquidity premium into question. Vayanos (1998), for instance, developed a theoretical model that suggests that illiquidity is not likely to have a major effect on stock prices. In a recent paper, Asparouhova et al. (2010) focused their attention on possible biases in tests of those studies in which an illiquidity premium was found. Based on the widely accepted assumption that illiquid stocks deserve a premium, the authors suspected that regressions were designed accordingly to find such a premium. Their results indicated that almost half of the premium estimates in the stock market were due to noise. They also added that “false” premiums are likely to exist in other asset classes as well.

Regarding hypothesis 5.3b, the present findings in the UK support the liquidity hypothesis. Once stocks are assigned to portfolios with similar characteristics, higher liquidity is found to be associated with higher one-year stock returns. From the critics’ point of view, the belief in an illiquidity premium that actually does not exist is justified by arguments in favour of a liquidity premium. This statement is rooted in the fixed-income markets, in which observations have shown that market participants are prone to shift their invested monies from less liquid to more liquid bonds. Of course, this so-called flight to quality is only justified in those cases in which the bonds have different risk characteristics attached to them. However, Longstaff (2004) noted that in recent times a phenomenon known as flight to liquidity has emerged, which challenges the efficient market hypothesis of Fama (1970). In an efficient market, two assets with similar future cash flows should have the same present value, as noted earlier. Longstaff (2004) found a liquidity premium of up to 15% for similar US treasury bonds, which he partly attributed to behavioural aspects such as consumer confidence. Here, too, the assumption of Asparouhova et al. (2010) that this behaviour is also present in other asset classes, such as the UK stock market, seems valid, although in the another direction. Moreover, if flight to liquidity is observable to such a large extent in one of the safest securities in the financial markets, it is by no means too far-fetched to assume that liquidity premiums should be observable in the stock markets as well. This would explain the strong response, especially in the low fquintiles, which is corroborated by the high t-statistics. If, for instance, market participants with a contrarian view invest in stocks with weak fundamentals, that is, with a low F-Score/F-rank, in anticipation of a turnaround, bad news is not conducive to triggering this turnaround. Investors who are caught wrong-footed would subsequently switch to higher-liquidity stocks with otherwise similar characteristics and would also be prepared to pay higher prices. This kind of investor behaviour can be inferred intuitively. Karolyi et al. (2012) referred to a scenario in which stock prices are declining or a danger of market illiquidity is looming. In these cases, investors are prepared to accept higher stock prices if this gives them the assurance to liquidate their portfolio at a reasonable cost if necessary. This interplay between the general market and a stock's individual liquidity on the one side and how it affects investors and stock prices on the other side is the above-mentioned concept of commonality. Karolyi et al. (2012) remarked that although commonality in stock markets has been proven empirically, little research has been undertaken on its drivers. Although answering this question is not the aim of this research, the behavioural aspect could still serve as an explanation for the key drivers of the F-Score/F-rank strategies.

Bali et al. (2014) presented a behavioural explanation for the liquidity premium in the cross-section. They agreed with Amihud and Mendelson (1986) on their first observation insofar as illiquidity is priced. However, with regard to the liquidity hypothesis, the increase in a firm's value appears not to be based on lower transaction costs, that is, a lower bid–ask spread, and therefore lower required rates of return, but rather on investors' under-reaction to liquidity shocks. In this context, positive (negative) shocks were defined as increases (decreases) in a stock's liquidity relative to its past 12-month average. After establishing long-short portfolios similar to the ones in this study, statistically and economically significant monthly returns between 0.70% and 1.20% were reported for US stocks for the period 1993 to 2010. Thereby, the variance of returns is due to the different liquidity measures used. These returns were achieved by going long (short) in stocks with positive (negative) liquidity shocks. In addition, the positive correlation between the liquidity shock and the stock returns was found not only to be contemporaneous but also to last for up to seven months. Interestingly, the results were robust after controlling for size, book-to-market, beta and other factors.

The reason for a longer transition from illiquid to liquid portfolios is possibly accompanied by investors' inattention (Bali et al., 2014) and herding behaviour (e.g. Ivković and Weisbenner, 2007; Ng and Wu, 2010), especially amongst stocks with lower institutional ownership. With regard to inattention, Hirshleifer et al. (2013) summarised the evidence of recent work on this topic and stated that particularly less salient public information is not reflected immediately in stock prices. If, by contrast, financial statement information, namely the F-Score/F-rank, is regarded as salient public information, the stock prices should adjust more quickly. However, this seems not to be the case, as shown by Choi and Sias (2012). According to the authors, financial strength information is gradually incorporated over time when sophisticated market participants, such as institutions, react first to changes in fundamentals. Following that, less sophisticated investors follow, which could be interpreted as a sign of herding behaviour. Under this assumption, liquidity as a proxy for investor behaviour is a key driver in explaining the higher returns of higher-liquidity portfolios for all fquintiles and F-strategies.

As described in section 3.5.2.4 regarding the distress risk puzzle, the turnover ratio can also be used to subdivide the sample into high-BM and low-BM firms. Applying the logic of Bulan et al. (2007) to the two-way ToR portfolios, this means that portfolios in the high-ToR (low-ToR) tercile are high-BM (low-BM) firms that are large (small), pay (do not pay) dividends and are financially healthy (distressed). In this scenario, a flight to quality, proxying for liquidity, corresponds to the research results of Næs et al. (2011). In their study, the authors observed that investors sell small firms when the market liquidity worsens. In other words, some investors abstain from investments before a recession altogether, while the remaining ones build portfolios that consist of larger and more liquid stocks. Here, too, liquidity as a proxy for size is driving the subsequent returns within the fquintiles.

In summary, liquidity is a key driver of portfolio returns for the F-Score/F-rank investment strategies. On the one hand, illiquid stocks are the main contributor to their overall success. Furthermore, the research findings in this chapter show that investors are willing to pay a premium for more liquid stocks on condition that stocks have a similar future outlook. The reason for this is a flight to quality, which is triggered by investors' fear of decreasing liquidity in financial markets. As a consequence, this fear translates into trading activity, which results in higher returns for stocks with higher liquidity. This process is time-dependent and therefore suggests that publicly available information is not incorporated into stock prices immediately: a contradiction to the efficient market hypothesis. Firm size plays a part in this transition process because it is positively correlated with liquidity. Another factor that most likely facilitates this process is herding behaviour whereby institutional investors begin the transitioning process, which is subsequently followed by the broad market.

5.8 Chapter summary

The aim of this chapter was to analyse the key drivers of the various investment strategies. Based on previous work in the literature, there is reason to assume that the concentration of anomalies in small firms is a large part of their overall success. For instance, Fama and French (2008) found a negative relationship between asset growth and average returns amongst smaller firms, which was not present for larger firms. Further, Ayers and Freeman (2000) documented size-based timing differences for industry information. Observing large firms' returns, a small-firm trading strategy returned abnormal returns of 4.5% over 12 months. Because of the various dimensions of the size–return relationship, firm size is regarded as a proxy for different anomalies. To isolate one specific aspect of firm size, this chapter documented the empirical research results of the F-Score and F-rank investment strategies in relation to information uncertainty and liquidity, which, in turn, can be regarded as a function of information availability and quality. Regarding information uncertainty, three different measures – the size, standard deviation and mean absolute deviation of the nine fundamental constituents – were explored. In each respective section, the chapter provided a discussion and links to the current literature. The main focus was thereby to evaluate how information uncertainty and liquidity drive stock returns. The interpretations were mostly based on investor behaviour.

The key findings are as follows. Information uncertainty is a determinant of one-year-ahead stock returns. In line with the findings of Zhang (2006), firms with higher uncertainty underperform their low-uncertainty counterparts when there is bad news or bad signals come from accounting information. However, this is also the case once good news hits the market. Therefore, the first and second hypotheses are only partly confirmed. The first hypothesis is directly adapted from Zhang (2006) to validate the original results in the UK. One possible reason for the deviating results between the original and the present study could be ascribed to positive skewness in the data set. Zhang (2006) only reported mean returns while omitting the median counterparts. The second hypothesis represents an extension insofar as not only information uncertainty itself is measured but also its respective variability. Thereby, higher volatilities are associated with higher information uncertainty in general. Still, the results of Zhang (2006) were not fully confirmed.

While the argument of skewness in the data is technical, a behavioural attempt to explain the differing research findings could be limits to arbitrage and the disposition effect, both of which are in favour of not fully efficient markets. This means that in general investors hold on to their losing positions while they reallocate their funds from sales of profitable positions to low-uncertainty stocks following good *and* bad news. Hence, prices are continuously in disequilibrium, as hypothesised by Kothari et al. (2010). Another insight is that firm size as an information uncertainty proxy is not as suitable as the SDF and MADF alternatives based on its statistical power, which agrees with similar observations by Zhang (2006).

Further key drivers of the investment strategies are liquidity-related. Cross-sectionally, illiquidity is an important factor, which means that the returns of a long-short portfolio that is based on the F-Score investment strategy are mainly owed to illiquid stocks. Consequently, the first part of the third hypothesis is validated. The second part of this hypothesis refers to the liquidity hypothesis, which states that more liquid stocks should outperform their less liquid peers conditional on similar firm characteristics. In this regard, the F-Score was utilised as a yardstick to group stocks with similar future return prospects. Indeed, the analyses revealed higher returns of more liquid stocks. One rational and one behavioural reason were given for this observation. On the one side, firm value increases as a consequence of the decrease in the cost of capital. On the other side, under-reaction in combination with herding by a group of investors seems to be causal. In this scenario, a first group of sophisticated investors, such as institutions, immediately switch to more liquid stocks. The resulting price increase is reinforced by less sophisticated investors, who follow this group. The research findings are in line with this observation.

In summary, therefore, the knowledge has been advanced in three ways. Firstly, market participants generally prefer to invest in companies that are characterised by a low level of uncertainty. This relation can be established because firms with lower information uncertainty tend to have higher subsequent returns. Secondly, the success of an investment strategy such as the F-Score and its alternatives is more closely related to liquidity than to size and seems able partly to explain its success. Lastly, market participants favour, ceteris paribus, more liquid stocks. However, these findings leave room for future research. One possible question arises in the context of commonality in liquidity. In addition to a stock's idiosyncratic liquidity, both the Amihud and the ToR measure are likely to capture market-wide liquidity factors as well. It would therefore be of interest to analyse the interplay of those two parameters with the F-Score separately to answer the question of which one of them has a stronger influence on its success.

The next and final empirical chapter focuses on the German stock market. It serves not only as a robustness check for the UK results but also contributes to the limited literature on the German stock market in general by extending prior research to examine how effective financial health strategies are in a very different institutional setting. Finally, it contrasts the results given the different legal systems in the two countries.

Chapter 6 Empirical Analysis and Results – Germany

6.1 Introduction

The purpose of the previous two empirical chapters was to extend the knowledge, to gain a deeper understanding of the functioning of the original F-Score and other financial health strategies in the UK and to create the basis for further research in the European context. This included the development and testing of alternative investment strategies as well as the scrutinising of its main drivers. The evidence suggested that the success of the presented methods is largely owed to small firms and firms with low BM ratios. Apart from that, the F-Score method is driven by low-uncertainty firms and firms with high liquidity. Because of the commonalities between the UK and the US regarding their market-oriented economies and their common law legal system, the success of the original F-Score strategy in the UK might not be regarded as surprising. As it is not known how these financial health strategies perform in a different institutional setting, this chapter aims to provide evidence on their effectiveness. In general, there is still a lack of evidence for other markets with respect to the original F-Score strategy. This is intriguing because the validity of Piotroski's (2000) US results has not been disproved yet (Richardson et al., 2010).

To narrow this research gap and to corroborate the previous findings in a different economic and legal environment, the following final empirical chapter of the thesis places the emphasis on the German stock market. It pursues three main objectives. Firstly, the results of Piotroski's (2000) original investment strategy are checked for robustness. This ensures the validity and reliability of this method in the European context. Secondly, it contributes to the literature on the German stock market, which, according to Amel-Zadeh (2011), has not been as well researched as its US and UK counterparts. As the German economy was the fifth-largest contributor to the world's gross domestic product in 2013, this may seem bewildering.[14] However, the reason for this is most likely rooted in the third and final point. La Porta et al. (1997) documented that, in comparison with common law countries such as the US and the UK, investors in code law countries such as Germany are faced with weaker investor protection. In addition, the capital markets in those countries are smaller. In summary,

[14] Source: International Monetary Fund: http://www.imf.org

this chapter contrasts and discusses the findings from the two markets and puts them into context with the relevant literature.

The rest of the chapter is organised as follows. Section 6.2 provides an overview of the peculiarities of the German economy, firms and legal environment. Based on these, testable hypotheses are developed. Following that, section 6.3 summarises the descriptive statistics and contrasts them with the UK and the US findings. It also presents the results of the original and ranked versions of the F-Score strategy. Sections 6.4 to 6.5 deal with the tests of the alternative methods. The results, including the BM ratio, size and alternative uncertainty measures, are presented in sections 6.6 to 6.9. Section 6.10 summarises and concludes the chapter. With regards to the methodology used, the reader is referred to chapter 3 and the appendix of this thesis.

6.2 Characteristics of the German economy

The application of the F-Score strategy and its alternatives to the German stock market is more than just a mere replication of the previous two chapters in a different environment. Of course, one aspect is to analyse their robustness so their successful application can be ensured independently of the geographical region and legal system of the stock market. Another aspect concerns the general access to financial information in bank-oriented economies as opposed to more market-oriented economies, such as the UK and the US. In any case, investors are better prepared to make decisions the more they know about the intricacies of the respective financial market in which they are investing. This appears to be even more appropriate for stock markets such as the one in Germany, which is deemed to be underdeveloped within the group of industrial nations (Audretsch and Elston, 1997). This phenomenon can be ascribed to the unique characteristics of the German economy. However, according to LaPorta et al. (1998), a more generalist cause also exists. The authors found a tendency of higher-developed markets in common law countries such as the UK than in code law countries such as Germany. Other research suggests that the German market is not underdeveloped but the set-up is different. This was observable in 2008 when governments tried to avoid the financial crisis and purchased large stakes in major commercial banks. This step was regarded less hostile in Germany compared to the UK (Andrianova et al., 2012). The following two subsections will deal with these differences separately.

6.2.1 The German Mittelstand

The distinctiveness of the German Mittelstand is reflected in the word itself, which is usually not translated into English. In fact, companies that fall under this definition are so important to the economy that a dedicated institute (IfM) for researching those companies was founded in 1957 on the initiative of Ludwig Erhard, father of the economic miracle (Wirtschaftswunder) and later German Chancellor.[15] According to their definition, equating Mittelstand with small and medium enterprises (SMEs) is not enough. The IfM postulates a qualitative requirement as well, according to which a company has to be owner-managed. Consequently, even large family-run businesses are defined as Mittelstand, whereas small non-owner-managed companies are not covered by this term.

In this context, it may be questioned whether the term Mittelstand in the narrow sense of the word, namely family-run and owned, is related to the stock market at all, in which companies are publicly owned and mostly not managed by the owner. However, as documented later in this chapter, firms in Germany have higher gearing than their UK counterparts, which signals that old habits die hard. This idea was indirectly confirmed by a study by Cheffins et al. (2013), who found that, despite the abolition of the so-called two-thirds rule[16] imposed by the London Stock Exchange until the 1950s, UK corporate governance is still characterised by the separation of ownership and control.

With regard to German firms, this is mainly not the case, though. The reason for that is likely to be rooted in the inherently different characteristics of the economies in the UK and Germany. Haberly (2013) analysed how the Anglo-American/liberal market economy style (LME) affected the coordinated market economies (CME) regarding institutional organisation. This question was raised in the 1990s due to a strategic shift of the focus of the main financial institutions, which was supported by the policy of the German Government, which started an extensive privatisation programme in the 1990s (Megginson and Netter, 2001). Because of the heavy reliance on banks, firms needed to prevent the possibility of hostile takeovers in their search for new capital. Swapping one financial institution for another appeared not to be a viable option, because, until more recently, the bank market was shared by only three dominant banks[17] (Kwok and Tadesse, 2006). Still, many German companies tried to fend off

[15] Source: Institut für Mittelstandsforschung (IfM) in Bonn: www.ifm-bonn.org
[16] To be listed on the LSE, the rule required two-thirds of the shares to be held by the public.
[17] Deutsche Bank, Commerzbank and Dresdner Bank (now merged with Commerzbank)

investors whose primary goal was continuous shareholder maximisation, which in their view threatened to undermine the stakeholder value, that is, the framework of a high-skill and high-quality production process.

Two options appeared to be practicable in this regard. Haberly (2013), for instance, investigated the search for what he called "patient capital" by three car manufacturers, Daimler, Volkswagen and Porsche. This kind of capital stands in contrast to maximising shareholder value as quickly as possible. The second option entails a limitation of outside shareholders' voting rights, as documented by Crane and Schaede (2005).

In this context, companies can choose the following three options. Firstly, votes can be restricted to a certain percentage irrespective of the percentage ownership of a shareholder. At the same time, this rule can be used to establish a blocking minority for one particular shareholder. A well-known case is the so-called Volkswagen law, which ensures that the German state of Lower Saxony has direct control over matters that would have repercussions for the employees. Secondly, shares with no voting rights can be issued. Examples include the car manufacturer Porsche, of which both the Porsche and the Piëch family own all of the common stock with voting rights.[18] A similar case is BMW, for which the Quandt family holds a combined 46.7%[19] of the company, therefore making a takeover attempt virtually impossible. Thirdly and lastly, every transfer of shares between investors requires the permission of the corporation. An example of this type of restriction is Munich Re, one of the world's largest reinsurers with one of the most successful investors, Warren Buffett, as one of its shareholders.[20]

Faccio and Lang (2002) documented the differences in corporate structure between the UK and Germany in more general terms. Regardless of firm size, 64.62% of publicly traded corporations are controlled by a family. This term implies an actual family or person or a firm that is not listed on a stock exchange. The control term refers to a controlling interest above the 20% threshold. In the UK, the same figure amounts to 23.68%. Once firm size was taken into consideration, the authors separated the largest 20 companies while splitting the remaining data set into two, defining firms as either medium or small. The results indicated that 15%, 75% and 81% of the large, medium and small firms, respectively, were family

[18] Source: Porsche SE Annual Report 2013
[19] Source: BMW AG Annual Report 2013
[20] Source: Munich Re AG Annual Report 2013

dominated. In the UK, family ownership came to 0.00%, 19.6% and 38.67% for the same size criteria. Interestingly, 12.5% of the large German firms had cross-holdings, which confirms the existence of the so-called “Deutschland AG”, freely interpreted as Germany plc. This term is used to represent the intertwined nature of these companies.

6.2.2 Investor protection in common law and code law countries

This subsection outlines the differences in the degree of investor protection with regard to a country’s legal system. Naturally, the question of whether the traditional business habits in a certain country are the result of legislation or whether legislation is the result of those habits can be asked. The answer to this question is of interest to investors because it has implications for the level of investor protection. As La Porta et al. (1999) found, civil law, as prevalent in Germany, tends to preserve the power of the state and does not put much emphasis on upholding private property rights, to the detriment of financial markets. An example was given earlier by the colloquially known “VW-Gesetz”, that is, the Volkswagen law. By contrast, common law countries such as the UK appear to be more adaptable to changes in societal needs, as observed by Kwok and Tadesse (2006).

In general, the comparison between specific laws regarding investor protection assumes that the general quality of the legal environment is equivalent in the countries under scrutiny, as proposed by La Porta et al. (1997, 1998). There are different aspects of assessing this quality in the literature, such as measuring financial transparency (Doidge et al., 2007) or the degree of corruption (Ball et al., 2008). Considering these quality standards, Hail and Leuz (2006) found that out of 40 international countries, the legal standards both in the UK and in Germany are similar and above the sample average at the same time. These results were later confirmed by Byard et al. (2011) for a sample of 20 European countries. Again, both legal systems were found to be on or above the median qualitative requirement. By contrast, the degree of investor protection itself was, amongst others, analysed more recently by Pevzner et al. (2013). Out of 25 international countries, the German (UK) law achieved a below (above) average ranking in this context.

6.2.3 Hypothesis development

This section develops the hypotheses, which are specifically tailored to the aforementioned characteristics of the German economy.

In general, the first part of the first hypothesis tests the success of the original F-Score strategy after its replication, similar to the test for the UK. As highlighted in the previous two subsections, investors are faced with a set of market characteristics that are dissimilar from the UK and the US. Despite these differences, the F-Score investment strategy should be universally applicable because the original idea is to provide a simple and easily implementable tool for stock investors to generate profits consistently. Regarding the characteristics of German firms and the German stock market, there are at least three reasons why the original F-Score strategy could be more successful than in the UK. Firstly, the German stock market is less liquid, even for stocks in the DAX 30, the leading stock market index containing the biggest thirty companies by market capitalisation (Buck and Shahrim, 2005). Consequently, stock prices should react more slowly to newly released financial statement information, which would reward investors who instantly trade on this information before others become aware of its implications. This also agrees with the research findings by Lesmond et al. (2004) together with Corwin and Schultz (2012), who revealed that stock illiquidity is positively correlated with subsequent stock returns. In addition, using the Amihud measure as an indicator of illiquidity risk, the stock returns in Germany were found to be significantly higher than those in the UK (Liang and Wei, 2012).

Secondly, investors are left with more ambiguity about financial information as a result of poorer accounting quality. The reason is most likely to be the opaqueness of the German corporate culture, according to Buck and Shahrim (2005). Closely linked to that might also be the degree of difference between the respective local GAAP and the IFRS. Daske et al. (2008), for instance, observed greater uniformity between the UK-GAAP and the IFRS than between the German-GAAP and the IFRS. In general, the IFRS was found to lead to better transparency of financial information, but only in those countries where legal enforcement is strong. Even though legal enforcement is comparable between the UK and Germany, although investor protection is not by contrast, the majority of observations in this study are prior to 2005, when the IFRS became mandatory. This higher level of ambiguity should lead to higher returns.

Thirdly, as Marginson and McAulay (2008) suggested, US company managers are more likely to engage in short-termism due to the transitory nature of the stock markets there as compared with Germany. More concentrated stockholding by a group or family generally encourages long-term orientation by managers and therefore increases a company's stock price (Bushee, 1998). This would mean that investors pay less attention to current financial statements, which would generate more opportunities to increase the long-term returns. For this reason, the first part of the first hypothesis is as follows:

Hypothesis 6.1: The F-Score investment strategy and its alternatives can successfully separate future winning from losing stocks regardless of the country's legal system.

Given that accounting data from 1992 to 2010 are analysed in this study, the majority of financial statements were not prepared under the IFRS before it became a legal requirement in 2005. Therefore, the F-Score strategy, which is based on the US-GAAP, might not be as successful under the German-GAAP (HGB).

One possible reason for this is the different levels of value relevance between those two accounting principles. Bartov et al. (2005) found that, although the quality of earnings is not affected by any regime, the US-GAAP provides generally better information to investors than the HGB. At the same time, no difference regarding value relevance was identifiable between the US-GAAP and the IFRS. Although the US-GAAP is not the same as the UK-GAAP, Cuijpers and Buijink (2005) assumed broad consistence between those two principles, as assumed in the present study. However, this observation might be outweighed by the close bank–firm relationship, which extends a firm's access to capital and therefore leads to higher profitability, as hypothesised by Agarwal and Elston (2001).

However, a close bank–firm relationship might also bear a negative ramification for outside investors that is related to the speed of information incorporation into stock prices. Ultimately, this would influence the success of investment strategies given the aforementioned different degrees of investor protection and differences in accessing financial information. This also agrees with the research conducted by Hou (2007), who highlighted the central question in financial economics of how information is incorporated into stock prices. According to the EMH, market prices update almost instantaneously on the arrival of new information.

Another possible mechanism that potentially speaks on this issue is the so-called lead–lag effect, whereby certain stocks adjust prices more slowly than others to the arrival of new information. According to Hou (2007), this effect is likely to be caused, inter alia, by asymmetric information. This asymmetry, in turn, is likely to be fostered by both the legal environment and the type of market orientation. In a code law/bank-oriented environment, banks hold a unique position insofar as they have access to information that shareholders do not have (Agarwal and Elston, 2001). Even in the case in which the same kind of information is shared, banks with close relationships with their clients are likely to access it earlier. As a consequence, investors could be under the impression that financial statements do not contain much new information, which would make investment strategies less successful. To put it in EMH terminology, the German stock market would indeed be semi-strong-form efficient because information travels quickly to banks first, which then trade on it as insiders.

Due to the institutional environment and the dominant role of banks in Germany, the second part of the first hypothesis is as follows.

Hypothesis 6.2: The speed of information incorporation into and reaction of stock prices is expected to be faster for the German stock market than the UK stock market.

The second hypothesis focuses on the success of the F-Score strategy conditional on a firm’s book-to-market affiliation, similar to Duong et al. (2014). As a secondary outcome, this hypothesis also aims to analyse whether the investment strategies are driven by low or high book-to-market firms. Regarding the primary focus, the test for the existence of a value premium in the German stock market can be regarded as an out-of-sample robustness check similar to the UK data, aiming to avoid data-mining bias. This is in line with the procedure suggested by Ang and Bekaert (2007) and should further contribute to the mounting evidence for the existence of a value premium outside the US. Artmann et al. (2012b) provided long-term results that confirmed the outperformance of high book-to-market firms in the period from 1963 to 2006 in Germany. Although the book-to-market ratio is one of the two main drivers of stock returns in the German stock market amongst other factors, such as momentum and asset growth, the authors provided no clear answer to its cause.

Piotroski and So (2012), in this regard, suggested a behavioural explanation for investors’ under-reaction to firm fundamentals. They made the point that investors in low-BM (high-BM) stocks discard worsening (improving) fundamentals due to their positive (negative) expectations of this particular stock. If an investment strategy such as the F-Score serves as a yardstick for the quality of firm fundamentals, then low-BM (high-BM) firms with low (high) F-Scores should be overvalued (undervalued) and provide worse (better) subsequent returns as a consequence. Conversely, this means that both low-BM and high-BM firms with high and low F-Scores should be priced appropriately. For these reasons, the second hypothesis is stated as follows.

Hypothesis 6.3: A value premium is present in the German stock market. Therefore, firms with higher book-to-market ratios outperform those with lower book-to-market ratios.

While one of the aims of the first hypothesis is to analyse the speed of information transmission to stock prices, the prime focus of the third hypothesis is to research how the differences in information quality, meaning good and bad news, are reflected in the stock prices. As a secondary result, the purpose of this hypothesis is to uncover whether the success of the various investment strategies is driven by small or large firms.

The basic concept regarding the prime focus of the hypothesis is that investors under-react to information. Firm value depends on how homogeneous investors react to information, which then has consequences for stock prices (Zhang, 2006). If investors have a clear perception of the cause (information) and what it means for the firm (stock price), then information uncertainty can be regarded as low and vice versa. Consequently, a more gradual price adjustment with higher subsequent returns can be expected in cases of high information uncertainty.

This, in general, contradicts the assumption of fast information processing according to the EMH and favours a more heterodox explanation, as proposed by behavioural finance. To measure the degree of information uncertainty, firm size and the volatility of financial statement items are utilised as proxies, similarly to the previous empirical chapter on the UK data. This is also in accordance with Zhang (2006). Because listed firms in Germany are larger than those in the UK and larger firms are associated with lower information uncertainty, less uncertainty is expected in the German stock market. This, in turn, has implications for the size effect (e.g. Heston et al., 1999). For instance, Amel-Zadeh (2011) highlighted that, although there is international evidence for this effect, its markedness might be dependent on the market structure and the stage of market development.

However, this dependency might be outweighed in those cases in which investors' trading behaviour is highly homogeneous. This might apply to the German stock market. The reason is that due to the larger firm size and international exposure of German firms, the German stock market is likely to resemble the ones in the UK and the US. For this reason, Zhang's (2006) initial hypothesis is upheld and stated as follows.

Hypothesis 6.4: Firms with higher information uncertainty, proxied by firm size and the volatility of financial statement items, provide higher (lower) subsequent returns following good (bad) news.

6.3 An overview of the German data set

In chapter four, table 4.1 documented the availability of accounting data for establishing the F-Score in the UK and Germany. Regarding the latter, 5,944 observations are available over the period 1992 to 2010. This number represents only approximately a quarter of the 20,053 UK observations within the same period and already confirms the research outcomes of La Porta et al. (1997) with respect to smaller stock markets in code law countries. Figure 4.1 serves as a visual interpretation of the same fact.

6.3.1 Descriptive statistics of German firms

Before reconstructing the various investment strategies in the German stock market, the descriptive statistics of the individual F-Score contributors are outlined in table 6.1. For instance, the average firm has market capitalisation of 1.9 billion euros, with total assets of 3.7 billion euros. In terms of profitability, the mean (median) return on assets amounts to 0.16% (2.73%), with a change in that figure of -7.82% (0.15%) per year. As for the UK firms, the remaining four columns present the standard deviation of each item, the number of those observations that feature a positive sign as well as the minimum and maximum values.

Table 6.1 Financial characteristics of German firms

Variable	***Obs.***	***Mean***	***Median***	***Std dev.***	***% positive***	***Min.***	***Max.***
ROA	5,945	.0016163	.0272971	.4593706	71.51	-18.17	23.57
CFO	5,945	.0036422	.0676565	3.000476	76.79	-226.06	13.00
ΔROA	5,945	-.0781931	.0015073	4.520867	52.30	-267.88	74.32
ACCRUAL	5,945	-.0020259	-.047531	2.963618	25.33	-4.10	226.16
ΔLEVERAGE	5,945	.124198	.0703787	0.183215	81.87	0	4.28
ΔLIQUIDITY	5,945	-.1938486	-.01	12.97	47.23	-779.84	250.78
ΔGPM	5,945	9.634123	-8.51E-07	737.10	40.29	-208.70	56,823
ΔTURN	5,945	-5.263847	.0031555	346.20	50.83	-26,595	31.79
BM	5,862	.6853576	.5555556	.892694	-	-25.00	14.29
TASSETS	5,945	3,714,583	178,819	1.72E+07	-	417	2.64E+08
Market Cap.	5,847	1,906,449	112,996	7,734,463	-	0	2.14E+08

Table 6.2 presents those figures that are of most interest in highlighting the main differences between UK and German firms. The first difference relates to firm size as measured by market capitalisation. Public listed companies are on average twice as large in Germany as in the UK, with values of close to 2 billion euros and 835 million GBP, respectively. Due to the daily varying exchange rate between the GBP and the euro, this finding should be regarded as a ballpark figure only. This is further confirmed by the total assets figure, which is approximately 4 times higher for German firms. However, a similar trend was documented by Leuz et al. (2003) for the median and by Hou et al. (2011) for the mean firm size. The reason for that is most likely to be the heavy reliance on banks by German firms in comparison with Anglo-Saxon countries with more market-based financial systems (Audretsch and Elston, 2002). In their study, the authors analysed the interplay between firm size and liquidity constraints. They found that smaller firms appeared to be less constrained but that medium-sized firms faced more of those constraints. This would suggest that fewer small-sized firms would consider it necessary to satisfy their financing needs through the capital markets but rather through their long-term banking partner. Once the company has grown to a medium size, it only then becomes necessary to raise equity capital through a public offering. Consequently, the amount of large firms listed in the stock market is higher than that in the UK. Overall, however, the combined market value of all firms is larger in the UK because more firms are listed on the stock market (Watanabe et al., 2013).

The second distinguishing factor between UK and German firms is that the accruals ratio, that is, the net income minus cash flow from operations scaled by total assets, is overwhelmingly negative for the latter. According to Sloan (1996), investors have difficulties in interpreting the two components of earnings, which are accruals on the one side and cash flow on the other side. He also found that investors who went long in the lowest decile accrual portfolio and shorted the highest decile portfolio enjoyed returns of 10.4% on average during the period from 1962 and 1991 in the US. Although the accruals variable is just one of the nine F-Score components in this research, it highlights the requirements for firms within their respective judicial area. Because the accrual ratios are much lower for German firms, this could be indicative of stricter accounting rules, especially before the compulsory introduction of the IFRS standards on 1 January 2005 (e.g. Hung and Subramanyiam, 2007). According to Guay

Table 6.2 Comparison of financial characteristics

Variable	*Obs.*	*Mean*	*Median*	*Std dev.*	*% positive*
Market Cap. (DE)	5,847	1,906	112,996	7,734,463	-
Market Cap. (UK)	19,485	835.65	42.03	5,141	-
ACCRUAL (DE)	5,945	-.0020259	-.047531	2.963618	25.33
ACCRUAL (UK)	20,053	-.0462963	-.04271	4.350192	73.49
TASSETS (DE)	5,945	3,714,583	178,819	1.72E+07	-
TASSETS (UK)	20,053	968.94	53,133	6,245,722	-
ΔLEVERAGE (DE)	5,945	.124198	.0703787	0.183215	81.87
ΔLEVERAGE (UK)	20,053	.1386251	.0617856	0.740685	66.44

and Verrecchia (2006), strict accounting systems are characterised by more timely recognition of losses than of gains. Kaserer and Klingler (2008) confirmed the US results for Germany insofar as earnings were less persistent with regard to the accruals than with regard to the cash flows of the previous financial year. Their sample covered the period between 1995 and 2002. This was especially observed amongst those German firms that prepared financial statements under either the IFRS or the US-GAAP and not the German-GAAP (HGB). Interestingly, the authors did not find an accruals anomaly, like Sloan (1996), before the voluntary adoption of the IFRS before the year 2000. The observations summarised in table 6.2 are consistent with Pincus et al. (2007), who conducted research on the accrual anomaly in 20 countries. They found that the likelihood of the anomaly increases in countries with a common law legal tradition, non-concentrated stock ownership and less strict rules for accrual accounting. Further, earnings management appeared to be causal for the existence of the accrual anomaly. Regarding the accrual ratio in this study, this would mean that earnings management is less pronounced in Germany than in the UK. Due to the negative relationship between net income and subsequent stock returns, German stock returns should be higher. In addition, the relatively low percentage of positive accrual ratios of 25.33% amongst German firms, as shown in table 6.2, indicates that in general this measure as part of the overall F-Score should be more reliable in forecasting stock returns.

The third and last distinctive point is the higher gearing of German companies. This is a consequence of the first point regarding firm size. As mentioned earlier, small firms tend to have easier access to outside capital due to their long-term relationship with their bank. Once the threshold to a medium-sized firm is exceeded, the firm would need to raise capital in an initial public offering, while part of the debt would still remain on the balance sheet. Because debt is a large part of the overall financing needs, the equity ratio, that is, the market capitalisation divided by the total assets, is also lower in comparison with UK firms. The respective ratios are approximately 51.3% (63.2%) in Germany and 86.2% (79.1%) based on the mean (median) version of numerator and denominator. However, the management decision about a company's capital structure is in accordance with the pecking order theory in both the UK and Germany. As documented by Antoniou et al. (2008), managers are reluctant to raise expensive equity capital unless no other option is available under this theory. As the best option is to generate funds internally, an inverse relationship between profitability and gearing was predicted and identified in both markets. In summary, the comparison and interpretation of the different financial characteristics between UK and German firms is in line with the previous findings in the literature.

The next table contrasts the one-year and two-year buy-and-hold stock returns of Germany (DE), the UK and the US and represents the equivalent of table 4.3 in the previous fourth chapter. Compared with the two Anglo-Saxon stock markets, both the negative and the positive returns of the two periods are less pronounced in Germany, which indicates a slightly smaller width of the distribution and therefore a smaller standard deviation. This is confirmed by the last column of table 6.3, which contains information about the standard deviation. In general, however, all three stock markets are in line with each other regarding their buy-and-hold returns.

Table 6.3 Comparison of buy-and-hold returns between German, UK and US firms

Returns	*Mean*	*10th pctile*	*25th pctile*	*Median*	*75th pctile*	*90th pctile*	*% positive*	*S.D.*
1 Y (DE)	5.23E-17	-0.491	-0.271	-0.046	0.191	0.483	0.450	0.493
1 Y (UK)	6.15E-17	-0.558	-0.306	-0.048	0.229	0.542	0.451	0.511
1 Y (US)	0.059	-0.560	-0.317	-0.061	0.255	0.708	0.437	-
2 Y (DE)	4.19E-17	-0.731	-0.442	-0.111	0.243	0.781	0.405	0.778
2 Y (UK)	-2.36E-17	-0.802	-0.496	-0.118	0.304	0.822	0.424	0.794
2 Y (US)	0.127	-0.872	-0.517	-0.111	0.394	1.205	0.432	-

Piotroski (2000) did not report this figure, but the results are in line with the findings by Dellas and Hess (2005). According to the authors, the stock returns in the period from 1980 to 1999 were significantly related to the degree of financial development. In total, 49 stock markets were analysed and higher stock returns were associated with higher volatility. As both Germany and the UK can be regarded as developed markets with a high-quality banking system, the standard deviation measure for the two periods is similar, as expected. By contrast, the standard deviation in emerging markets was found to be at least twice as high in the majority of cases. With regard to the overall return performance of the German stock market, less than half of the tradable stocks show positive returns after one year. This is also valid for the two-year investment horizon and similar to the UK findings as well as the US findings originally documented by Piotroski (2000).

6.3.2 Correlation matrix and basic F-Score strategy in Germany

Similar to the analysis of the UK stock market, the next table presents the Spearman correlation matrix. Of particular interest is the relationship between the composite F-Score and the winsorised and non-winsorised one-year and two-year stock returns, which are highlighted in bold at the bottom of table 6.4. The results are highly significant at the 1% level for all of the four combinations, with only very little difference between the winsorised and the non-winsorised version of the returns. As outlined in chapter 4, each of the nine variables is assigned to one of the three dimensions that are used to evaluate a firm’s financial health. In general, all nine variables contribute significantly to the overall F-Score, as can be seen from the bottom row of the table. However, similar to the UK, accruals are of least importance to the overall F-Score amongst the four profitability F-components, although they are stronger than in the UK.

However, the picture looks different once the correlations between the individual variables and the four return measures are established. Amongst the three financial health dimensions, all the variables within the profitability part are highly significant. For instance, the correlation between the cash flow from operations (CFO) and the subsequent one-year winsorised stock return amounts to 0.151. Across the profitability measures and in total, both return on assets (ROA) and CFO feature the highest stock return correlation, indicating that they are crucial to the success of the F-Score investment strategy.

The remaining two dimensions, that is, leverage, liquidity and source of funds and operating efficiency, show much weaker results within the same setting. In both cases, less than 50% of the variables in those two dimensions provide statistically significant results. In other words, only the issuance of equity capital (EQO) and the change in asset turnover (ΔTRN) are relevant to future stock returns in most cases. Investors appear not to pay much attention to changes in the level of gearing (ΔLEV), liquidity (ΔLIQ) and gross profit margin (ΔGPM).

6.3.3 Comparison of correlation matrices between Germany, the UK and the US

The next section contrasts the results of the Spearman matrix regarding the German, the UK and finally the US stock market to validate the findings of the original study. Generally, the profitability dimension of both European markets includes those variables that contribute the most to the F-Score and that are highly correlated with future stock returns at the same time. Both ROA and CFO are indispensable financial statement items that represent the key drivers of the investment strategy. Similar findings were documented by Piotroski (2000). The remaining two profitability measures, ΔROA and ACC, are significant in Germany and the UK, albeit to a much lesser extent than the first two.

Interestingly, this changes once the focus shifts to the second dimension of leverage, liquidity and source of funds. In this scenario, the change in liquidity (ΔLIQ) variable is significant in the UK but not in Germany. The same applies to the change in financial gearing (ΔLEV) measure, which is significant for the two-year UK stock returns only but insignificant in Germany. However, investors in both markets seem to pay close attention to a firm's propensity to raise equity capital. Considering the correlation figures, this is approximately twice as important to the market participants in Germany as to those in the UK. By way of illustration, the correlation between the EQO variable and the one-year-ahead stock return amounts to 0.070 (0.033) for Germany (UK) and 0.041 for the US.

A glance at the last dimension, representing operating efficiency, reveals that for both European stock markets the change in the gross profit margin (ΔGPM) does not play any role in determining the subsequent one-year and two-year returns. In fact, the correlation is even negative in the case of the UK and, apart from that, this F-component generally has values close to zero. Regarding the US data, ΔGPM has higher correlation values and the fifth-highest correlation for both the one-year and the two-year stock returns. The second and last component in this dimension is the change in asset turnover (ΔTRN). This variable is significant in three out of four cases in Germany. For instance, the correlation with the winsorised one-year return amounts to 0.038. However, the same variable is without

Table 6.4 Spearman correlation matrix between stock returns, F-components and F-Score[a]

Variable	***I) Profitability***				***II) Leverage, liquidity and source of funds***			***III) Operating efficiency***		***Returns***				***F-Score***
	ROA	CFO	ΔROA	ACC	ΔLEV	ΔLIQ	EQO	ΔGPM	ΔTRN	RET	W_RET	RET2	W_RET2	F-Score
ROA	1	-	-	-	-	-	-	-	-	-	-	-	-	-
CFO	0.410***	1	-	-	-	-	-	-	-	-	-	-	-	-
ΔROA	0.248***	0.082***	1	-	-	-	-	-	-	-	-	-	-	-
ACC	-0.132***	0.414***	-0.083***	1	-	-	-	-	-	-	-	-	-	-
ΔLEV	0.035	0.001	0.031	-0.011	1	-	-	-	-	-	-	-	-	-
ΔLIQ	0.113***	0.018	0.143***	-0.076***	-0.056***	1	-	-	-	-	-	-	-	-
EQO	0.025	0.104***	-0.023	0.069***	0.014	-0.074***	1	-	-	-	-	-	-	-
ΔGPM	-0.081***	-0.001	0.121***	0.051***	0.002	0.063***	0.075***	1	-	-	-	-	-	-
ΔTRN	0.054***	0.018	0.292***	-0.005	-0.060***	0.001	-0.062***	-0.044***	1	-	-	-	-	-
RET	0.132***	0.150***	0.044***	0.065***	-0.002	0.025	0.069***	0.015	0.040***	1	-	-	-	-
W_RET	0.131***	0.151***	0.043***	0.067***	-0.002	0.026	0.070***	0.016	0.038***	0.999***	1	-	-	-
RET2	0.170***	0.178***	0.051***	0.083***	0.011	0.026	0.074***	0.009	0.036***	0.750***	0.751***	1	-	-
W_RET2	0.166***	0.180***	0.048***	0.088***	0.011	0.026	0.077***	0.013	0.032	0.750***	0.751***	0.999***	1	-
F-Score	0.460***	0.528***	0.551***	0.305***	0.266***	0.339***	0.308***	0.346***	0.354***	**0.147*****	**0.147*****	**0.172*****	**0.173*****	1

[a] The nine F-components are correlated between each other, one-year and two-year returns (RET and RET2), one-year and two-year winsorised returns (W_RET and W_RET2) and the F-Score. All the F-components are correlated using their binary score of either 1 or 0 representing good or bad financial performance.

significance in the UK. Piotroski (2000) reported the second-lowest and lowest correlation values for the same variable for the one-year and two-year investment horizon.

To summarise the results of all three stock markets, table 6.5 presents the nine F-components ranked according to their influence on the following one-year winsorised stock returns.

Table 6.5 Correlation between F-components and one-year winsorised returns

Germany		***UK***		***US***	
Variable	ρ	Variable	ρ	Variable	ρ
CFO (I)	0.151	CFO (I)	0.163	CFO (I)	0.096
ROA (I)	0.131	ROA (I)	0.144	ROA (I)	0.086
EQO (II)	0.070	ACC (I)	0.054	ΔLEV (II)	0.055
ACC (I)	0.067	EQO (II)	0.033	ACC (I)	0.053
ΔROA (I)	0.043	ΔROA (I)	0.023	ΔGPM (III)	0.042
ΔTRN (III)	0.038	ΔLIQ (II)	0.020	EQO (II)	0.041
ΔLIQ (II)	0.026	ΔLEV (II)	0.018	ΔROA (I)	0.037
ΔGPM (III)	0.016	ΔTRN (III)	0.014	ΔTRN (III)	0.034
ΔLEV (II)	-0.002	ΔGPM (III)	-0.009	ΔLIQ (II)	0.032
F-Score	**0.147**	**F-Score**	**0.130**	**F-Score**	**0.121**

As detailed earlier, each variable is part of one out of three dimensions that assesses one specific financial aspect of a company. The association with this dimension is indicated in parentheses. The variable with the highest correlation, represented by the Greek letter ρ, is ranked first. As can be seen, profitability measures play a vital part in the success of the F-Score investment strategy in Germany and the UK. These financial statement measures are ranked consistently within the top five F-components. By contrast, measures of operating efficiency can mostly be found at the bottom of the table. The results for the US are not as clear-cut as for the European market, which indicates that a broader range of F-components is causal for the subsequent returns. Generally, however, profitability measures play a dominant role. Despite the focus on the profitability measures, the F-Score correlation is highest in Germany followed by the UK and significant similar to the US. Expressed as R^2, more variance is shared between the F-Score and the stock returns in Germany than in the UK and the US. In summary, therefore, the results confirm that the F-Score investment strategy works in Germany as well. This represents a firm grounding on which to proceed with the more detailed analysis, which follows next.

6.3.4 Analysis of one- and two-year buy-and-hold returns in Germany

Analogous to table 4.5 for the UK, table 6.6 contains the German one-year and two-year buy-and-hold returns of the 5,945 sample firms. These are subdivided vertically into ten portfolios according to their overall F-Score. Horizontally, the columns include the mean, the five percentiles, the percentage of positive stock returns and the amount of observations for each of the ten portfolios. From the last column of the table, it is evident that, compared with the UK, the German data set is tilted towards higher F-Scores. This, in turn, is directly related to the firm years of the F-Score, as presented in table 3.4, and highlights the need to construct the fquintiles in a slightly different way for the German data. To obtain balanced fquintiles, that is, fquintiles that ideally contain an equal amount of stocks in the lowest/highest and low/high fquintiles, firms with an F-Score of up to 3 (4) are assigned to the lowest (low) fquintile. This is in contrast to the UK, for which the lowest (low) fquintile consists of stocks with a maximum F-Score of 2 (3). The middle fquintile thereby serves as a buffer and is not subject to further analysis.

As before, the structure of table 6.6 is kept consistent. The mean returns in panel A document an improvement from the lowest to the highest F-Score firms. Winsorised returns are shown in italics below the respective non-winsorised figures. However, neither version of the stock returns improves in a linear manner in comparison with the UK, as found in chapter 4. The most likely reason for this can be attributed to the overall smaller amount of observations and the associated smaller amount of firms within each of the ten portfolios. For instance, the seven firms in the lowest F-Score portfolio return a negative 29.7% after one year. Firms with an F-Score of 1 return 11.7% on average, but the 182 firms in the next best portfolio return minus 13.3%. The extent of this high volatility subsequently decreases because the amount of observations increases. Evidence for the better performance of the highest fquintile is presented in the last two rows of panel A. The mean return of a portfolio consisting of firms with an F-Score of either 8 or 9 is significantly different from the mean return of a portfolio that includes the entire sample. The respective t-statistics are reported in the penultimate row of panel A and are significant at the 5% and 1% level depending on the non-winsorised and winsorised return version, respectively. Even better results are reported where the average returns of the highest (H*) and lowest (L*) F-Score portfolios are analysed.

The above-mentioned process was replicated for the two-year buy-and-hold returns and the results are presented in panel B. It is observable that the efficiency of the F-Score strategy weakens once the investment horizon is extended. This is particularly the case once the returns of the highest F-Score portfolio are compared with the performance of the entire data set. The non-winsorised t-statistics are insignificant, while the winsorised results are only significant at the 10% level, although very marginally. The picture improves dramatically once the highest and lowest portfolios are compared. In this scenario, both versions of the stock returns are highly significant.

Overall, the one-year performance of the F-Score strategy in Germany is comparable to that of the UK. This means that the H*–All as well as the H*–L* return differentials are at least significant at the 5% level. However, the two-year results are weaker with regard to the H*–All differences in the two-year environment, with only the winsorised differences being significant at the 10% level. The reason for this might be the smaller sample size, whereby outliers exert an unduly large influence on the means. A good example is the mean return of 11.9% for firms with an F-Score of 1 in panel B of table 6.6. This return is based on 43 observations only and is considerably higher than in the UK, where the mean return of 197 firms amounts to -12.6% (table 4.5). As the original intention by Piotroski (2000) was to separate winning from losing stocks, the results of the differences between the mean portfolio returns of the highest/lowest (H*–L*) portfolios are considered as the decisive factor. This is of great interest particularly for investors who are interested in a long/short investment strategy. In this regard, the German results are also comparable to the US. The respective US t-statistics for the one-year (two-year) winsorised returns amount to 5.590 (5.749).

Table 6.6 One-year and two-year buy-and-hold returns of F-Score portfolios[a]

Panel A: One-year buy-and-hold returns								
	Mean	***10%***	***25%***	***Median***	***75%***	***90%***	***% positive***	***Obs.***
All firms	-8.6E-11 *2.4E-10*	-0.500 *-0.491*	-0.283 *-0.271*	-0.056 *-0.046*	0.182 *0.191*	0.474 *0.483*	0.440	5,945
F-Score								
0	-0.299 *-0.297*	-0.747 *-0.745*	-0.492 *-0.477*	-0.364 *-0.364*	-0.027 *-0.024*	0.290 *0.290*	0.143	7
1	0.026 *0.117*	-0.881 *-0.875*	-0.637 *-0.632*	-0.153 *-0.135*	0.267 *0.281*	0.660 *0.677*	0.395	43
2	-0.131 *-0.133*	-0.685 *-0.686*	-0.483 *-0.472*	-0.243 *-0.228*	0.069 *0.095*	0.526 *0.527*	0.302	182
3	-0.065 *-0.064*	-0.707 *-0.694*	-0.458 *-0.448*	-0.179 *-0.169*	0.124 *0.130*	0.547 *0.555*	0.353	529
4	-0.015 *-0.027*	-0.563 *-0.550*	-0.334 *-0.325*	-0.087 *-0.080*	0.180 *0.189*	0.462 *0.483*	0.415	892
5	-0.022 *-0.025*	-0.520 *-0.515*	-0.307 *-0.295*	-0.087 *-0.078*	0.163 *0.171*	0.463 *0.467*	0.408	1,317
6	-0.002 *0.008*	-0.450 *-0.443*	-0.248 *-0.234*	-0.021 *-0.012*	0.175 *0.183*	0.452 *0.457*	0.467	1,298
7	0.064 *0.065*	-0.372 *-0.361*	-0.200 *-0.189*	-0.000 *-0.013*	0.202 *0.211*	0.480 *0.487*	0.499	1,036
8	0.048 *0.059*	-0.311 *-0.302*	-0.180 *-0.168*	0.006 *0.017*	0.211 *0.219*	0.461 *0.473*	0.500	527
9	0.100 *0.077*	-0.273 *-0.269*	-0.180 *-0.163*	-0.002 *0.004*	0.187 *0.203*	0.479 *0.483*	0.559	114
H*–All t-statistic *(p-value)*	2.243** *(0.025)* 3.089*** *(0.002)*	-	-	-	-	-	-	-
H*–L* t-statistic *(p-value)*	4.077*** *(0.000)* 4.768*** *(0.000)*	-	-	-	-	-	-	-

Panel B: Two-year buy-and-hold returns								
	Mean	***10%***	***25%***	***Median***	***75%***	***90%***	***% positive***	***Obs.***
All firms	-5.9E-11 *-1.3E-09*	-0.766 *-0.742*	-0.455 *-0.434*	-0.124 *-0.105*	0.226 *0.241*	0.732 *0.758*	0.535	5,615
F-Score								
0	-0.130 *-0.115*	-1.231 *-1.210*	-1.230 *-1.172*	-0.519 *-0.516*	0.514 *0.535*	1.974 *1.974*	0.429	7
1	0.119 *-0.251*	-1.172 *-1.114*	-0.863 *-0.841*	-0.498 *-0.498*	-0.157 *-0.137*	0.343 *0.343*	0.186	43
2	-0.216 *-0.204*	-0.986 *-0.983*	-0.706 *-0.685*	-0.415 *-0.386*	0.009 *0.011*	0.512 *0.512*	0.260	173
3	-0.139 *-0.137*	-0.974 *-0.954*	-0.700 *-0.672*	-0.325 *-0.301*	0.066 *0.078*	0.936 *0.953*	0.288	517
4	-0.027 *-0.033*	-0.809 *-0.801*	-0.537 *-0.516*	-0.186 *-0.158*	0.206 *0.231*	0.788 *0.815*	0.369	862
5	-0.028 *-0.018*	-0.770 *-0.758*	-0.475 *-0.459*	-0.142 *-0.134*	0.189 *0.208*	0.667 *0.679*	0.388	1,274
6	0.043 *0.034*	-0.638 *-0.624*	-0.386 *-0.368*	-0.087 *-0.074*	0.245 *0.263*	0.701 *0.744*	0.426	1,209
7	0.095 *0.101*	-0.602 *-0.568*	-0.332 *-0.326*	-0.040 *-0.034*	0.324 *0.338*	0.839 *0.854*	0.459	940
8	0.050 *0.064*	-0.543 *-0.525*	-0.304 *-0.286*	-0.039 *-0.022*	0.287 *0.296*	0.679 *0.695*	0.474	475
9	-0.001 *0.013*	-0.570 *-0.557*	-0.370 *-0.360*	-0.067 *-0.067*	0.236 *0.236*	0.680 *0.702*	0.426	115
H*–All t-statistic *(p-value)*	0.946 *(0.345)* 1.654* *(0.098)*	-	-	-	-	-	-	-
H*–L* t-statistic *(p-value)*	3.210*** *(0.001)* 4.966*** *(0.000)*	-	-	-	-	-	-	-

[a] Values in the upper (lower) half of each cell represent non-winsorised (winsorised) one-year and two-year returns.

6.3.5 One-way portfolios with the original F-Score

Having presented the descriptive statistics of the F-Score strategy with regard to the German stock market, the analysis now moves on to provide a more thorough analysis of this method. This ensures the comparability with the UK findings and simulates ex ante the possible returns that investors could have expected from the strategy during the period from 1992 to 2010. The first step is the evaluation of the one-way portfolios utilising the original F-Score method. It should be noted that the structure of the one-way and later the two-way table corresponds to that in the previous chapters. Table 6.7 contains the one-way results of the investment strategy, which suggest that a distinction between outperformers and underperformers can be achieved reliably.

Table 6.7 One-way F-Score portfolios

	F-Score				*t-test*		*MWW*		*median*	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.078	-0.044	0.062	0.057	0.11*** *(4.459)*	0.14*** *(4.077)*	*(10.67)****	*(9.35)****	-	-
median	-0.198	-0.136	0.002	0.002	-	-	-	-	0.14*** *(74.48)*	0.20*** *(65.53)*
winsd.										
mean	-0.079	-0.051	0.064	0.062	0.12*** *(6.274)*	0.14*** *(4.768)*	*(10.69)****	*(9.36)****	-	-
median	-0.190	-0.128	0.014	0.016	-	-	-	-	0.14*** *(76.89)*	0.21*** *(67.28)*

Stars indicate statistical significance as follows: *** = 1%, ** = 5% and * = 10%.

Portfolios consisting of high (highest) F-Score firms consistently outperform their low (lowest) complement with highly significant t-statistics throughout. This is observable for both the parametric and the non-parametric tests. For instance, the mean return difference of the high/low portfolio at the bottom of the table is 11.5 percentage points (rounded to 12), with t-statistics of 6.274. Assuming a long-only investment style, the high (H) ranked portfolios perform at least as well as the highest (H*) ranked ones in the majority of cases. The only case of underperformance can be found in the winsorised median returns, in which the H (H*) portfolio returns 1.4% (1.6%). In a long/short scenario, however, going long with the H* portfolio and shorting the L* portfolio would yield a higher return (14%) than applying the same strategy to the H/L portfolio (11%).

In summary, compared with the UK, the return differentials in Germany are less pronounced and feature lower t-statistics at the same time. This applies particularly to the median portfolio returns. Due to the strong statistical significance at the 1% level, it is recommendable to use it in the German stock market.

6.3.6 One-way portfolios with an alternative: The F-rank approach

In the next step, the first alternative to the original F-Score is presented, namely the F-rank approach. Table 6.8 summarises the respective results in two panels. The portfolio returns in both panels were calculated as usual. However, the F-ranks were calculated using either the average of the nine individual Fi-ranks (panel A) or its median (panel B). For a detailed description of the computation of the Fi-ranks, the reader should refer to section 3.5.1.5 of this thesis.

Table 6.8 One-way F-rank portfolios

A	***F-rank (mean)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.081	-0.027	0.045	0.065	0.07*** *(2.985)*	0.15*** *(3.819)*	*(7.39)****	*(7.41)****	-	-
median	-0.206	-0.104	-0.010	0.002	-	-	-	-	0.09*** *(37.07)*	0.21*** *(51.25)*
winsd.										
mean	-0.081	-0.034	0.040	0.067	0.07*** *(4.224)*	0.15*** *(4.268)*	*(7.40)****	*(7.41)****	-	-
median	-0.196	-0.095	-0.000	0.008	-	-	-	-	0.09*** *(36.26)*	0.20*** *(52.92)*
B	***F-rank (median)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.094	-0.051	0.036	0.078	0.09*** *(3.547)*	0.17*** *(3.568)*	*(8.07)****	*(6.83)****	-	-
median	-0.224	-0.137	-0.023	-0.005	-	-	-	-	0.11*** *(52.22)*	0.22*** *(44.85)*
winsd.										
mean	-0.093	-0.058	0.032	0.057	0.09*** *(5.005)*	0.15*** *(4.115)*	*(8.08)****	*(6.82)****	-	-
median	-0.214	-0.127	-0.015	0.004	-	-	-	-	0.11*** *(51.25)*	0.22*** *(41.81)*

From the above table, it becomes clear that, due to the highly significant test statistics, the F-rank method is coequal to the original F-Score strategy. At second glance, however, some differences can be identified. Firstly, panels A and B are compared with each other. Regarding the return differentials that appear in the t-test column, the median F-rank provides investors with higher profits considering a long/short strategy, although this is paid for by lower statistical significance. The non-winsorised H*–L* difference in the case of the median F-rank in panel B, for example, amounts to some 17%, while its mean F-rank counterpart in panel A only returns less than 15%. However, the respective t-statistics for the latter are 3.819 and for the former 3.568. This also applies to the winsorised portfolio returns and is confirmed by the lower MWW and median test results in panel B.

Secondly, the F-rank strategy is compared in more detail with the original F-Score method. It is notable that the t-statistics of the parametric and non-parametric tests are consistently lower than those for the F-Score return differences in table 6.7, which indicates lower reliability of the F-rank method. Besides, especially the H–L percentage point differences are lower in all the cases as well. The H*–L* differentials are on a par with the F-Score strategy. In overall terms, therefore, the F-rank approach has some drawbacks but is still a credible alternative for investors.

Thirdly and lastly, the like-for-like comparison of the F-rank strategy in Germany with the UK reveals stronger t-statistics for the UK in general. However, on the basis of the absolute return differentials, no definite answer can be given. In total, the use of the strategy yields slightly better results for the UK.

6.4 The first alternative combination measures for the F-rank: An overview

Similar to the fourth chapter, which partly deals with the UK results of the mean squared forecast errors (DMSFEs), the following section and subsections of this chapter highlight the respective results from the German stock market. As before, the detailed procedure for computing the relevant weighted DMSFEs can be found in sections 3.5.1.6. The focus will thereby mainly rest on the discussion of the differences between Germany and the UK. It should be noted that the expected sign of the horizontal return differences is negative. This makes sense because stocks with lower forecast errors should have higher returns.

6.4.2 Combination measures for the F-rank: The weighted DMSFE

To complete the presentation of the one-way DMSFE results, table 6.10 highlights the findings of the two weighted versions of this approach. Similar to table 4.8 in the previous chapter, table 6.9 presents the yearly composition of fquintiles from 1993 to 2010 in Germany. There are differences in the way in which the respective cut-off points were calculated. The lowest and highest fquintiles now contain 10% of the data instead of 5%. Nevertheless, the cut-off points for the remaining fquintiles (L, M and H) are kept at 30%, 40% and 30%. As usual, the lowest and highest fquintiles are included in the low and high fquintiles.

Table 6.9 Overview of DMSFE portfolios by year

DMSFE	L*	L	M	H	H*	Total (L, M, H)
1993	0	1	0	0	0	1
1994	5	15	19	15	5	49
1995	12	35	46	35	12	116
1996	16	48	64	48	16	160
1997	21	62	82	62	21	206
1998	25	73	98	73	25	244
1999	31	93	124	93	31	310
2000	33	98	129	98	33	325
2001	40	119	158	119	40	396
2002	38	113	151	113	38	377
2003	40	119	158	119	40	396
2004	42	124	164	124	42	412
2005	46	136	181	136	46	453
2006	49	146	193	146	49	485
2007	49	146	193	146	49	485
2008	48	144	191	144	48	479
2009	49	147	196	147	49	490
2010	56	169	224	168	56	561
						5,945

The winsorised weighted version is included as a robustness check and generally yields similar results. Hence, only panel A will be discussed in the further course of this subsection.

Table 6.10 One-way weighted DMSFE portfolios

A	***DMSFE (weighted)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	0.074	0.049	-0.037	-0.090	-.086*** *(-3.559)*	-.164*** *(-4.186)*	*(-8.65)****	*(-7.83)****	-	-
median	0.002	-0.009	-0.122	-0.219	-	-	-	-	-0.113*** *(53.19)*	-0.221*** *(58.08)*
winsd.										
mean	0.075	0.045	-0.045	-0.090	-.089*** *(-5.054)*	-.164*** *(-4.647)*	*(-8.66)****	*(-7.84)****	-	-
median	0.010	0.001	-0.113	-0.209	-	-	-	-	-0.114*** *(52.22)*	-0.218*** *(58.08)*
B	***DMSFE (winsorised weighted)***				***t-test***		***MWW***		***median***	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	0.074	0.049	-0.037	-0.091	-.086*** *(-3.541)*	-.165*** *(-4.213)*	*(-8.59)****	*(-7.86)****	-	-
median	0.002	-0.009	-0.121	-0.219	-	-	-	-	-0.112*** *(51.25)*	-0.221*** *(58.08)*
winsd.										
mean	0.075	0.044	-0.045	-0.091	-.089*** *(-5.030)*	-.166*** *(-4.677)*	*(-8.60)****	*(-7.86)****	-	-
median	0.010	0.001	-0.111	-0.209	-	-	-	-	-0.112*** *(51.25)*	-0.218*** *(58.08)*

Unlike the UK findings, the weighting procedure does not dramatically improve the DMSFE strategy. A direct comparison of the results between the two stock markets under scrutiny shows that weighting the forecast errors is more successful in Germany than in the UK. Although for both markets, the non-parametric tests, namely the MWW and median test, provide highly significant outcomes, this is not the case for the parametric t-test. In this case, the UK results are not significant regarding the H*–L* differential of the raw returns (shown in panel A of table 4.9). Consequently, this increases the likelihood of this strategy working more reliably on German stocks.

Of interest is also the performance comparison between the F-Score, the F-rank and the weighted DMSFE F-rank method. A detailed overview is provided in table 6.12 after the presentation of the results of the one-way $k = 2$ and $k = 3$ clusters in the following section.

6.5 The second alternative combination measures for the F-rank: An overview

Unlike the presentation of the cluster results for the UK in chapter four, this section summarises the respective $k = 2$ and $k = 3$ results for the German stock market to keep the thesis structure as concise and clear as possible. Before shifting the focus to the two-way portfolio results, a detailed table will provide a summary of all the investment strategies in the one-way scenario for Germany and contrast it with its UK counterpart. In line with the logic of the F-Score, the expected sign of the horizontal return differentials is positive. This means that the higher the c-score, the higher the subsequent return.

6.5.1 Combination measures for the F-rank: Clustering with $k = 2$ and $k = 3$

The next table is subdivided into panel A, which represents the portfolio returns according to the $k = 2$ clustering approach, and panel B, which contains the respective results of the $k = 3$ version. Section 3.5.1.8 serves as a basis for the calculation of the c-score. It is assumed that higher c-scores are accompanied by higher future portfolio returns. As can be seen from both panels, all the test statistics are highly significant at the 1% level, although the values are mostly higher for the $k = 2$ method in panel A. Furthermore, in terms of both absolute mean returns and mean return differentials, the $k = 2$ method performs better. The highest (H*) $k = 2$ ($k = 3$) non-winsorised c-score portfolio, for example, returns 7.0% (5.9%) on average, with a return difference (H*–L*) of 13.1% (11.3%). However, the results of the median return differences indicate that investors benefit slightly more from the $k = 3$ cluster method.

Table 6.11 One-way clustered portfolios ($k = 2$ and $k = 3$)

A	*F-rank (c-score, k = 2)*				*t-test*		*MWW*		*median*	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.061	-0.050	0.047	0.070	.097*** *(4.346)*	.131*** *(3.182)*	*(9.76)****	*(6.43)****	-	-
median	-0.183	-0.130	-0.001	-0.008	-	-	-	-	0.13*** *(76.55)*	0.18*** *(33.33)*
winsd.										
mean	-0.062	-0.055	0.050	0.065	.105*** *(6.059)*	.127*** *(3.549)*	*(9.76)****	*(6.43)****	-	-
median	-0.174	-0.123	0.010	0.001	-	-	-	-	0.13*** *(75.97)*	0.18*** *(33.33)*
B	*F-rank (c-score, k = 3)*				*t-test*		*MWW*		*median*	
Returns	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
mean	-0.054	-0.044	0.041	0.059	.085*** *(3.560)*	.113*** *(2.895)*	*(9.24)****	*(7.64)****	-	-
median	-0.192	-0.136	-0.007	0.004	-	-	-	-	0.14*** *(69.13)*	0.20*** *(58.08)*
winsd.										
mean	-0.058	-0.051	0.040	0.061	.091*** *(5.101)*	.119*** *(3.442)*	*(9.24)****	*(7.64)****	-	-
median	-0.180	-0.127	0.003	0.016	-	-	-	-	0.13*** *(68.02)*	0.20*** *(56.33)*

In summary, both cluster approaches deliver solid results in the German stock market and should therefore be considered as a practical tool to enhance portfolio returns. This applies particularly to market participants with a long/short investment strategy.

6.5.2 A comparison between one-way strategies in Germany and the UK

As a conclusion of this section of this chapter, which presented the portfolio results of the one-way setting, a more detailed comparison between the German and the UK stock market will be provided next. To this end, table 6.12 summarises the one-way results at a glance. All nine approaches are ranked from best (1) to worst performing (6).

Table 6.12 One-way strategies in Germany and the UK

A		**Germany (one-way)**			
***Rank*[1]**	***Overall***	***Long-only (mean)***		***Long-only (median)***	
		H	H*	H	H*
1	wDMSFE	F-Score (6.4%)	wDMSFE (7.5%)	F-Score (1.4%)	F-Score (1.6%)
2	F-Score	cluster2 (5.0%)	mF-rank (6.7%)	cluster2 (1.0%)	cluster3 (1.6%)
3	mF-rank	wDMSFE (4.5%)	cluster2 (6.5%)	cluster3 (0.0%)	wDMSFE (1.0%)
4	mdF-rank	cluster3 (4.0%)	F-Score (6.2%)	wDMSFE (0.0%)	mF-rank (0.8%)
4	cluster2	mF-rank (4.0%)	cluster3 (6.1%)	mF-rank (0.0%)	mdF-rank (0.4%)
6	cluster3	mdF-rank (3.2%)	mdF-rank (5.7%)	mdF-rank (-1.5%)	cluster2 (0.0%)
B		**UK (one-way)**			
Rank	***Overall***	***Long-only (mean)***		***Long-only (median)***	
		H	H*	H	H*
1	F-Score	F-Score (5.4%)	F-Score (7.8%)	F-Score (1.3%)	F-Score (2.6%)
2	cluster2	cluster3 (3.7%)	mF-rank (4.8%)	cluster2 (-0.4%)	cluster3 (0.6%)
3	mF-rank	cluster2 (3.4%)	cluster3 (4.3%)	cluster3 (-0.6%)	cluster2 (-0.9%)
4	cluster3	mF-rank (3.1%)	wDMSFE (4.1%)	wDMSFE (-1.9%)	wDMSFE (-1.0%)
5	wDMSFE	wDMSFE (3.0%)	cluster2 (3.6%)	mF-rank (-2.0%)	mF-rank (-1.0%)
6	mdF-rank	mdF-rank (2.9%)	mdF-rank (1.2%)	mdF-rank (-2.6%)	mdF-rank (-9.1%)

[1] F-Score: *Original F-Score*; cluster2: *k = 2 clustering*; cluster3: *k = 3 clustering*; mF-rank: *mean F-rank*; mdF-rank: *median F-rank*; wDMSFE: *weighted DMSFE*.

Panel A shows the German results and panel B the ones from the UK stock market. Each investment strategy is evaluated according to three criteria. Firstly, the overall column determines which method has i) the best long-only mean results, ii) the highest statistical H*–L* significance and iii) the largest H*–L* return differentials of the average winsorised portfolio returns. The best performance with regard to points ii) and iii) means whichever strategy provides investors with the highest return differential between the highest and the lowest fquintile (H*–L*).

Secondly and thirdly, both the average and the median portfolio returns are shown separately with respect to their high (H) and highest (H*) fquintile affiliation. Their respective one-year returns are highlighted in parentheses and decrease from the top to the bottom of the table. Apart from the original F-Score strategy, there are five additional alternatives, for which the abbreviations in the table are clarified in the footnote.

Regarding panel A, the F-Score is ranked second overall but shows strong results in the long-only environment, which speaks for its high reliability in the German stock market. Both the mean and the median F-rank strategy appear to be suboptimal substitutes for the F-Score. This shows that, against the initial assumption, a more detailed breakdown of firm information does not lead to superior results. Improvements can only be achieved by applying the more complex forecasting and weighting methods. This is particularly interesting for long-only investors, who can increase their returns using the wDMSFE strategy (7.5%) compared with the F-Score (6.2%) in the H* fquintile or mean returns. However, in terms of consistency and reliability, the weighted DMSFE and F-Score provide the best tools for long-only and long/short returns as well as high significance. Nevertheless, in consideration of the additional effort that needs to be spent on establishing the weights, the F-Score approach appears to be more practical. Regarding the remaining methods, the results are generally sobering. The median DMSFE features the worst results in all the instances. However, it is noteworthy that at least all the mean long-only H and H*-fquintiles in the third and fourth columns produce positive returns for investors under the one-year investment horizon.

These findings are not entirely comparable to the UK results demonstrated in panel B. Here, the F-Score is unmatched as the best-performing strategy throughout and yields strong positive returns in all the long-only cases. In addition, the previously overall best strategy now ranks last among the more complex strategies. While the cluster2 strategy is ranked second overall, it is not clearly discernible which strategy should be preferred in the long-only cases. Even the simpler mean F-rank strategy, which delivers positive results in Germany throughout, now generates returns of -2.0% (-1.0%) for the median H (H*) long-only portfolios. Despite the fact that also all of the mean long-only returns in the high (H) fquintile are positive, they are less pronounced at the same time.

In conclusion, it can be said that market participants would be well advised to consider the weighted DMSFE strategy as well as the original F-Score in Germany. For the UK, investors should prefer the F-Score approach in view of the inconsistent return patterns of the alternatives. As stated earlier, it is noteworthy that the relatively simpler F-Score method is able to keep pace with the more complex strategies or even outperform them. Its easier implementation makes it a very accessible and appealing method for investors. However, as test results in table 6.13 show, the differences between the German and UK mean returns are mostly not statistically significant. This table is related to table 6.12 above and is the outcome of comparing long-only portfolios in the one-way environment between the two stock markets. The only significant differences appear once the F-Score and both variants of the cluster approach are employed. Further, they are only relevant with regard to the low fquintile (F-Score) and the lowest fquintiles ($k = 2$ and $k = 3$) whereby the German market performs better in all three scenarios.

Table 6.13 Test results of return differences between Germany and the UK

A	***F-Score, long-only (mean)***			
Country	L*	L	H	H*
DE	-0.079	-0.051	0.064	0.062
UK	-0.142	-0.104	0.054	0.078
Diff.	.064	.053***	.010	-.016
t-test	*(2.14)*	*(3.01)*	*(0.78)*	*(-0.74)*
B	***Cluster2 (k = 2), long-only (mean)***			
Country	L*	L	H	H*
DE	-0.062	-0.055	0.050	0.065
UK	-0.149	-0.054	0.034	0.036
Diff.	.087**	-.002	.016	.029
t-test	*(2.48)*	*(-0.13)*	*(1.28)*	*(1.07)*
C	***Cluster3 (k = 3), long-only (mean)***			
Country	L*	L	H	H*
DE	-0.058	-0.051	0.040	0.061
UK	-0.127	-0.054	0.037	0.043
Diff.	.069*	.004	.004	.018
t-test	*(1.92)*	*(0.23)*	*(0.29)*	*(0.76)*
D	***wDMSFE, long-only (mean)***			
Country	L*	L	H	H*
DE	0.075	0.045	-0.045	-0.090
UK	0.041	0.030	-0.064	-0.136
Diff.	.034	.015	.019	.046
t-test	*(1.25)*	*(1.15)*	*(1.27)*	*(1.33)*
E	***mF-rank, long-only (mean)***			
Country	L*	L	H	H*
DE	-0.081	-0.034	0.040	0.067
UK	-0.129	-0.058	0.031	0.048
Diff.	.049	.023	.009	.019
t-test	*(1.42)*	*(1.53)*	*(0.68)*	*(0.72)*
F	***mdF-rank, long-only (mean)***			
Country	L*	L	H	H*
DE	-0.093	-0.058	0.032	0.057
UK	-0.147	-0.062	0.029	0.012
Diff.	.054	.004	.003	.045
t-test	*(1.53)*	*(0.23)*	*(0.20)*	*(1.44)*

These results generally contradict the findings of Liang and Wei (2012), as mentioned earlier, who found that stock returns are higher in Germany than in the UK. However, better performance of the German stock market might still have ramifications on investor portfolios because they appear to be economically significant.

6.6 Empirical results in the two-way environment

The remaining part of this chapter deals with the following two topics. On the one hand, the two-way results of the German stock market are presented as for the UK in the preceding two empirical chapters. Thereby, both the book-to-market ratio (BM) and the firm size serve as the second dimension. After a brief summary of the results in table format, the main focus lies on the comparison with the respective UK findings. On the other hand, the results are discussed against the background of the literature with a special emphasis on the differing legal systems in Germany and the UK. Further, links are established with regard to the characteristics of the German economy, which contains a larger amount of middle-sized companies known as “Mittelstand”. It should be noted that, although the non-winsorised portfolio returns are presented for the F-Score strategy in the subsequent tables, only the winsorised returns are discussed. The reason for this decision is succinctness and because the trends of those returns correspond to each other.

6.6.1 The F-Score and the book-to-market ratio

To begin with, table 6.13 shows the results of the original F-Score investment strategy dependent on a firm’s BM ratio. Panels A and B show the average and median portfolio returns, respectively. As can be seen from both panels, the results of the horizontal return differentials are at least significant at the 10% level and confirm the underlying assumption of Piotroski (2000) that higher F-Score firms have higher stock returns.

Vertically, panel A reveals that firms with a high BM ratio perform best, except for the portfolio containing the highest F-Score stocks. For the median portfolio returns, only the lowest- and low-ranked portfolios confirm studies that find a positive relationship between stock returns and high BM ratios (e.g. Chen, 2011).

The corresponding UK results can be found in table 4.12. It is apparent that the original strategy produces mostly larger return differentials both horizontally and vertically, accompanied by higher statistical significance. However, the overall results are generally confirmed by the German data. This, in turn, validates the better performance of high-BM stocks on the one side and asserts that the F-Score approach works in a code law environment on the other side. In both countries, but especially in Germany, horizontal hedge portfolios are driven by low-BM firms.

Table 6.14 Two-way F-Score portfolios

A	*F-Score (mean returns)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.147	-0.130	0.032	0.028	.162*** *(5.18)*	.175*** *(3.28)*	*(7.76)****	*(6.86)****	-	-
H	-0.004	0.035	0.126	0.098	.091 *(1.59)*	.102 *(1.48)*	*(5.11)****	*(4.43)****	-	-
H–L	.143**	.165***	.094**	.070						
t-test	*(2.14)*	*(3.14)*	*(2.54)*	*(1.43)*						
MWW	*(3.03)****	*(3.85)****	*(1.67)**	*(0.33)*						
winsd.										
L	-0.146	-0.125	0.040	0.037	.165*** *(5.55)*	.183*** *(3.72)*	*(7.77)****	*(6.87)****	-	-
H	-0.014	0.004	0.113	0.093	.109*** *(2.96)*	.107* *(1.87)*	*(5.12)****	*(4.44)****	-	-
H–L	.132**	.129***	.073**	.056						
t-test	*(2.23)*	*(3.59)*	*(2.41)*	*(1.43)*						
MWW	*(3.01)****	*(3.85)****	*(1.66)**	*(0.30)*						
B	*F-Score (median returns)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.278	-0.209	-0.004	0.003	-	-	-	-	.205*** *(44.38)*	.281*** *(40.60)*
H	-0.159	-0.116	0.013	0.006	-	-	-	-	.129*** *(14.25)*	.165*** *(21.78)*
H–L	.119*** *(9.79)*	.093*** *(11.06)*	.017 *(0.30)*	.003 *(-0.04)*						
winsd.										
L	-0.260	-0.196	0.010	0.017	-	-	-	-	.206*** *(47.01)*	.277*** *(40.60)*
H	-0.154	-0.107	0.025	0.025	-	-	-	-	.132*** *(15.16)*	.179*** *(21.78)*
H–L	.106*** *(9.79)*	.089*** *(10.29)*	.015 *(0.30)*	.008 *(-0.04)*						

6.6.2 The F-rank and the book-to-market ratio

Unlike in chapter four for the UK, this section combines the exposition of the results after applying the mean and median F-rank strategies to the German data set. To keep the tables clear, the results presented in all of the remaining tables in this chapter are based on the winsorised returns only, meaning that the presentation of the non-winsorised returns is waived.

The respective equivalents to table 6.14 below are tables 4.13 and 4.14. Each panel represents one strategy for which the upper (lower) part contains the mean (median) returns indicated by the respective statistical test results in the last three columns. According to the horizontal parametric and non-parametric test outcomes, both strategies appear to work well in the

German stock market. By contrast, this is also the case regarding the BM terciles, at least for the mean portfolio returns of the two methods. For instance, the return differential of the high (H) mean (median) F-rank fquintile amounts to 7.6 (7.9) percentage points and is significant at the 5% level according to the t-test.

Table 6.15 Two-way mean/median F-rank portfolios

A	***F-rank (mean)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.124	-0.106	0.018	0.028	.124*** *(4.38)*	.152*** *(3.02)*	*(6.33)****	*(5.20)****	-	-
H	-0.046	0.032	0.094	0.142	.062* *(1.68)*	.188** *(2.55)*	*(3.31)****	*(3.91)****	-	-
H–L	.078	.138***	.076**	.114**						
t-test	*(1.20)*	*(4.05)*	*(2.47)*	*(2.01)*						
MWW	*(1.23)*	*(4.12)****	*(1.87)**	*(1.33)*						
L	-0.240	-0.170	-0.008	-0.020	-	-	-	-	.162*** (29.61)	.220*** (22.08)
H	-0.183	-0.057	0.025	0.033	-	-	-	-	.082*** (7.34)	.216*** (18.56)
H–L	.057 *(2.06)*	.113*** *(8.12)*	.033 *(1.58)*	.053 *(0.97)*						
B	***F-rank (median)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.151	-0.126	0.003	0.042	.129*** *(4.52)*	.193*** *(3.52)*	*(6.24)****	*(5.02)****	-	-
H	-0.021	-0.011	0.081	0.110	.092** *(2.51)*	.131* *(1.76)*	*(3.91)****	*(2.78)***	-	-
H–L	.130**	.115***	.078**	.068						
t-test	*(2.04)*	*(3.42)*	*(2.50)*	*(1.04)*						
MWW	*(2.20)***	*(3.48)****	*(1.73)**	*(1.33)*						
L	-0.257	-0.195	-0.042	-0.039	-	-	-	-	.153*** (27.61)	.218*** (23.21)
H	-0.177	-0.114	-0.010	0.035	-	-	-	-	.104*** (18.91)	.212*** (12.80)
H–L	.080** *(4.13)*	.081*** *(7.40)*	.032 *(0.46)*	.074 *(0.91)*						

6.7 Two-way alternative combination measures: An overview

The following two subsections comprise two result tables for the DMSFE and cluster methods. This structure is maintained for the remainder of the chapter and also applies to the other two second dimensions, that is, firm size and information uncertainty, which are presented in sections 6.7 and 6.8.

6.7.1 Two-way alternative combination measures: The weighted DMSFE

A brief overview of the findings suggests that investors are likely to profit only partly from the weighted DMSFE strategy in Germany. As can be seen from the horizontal return differentials in both panel A and panel B of table 6.15, significant parametric results are only achievable amongst the low-BM firms. For instance, a long/short strategy applied to the H*–L* portfolios generates a return of 17.3% (panel A.1), which is significant at the 1% level. This stands in opposition to the MWW test results, which are highly significant throughout. With respect to the median differences, the results for the high-BM fquintile are weak as well. Panel A.2 documents a differential of 5.8% between the H and the L fquintile with insignificant test statistics of 2.27. Therefore, compared with the UK (table 4.15), this strategy generates inferior returns. However, vertical differentials produce solid returns, as shown in panel A.1. With the exception of the L* fquintile, all the other fquintiles are significant. Thereby, the highest return can be achieved between high-BM and low-BM firms within the H fquintile, with a significant 15.6% annually.

In this scenario, the picture is reversed in comparison with the UK. Investors enjoy higher returns throughout. This could be regarded as an argument against the initial assumption that financial statements are not the prime method of mitigating information asymmetries in bank-oriented economies. However, while the median vertical H and H* differentials in panel A.2 are higher and equally significant as those in the UK, they become insignificant once the focus shifts to the L and L* fquintiles. For completeness, the results of the winsorised weighted method were added in panels B.1 and B.2 but are analogous to those described before.

Table 6.16 Two-way weighted DMSFE portfolios

A.1	*F-rank (weighted)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	0.052	0.027	-0.085	-0.121	-.112*** *(-3.88)*	-.173*** *(-3.36)*	*(-5.75)****	*(-5.11)****	-	-
H	0.111	0.100	0.071	0.030	-.026 *(-0.72)*	-.081 *(-1.14)*	*(-2.54)****	*(-2.31)****	-	-
H–L	.059	.073**	.156***	.151**						
t-test	*(1.05)*	*(2.37)*	*(4.45)*	*(2.35)*						
MWW	*(0.48)*	*(1.93)**	*(4.63)****	*(2.74)****						
A.2	*F-rank (weighted)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.020	-0.006	-0.150	-0.248	-	-	-	-	-.144*** *(20.74)*	-.228*** *(19.57)*
H	0.025	0.033	-0.025	-0.077	-	-	-	-	-.058 *(2.27)*	-.102** *(4.75)*
H–L	.045 *(0.35)*	.039 *(1.47)*	.125*** *(9.82)*	0.171*** *(9.76)*						
B.1	*F-rank (winsorised weighted)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	0.049	0.027	-0.086	-0.122	-.113*** *(-3.93)*	-.171*** *(-3.32)*	*(-5.79)****	*(-5.08)****	-	-
H	0.111	0.100	0.071	0.030	-.029 *(-0.71)*	-.081 *(-1.14)*	*(-2.54)***	*(-2.31)*	-	-
H–L	.062	.073**	.157***	.152**						
t-test	*(1.10)*	*(2.36)*	*(4.48)*	*(2.37)*						
MWW	*(0.77)*	*(1.91)**	*(4.67)****	*(2.77)****						
B.2	*F-rank (winsorised weighted)*				*t-test*		*MWW*		*median*	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.022	-0.006	-0.153	-0.257	-	-	-	-	-.147*** *(21.02)*	-.235*** *(18.29)*
H	0.025	0.032	-0.025	-0.077	-	-	-	-	-.057 *(2.37)*	-.102** *(4.75)*
H–L	.047 *(0.40)*	.038 *(1.39)*	.128*** *(9.61)*	.180*** *(11.06)*						

6.7.2 Two-way alternative combination measures: Clustering

Conclusively, this paragraph utilises the BM ratio as the second dimension in the construction of the cluster fquintiles. For this purpose, table 6.16 illustrates the findings after the application of the $k = 2$ and $k = 3$ cluster methods.

Using one of those methods greatly improves the results in comparison with the DMSFE approach in the previous section. This holds true for both the horizontal and the vertical test statistics, which are now significant in the majority of cases. Particularly good results are obtained when the nine Fi-ranks are assigned to three clusters ($k = 3$), as shown in panel B.

A comparison between the two stock markets shows that typically the cluster strategies are more successful in the UK both statistically and in return terms, at least in the case of the mean portfolio returns. Regarding the median returns, not much difference is identifiable between the two. A similarity between the two markets is that the $k = 3$ clusters also perform relatively better in Germany. The subdivision of the Fi-ranks into $k = 3$ rather than $k = 2$ clusters requires less effort than the potential return benefits for investors. Consequently, this strategy is advisable both for the German and for the UK stock markets.

Table 6.17 Two-way clustered portfolios

A.1	***F-rank (c-score, k = 2)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.133	-0.114	0.030	0.042	.144*** *(5.16)*	.175*** *(3.42)*	*(7.48)****	*(5.70)****	-	-
H	0.017	0.000	0.094	0.086	.094*** *(2.71)*	.069 *(0.95)*	*(4.69)****	*(2.35)***	-	-
H–L	.150**	.114***	.063**	.044						
t-test	*(2.33)*	*(3.52)*	*(2.17)*	*(0.79)*						
MWW	*(3.14)****	*(4.26)****	*(2.23)***	*(1.15)*						
A.2	***F-rank (c-score, k = 2)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.257	-0.195	-0.004	-0.025	-	-	-	-	.191*** *(43.98)*	.232*** *(27.83)*
H	-0.077	-0.065	0.037	0.025	-	-	-	-	.102*** *(14.69)*	.102* *(3.01)*
H–L	.180*** *(15.22)*	.130*** *(15.71)*	.041 *(2.68)*	0.050 *(1.08)*						
B.1	***F-rank (c-score, k = 3)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.136	-0.111	0.011	0.049	.122*** *(4.45)*	.185*** *(3.91)*	*(6.47)****	*(5.98)****	-	-
H	0.072	0.016	0.097	0.083	.081** *(2.10)*	.011 *(0.13)*	*(4.84)****	*(2.45)***	-	-
H–L	.208***	.127***	.086***	.034						
t-test	*(3.14)*	*(3.76)*	*(2.82)*	*(0.63)*						
MWW	*(3.07)****	*(3.59)****	*(2.96)****	*(0.75)*						
B.2	***F-rank (c-score, k = 3)***				***t-test***		***MWW***		***median***	
BM	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.239	-0.189	-0.025	0.012	-	-	-	-	.164*** *(33.31)*	.251*** *(32.54)*
H	-0.114	-0.094	0.035	0.024	-	-	-	-	.129*** *(18.83)*	.138** *(7.51)*
H–L	.125*** *(7.73)*	.095*** *(9.47)*	.060** *(4.24)*	0.012 *(0.19)*						

6.7.3 Summary and comparison of the results

Similar to the one-way presentation of the F-Score and its alternative combination methods, table 6.17 summarises and contrasts the two-way methods with regard to the book-to-market ratio in Germany and the UK. First of all, the horizontal return differences are evaluated on statistical and percentage grounds and ranked in the column labelled “overall”. The absolute one-year returns are presented for the mean and median returns in the subsequent four columns.

From the table it is apparent that the original F-Score is, with one exception, ranked first both in Germany and in the UK in the two-way environment, closely followed by both of the cluster methods. This corresponds mostly to the one-way findings, which were presented earlier in table 6.12. Further, it becomes apparent that the original F-Score should be the preferred choice of long-only investors in both the UK and Germany. The overall ranking was established by taking into account i) the cumulative percentage mean and median returns of the H and H* fquintiles, ii) their cumulative return differentials as well as iii) their cumulative t-statistics. The usually higher absolute one-year returns in Germany compared with the UK are remarkable. For instance, the high (H) fquintile of the F-Score strategy returns on average 15.3% per annum compared with 8.6% in the UK, which is nearly twice as much. Those higher returns are even more pronounced than in the one-way scenario.

Table 6.18 Comparison of two-way strategies in Germany and the UK (horizontally)

A	***BM***	**Germany (two-way)**			
Rank	***Overall***	***Long-only (mean)***		***Long-only (median)***	
		H	H*	H	H*
1	F-Score	F-Score (15.3%)	mF-rank (17.0%)	F-Score (3.5%)	F-Score (4.2%)
2	mF-rank	wDMSFE (12.7%)	wDMSFE (16.3%)	cluster2 (3.3%)	cluster3 (3.6%)
3	cluster3	cluster2 (12.4%)	mdF-rank (15.2%)	wDMSFE (2.7%)	mF-rank (1.3%)
4	cluster2	mF-rank (11.2%)	cluster3 (13.2%)	mF-rank (1.7%)	wDMSFE (0.0%)
5	mdF-rank	cluster3 (10.8%)	F-Score (13.0%)	cluster3 (1.0%)	cluster2 (0.0%)
6	wDMSFE	mdF-rank (8.4%)	cluster2 (12.8%)	mdF-rank (-5.2%)	mdF-rank (-0.4%)
B	***BM***	**UK (two-way)**			
Rank	***Overall***	***Long-only (mean)***		***Long-only (median)***	
		H	H*	H	H*
1	F-Score	F-Score (8.6%)	F-Score (12.9%)	F-Score (2.1%)	F-Score (6.4%)
2	cluster2	cluster3 (7.7%)	mF-rank (7.8%)	cluster3 (0.1%)	cluster3 (-0.5%)
3	cluster3	cluster2 (6.9%)	wDMSFE (5.2%)	wDMSFE (-0.0%)	cluster2 (-4.2%)
4	wDMSFE	mdF-rank (5.3%)	cluster3 (5.1%)	cluster2 (-0.6%)	mF-rank (-4.2%)
5	mF-rank	mF-rank (4.8%)	cluster2 (3.4%)	mF-rank (-4.2%)	wDMSFE (-4.4%)
6	mdF-rank	wDMSFE (4.2%)	mdF-rank (-2.1%)	mdF-rank (-5.4%)	mdF-rank (-22.0%)

Moreover, the amount of successful strategies is higher in Germany with regard to the median returns. For instance, both *k = 2* cluster strategies generate negative returns of -0.6% (H) and -4.2% (H*), respectively, in the UK, while they are positive in Germany. The same applies to the weighted DMSFE approach. Therefore, the margin of error is lower in the UK.

Finally, investors should also be interested in the performance of the strategies within each fquintile, that is, vertically. For this reason, table 6.18 provides a succinct overview.

Table 6.19 Comparison of two-way strategies in Germany and the UK (vertically)

Rank	***Germany***	***Rank***	***United Kingdom***
1	cluster3	1	mF-rank
2	cluster2	2	cluster3
2	wDMSFE	3	cluster2
4	F-Score	4	F-Score
5	mF-rank	5	wDMSFE
5	mdF-rank	6	mdF-rank

Similarly to the horizontal results, the table was constructed by evaluating the cumulative mean and median return differences for the H and H* portfolios and their statistical significance. Interestingly, the original F-Score approach is ranked fourth in this scenario in both the UK and Germany. The results could be interpreted as follows. In the UK, financial statements contain more information than can be captured by the bivariate nature of the F-Score. For this reason, the mean F-rank approach as a continuous indicator seems more suitable. However, the two measures are similar insofar as they weight information equally.

This is opposed to the German market, where the more complex strategies are ranked highest. A reason for this might be the bank-oriented characteristic of the economy. As insiders have already extracted some of the material information, the removal of noise appears to be a better strategy in generating returns. In general, therefore, the extension of Piotroski's (2000) strategy from only high-BM firms to the entire universe of BM firms has failed to have the desired effect. Value investors, who speculate on the outperformance of high-BM firms, should attach priority to the alternative strategies.

6.8 Size as an information uncertainty proxy: Overview

The next section will cover the investment strategies with regard to the firm size terciles. Recall that this is necessary to identify an underlying size effect, which is proxying for information uncertainty, in the stock market first. On this basis, it would then be possible to evaluate whether portfolios consisting of high information uncertainty stocks do actually perform better (worse) following good (bad) news, as proposed by Zhang (2006). As mentioned before, size was not found to be a particularly useful substitute for information uncertainty (e.g. Zhang, 2006; Duong et al., 2012). Further, this section carries out a cross-check with the UK data to determine whether similar results are obtained for the German stock market.

6.8.1 Size, the F-Score/F-rank and combinations of the F-rank

As with the presentation of the two-way results for the BM, this section highlights the main findings for firm size and the F-Score strategy as well as its variant derivatives. A more thorough discussion is provided at the end of this chapter. As usual, table 6.19 contains the F-Score and mean/median F-rank results. While the former strategy is able to distinguish between the outperforming and the underperforming stocks (horizontally), the vertical t-test shows no indication of the existence of a size effect according to panel A.1. The respective non-parametric results are likewise weak and only significant for the low and lowest F-Score fquintiles.

Equally, both of the F-rank combination methods deliver strong results horizontally, although they are slightly inferior to the F-Score in general. In terms of the size effect, the t-test results confirm the findings from the F-Score strategy. However, in the non-parametric test environment, all the portfolio differentials in panel B and three out of the four in panel C are significant, which actually speaks for a size effect.

Table 6.20 Two-way F-Score/F-rank-size portfolios

A.1	***F-Score (mean returns)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.057	-0.037	0.087	0.073	.124*** *(3.22)*	.130** *(2.35)*	*(5.86)****	*(5.79)****	-	-
L	-0.053	-0.032	0.056	0.030	.088*** *(3.57)*	.083* *(1.73)*	*(4.71)****	*(3.38)****	-	-
L–S	.004	.005	-.031	-.043						
t-test	*(0.06)*	*(0.14)*	*(-1.12)*	*(-1.29)*						
MWW	*(1.16)*	*(2.52)***	*(0.98)*	*(-0.93)*						
A.2	***F-Score (median returns)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.212	-0.168	-0.001	0.025	-	-	-	-	.167*** *(32.42)*	.237*** *(30.20)*
L	-0.127	-0.068	0.018	-0.007	-	-	-	-	.086*** *(6.96)*	.120*** *(7.35)*
L–S	.085** *(4.14)*	.100*** *(9.55)*	.017 *(0.20)*	-0.032 *(-0.35)*						
B.1	***F-rank (mean)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.048	-0.011	0.026	0.019	.037 *(1.04)*	.067 *(1.11)*	*(3.19)****	*(3.39)****	-	-
L	-0.035	-0.020	0.037	0.088	.057** *(2.47)*	.123** *(2.33)*	*(2.69)****	*(2.75)****	-	-
L–S	.013	-.009	.011	.069						
t-test	*(0.17)*	*(-0.24)*	*(0.38)*	*(1.37)*						
MWW	*(1.95)**	*(3.13)****	*(2.43)***	*(2.25)***						
B.2	***F-rank (mean)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.227	-0.128	-0.041	-0.019	-	-	-	-	.087*** *(9.64)*	.208*** *(15.64)*
L	-0.096	-0.042	0.009	0.024	-	-	-	-	.051** *(5.99)*	.120** *(4.51)*
L–S	.131*** *(11.81)*	.086*** *(7.74)*	.050** *(6.35)*	0.043** *(3.19)*						
C.1	***F-rank (median)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.038	-0.036	0.015	0.028	.051 *(1.52)*	.066 *(1.16)*	*(3.57)****	*(2.93)****	-	-
L	-0.170	-0.056	0.036	0.090	.092*** *(3.49)*	.260*** *(3.64)*	*(3.96)****	*(4.08)***	-	-
L–S	-.132	-.020	.021	.062						
t-test	*(-1.40)*	*(-0.52)*	*(0.72)*	*(0.99)*						
MWW	*(-0.88)*	*(2.15)***	*(2.92)****	*(1.98)***						
C.2	***F-rank (median)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.208	-0.151	-0.059	-0.026	-	-	-	-	.092*** *(12.27)*	.182*** *(12.39)*
L	-0.158	-0.079	0.002	0.031	-	-	-	-	.081*** *(9.12)*	.189*** *(10.73)*
L–S	.050 *(1.09)*	.072** *(5.70)*	.061** *(5.97)*	0.057** *(4.32)*						

Overall, therefore, the findings are twofold. On the one hand, the F-Score strategy specifically and its alternatives in the majority of cases prove once more that they are able to isolate winning stocks from the entire stock universe in Germany. On the other hand, the success of these strategies seems to be independent of a firm’s size given the inconclusive vertical test statistics.

6.8.2 Size and combinations of DMSFE

The following table provides an overview of the results of the weighted DMSFE combination methods. Its respective counterpart for the UK stock market is table 5.2 in the previous chapter. Here, too, the median portfolio returns are omitted because they very much resemble the mean version of those results. Again, the expected sign regarding the horizontal differentials is negative, as expected, because lower forecast errors are assumed to be accompanied by higher returns.

Table 6.21 Two-way weighted DMSFE portfolios

A	***F-rank (weighted), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	0.037	0.046	0.019	-0.004	-.027 *(-0.75)*	-.041 *(-0.66)*	*(-3.20)****	*(-2.92)****	-	-
B	0.110	0.049	-0.013	-0.031	-.062** *(-2.47)*	-.141** *(-2.28)*	*(-2.60)****	*(-2.43)***	-	-
B–S	.073	.003	-.032	-.027						
t-test	*(1.34)*	*(0.10)*	*(-0.88)*	*(-0.32)*						
MWW	*(2.10)***	*(1.68)**	*(2.30)***	*(1.09)*						
B	***F-rank (winsd weighted), mean returns***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	0.039	0.046	0.020	-0.005	-.026 *(-0.71)*	-.044 *(-0.72)*	*(-3.19)****	*(-2.96)****	-	-
B	0.110	0.050	-0.014	-0.039	-.064** *(-2.55)*	-.149** *(-2.41)*	*(-2.70)****	*(-2.57)***	-	-
B–S	.071	.004	-.034	-.034						
t-test	*(1.30)*	*(0.16)*	*(-0.93)*	*(-0.41)*						
MWW	*(2.03)***	*(1.74)**	*(2.26)***	*(0.98)*						

From table 6.20 above it becomes obvious that the horizontal return differences are weak, especially in the small-firm tercile. Although firms with low forecast errors outperform, the differences of 2.7% (H) and 4.1% (H*), respectively, are insignificant as a result of the t-test. This was not the case in the UK analysis, as the results were significant at least at the 10% level as well as more robust considering the MWW test outcomes. Regarding the vertical return differences, the situation improves slightly, although this is exclusively attributable to the significant MWW test statistics throughout apart from the highest (H*) fquintile, which is insignificant. The results of the t-test remain very weak. Thus, it is doubtable that a size effect is existent in the German stock market, at least as far the findings of the weighted DMSFE approach are concerned. It therefore appears to be justifiable to state that this approach should not be used by investors. This is especially true if they intend to categorise firms by size (vertically) and are not solely interested in the differentiation of prospective future winners and losers (horizontally).

6.8.3 Size and cluster combinations

The presentation of the cluster method results concludes this section on firm size, which was used as a proxy for information uncertainty. Thereby, table 6.21 below complements the similar analysis in the UK stock market, the findings of which were highlighted in table 5.3.

Table 6.22 Two-way clustered-size portfolios

A.1	***F-rank (c-score, k = 2)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.033	-0.032	0.057	0.054	.089*** (2.63)	.087 (1.41)	(5.15)***	(3.20)***	-	-
L	-0.102	-0.055	0.035	0.070	.090*** (3.52)	.172*** (2.89)	(3.41)***	(3.02)***	-	-
L–S	-.069	-.023	-.022	.016						
t-test	(-0.70)	(-0.59)	(-0.85)	(0.31)						
MWW	(0.31)	(1.84)*	(0.26)	(1.75)*						
A.2	***F-rank (c-score, k = 2)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.187	-0.141	-0.009	-0.054	-	-	-	-	.132*** (28.38)	.133*** (10.17)
L	-0.079	-0.060	0.006	0.026	-	-	-	-	.066*** (7.45)	.105** (3.96)
L–S	.108 (1.60)	.081** (6.28)	.015 (0.19)	0.080** (4.58)						
B.1	***F-rank (c-score, k = 3)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	0.006	-0.029	0.046	0.065	.075** (2.12)	.059 (0.92)	(4.85)***	(4.00)***	-	-
L	-0.151	-0.064	0.034	0.048	.098*** (3.65)	.199*** (3.34)	(4.21)***	(3.40)***	-	-
L–S	-.156	-.035	-.012	-.017						
t-test	(-1.51)	(-0.90)	(-0.42)	(-0.37)						
MWW	(-0.47)	(1.50)	(0.60)	(-0.31)						
B.2	***F-rank (c-score, k = 3)***				***t-test***		***MWW***		***median***	
Size	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
S	-0.158	-0.146	-0.017	0.011	-	-	-	-	.129*** (33.64)	.169*** (17.44)
L	-0.131	-0.103	0.007	0.016	-	-	-	-	.110*** (13.06)	.147*** (8.87)
L–S	.027 (-0.08)	.043 (4.75)	.024 (0.30)	0.005 (-0.09)						

Based on the test outcomes in the horizontal direction, both cluster methods deliver solid return differentials, which increase the likelihood of consistent positive returns for investors in a long/short scenario. Likewise, positive returns can be expected in absolute terms at least for the average portfolio returns shown in panels A.1 and B.1. Apart from a 7.0% (3.5%) return in the highest (high) c-score fquintile in panel A.1, the highest c-score fquintiles outperform their high counterpart as expected. Again with one exception in panel A.2, this is also the case for all the portfolios for which the median returns were calculated.

Turning the attention to the return differences between large and small firms reveals a rather mixed picture. It is clear at first glance that, if at all, a size effect on the German stock market can only be proven by using the $k = 2$ cluster approach. Further, as panels A.1 and A.2 suggest, the effect is most likely to be found within the L and H* c-score fquintiles. However, this finding is partly inconsistent with the UK data set, in which the L and H fquintiles documented this effect. No size effect at all is detectable once the $k = 3$ method is employed. Apart from that, the analysis of the data allows the same conclusion as for the UK. Firstly, clustering works on the German stock exchanges irrespective of firm size. Secondly, the findings do not provide clear-cut proof of the existence of a size effect. Thirdly and lastly, if a size effect does indeed exist, large firms outperform their smaller competitors in general. As a result of the last outcome, the information uncertainty hypothesis of Zhang (2006) remains unconfirmed.

6.9 Alternative IU measures: The SDF and MDF

In the following section, firm size as a measure of information uncertainty is replaced by the standard deviation of the F-components/Fi-ranks (SDF) as well as their mean absolute deviation (MADF). The theoretical underpinnings of this method can be referred to in section 5.4. As before, the results are presented in table format; the equivalent UK findings can be found in table 5.4 (SDF) and table 5.5 (MADF), respectively. It should be noted that both the DMSFE and the cluster method were omitted deliberately. The reason for this is based on the nature of the SDF/MADF. As a more direct measure of information uncertainty, the focus shifts from determining future winners by reducing the forecast errors to the impact on portfolio returns caused by the volatility of fundamental data. In other words, the analysis is more concerned with the status quo than with finding the optimal investment tool.

6.9.1 The SDF and the German stock market

This subsection features the SDF as a second dimension and continues with the presentation of the respective results for Germany in table 6.22.

Table 6.23 Two-way SDF portfolios

A.1	*F-Score*				*t-test*		*MWW*		*median*	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.118	-0.032	0.061	0.026	.093*** *(3.55)*	.144*** *(3.70)*	*(4.34)****	*(4.88)****	-	-
H	-0.059	-0.069	0.083	0.144	.152*** *(3.28)*	.203** *(2.11)*	*(5.05)****	*(4.71)***	-	-
H–L	.059	-.037	.022	.118**						
t-test	*(0.92)*	*(-0.98)*	*(0.68)*	*(2.43)*						
MWW	*(-0.05)*	*(-3.22)****	*(-1.34)*	*(2.22)***						
A.2	*F-Score*				*t-test*		*MWW*		*median*	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.164	-0.063	0.015	-0.007	-	-	-	-	.078*** *(8.15)*	.157*** *(13.75)*
H	-0.198	-0.177	-0.009	0.066	-	-	-	-	.168*** *(23.62)*	.264*** *(36.70)*
H–L	-.034 *(-1.17)*	-.114*** *(-12.73)*	-.024 *(-0.40)*	0.073 *(3.00)**						
B.1	*F-rank (mean)*				*t-test*		*MWW*		*median*	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.140	-0.019	0.052	0.078	.071*** *(2.93)*	.218*** *(4.49)*	*(3.96)****	*(5.71)****	-	-
H	-0.020	-0.043	0.014	0.023	.057 *(1.49)*	.043 *(0.56)*	*(3.01)****	*(2.38)***	-	-
H–L	.012*	-.024	-.038	-.055						
t-test	*(1.71)*	*(-0.71)*	*(-1.31)*	*(-1.00)*						
MWW	*(0.08)*	*(-3.37)****	*(-3.99)****	*(-2.42)***						
B.2	*F-rank (mean)*				*t-test*		*MWW*		*median*	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.153	-0.042	0.016	0.016	-	-	-	-	.058*** *(11.33)*	.169*** *(21.94)*
H	-0.227	-0.150	-0.054	-0.071	-	-	-	-	.096*** *(9.17)*	.156*** *(12.59)*
H–L	-.074 *(-2.39)*	-.108*** *(-11.98)*	-.070*** *(-7.29)*	-0.087 *(-2.36)*						
C.1	*F-rank (median)*				*t-test*		*MWW*		*median*	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.255	-0.051	0.056	0.112	.107*** *(3.50)*	.367*** *(5.37)*	*(4.75)****	*(5.84)****	-	-
H	-0.051	-0.065	-0.003	0.018	.062* *(1.93)*	.069 *(1.28)*	*(3.33)****	*(2.93)***	-	-
H–L	.204***	-.014	-.059*	-.094						
t-test	*(2.47)*	*(-0.39)*	*(-1.95)*	*(-1.39)*						
MWW	*(2.27)***	*(-2.08)***	*(-4.71)****	*(-3.32)****						
C.2	*F-rank (median)*				*t-test*		*MWW*		*median*	
SDF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.271	-0.090	0.016	0.040	-	-	-	-	.106*** *(14.23)*	.311*** *(23.08)*
H	-0.201	-0.159	-0.082	-0.055	-	-	-	-	.077*** *(9.34)*	.146*** *(11.11)*
H–L	.070 *(0.92)*	-.069** *(-4.81)*	-.098*** *(-12.84)*	-.095** *(-6.56)*						

The evaluation of the horizontal return differences once again shows a strong performance of the F-Score strategy both in percentage and in statistical terms. This also applies in relation to the mean/median F-rank methods described in panels B and C. In particular, the mean version of the F-rank exposes some weakness within the high-SDF tercile of panel B.1. More precisely, none of the return differences of 5.7 and 4.3 percentage points of the H–L and H*–L* are significantly different from zero. All in all, however, the results appear not to be affected by the change in the second dimension and provide reasonably good results to be considered by investors.

Regarding the vertical differentials, the conclusion is less clear. It should be recalled that it is expected that a higher SDF, that is, information uncertainty, should be accompanied by lower (higher) one-year returns following bad (good) news.[21] The term "news" is reflected by the fquintiles of the three investment strategies. From the test results it can be said that there is general agreement on the poorer performance of high-uncertainty portfolios within the L fquintile, which is in accordance with Zhang's (2006) hypothesis. However, confirmation is only provided by the non-parametric tests. Regarding the high fquintile, only the two F-rank strategies provide significant results. These are opposed to what is expected, though. Hence, similar to the UK, investors in Germany likewise prefer to hold low-uncertainty portfolios even though good company news is announced. Overall, the test results are not as strong as for the UK data.

Quite unusually, the t-test of the highest (H*) F-Score fquintiles in panel A indicates better performance of the high-uncertainty stocks, as predicted by Zhang (2006). This is also corroborated by the MWW and median test outcomes. However, the comparison with the other two strategies contradicts this observation. The highest fquintile still provides better returns for low-uncertainty stocks in those two scenarios. Another abnormality is detectable in panels B.1 and C.1. The portfolios in the high-SDF tercile of the lowest (L*) fquintile actually outperform their low-tercile competitors, confirmed by significant t-test findings.

Given these mixed results, a conclusion might be deemed somewhat arbitrary. However, in the majority of cases, the stock portfolios with high information uncertainty underperform their low-uncertainty peers. Therefore, this trend should be regarded as dominant in the German stock market.

[21] The same applies to the MADF measure presented in section 6.9.2.

6.9.2 The MADF and the German stock market

In light of the mixed vertical SDF results shown previously, the alternative MADF approach could possibly provide a clearer picture. However, the first stage of the analysis is concerned with the performance of the horizontal return differences. In short, no new insights can be gained, which means that the MADF broadly agrees with the SDF findings as documented in table 6.23 below.

Stronger disparities between the two methods come to light with respect to the vertical differences. Unlike in the SDF case, high-uncertainty portfolios do not perform better in the highest F-Score fquintiles of panel A. Although the absolute return for the high-MADF (low-MADF) tercile in panel A.1 is 6.9% (2.9%), the difference is insignificant. Both the low and the high F-Score fquintiles still highlight a preference for low-uncertainty stocks. Combined with the strong horizontal results, the F-Score method as presented in panel A.1 appears to be best suited to isolating future winners and to abstaining from stocks with high information uncertainty. The results of the median F-rank approach in panel C.1, however, are not as hypothesised by Zhang (2006). In particular, the lowest (L*) F-rank fquintile contains the high-MADF tercile, which should normally underperform its low-tercile counterpart. Due to this one-time occurrence, it is likely that this might be a statistical fluke.

In general, however, the two-way MADF portfolio results indicate that that all three of the investment strategies are useful tools for generating solid investment returns in practice. Besides, they demonstrate a preference by investors for low-uncertainty portfolios, although this is mostly restricted to the low/high (L/H) fquintiles. Lastly, they corroborate the earlier findings in the SDF environment.

Table 6.24 Two-way MADF portfolios

A.1	***F-Score***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.116	-0.047	0.054	0.029	.101*** *(3.87)*	.145*** *(3.84)*	*(4.80)****	*(4.79)****	-	-
H	-0.093	-0.082	0.069	0.069	.151*** *(3.42)*	.162* *(1.76)*	*(4.40)****	*(4.01)****	-	-
H–L	.023	-.035	.015	.040						
t-test	*(0.40)*	*(-1.01)*	*(0.46)*	*(0.83)*						
MWW	*(-0.55)*	*(-2.65)****	*(-1.88)**	*(0.97)*						
A.2	***F-Score***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.164	-0.076	0.006	-0.012	-	-	-	-	.082*** *(9.82)*	.152*** *(12.63)*
H	-0.202	-0.164	-0.049	0.051	-	-	-	-	.115*** *(13.31)*	.253*** *(21.56)*
H–L	-.038 *(-2.04)*	-.088*** *(-7.14)*	-.055 *(-1.80)*	0.063 *(0.49)*						
B.1	***F-rank (mean)***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.139	-0.036	0.045	0.064	.081*** *(3.33)*	.203*** *(4.50)*	*(4.22)****	*(5.45)****	-	-
H	-0.045	-0.066	0.007	0.009	.073** *(2.03)*	.054 *(0.73)*	*(3.26)****	*(2.07)***	-	-
H–L	.094	-.030	-.038	-.055						
t-test	*(1.53)*	*(-0.94)*	*(-1.30)*	*(-1.02)*						
MWW	*(0.20)*	*(-2.99)****	*(-3.56)****	*(-1.99)***						
B.2	***F-rank (mean)***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.158	-0.060	0.011	-0.003	-	-	-	-	.071*** *(10.58)*	.155*** *(20.66)*
H	-0.211	-0.151	-0.056	-0.071	-	-	-	-	.095*** *(9.78)*	.140*** *(6.94)*
H–L	-.053 *(-1.04)*	-.091*** *(-9.19)*	-.067** *(-5.82)*	-0.068 *(-0.30)*						
C.1	***F-rank (median)***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.252	-0.068	0.046	0.087	.114*** *(3.80)*	.339*** *(5.45)*	*(4.79)****	*(5.81)****	-	-
H	-0.074	-0.080	-0.006	0.009	.074** *(2.43)*	.083 *(1.62)*	*(3.40)****	*(2.87)***	-	-
H–L	.178***	-.012	-.052*	-.078						
t-test	*(2.61)*	*(-0.36)*	*(-1.70)*	*(-1.19)*						
MWW	*(2.30)***	*(-1.68)**	*(-3.96)****	*(-2.77)***						
C.2	***F-rank (median)***				***t-test***		***MWW***		***median***	
MADF	L*	L	H	H*	H–L	H*–L*	H–L	H*–L*	H–L	H*–L*
L	-0.252	-0.100	-0.004	0.016	-	-	-	-	.096*** *(14.88)*	.268*** *(29.24)*
H	-0.201	-0.154	-0.082	-0.063	-	-	-	-	.072*** *(9.08)*	.138*** *(10.39)*
H–L	.051 *(1.98)*	-.054* *(-2.80)*	-.078*** *(-8.91)*	-.079 *(-1.90)*						

6.9.3 Summary and comparison of IU strategies in Germany

Following the structure of the two-way portfolios, in which the BM ratio was utilised as a second dimension, this part contrasts the nine investment strategies with regard to information uncertainty. Table 6.24 evaluates and ranks the performance of the horizontal return differentials, taking into consideration a firm's size and the volatility of its accounting data (SDF and MADF). At the bottom of the table, an explanation of the shortened investment method's name is provided, which is preceded by one of the three IU measures.

Table 6.25 Comparison of two-way strategies in Germany and the UK (horizontally)

A	***IU***	**Germany (two-way)**			
Rank[1]	***Overall***	***Long-only (mean)***		***Long-only (median)***	
		H	H*	H	H*
1	sdf-F	sdf-F (14.4%)	sdf-F (17.0%)	size-F (1.7%)	sdf-F (5.9%)
2	size-F	size-F (14.3%)	sdf-rank(md) (13.0%)	sdf-F (0.6%)	madf-F (3.9%)
3	madf-F	madf-F (12.3%)	size-rank(md) (11.8%)	size-rank(m) (-3.2%)	size-F (1.8%)
4	sdf-rank(md)	sdf-rank(m) (6.6%)	size-rank(m) (10.7%)	sdf-rank(m) (-3.8%)	size-rank(m) (0.5%)
5	madf-rank(md)	size-rank(m) (6.3%)	size-F (10.3%)	madf-F (-4.3%)	size-rank(md) (0.5%)
6	size-rank(md)	sdf-rank(md) (5.3%)	sdf-rank(m) (10.1%)	madf-rank(m) (-4.5%)	sdf-rank(md) (-1.5%)
7	sdf-rank(m)	madf-rank(m) (5.2%)	madf-F (9.8%)	size-rank(md) (-5.7%)	madf-rank(md) (-4.7%)
8	madf-rank(m)	size-rank(md) (5.1%)	madf-rank(md) (9.6%)	sdf-rank(md) (-6.6%)	sdf-rank(m) (-5.5%)
9	size-rank(m)	madf-rank(md) (4.0%)	madf-rank(m) (7.3%)	madf-rank(md) (-8.6%)	madf-rank(m) (-7.4%)
B	***IU***	**UK (two-way)**			
Rank	***Overall***	***Long-only (mean)***		***Long-only (median)***	
		H	H*	H	H*
1	madf-F	size-F (11.8%)	madf-F (23.1%)	size-F (2.1%)	madf-F (8.2%)
2	size-F	madf-F (10.8%)	sdf-F (16.6%)	madf-F (-1.8%)	sdf-F (5.6%)
3	sdf-F	sdf-F (8.8%)	size-F (15.3%)	sdf-F (-2.4%)	size-F (5.4%)
4	madf-rank(md)	size-rank(md) (8.1%)	size-rank(m) (13.4%)	size-rank(m) (-3.4%)	size-rank(m) (1.1%)
5	madf-rank(m)	size-rank(m) (7.4%)	sdf-rank(md) (10.8%)	size-rank(md) (-4.3%)	sdf-rank(m) (-8.4%)
6	size-rank(m)	madf-rank(md) (5.8%)	madf-rank(md) (10.1%)	madf-rank(m) (-5.8%)	madf-rank(m) (-8.9%)
7	sdf-rank(md)	madf-rank(m) (5.7%)	sdf-rank(m) (8.4%)	sdf-rank(m) (-6.2%)	madf-rank(md) (-9.9%)
8	sdf-rank(m)	sdf-rank(m) (5.0%)	madf-rank(m) (7.5%)	madf-rank(md) (-6.7%)	sdf-rank(md) (-10.9%)
9	size-rank(md)	sdf-rank(md) (4.3%)	size-rank(md) (5.3%)	sdf-rank(md) (-7.3%)	size-rank(md) (-13.8%)

[1] F: *original F-Score*; rank(m): *mean F-rank*; rank(md): *median F-rank*

In both markets, the F-Score method works best overall and therefore should be especially appealing to long/short investors. The further positions are occupied by the median F-rank approach and the mean F-rank at the bottom of the ranking in Germany. The results from the UK are not definite.

For long-only investors, the F-Score approach continues to dominate clearly in the UK irrespective of either the mean or the median construction of the portfolio returns. This is mostly true in Germany as well, although the results indicate some weakness in the H* mean and H median portfolios. In other words, by using the F-Score method, investors are most likely to be insulated from pronounced losses, regardless of the method used to measure information uncertainty. Consequently, the use of either of the alternative F-rank strategies is highly likely to result in inferior returns, although neither of the two dominates the other one.

While table 6.24 documents the performance of investment strategies, it is also of interest which of the three second-dimension measures is the best proxy for information uncertainty. This has both theoretical and practical ramifications. On the one hand, the analysis sets out whether information uncertainty is not better quantified by more direct measures compared with a firm's size. The reason is that both the SDF and the MADF measure emerge from the specific investment strategy and therefore are more customised to it. By contrast, firm size is a datum that is not immediately aligned with a firm's fundamental data.

On the other hand, practising investors should be interested in knowing which of the proxies enables them either to reduce the information uncertainty or to increase their portfolio returns. For this reason, table 6.25 provides an insight into the ability of each of those measures to generate a quantifiable gap between the high-IU and the low-IU tercile for each of the F-Score/F-rank fquintiles.

Table 6.26 Comparison of two-way strategies in Germany and the UK (vertically)

Rank	***Germany***	***United Kingdom***
1	size-rank(m)	sdf-rank(m)
2	sdf-rank(m)	sdf-rank(md)
3	size-rank(md)	madf-rank(m)
4	madf-rank(m)	size-rank(m)
5	size-F	madf-rank(md)
6	sdf-rank(md)	size-rank(md)
7	madf-F	sdf-F
8	madf-rank(md)	madf-F
9	sdf-F	size-F

The ranking suggests that firm size is indeed a viable proxy for information uncertainty, at least in the German stock market. For each of the three trading strategies, size ranks among the top five. Regarding the SDF and MADF, again the picture is mixed and neither proxy is dominated by the other. The UK data, however, reveal that the SDF might be a more suitable measure as it occupies the top two ranks, followed by the MADF. Only one strategy in combination with size can be found amongst the top five. Overall, therefore, it can be concluded that different markets require different information uncertainty proxies as well.

6.10 Chapter summary and conclusion

As the F-Score strategy has not been studied in bank oriented, code law institutional settings in general and, more specifically, in Germany before, the intention of this chapter was to narrow this research gap. In addition, it aimed to contribute further to the relatively scarce literature on the German stock market by dealing with two other main objectives. On the one hand, it tested the robustness of the various investment strategies in a different geographical location. On the other hand, those strategies were analysed in view of the different corporate culture, which is more focused on fulfilling the requirements of stakeholders than on generating shareholder value, as well as the generally lower degree of investor protection in Germany. For this reason, five testable hypotheses were developed. To improve the readability, they are restated again and discussed in the light of the research outcomes.

Hypothesis 6.1: The F-Score investment strategy and its alternatives can successfully separate future winning from losing stocks regardless of the country's legal system.

This hypothesis can generally be confirmed for the German stock market. Apart from the very weak results for the DMSFE approach shown in table 6.20, all of the remaining strategies are able to generate satisfactory investment returns in the one-way and two-way environments. The results were evaluated according to both the absolute returns of the high/highest fquintiles and the horizontal return differences. Once the one-way portfolio return differentials are inspected in isolation, that is, without any second dimension, such as size, book-to-market and so on, and similar to the original study by Piotroski (2000), the strategy clearly works better in the UK. This means that the magnitude of both the return differentials and the test statistics is greater. It can therefore be concluded that financial statements prepared under the UK-GAAP, which is similar to the US-GAAP, indeed contain more value-relevant information for investors. Although the preparation of financial statements under the IFRS has been compulsory from 2005 onwards, which supposedly should have improved the value relevance of German firm fundamentals, this assumption is not confirmed by the results. In other words, although the F-Score strategy and its alternatives generate significant return differentials, they are still less pronounced than in the UK. A reason for this was provided by Haller and Wehrfritz (2013). The authors found that, despite adopting the IFRS from 2005, the accounting policies required by the national rules are still retained, which suggests that true harmonisation of the accounting standards is still pending.

Hypothesis 6.2: The speed of information incorporation into and reaction of stock prices is expected to be faster for the German stock market than the UK stock market.

With few exemptions, the German one-year stock returns are higher than those of the UK. This can be derived from the analysis of the absolute one-year returns of both the one-way and the two-way investment strategies, in which BM is included as the second dimension. Although the return differences between the two stock markets are not significant in the one-way scenario, this outcome speaks for the assumption of asymmetric information diffusion, which implies that stock prices update much faster in Germany. A reason for this might be the unique position of banks, as suggested by Agarwal and Elston (2001). This kind of immediate access to financial information about a company is likely to be followed by insider trading activities. In fact, Dittmann et al. (2010) found rather that monitoring company management banks are inclined to pursue their own interests. In addition, Betzer and Theissen (2009) analysed insider transactions in German firms that would have been illegal in the UK and found that these trades significantly affected the stock prices, especially ahead of quarterly earnings announcements. In summary, therefore, higher stock prices are likely to be a combination of relatively lax investor protection, asymmetric information and if not illegal at least unethical trading on this information.

However, faster information processing could also result in less profitable investment strategies, since relevant information could already have been incorporated by insiders prior to the strategy being implemented by the general investment community. As this is not the case in this study, this supports the assumption of not entirely efficient markets. After insiders have built their positions, the wider investment community follows after accounting information has been officially published and fuels disequilibrium. Even if these investors are considered rational, the disequilibrium would become larger if there is no incentive to invest contrary to the ongoing trend.

Hypothesis 6.3: A value premium is present in the German stock market. Therefore, firms with higher book-to-market ratios outperform those with lower book-to-market ratios.

To test this hypothesis, the data were divided into book-to-market terciles to analyse which kinds of stocks are the key drivers of the investment strategies. The results show that those drivers are low book-to-market firms in the vast majority of cases, implying that the original study by Piotrsoki (2000), who focused on high-BM stocks only, is self-limiting. In this regard, the results in this thesis correspond to those of Duong et al. (2014), who applied the F-Score strategy to the entire stock universe of the UK. In addition, the outcomes are very similar to those of the UK.

The further analysis of the differences between the high-BM and low-BM firms within the same fquintile, that is, vertically, is generally in accordance with Piotroski and So (2012). The authors provided a behavioural view on the cause of the outperformance of high-BM stocks, which can explain the findings made in this study. For instance, while high-BM firms in the low fquintile are priced correctly, mispricing, that is, overvaluation, occurs in the same fquintile for low-BM firms. Similarly, high-BM firms in the high fquintile are undervalued, whereas low-BM firms within this fquintile are priced correctly. Of course, this applies analogously to the lowest and highest fquintiles. For this reason, the stock prices should revert and ultimately align with the reality of firm fundamentals once investors realise that mispricing exists. These findings are very similar to those of the UK.

Hypothesis 6.4: Firms with higher information uncertainty, proxied by firm size and the volatility of financial statement items, provide higher (lower) subsequent returns following good (bad) news.

This relationship, which was initially tested by Zhang (2006), holds only partly true. Similar to the UK, firms with higher information uncertainty, that is, either smaller firms or firms within higher SDF/MADF terciles, generate lower returns following bad news. However, those firms do not provide investors with higher returns following good news. Here, too, the reversal of the size effect might play a huge part. Similarly to Dimson and Marsh (1999), Schrimpf et al. (2007) found a reverse size effect in Germany for the period from 1969 to 2002. In accordance with this study, Artmann et al. (2012) found no size effect for the period 1960 to 2006 and a reverse size effect between 1997 and 2006. This is further underlined by the alternative two measures of the SDF and MADF, which point in a similar direction.

Regarding the secondary focus of this hypothesis, to analyse the key drivers of the investment strategies, no definite conclusion can be made. Unlike in the UK, where small firms are the drivers, this is not necessarily the case in the German stock market. The reason most likely concerns the fact that German listed firms are usually large.

The analysis of the German stock market data has opened potential routes for further research. For instance, while the present findings are the result of a more general approach to stock market investing, the data could be further subdivided into a group of family-run and non-family-run businesses. This process would generate an LME and a CME group of firms, the returns of which could be compared in the long run. Likewise, it would also represent a direct comparison of the success between the shareholder-oriented and the stakeholder-oriented approach.

Further research could be undertaken with regard to the stock market performance of family-run businesses in Germany and their non-family-run international competitors within the same industry. Adding different variables that represent country-specific characteristics of the respective legal system could help to determine the key drivers of stock returns.

Chapter 7 Conclusion and Limitations

7.1 Introduction

The French mathematician, physicist and philosopher Blaise Pascal was quoted in Warren Buffett's shareholder letter of 1982[22] with the following words:

"It has struck me that all men's misfortunes spring from the single cause that they are unable to stay quietly in one room."

This quote can be applied to both the financial markets and the academic discipline of finance, which is built on the two pillars of the efficient market hypothesis on the one hand and the emerging newer strand of behavioural finance on the other hand. Regarding financial markets, this quote summarises the inability of most investors to generate alpha, that is, higher returns over time than the overall market. Regarding the academic literature, it basically condenses the opposition between the EMH and the BF school of thought.

From a "natural science" perspective, it would seem justified to theorise about the possible outcomes once the investor hypothetically leaves the room, that is, takes an active part in the financial markets. In this scenario, EMH and BF adherents would be most likely to concur on the development and use of mathematical methods that would model these hypothetical outcomes. However, as the investor decides actually to leave the room, researchers from the "natural sciences" would quickly realise that the investor's market actions cannot be controlled for as reliably as before in the hypothetical test environment. This, again, coincides with the beliefs of most BF advocates and is the underlying assumption behind this research. By contrast, this would not necessarily pose a problem for EMH followers.

Because adherents to the EMH neither follow the strict requirements of the "natural sciences" nor widely accept the propositions of behavioural finance, this "physics envy" ultimately results in discrepancies between the modelled outcome and the actual outcome, which are ultimately caused by the investor's behaviour. The question is, of course, whether the degree of discrepancy between theory and reality is caused by an incomplete or even ill-conceived

[22] Source: Berkshire Hathaway shareholder letter of 1982 (http://www.berkshirehathaway.com)

model or because reality simply cannot be modelled perfectly. However, the EMH and the BF share similarities. The intention of these two strands in the literature is to understand the dynamics of financial markets to provide practitioners such as standard setters with the necessary tools for preventing financial crises. Although this aim implicates a forward-looking stance, researchers in finance are mostly limited to using historical data. This argument is used as a valid point of criticism of behavioural finance by EMH adherents. In their view, it is too short-sighted to explain the consequences of investor behaviour with the benefit of hindsight, because the explanation will always fit the data. However, this seems not to be too dissimilar from the more formal approach.

In consideration of the arguments presented above, the topic of the thesis was chosen to meet the following three requirements. Firstly, the aim was to validate a pragmatic investment approach, such as the F-Score strategy. This was achieved by applying this strategy to different geographical locations, the UK and Germany, and to different time periods. Therefore, this approach features the behavioural part of the research as its assumptions allow for variation in the results. At the same time, however, the research intended to adhere to scientific rigour, which satisfies the second requirement. This was achieved by ensuring the reproducibility of the original findings by Piotroski (2000) in the European context using similar methods. Further, the analysis was extended to alternative versions of the original strategy and the underlying rationale of this approach was guided by theory, as required by Richardson et al. (2010). Finally, amongst the possible strategies, investors were provided with the most practical one, assuring that the theory is directly linked and thus relevant to the reality of the stock market.

7.2 Research scope

In general, the research sought to discover the following five main aspects. Firstly, it aimed to test the success of an easily implementable financial health strategy, known as the F-Score, in two European stock markets. This part also served as a robustness check for the earlier findings in the US and as a basis on which to build further.

Secondly, the aim was to extend this knowledge by finding alternatives to the F-Score.[23] The search for alternatives was motivated by the binary nature of the original F-Score strategy, which effectively ignores a large amount of useful information that could help to enhance the strategy's performance. Moreover, the nine accounting signals are weighted equally under the F-Score strategy. In contrast, the alternative strategies allow the weights to fluctuate year after year, which may lead to better results. Finally, the proportion of firms in each of the portfolios was held constant each year. This is not the case for the F-Score, which might have an undue influence on the overall returns in the case of both overly positive and overly negative years. The alternatives were constructed by borrowing more advanced methods from the forecasting literature in economics to gain the possibility of improving the investment returns at the same time.

Thirdly, the potential key drivers of the original and alternative strategies were assessed by the introduction of a second dimension. This second dimension included the book-to-market ratio, firm size as a proxy for information uncertainty and liquidity. With regard to size, the analysis then introduced various alternatives[24] to achieve a better approximation of the concept of information uncertainty. The inclusion of firm size was an important factor in the analysis as there is a concentration of anomalies in small stocks. This includes the size effect (Banz, 1981), asset growth (Fama and French, 2008) and size-based timing differences (Ayers and Freeman, 2000). Regarding liquidity, another two measures were introduced that capture either the price impact dimension or the trading quantity dimension of a stock. This was motivated by the seminal paper by Amihud and Mendelsohn (1986), who found evidence of a positive relationship between average returns and stock liquidity. However, as there is no consensus amongst academics on this topic even after almost thirty years of research, the concept of liquidity in conjunction with a financial health investment strategy seemed to be particularly promising and necessary to narrow some of the present research gaps.

Fourthly, the test results were assessed against the background of different legal, accounting, country-, stock market- and firm-specific characteristics. This is of interest not only because of the generally limited research on the German stock market but also to verify whether the success of an easily implementable financial health investment strategy is reproducible in a wider context. In this regard, the German economy provides the ideal conditions for deeper

[23] These included the F-rank, DMSFE and clusters.
[24] These included the SDF and MADF as fundamentals-based IU measures for the UK and Germany and Amihud and ToR as liquidity-based IU measures for the UK.

analysis. The reason is that, comparable to both the UK and the US, it is amongst the top five contributors to the worldwide gross domestic product and is politically stable. However, the difference is that it is based on a code law legal environment and has a relatively underdeveloped stock market. Hence, if the investment strategies were as successful as in the UK and the US, this would mean that its success is more due to investors' common behavioural traits than to systemic ones. In general, the results are similar to those of the UK and are in favour of these assumptions. To wit, the respective strategies are successful irrespective of systemic differences.

Finally, the aggregate findings were meant to contribute to the literature on capital markets research and to the literature on fundamental analysis and valuation.

7.3 Research design, implementation and boundaries

As noted in the introductory section of this chapter, this research was not intended to replicate the "natural science" approach. In philosophical terms, this means that it relaxed the strictness of the test environment insofar as it acknowledged the possibility of multiple realities that emerge depending on the chosen research method. In other words, instead of assuming a direct relationship between cause and effect, a probability is rather attached to the outcome.

To conduct the analysis, the necessary secondary data were obtained from Datastream. In general terms, these included all the relevant financial statement information needed for the computation of the nine F-Score components. Moreover, stock price data were acquired, which were crucial for evaluating the performance of each of the aforementioned investment strategies.

The amount of data was intentionally limited by the geographical region and time period. Regarding the former limit, the UK stock market was chosen for the analysis because London is the home of one of the most developed and internationally relevant financial centres. Therefore, investors should be particularly interested in its success there. By contrast, Germany was chosen because of its importance to the global GDP and because it has a code law system, which became relevant later in the analysis. Regarding the time horizon, the period from 1992 to 2010 was chosen for two reasons. On the one side, the data availability turned out to be problematic before the year 1992, in particular for Germany. On the other side, a reasonable balance was aimed for between the pre-IFRS and post-IFRS era. Since the IFRS became compulsory from 2005, research outcomes based on earlier start dates would tend to become less reliable because of inconsistent and superseded accounting principles.

7.4 Research questions and factual conclusions

The aforementioned motives and rationale for this research led to the formulation of the following four research questions:

i. To what extent can the success of a simple fundamentals-based investment strategy be replicated in the stock markets of the UK and Germany?
ii. Are there alternative strategies that can be used by investors to achieve comparable results and which one is the best?
iii. How do the book-to-market ratio, different kinds of information uncertainty measures and liquidity drive the success of those strategies?
iv. Does the country-specific legal system have an influence on the success of the various investment strategies?

Regarding the first research question, the success of the original F-Score strategy was empirically confirmed outside the US. The evaluation of one-way portfolios, that is, absolute returns and returns in a long/short scenario, led to this conclusion. Having replicated the original strategy, the answer to the second question was dependent on the respective investment strategy under scrutiny. Generally speaking, clustering the forecast errors should be the next best method of choice for investors followed by the F-rank and DMSFE methods. However, the additional time and effort necessary for the computation of the ranks and forecast errors do not stand in reasonable relation to the subsequent results. Consequently, the original F-Score method is regarded as the best and should therefore be recommended to practitioners.

The answer to the third question is twofold. Regarding the book-to-market ratio, the evidence shows that, in the vast majority of cases, all of the investment strategies work best amongst low-BM firms both in the UK and in Germany. However, the degree of the success depends on the specific method. In general, the F-Score, F-rank and cluster methods work better than all of the DMSFE combinations. This conclusion is derived from the ability of those methods to generate higher return differences within the group of low-BM firms. It should be noted that this is independent of the secondary research findings, which confirm that high-BM firms usually perform better in absolute terms.

The analysis of firm size as an information uncertainty proxy shows that the success of the investment strategies is owed to small firms in the UK. Again, this is based on the level of portfolio return differences within each of the size terciles, which is highest for small firms. However, the German findings indicate that the returns are driven more by larger firms. It can reasonably be assumed that this is due to the overall larger market capitalisation of German firms in comparison with that of UK firms. As a secondary finding, it can be stated that larger firms perform better both in the UK and in Germany, although this result is not as clear as for the UK.

A summary of the results of all the investment strategies with regard to the alternative two information uncertainty measures, the SDF and MADF, provides no clear results for both countries. However, once the original F-Score method is viewed in isolation, there is very strong evidence that high-uncertainty firms are the main contributors to its success. Success in this regard means the ability to generate the highest possible horizontal return differentials, which is of interest for long/short investors. By contrast, the F-rank results are very mixed. According to these methods, low-uncertainty firms actually contribute more in most of the cases.

For the UK, the addition of two liquidity measures, i.e. Amihud and ToR, tested the influence of a stock’s liquidity on the success of the strategies. In almost all the cases, higher returns can be generated from portfolios that contain stocks with low liquidity. Again for the UK only, the results can be used further to determine whether the information uncertainty measures or the liquidity measures have more influence on the results in general. Judging from the amplitude of horizontal return differentials, both IU and liquidity are equal contributors to a strategy. Further still, the IU and liquidity measures can be examined in isolation to analyse which of the respective two measures has more impact. Amongst the two IU measures, the MADF clearly has the better ability to generate larger return differentials. Regarding liquidity, the same holds true for the ToR measure.

Because no liquidity measures were generated for Germany, it can only be assessed which of the two IU measures has a greater impact on a strategy's success. Interestingly, the SDF can provide investors with higher return differentials, which contrasts the UK results. Finally, as a secondary outcome, stocks with low levels of information uncertainty and high levels of liquidity are preferred investment targets because of their higher subsequent absolute returns. This is particularly interesting for long-only investors.

The final hypothesis dealt with the legal environment and its influence on the success of the investment strategies in the UK and Germany. Generally, the findings did not suggest that a different system of justice has a great impact on the functioning of the various strategies. This finding is even more interesting considering the weaker investor protection in Germany. However, investors generally enjoy higher stock returns in Germany than in the UK.

7.5 Conceptual conclusions and the contribution to knowledge

The basic idea of Piotroski's (2000) investment strategy is to provide investors with an easily implementable tool for the selection of potentially well-performing stocks from the entirety of the stock universe. This concept, however, is faced with differing assumptions about its practicability in the academic literature. In short, market efficiency states that persistent above-average market returns are unattainable because relevant information is immediately reflected in stock prices. By contrast, advocates of behavioural finance counter that inefficiencies exist due to the irrationality of market participants.

The research findings bear implications for the theory of efficient markets. As the success of the F-Score strategy in particular has shown in the UK and Germany, the study provides support for the behavioural finance theory. This implies that investors apply similar heuristics and are subject to cognitive biases irrespective of their geographical location, which ultimately contributes to the generalisability of the BF theory. Besides, this study narrows the literature gap by empirically testing a composite measure that is based on financial statement items in the stock markets of the UK and Germany. To ensure the comparability and reliability of the F-Score results between the US and the European stock market, the study followed the parameters preset by Piotroski (2000). Hence, the research outcomes are both guided by theory (Richardson et al., 2010) and not a result of in-sample fitting, as argued by Greig (1992).

The purpose of the second research question was twofold. On the one hand, alternatives to the original F-Score approach were established with the intention of corroborating the initial findings of exploitable market inefficiencies and hence to provide further support for the theory of behavioural finance. On the other hand, new methods were used to accomplish this task. These methods were borrowed from the economics literature and applied to the concept of a simple fundamentals-based investment strategy. This idea builds directly on the observation by Rapach et al. (2010), who found that forecast combination methods are rare in the finance literature despite their increasing popularity in the field of economics. The research results in this thesis generally agree with Rapach et al. (2010) insofar as the more elaborate forecasting methods perform relatively poorly compared with the simpler ones. This is especially interesting as cross-sectional rather than time-series information, which has almost exclusively been used previously (e.g. Rapach et al., 2010), was used for the forecast combinations. However, even the simplest methods underperform the F-Score in aggregate. It can be argued whether the F-Score is indeed the simplest investment strategy amongst those tested in this study. Taking into consideration that it can be applied instantly with no historical stock return data, no need for calculating ranks and a binary nature, there is much support for this conclusion. Consequently, this would be in line with the assumption made by Rapach et al. (2010).

If, on the other hand, this is not the case, the robustness of the F-Score remains a puzzle, as noted by Richardson et al. (2010). This, in turn, would further add to the findings related to the first research question. As the authors argued, the two reasons for the strategy’s success are most likely to be in-sample fitting and the lack of external validity, that is, the omission of a holdout sample. This research has shown, however, that both arguments can be countered. In-sample fitting is not an issue because the same parameters are applied to the UK and German observations without mining the data first. As a consequence, no holdout sample is necessary because the Piotroski (2000) study already serves this purpose.

The third research question targets the specific conundrum of the BM effect (e.g. Lakonishok et al., 1994) and the size effect (e.g. Banz, 1981) and their interplay with the respective investment strategies. Regarding the former, two explanations usually exist for the outperformance of high-BM firms, as summarised by Ali et al. (2003). On the one hand, this difference can be explained by the higher compensation for taking on higher risk, the EMH view. On the other hand, differences occur due to mispricing, the BF view. Aside from those two views, it is arguable why Piotroski (2000) analysed only high-BM firms. The research findings of the present study provide sufficient evidence to conclude that investment strategies based on a composite measure are able to generate sizeable returns regardless of a firm’s BM ratio. Hence, this represents an extension of the original study’s scope. In fact, long/short returns from low-BM firms contribute even more to the overall success of strategies in the majority of cases. Furthermore, it confirms similar findings by Duong et al. (2014) for the UK stock market, for which the analysis was guided by similar assumptions. As a secondary finding, however, long-only investors enjoy higher profits from high-BM portfolios, which confirms the BM effect in general.

Regarding firm size, the present study analysed the key drivers of the original F-Score strategy similarly to the BM ratio. In addition, it aimed to contribute to the effort to uncover the economic rationales behind the existence of a size effect. Contrary to the BM ratio, behavioural finance is relatively silent regarding the grounds for this very size effect (Amel-Zadeh, 2011). With respect to the first point, the analyses obtained opposing results. While the returns from the investment strategies are driven by smaller firms in the UK, the contrary is observable in Germany. However, this does not limit the success of the respective strategies altogether and thus further evidences the robustness of the original concept of this straightforward accounting-based investment tool. Regarding the second point, the methods used for scrutinising the size effects were borrowed from Zhang (2006). In addition, new measures of information uncertainty were implemented in this thesis. Similarly to his study, this research aimed to provide a behavioural explanation for the size anomaly. While EMH adherents ascribe this effect to higher levels of risk (e.g. Roll, 1981), behaviouralists suggest that investors under-react to new information depending on a stock’s degree of information uncertainty. This implies that the speed of news absorption into stock prices varies and profit opportunities arise. Size, in this regard, is used as a measure of this uncertainty rather than representing a measure of risk. The underlying concept is that, due to lower disclosure requirements for small firms, for instance, those firms bear higher uncertainty. However, the test results of this study do not point to this conclusion but are rather in line with Dimson and Marsh (1999), who presented evidence for a reversed size effect, specifically higher returns for bigger companies. Zhang (2006) picked up on this evidence and concluded that size is not a risk measure but rather an indicator of information uncertainty. Because the market reaction is relatively complete for low-uncertainty stocks, high-uncertainty stocks should generate better (worse) returns following good (bad) news. However, the research results for the UK and Germany only partly confirm the findings of Zhang (2006). While high-uncertainty firms do indeed produce lower returns in the case of bad news, they still do so in the case of good news.

Two interpretations are possible. On the one hand, given that low-uncertainty stocks generally produce higher returns, it is likely that different degrees of information completeness between the good and the bad news stocks exist. From this it would follow that an investor’s surprise from bad news is greater than that from good news. This is corroborated by usually stronger test statistics within the group of bad news firms. As a result, investors are not sufficiently surprised to induce them to buy those stocks; rather, they prefer to stick with what they have already.

On the other hand, the behavioural finance concept of herding could provide an answer. On a micro-data level from Finland, Kaustia and Knüpfer (2012) found a positive correlation between returns in one particular neighbourhood and new investors from the same neighbourhood who enter the market. Due to the uniqueness of the official ownership registry in Finland, which records daily transactions, the data set is highly reliable, thus providing a solid foundation for interpreting the findings from this study. Due to the monotonic and positive correlation between participation rates and returns, the results should be stronger the higher the overall market participation is. Considering the more market-oriented UK economy, this assumption is clearly substantiated by the stronger test statistics in the UK compared with Germany. Generally, however, investors in both countries prefer low-uncertainty stocks. These findings are further confirmed by the alternative two uncertainty measures, namely the SDF and MADF.

Regarding the liquidity measures in the UK, the success of investment strategies is almost exclusively dependent on low-liquidity stocks. In other words, once the investment strategies are applied to the subset of low-liquidity stocks, the return differentials between high-scored and low-scored or ranked stocks are higher than the ones in the high-liquidity subset. As low liquidity is conceptually tantamount to higher uncertainty, similar to small firm size, it is not too surprising that these outcomes coincide. However, the addition of a new dimension, in this case liquidity, does not diminish the effectiveness of Piotroski's (2000) basic idea and thus provides further proof of its functioning. Similar to the investors' desire to hold good-quality stocks within each of the liquidity subsets, they usually aim to hold stocks with higher levels of liquidity within each of the measures of stock quality. Again, this confirms the previous observations from the information uncertainty measures.

In principle, the same two interpretations of information completeness and herding are possible and are therefore not discussed further. In general, however, a positive relation between a stock's return and its illiquidity, as documented by Amihud (2002), cannot be confirmed, as the results from the stock markets of the UK and Germany show. The conception of a higher reward for taking on higher risk in the form of not being able to buy or sell a specific stock agrees with the EMH. However, this holds true only if the stocks are characterised by liquidity features alone. Once a basis for evaluating the quality of two stocks is created, such as the F-Score, it could be objected that a premium for the more illiquid stock is unjustified. The reason is that the two stocks have a similar probability of generating similar returns, assuming that they have the same quality features. Following this interpretation, it would be conceivable that the stock with the higher liquidity could, if at all, be selling at a premium. This assumption was confirmed by Fang et al. (2009), who also pointed out that the empirical research on this matter is very limited. Thus, the results of this study contribute to narrowing this gap. The authors used the market-to-book ratio to show that high-liquidity stocks perform better and used three alternative liquidity proxies as robustness checks, one of which was the Amihud (2002) measure. In summary, their study documented a positive relationship between liquidity and firm performance through higher operating profitability.

While the previous conclusion is model-dependent, specifically on the so-called stock price feedback models (e.g. Subrahmanyam and Titman, 2001; Khanna and Sonti, 2004), Beber et al. (2009) reached a more generic conclusion. Analysing the euro-area bond markets, they investigated whether bond investors demand credit quality or liquidity both in normal and in distressed market states. While the findings suggested that the requirements for credit quality and liquidity are equal, the preference is given to liquidity in times of high market uncertainty. These results are in accordance with the findings of the present study. This is because, within each group of stocks with the same quality features, the assessment of liquidity should not make a difference. However, as the data cover a period of nearly 20 years, the outperformance of higher-liquidity stocks suggests that investors preferred to buy these stocks during times of market uncertainty, which yields measurable results in aggregate. The findings further suggest that this effect is stronger amongst low-quality stocks.

Finally, the last research question aimed to analyse the interplay between investor protection rights and stock return performance. The degree of investor protection, in turn, is directly related to the legal environment of the UK and Germany. Usually, common law countries, such as the UK, place more emphasis on protection than their code law counterparts, such as Germany. However, this legal tradition is most likely to determine the shareholder structure of a company. This concept is related to agency theory, which analyses the relationship between shareholders and their agents who are in charge of managing the firm (Fama and Jensen, 1983). Because the problems between shareholders and agents increase the lower the level of investor protection rights, Burkart et al. (2003) recommended family ownership in these cases to align the interests with those of shareholders. While firms in the UK are mostly widely held, family control is much more common in German firms (Faccio and Lang, 2002). Therefore, it has to be answered first whether family-controlled or non-family-controlled businesses perform better considering the potential conflicts between shareholders and agents. The present study reports better overall returns in Germany, which are consistent with the earlier research by Anderson and Reeb (2003) for the US and Barontini and Caprio (2006) for Europe minus the UK. Because more firms in Germany than in the UK are family-owned according to Faccio and Lang (2002) and those firms perform better in general, the returns in the German stock market should be higher. This concept was substantiated in view of the investment strategies' return performances and also agrees with the findings by Maury (2006).

In summary, this subsection provided a conceptual conclusion and pointed to gaps in the academic literature. These gaps were narrowed first by replicating an investment strategy found to be successful in the US in the European context, specifically for the stock markets of the UK and Germany. Second, the basic concept of the original strategy was extended. For this reason, various alternatives were developed and tested following the same methods as previously to ensure consistence and comparability of the research findings. Third, the key drivers of these strategies were analysed and put into a behavioural context. Fourth and last, the performance of the strategies was tested against the background of different legal systems. Overall, therefore, this research satisfies the requirements for a modest contribution to knowledge.

7.6 Limitations of the study

Apart from the contributions made by this study, some major limitations are addressed in this section. Firstly, the study focused only on two European economies, which have a series of distinctive features. For instance, the UK is not part of the euro area and fosters a common law legal system. By contrast, German companies tend to be intertwined insofar as there are substantial cross-holdings amongst them. Further, a German-origin code law system is in place as compared with a French- or Scandinavian-origin law system, which is used in the majority of euro area countries. For these reasons, the research results of this study should be interpreted with caution. Secondly, similar to Piotroski's (2000) original study, the existence of a data-mining bias cannot be ruled out entirely. This problem occurs because the nine financial statement items have not been actively developed as part of this study but rather adopted from previously published results. Consequently, it is possible that this bias has been carried forward to this study. Thirdly, all of the variables used in the analyses could be subject to measurement error. In other words, they might not adequately measure what they are supposed to measure, thus inducing a bias into the outcomes with the effect of overall lower reliability. Finally, the generalisability of the research results can be critically examined. Although, due to the choice of a deductive research paradigm, the results should be both reliable and generalisable, the research outcomes should still be viewed cautiously as outlined previously. However, this does not affect the European economies so much due to the process of harmonising financial markets and reporting standards but rather emerging and even frontier markets.

7.7 Recommendations for future research

In view of the limitations highlighted above, this final section suggests potential avenues for future research. First of all, a more qualitative approach could help to analyse the characteristics of high-returning firms in more depth. This would complement the quantitative approach that was used in this study and would also shed light on the "soft skills" of companies as potential drivers of their stock price performance. Conceptually, this process agrees with practising investors such as Warren Buffett, who also looks behind the numbers. For instance, in his earlier shareholder letters, he actively sought businesses that were easy to understand and had a trustworthy and apt management team in place.[25] Following that, these traits could be compared with the highest-scoring and lowest-scoring firms from the previous investment strategies and measured. Similar to the ranking process in this study, the qualitative components could include the length of management tenure, firm age, number of direct competitors, global reach, social accountability and corporate greenwashing. As a result, these additional measures would expand the initial strategies by adding a qualitative component.

Additionally, a more detailed analysis could be conducted across different industries. The research by Hou and Robinson (2006) showed that firms in competitive industries earn higher stock returns. This finding seems surprising as industries with an oligopolistic structure, for instance, usually aim to maximise their profits, which should result in higher stock prices. It would therefore be interesting to research how the success of an investment strategy is affected by the structure of a specific industry's market, that is, the degree of industry concentration, maturity and barriers to entry.

Furthermore, primary data could be collected directly from a brokerage firm. The idea would be to access trade tickets and create groups of both persistently successful and persistently unsuccessful retail traders. After the analysis of the data, the main characteristics of the stocks within the two groups could be compared with those of the high-ranking and low-ranking stocks that were used in this study. Finally, the stock return performances of the two approaches would be compared. To add another dimension, traders could be interviewed to gain an understanding of what else they consider important during their investment decision process.

Finally, the analysis could be further extended to countries which have different characteristics to those already studied. This is of particular interest as the success of investment strategies in the US, the UK and Germany might be due to the strong economic and political links between these countries, which might fuel the data-mining bias mentioned earlier. For this reason, this bias could be potentially reduced if stock markets with much weaker economic and political ties were analysed.

[25] Source: Berkshire Hathaway (http://www.berkshirehathaway.com/letters/1991.html)

Appendix

<u>Firm selection (**section 3.5.1.1**):</u>

- Firms with GBP (UK) as their reporting currency are included
- The following variables were downloaded from Datastream for the period 31.12.1990 to 31.12.2011:

A.1 Variable definitions

Variable	*Description*
FYE (D)	financial year end [DS code: WC05350]
MP-FYE (D)	market price at the end of June [DS code: RI] (30/06/1991–30/06/2012)
BV-FYE (D)	book value at the financial year end [DS code: WC05491]
BM-FYE (C)	book value of outstanding shares (= BV-FYE / MP-FYE)
PB-FYE (D)	price to book value at the financial year end [DS code: PTBV]
NIN (D)	net income before extra items and dividends [DS code: WC01551]
TAS (D)	total assets [DS code: WC02999]
ATAS (C)	average total assets ($ATAS_t = (TAS_t + TAS_{t-1}) / 2$))
CFO (D)	cash flow from operating activities [DS code: WC04860]
LTD (D)	long-term debt [DS code: WC03251]
CRA (D)	current ratio [DS code: WC08106]
EQO (D)	net proceeds from equity offerings [DS code: WC04251]
GPM (D)	gross profit margin [DS code: WC08306]
TSA (D)	net sales [DS code: WC01001]
CAP (D)	market capitalisation [DS code: WC08002]
	(D) = downloaded; (C) = computed

- Firms with missing data were removed
- This process yields 3,089 firms with complete data, i.e. with all 9 variables necessary for constructing the F-Score

<u>Benchmark definition (**section 3.5.1.3**):</u>

- Individual stock returns are downloaded and winsorised at the 0.5% level using Stata
- Market-adjusted returns are computed only for firms with complete F-Score data
- The equally weighted benchmark index is calculated only from those firms with complete F-Score data
- Firm-specific returns are computed on 30 June each year

- The one year market-adjusted buy and hold return (BHAR) is therefore:
  ➔ BHAR = Individual stock return – benchmark return
- Portfolios are constructed at ***t*** using the F-Score data from ***t - 1*** measured against market-adjusted returns at ***t + 1***

Portfolio construction (**section 3.5.1.4**):

- F-Scores are assigned to individual firms for the period 1992 to 2010
- This process results in 20,053 firm years
- Firm years are defined as cumulative amount of firms with complete data
- Fquintiles are constructed as follows:
  - L=30%, M=40% and H=30%
  - L*=5% and H*=5%
- L*/H* are included in L/H
- It is assumed that firms with higher F-Score perform better than low F-Score firms

Portfolio construction with alternatives (**section 3.5.1.5**):

- Each of the nine individual F-Score components is ranked as follows:

$$F_{i,srank} = \frac{(F_{i,rank} - 1)}{(F_{max} - 1)} \tag{1}$$

- This generates nine Fi-ranks per firm
- A continuous value between 0 and 1 is attached to each Fi-rank
- Ties are retained
- Fi-ranks are aggregated
- Both the mean and median are computed from the aggregate value resulting in the F-rank

Portfolio construction with DMSFE (**section 3.5.1.6**):

- Forecast errors (DMSFE) are calculated as follows:
  ➔ $DMSFE_i = (r_t – \text{Fi-rank}_{i,t-1})^2$
- This calculates the difference between the ranked and standardised stock return (*r*) at ***t*** and the Fi-rank at ***t – 1*** (r is ranked and standardised according to equation (1))
- Therefore, Fi-ranks are used as nine individual forecasts
- The lower the DMSFE, the higher the forecast accuracy
- Portfolios are constructed at ***t + 1*** based on DMSFE at ***t***
- Weights are attached to each of the forecasts as follows:

$$\phi_{j,t} = \frac{\sum_{i=1}^{N} \left(r_t - F\text{-}rank_{i,t-1}\right)^2}{N} \tag{2}$$

where

$$\omega_{i,t} = \Phi_{j,t}^{-1} \Big/ \sum_{j=1}^{9} \Phi_{j,t}^{-1} \tag{3}$$

- The best three, the next best three and the worst three forecasts are summarised
- Forecasts, i.e. terciles, are labelled as high, middle or low
- The mean, i.e. weight, of each of the terciles is calculated
- This average weight is multiplied by those Fi-ranks that are part of the respective tercile

Portfolio construction with clusters (**section 3.5.1.7**):

- Similar to DMSFE
- Only the three forecasts with the lowest forecast errors are considered ($k = 3$)
- Only the five forecasts with the lowest forecast errors are considered ($k = 2$)

Inclusion of a second dimension: BM and size (**sections 3.5.1.8 and 3.5.1.9**)

- Values for BM and size (market capitalisation) are downloaded from Datastream
- Firms are assigned to one of three BM/size terciles

Information uncertainty proxies: SDF and MADF (**section 3.5.2.2**)

- Both measures use the F-rank as input variable
- Both measures are constructed as terciles as they are used in the second dimension
- SDF is calculated as follows:

$$SDF_{i,t} = \sqrt{\frac{\sum(F\text{-}rank_{i,t} - \overline{F\text{-}rank}_t)^2}{(n-1)}} \tag{4}$$

- MADF is calculated as follows:

$$MADF_{i,t} = \frac{\sum |F\text{-}rank_{i,t} - \overline{F\text{-}rank}_t|}{n} \tag{5}$$

<u>Causality of investment performance: The Amihud and ToR liquidity measure (**sections 3.5.2.3 and 3.5.2.4**)</u>

- Amihud measures the price impact dimension using the following equation:

$$LIQ_{i,t} = 1 \,/\, D_{i,t} \sum_{\mathrm{t}}^{\mathrm{D_{i,t}}} |\mathrm{R_{i,t,d}}| \,/\, \mathrm{VOLP_{i,t,d}} \tag{6}$$

- VOLP equals the price times volume on any given day
- Both variables were downloaded from Datastream
- *D* represents the number of days in one specific year for each firm

- ToR measures the trading quantity dimension as follows:

$$ToR_{i,t} = 1 \,/\, D_{i,t} \sum_{\mathrm{t}}^{\mathrm{D_{i,t}}} \mathrm{VOL_{i,t,d}} \,/\, \mathrm{NOSH_{i,t,d}} \tag{7}$$

- NOSH measures the number of shares outstanding on any given day
- VOL measures the turnover by volume on any given day
- Both variables were downloaded using Datastream
- Both Amihud and ToR are used as the second dimension

<u>Parametric and non-parametric tests (**section 3.6.3**)</u>

- Parametric unpaired t-tests are performed on the equality of means using Stata
- The test for $\mu = \mu_0$ for unknown σ is given by:

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \tag{8}$$

- The statistical distribution follows a Student's *t* with *n* – 1 degrees of freedom

- The non-parametric Mann-Whitney-Wilcoxon (MWW) test is used
- This test is implemented in Stata as the ranksum test
- We test the null hypothesis that $X_1 \sim X_2$
- We have a sample of size n_1 from X_1 and another of size n_2 from X_2
- The Wilcoxon test statistic is the sum of the ranks for the observations in the first sample and is given by:

$$T = \sum_{i=1}^{n_1} R_{1i} \tag{9}$$

- The Mann-Whitney test statistic is the number of pairs (X_{1i}, X_{2j}) such that $X_{1i} > X_{2j}$ given by:

$$U = T - \frac{n_1(n_1 + 1)}{2} \tag{10}$$

- The probability is given by:

$$p = \frac{U}{n_1 n_2} \tag{11}$$

- The non-parametric median test analyses the likelihood that two samples stem from populations with the same median
- The null hypothesis is that samples are drawn from populations with the same median
- The test is performed using the median command in Stata

Bibliography

Abarbanell, J. and Bushee, B. (1997). "Fundamental Analysis, Future Earnings, and Stock Prices." *Journal of Accounting Research* 35 (1): 1 – 24.

Abarbanell, J. and Bushee, B. (1998). "Abnormal Returns to a Fundamental Analysis Strategy." Accoun*ting Review*, 73 (1): 19 – 45.

Acharya, V. V. and Pedersen, L. H. (2005). "Asset Pricing with Liquidity Risk." *Journal of Financial Economics*, 77 (2): 375 – 410.

Agarwal, R. and Elston, J. A. (2001). "Bank-Firm Relationships, Financing and Firm Performance in Germany." *Economics Letters*, 72 (2): 225 – 232.

Aghion, P., Blundell, R., Griffith, R., Howitt, P. and Prantl, S. (2009). "The Effects of Entry on Incumbent Innovation and Productivity." *Review of Economics and Statistics*, 91 (1): 20 – 32.

Ali, A., Hwang, L.-S. and Trombley, M. A. (2003). "Arbitrage Risk and the Book-To-Market Anomaly." *Journal of Financial Economics*, 69 (2): 355 – 373.

Amel-Zadeh, A. (2011). "The Return of the Size Anomaly: Evidence from the German Stock Market." *European Financial Management,* 17 (1): 145 – 182.

Amihud, Y. (2002). "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5 (1): 31 – 56.

Amihud, Y. and Mendelson, H. (1986). "Asset Pricing and the Bid-Ask Spread." *Journal of Financial Economics*, 17 (2): 223 – 249.

Anderson, E. W., Ghysels, E. and Juergens, J. L. (2009). "The Impact of Risk and Uncertainty on Expected Returns." *Journal of Financial Economics*, 94 (2): 233 – 263.

Anderson, R. C. and Reeb, D. M. (2003). "Founding-Family Ownership and Firm Performance: Evidence from the S&P 500." *The Journal of Finance*, 58 (3): 1301 – 1328.

Andrianova, S., Demetriades, P. and Shortland, A. (2012). "Government Ownership of Banks, Institutions and Economic Growth", *Economica,* 79 (315): 449 – 469.

Andrikopoulos, P., Daynes, A., Latimer, D. and Pagas, P. (2008). "Size Effect, Methodological Issues and 'Risk-to-Default': Evidence from the UK Stock Market." *European Journal of Finance*, 14 (3-4): 299 – 314.

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

Ang, A. and Chen, J. (2007). "CAPM Over the Long Run: 1926 – 2001." *Journal of Empirical Finance*, 14 (1): 1 – 40.

Antoniou, A., Guney, Y. and Paudyal, K. (2008). “The Determinants of Capital Structure: Capital Market-Oriented versus Bank-Oriented Institutions.” *Journal of Financial and Quantitative Analysis,* 43 (1): 59 – 92.

Aiolfi, M. and Timmermann, A. (2006). “Persistence in Forecasting Performance and Conditional Combination Strategies.” *Journal of Econometrics,* 135 (1-2): 31 – 53.

Artmann, S., Finter, P. and Kempf, A. (2012b). “Determinants of Expected Stock Returns: Large Sample Evidence from the German Market.” *Journal of Business Finance and Accounting*, 39: 5 – 6.

Artmann, S, Finter, P., Kempf, A., Koch, S. and Theissen, E. (2012). “The Cross-Section of German Stock Returns: New Data and New Evidence.” *Schmalenbach Business Review,* 64 (1): 20 – 43.

Asgharian, H. and Hansson, B. (2010). “Book-to-Market and Size Effects: Compensations for Risks or Outcomes of Market Inefficiencies?” *European Journal of Finance,* 16 (2): 119 – 136.

Asness, C. S., Moskowitz, T. J. and Pedersen, L. H. (2013). “Value and Momentum Everywhere.” *The Journal of Finance,* 68 (3): 929 – 985.

Asparouhova, E., Bessembinder, H. and Kalcheva, I. (2010). “Liquidity Biases in Asset Pricing Tests.” *Journal of Financial Economics,* 96 (2): 215 – 237.

Audretsch, D.B. and Elston, J. A. (1997). “Financing the German Mittelstand.” *Small Business Economics*, 9 (2): 97 – 110.

Audretsch, D. B. and Elston, J. A. (2002). “Does Firm Size Matter? Evidence on the Impact of Liquidity Constraints on Firm Investment Behavior in Germany.” *International Journal of Industrial Organization*, 20 (1): 1 – 17.

Baker, H. K., Nofsinger, J. R. and Weaver, D. G. (2002). “International Cross-Listing and Visibility.” *Journal of Financial and Quantitative Analysis,* 37 (3): 495 – 521.

Baker, M. and Wurgler, J. (2006). “Investor Sentiment and the Cross-Section of Stock Returns.” *The Journal of Finance,* 61 (4): 1645 – 1680.

Bali, T. G., Peng, L., Shen, Y. and Tang, Y. (2014). “Liquidity Shocks and Stock Market Reactions.” *Review of Financial Studies,* 27 (5): 1434 – 1485.

Ball, R. and Brown, P. (1968). “An Empirical Evaluation of Accounting Income Numbers.” *Journal of Accounting Research,* 6 (Autumn): 159 – 178.

Ball, R., Robin, A. and Sadka, G. (2008). “Is Financial Reporting Shaped by Equity Markets or by Debt Markets? An International Study of Timeliness and Conservatism.” *Review of Accounting Studies,* 13 (2-3): 168 – 205.

Banerjee, S., Gatchev, V. A. and Spindt, P. A. (2007). “Stock Market Liquidity and Firm Dividend Policy.” *Journal of Financial and Quantitative Analysis*, 42 (2): 369 – 397.

Banz, R. (1981). “The Relationship between Return and Market Value of Common Stocks.” *Journal of Financial Economics*, 9 (1): 3 – 18.

Barber, B. M., Lee, Y. T., Liu, Y. J. and Odean, T. (2009b). “Just How Much Do Individual Investors Lose by Trading?” *Review of Financial Studies,* 22 (2): 609 – 632.

Barber, B. M., Lehavy, R., McNichols, M., & Trueman, B. (2001). “Can Investors Profit From the Prophets? Security Analyst Recommendations and Stock Returns.” *The Journal of Finance,* 56 (2): 531 – 563.

Barber, B., Lehavy, R., McNichols, M., and Trueman, B. (2003). “Reassessing the Returns to Analysts’ Stock Recommendations.” *Financial Analysts Journal,* 59 (2): 88 – 96.

Barber, B. M. and Lyon, J. D. (1997). “Detecting Long-Run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics.” *Journal of Financial Economics,* 43 (3): 341 – 372.

Barber, B. M. and Odean, T. (2000). “Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors.” *The Journal of Finance,* 55 (2): 773 – 806.

Barber, B. M. and Odean, T. (2001). “Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment.” *Quarterly Journal of Economics,* 116 (1): 261 – 292.

Barber, B. M., Odean, T. and Zhu, N. (2009). “Systematic Noise.” *Journal of Financial Markets,* 12 (4): 547 – 569.

Barberis, N. (2013). “The Psychology of Tail Events: Progress and Challenges.” *American Economic Review,* 103 (3): 611 – 616.

Barberis, N. and Thaler, R. H. (2005). “A Survey of Behavioral Finance.” in Advances in Behavioral Finance , 2: 1 – 65. New York, NY: Princeton University Press.

Barberis, N. and Xiong, W. (2009). “What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation.” *The Journal of Finance,* 64 (2): 751 – 784.

Barontini, R. and Caprio, L. (2006). “The Effect of Family Control on Firm Value and Performance: Evidence from Continental Europe.” *European Financial Management*, 12 (5): 689 – 723.

Barth, M. E., Landsman, W. R. and Lang, M. H. (2008). “International Accounting Standards and Accounting Quality.” *Journal of Accounting Research*, 46 (3): 467 – 498.

Barton, J., Hansen, T. B. and Pownall, G. (2010). “Which Performance Meassures Do Investors Around The World Value the Most - and Why?” *The Accounting Review*, 85 (3): 753 – 789.

Bartov, E., Goldberg, S. R. and Kim, M. (2005). “Comparative Value Relevance Among German, U.S., and International Accounting Standards: A German Stock Market Perspective.” *Journal of Accounting, Auditing and Finance,* 20 (2): 95 – 119.

Bassi, A., Colacito, R. and Fulghieri, P. (2013). "'O Sole Mio: An Experimental Analysis of Weather and Risk Attitudes in Financial Decisions." *Review of Financial Studies*, 26 (7): 1824 – 1852.

Basu, S. (1983). "The Relationship between Earnings' Yield, Market Value and Return for NYSE Common Stocks: Further Evidence" *Journal of Financial Economics,* 12 (1): 129 – 156.

Bates, J. M. and Granger, C. W. J. (1969). "The Combination of Forecasts." *Operational Research Quarterly,* 20: 451 – 68.

Beaver, W. H. (1981). "Market Efficiency." *The Accounting Review,* 56 (1): 23 – 37.

Beber, A., Brandt, M. W. and Kavajecz, K. A. (2009). "Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market." *Review of Financial Studies,* 22 (3): 925 – 957.

Bekaert, G., Harvey, C. R. and Lundblad, C. (2007). "Liquidity and Expected Returns: Lessons from Emerging Markets." *Review of Financial Studies,* 20 (6): 1783 – 1831.

Benartzi, S. and Thaler, R. H. (2001). "Naïve Diversification Strategies in Defined Contribution Saving Plans." *American Economic Review,* 91 (1): 79 – 98.

Benartzi, S. and Thaler, R. H. (2007). "Heuristics and Biases in Retirement Savings Behavior." *Journal of Economic Perspectives,* 21 (3): 81 – 104.

Ben-David, I. and Hirshleifer, D. (2012). "Are Investors Really Reluctant to Realize Their Losses? Trading Responses to Past Returns and the Disposition Effect." *Review of Financial Studies,* 25 (8): 2485 – 2532.

Beneish, M. D., Lee, C. M. C. and Tarpley, R. L. (2001). "Contextual Fundamental Analysis Through the Prediction of Extreme Returns." *Review of Accounting Studies,* 6 (2-3): 191 – 195.

Bennis, W. G. and O'Toole, J. (2005). "How Business Schools Lost Their Way." *Harvard Business Review,* 83 (5): 96 – 104.

Betzer, A. and Theissen, E. (2009). "Insider Trading and Corporate Governance: The Case of Germany." *European Financial Management,* 15 (2): 402 – 429.

Bhattacharya, N. (2001). "Investors' Trade Size and Trading Responses around Earnings Announcements: An Empirical Investigation." *The Accounting Review,* 76 (2): 221 – 244.

Bhattacharya, U. and Daouk, H. (2002). "The World Price of Insider Trading." *The Journal of Finance,* 57 (1): 75 – 108.

Biktimirov, E. N. and Li, B. (2014). "Asymmetric Stock Price and Liquidity Responses to Changes in the FTSE SmallCap Index." *Review of Quantitative Finance and Accounting,* 42 (1): 95 – 122.

Black, F. (1986). "Noise." *The Journal of Finance,* 41 (3): 529 – 543.

Black, F. (1993). “Beta and Return.” *Journal of Portfolio Management,* 20 (1).

Black, F. and Scholes, M. (1973). “The Pricing of Options and Corporate Liabilities.” *Journal of Political Economy,* 3: 637 – 654.

Blumberg, B., Cooper, D. R. and Schindler, P. S. (2008). “Business Research Methods.” 2. McGraw-Hill Higher Education.

Börsch, A. (2004). “What Happens After Privatization? Globalization, Corporate Governance and Adjustment at British Telecom and Deutsche Telekom.” *Journal of European Public Policy,* 11 (4): 593 – 612.

Bortolotti, B., De Jong, F., Nicodano, G. and Schindele, I. (2007). “Privatization and Stock Market Liquidity.” *Journal of Banking and Finance*, 31 (2): 297 – 316.

Boutchkova, M. K. and Megginson, W. L. (2000). “Privatization and the Rise of Global Capital Markets.” *Financial Management,* 29 (4): 31 – 75.

Bower, D. H, Bower, R. S. and Logue, D. E. (1984). “Arbitrage Pricing Theory and Utility Stock Returns.” *The Journal of Finance,* 39 (4): 1041 – 1054.

Bradshaw, M. T. (2004). How do Analysts Use Their Earnings Forecasts in Generating Stock Recommendations?” *The Accounting Review,* 79 (1): 25 – 50.

Brennan, M. J. and Subrahmanyam, A. (1996). “Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns.” *Journal of Financial Economics,* 41 (3): 441 – 464.

Buck, T. and Shahrim, A. (2005). “The Translation of Corporate Governance Changes Across National Cultures: The Case of Germany.” *Journal of International Business Studies*, 36 (1): 42 – 61.

Buckley, A. (1999). “An Introduction to Security Returns.” *The European Journal of Finance,* 5 (3): 165 – 180.

Bulan, L., Subramanian, N. and Tanlu, L. (2007). “On the Timing of Dividend Initiations.” *Financial Management*, 36 (4): 31 – 65.

Burger, M. J. C. (2008). “Towards a Framework for the Elicitation of Dilemmas.” *Quality & Quantity*, 42 (4): 541 – 562.

Burkart, M., Panunzi, F. and Shleifer, A. (2003). “Family Firms.” *The Journal of Finance*, 58 (5): 2167 – 2201.

Bushee, B. J. (1998). “The Influence of Institutional Investors on Myopic R&D Investment Behavior.” *The Accounting Review*, 73 (3): 305 – 333.

Byard, D., Li, Y. and Yu, Y. (2011). “The Effect of Mandatory IFRS Adoption on Financial Analysts’ Information Environment.” *Journal of Accounting Research,* 49 (1): 69 – 96.

Chan, J. S. P., Hong, D. and Subrahmanyam, M. G. (2008). “A Tale of Two Prices: Liquidity and Asset Prices in Multiple Markets.” *Journal of Banking and Finance,* 32 (6): 947 – 960.

Chan, L. K. C. and Lakonishok, J. (2004). “Value and Growth Investing: Review and Update.” *Financial Analysts Journal,* 60 (1): 71 – 86.

Cheffins, B. R., Koustas, D. K. and Chambers, D. (2013). “Ownership Dispersion and the London Stock Exchange’s ‘Two-Thirds Rule’: An Empirical Test.” *Business History,* 55 (4): 667 – 690.

Chen, H. J. (2011). “Firm Life Expectancy and the Heterogeneity of the Book-to-Market Effect.” *Journal of Financial Economics,* 100 (2): 402 – 423.

Chen, H. L. (2006). “On Russell Index Reconstitution.” *Review of Quantitative Finance and Accounting,* 26 (4): 409 – 430.

Chen, N. and Zhang, F. (1998). “Risk and Return of Value Stocks.” *Journal of Business,* 71 (4): 501 – 535.

Chevalier, J. and Ellison, G. (1999). “Career Concerns of Mutual Fund Managers.” *Quarterly Journal of Economics,* 114 (2): 389 – 432.

Cho, Y. H., Linton, O. and Whang, Y. J. (2007). “Are There Monday Effects in Stock Returns: A Stochastic Dominance Approach.” *Journal of Empirical Finance,* 14 (5): 736 – 755.

Choi, N. Y. and Sias, R. W. (2012). “Why Does Financial Strength Forecast Stock Returns? Evidence from Subsequent Demand by Institutional Investors.” *Review of Financial Studies,* 25 (5): 1550 – 1587.

Chordia, T., Roll, R. and Subrahmanyam, A. (2000). “Evidence on the Speed of Convergence to Market Efficiency.” *Journal of Financial Economics,* 76 (2): 271 – 292.

Chordia, T., Roll, R. and Subrahmanyam, A. (2000). “Commonality in Liquidity.” *Journal of Financial Economics,* 56 (1): 3 – 28.

Clement, M. B. (1999). “Analyst Forecast Accuracy: Do Ability, Resources, and Portfolio Complexity Matter?” *Journal of Accounting and Economics,* 27 (3): 285 – 303.

Clements, M. P., Franses, P. H. and Swanson, N. R. (2004). “Forecasting Economic and Financial Time-Series with Non-Linear Models.” *International Journal of Forecasting,* 20 (2): 169 – 183.

Clubb, C. and Wu, G. L. (2014). “Earnings Volatility and Earnings Prediction: Analysis and UK Evidence.” *Journal of Business Finance & Accounting,* 41 (1-2): 53 – 72.

Colliard, J. E. and Foucault, T. (2012). “Trading Fees and Efficiency in Limit Order Markets.” *Review of Financial Studies,* 25 (11): 3389 – 3421.

Copeland, T. and Mayers, D. (1982). “The Value Line Enigma (1965–1978): A Case Study of Performance Evaluation Issues.” *Journal of Financial Economics,* 10 (3): 289 – 321.

Corwin, S. A. and Schultz, P. (2012). “A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices.” *The Journal of Finance*, 67 (2): 719 – 760.

Crane, D. B. and Schaede, U. (2005). “Functional Change and Bank Strategy in German Corporate Governance.” *International Review of Law and Economics*, 25 (4): 513 – 540.

Creswell, J. W. (2013). “Qualitative Inquiry & Research Design: Choosing Among Five Approaches.” London: Sage.

Cuijpers, R. and Buijing, W. (2005). “Voluntary Adoption of Non-Local GAAP in the European Union: A Study of Determinants and Consequences.” *European Accounting Review*, 14 (3): 487 – 524.

Dangl, T. and Hailing, M. (2012). “Predictive Regressions with Time-Varying Coefficients.” *Journal of Financial Economics*, 106 (1): 157 – 181.

Daske, H. and Gebhardt, G. (2006). “International Financial Reporting Standards and Experts’ Perceptions of Disclosure Quality.” *Abacus*, 42 (3-4): 461 – 498.

Daske, H., Hail, L., Leuz, C. and Verdi, R. (2008). “Mandatory IFRS Reporting Around the World: Early Evidence on the Economic Consequences.” *Journal of Accounting Research*, 46 (5): 1085 – 1142.

Datar, V. T. (2001). “Impact of Liquidity on Premia/Discounts in Closed-End Funds.” *Quarterly Review of Economics and Finance*, 41 (1): 119 – 135.

Datar, V. T., Naik, N. Y. and Radcliffe, R. (1998). “Liquidity and Stock Market Returns: An Alternative Test.” *Journal of Financial Markets*, 1 (2): 203 – 219.

Dawes, R. M. (1979). “The Robust Beauty of Improper Linear Models in Decision Making.” *American Psychologist*, 34 (7): 571 – 582.

De Bondt, W. F. M. and Thaler, R. H. (1985). “Does the Stock Market Overreact?” *The Journal of Finance*, 40 (3): 793 – 805.

De Bondt, W. F. M. and Thaler, R. H. (1987). “Further Evidence on Investor Overreaction and Stock Market Seasonality.” *The Journal of Finance*, 42 (3), 105 – 137.

Dechow, P. M., Richardson, S. A. and Sloan, R. G. (2008). “The Persistence and Pricing of the Cash Component of Earnings.” *Journal of Accounting Research*, 46 (3): 537 – 566.

Dellas, H. and Hess, M. (2005). “Financial Development and Stock Returns: A Cross-Country Analysis.” *Journal of International Money and Finance*, 24 (6): 891 – 912.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990). “Noise Trader Risk in Financial Markets.” *Journal of Political Economy*, 98 (4): 703 – 738.

Dempsey, M. (2013). “The CAPM: A Case of Elegance is for Tailors?” *Abacus*, 49 Issue Supplement S1: 82 – 87.

Dempsey, M. (2013a). “The Capital Asset Pricing Model (CAPM): The History of a Failed Revolutionary Idea in Finance?” *Abacus*, 49 Issue Supplement S1: 7 – 23.

Denis, D. J. and Osobov, I. (2008). “Why Do Firms Pay Dividends? International Evidence on the Determinants of Dividend Policy.” *Journal of Financial Economics*, 89 (1): 62 – 82.

Denscombe, M. (2002). “Ground Rules for Good Research.” Buckingham: Open University Press.

Denzin, N. K. and Lincoln, Y. S. (2005). “The SAGE Handbook of Qualitative Research.” *Third Edition*, London: Sage

Deutsch, M., Granger, C. W. J. and Teräsvirta, T. (1994). “The Combination of Forecasts Using Changing Weights.” *International Journal of Forecasting*, 10 (1): 47 – 57.

Dichev, I. D. (1998). “Is the Risk of Bankruptcy a Systematic Risk?”. *The Journal of Finance*, 53 (3): 1131 – 1147.

Dichev, I. D. and Tang, V. W. (2009). “Earnings Volatility and Earnings Predictability.” *Journal of Accounting & Economics*, 47 (1-2): 160 – 181.

Dimson, E. and Marsh, P. (1999). “Murphy’s Law and Market Anomalies.” *Journal of Portfolio Management*, 25: 53 – 69.

Dimson, E., Nagel, S. and Quigley, G. (2003). “Capturing the Value Premium in the United Kingdom.” *Financial Analysts Journal*, 59 (6): 35 – 45.

Dittmann, I, Maug, E. and Schneider, C. (2010). “Bankers on the Boards of German Firms: What They Do, What They Are Worth, and Why They Are (Still) There.” *Review of Finance*, 14 (1): 35 – 71.

Djankov, S., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A. (2008). “The Law and Economics of Self-Dealing.” *Journal of Financial Economics*, 88 (3): 430 – 465.

Doidge, C., Karolyi, G. A. and Stulz, R. M. (2007). “Why Do Countries Matter So Much For Corporate Governance.” *Journal of Financial Economics*, 86 (1): 1 – 39.

Doidge, C., Karolyi, G. A. and Stulz, R. M. (2009). “Has New York Become Less Competitive than London in Global Markets? Evaluating Foreign Listing Choices Over Time.” *Journal of Financial Economics*, 91 (3): 253 – 277.

Døskeland, T. M. and Hvide, H. K. (2011). “Do Individual Investors Have Asymmetric Information Based on Work Experience?” *The Journal of Finance*, 66 (3): 1011 – 1041.

Drake, M. S., Rees, L. and Swanson, E. (2009). “Should Investors Follow the Prophets or the Bears? Evidence on the Use of Public Information by Analysts and Short Sellers.” *The Accounting Review*, 86 (1): 101 – 130.

Duong, C., Pescetto, G. and Santamaria, D. (2014). “How Value-Glamour Investors Use Financial Information: UK Evidence of Investors’ Confirmation Bias.” *The European Journal of Finance*, 20 (6): 524 – 549.

Edmans, A. (2011). “Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices.” *Journal of Financial Economics*, 101 (3): 621 – 640.

Ellsberg, D. (1961). “Risk, Ambiguity, and the Savage Axioms.” *The Quarterly Journal of Economics*, 75 (4): 643 – 669.

Erwin, G. R. and Miller, J. M. (1998). “The Liquidity Effects Associated with Addition of a Stock to the S&P500 Index: Evidence from Bid/Ask Spreads.” *Financial Review*, 33 (1): 131 – 146.

Faccio, M. and Lang, L. H. P. (2002). “The Ultimate Ownership of Western European Corporations.” *Journal of Financial Economics*, 65 (3): 365 – 395.

Fama, E. F. (1970). “Efficient Capital Markets: A Review of Theory and Empirical Work” *The Journal of Finance*, 25 (2): 383 – 417.

Fama, E. F. (1991). “Efficient Capital Markets: II.” *The Journal of Finance*, 46 (5): 1575 – 1617.

Fama, E. F. and French, K. R. (1988). “Permanent and Temporary Components of Stock Prices.” *Journal of Political Economy*, 96: 246 – 273.

Fama, E. F. and French, K. R. (1992). “The Cross-Section of Expected Stock Returns.” *The Journal of Finance*, 47 (2): 427 – 465.

Fama, E. F. and French, K. R. (1993). “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics*, 33: 3 – 56.

Fama, E. F. and French, K. R. (1995). “Size and Book-to-Market Factors in Earnings and Returns.” *The Journal of Finance*, 50 (1): 131 – 155.

Fama, E. F. and French, K. R. (1997). “Industry Costs of Equity.” *Journal of Financial Economics*, 43: 153 – 193.

Fama, E. F. and French, K. R. (2006). “The Value Premium and the CAPM.” *The Journal of Finance*, 61 (5): 2163 – 2185.

Fama, E. F. and French, K. R. (2006b). “Profitability, Investment and Average Returns.” *Journal of Financial Economics*, 82 (3): 491 – 518.

Fama, E. F. and French, K. R. (2008). “Dissecting Anomalies.” *The Journal of Finance*, 63 (4): 1653 – 1678.

Fama, E. F. and French, K. R. (2012). “Size, Value, and Momentum in International Stock Returns.” *The Journal of Finance*, 105 (3): 457 – 472.

Fama, E. F. and Jensen, M. C. (1983). “Separation of Ownership and Control.” *Journal of Law and Economics*, 26 (2): 301 – 325.

Fang, V. W., Noe, T. H. and Tice, S. (2009). “Stock Market Liquidity and Firm Value.” *Journal of Financial Economics*, 94 (1): 150 – 169.

Ferraro, F., Pfeffer, J. and Sutton, R. I. (2005). "Economics Language and Assumptions: How Theories Can Become Self-Fulfilling." *Academy of Management Review,* 30 (1): 8 – 24.

Finster, M. and Wood, M. (2005). "The Apgar Score Has Survived the Test of Time." *Anesthesiology,* 102 (4): 855 – 857.

Fisher, R. (1920). "A Mathematical Examination of the Methods of Determining the Accuracy of Observation by the Mean Error and the Mean Square Error." *Monthly Notes of the Royal Astronomical Soicety*, 80: 758 – 770.

Friedman, M. (1953). "The Methodology of Positive Economics." *Essays in Positive Economics*. University of Chicago Press.

Friedman, M. and Friedman, R. (1990). "Free to Choose: A Personal Statement." First Harvest edition.

Gleason, C. A. and Lee, C. M. C. (2003). "Analyst Forecast Revisions and Market Price Discovery." *The Accounting Review,* 78 (1): 193 – 225.

Gorard, S. (2005). "Revisiting a 90-Year-Old Debate: The Advantages of the Mean Deviation." *British Journal of Educational Studies,* 53 (4): 417 – 430.

Graham, B. (2005). "The Intelligent Investor." Special Edition, New York: HarperCollins.

Graham, J. R. and Harvey, C. R. (2001). "The Theory and Practice of Corporate Finance: Evidence from the Field." *Journal of Financial Economics,* 60 (2-3): 187 – 243.

Gregory, A., Tharyan, R. and Tonks, I. (2013). "More Than Just Contrarians: Insider Trading in Glamour and Value Firms." *European Financial Management,* 19 (4): 747 – 774.

Greig, A. C. (1992). "Fundamental Analysis and Subsequent Stock Returns." *Journal of Accounting and Economics,* 15 (2-3): 413 – 442.

Griffin, D. and Tversky, A. (1992). "The Weighing of Evidence and the Determinants of Confidence." *Cognitive Psychology,* 24 (3): 411 – 435.

Grossman, S. J. (1976). "On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information." *The Journal of Finance,* 31 (2): 573 – 584.

Grossman, S. J. and Stiglitz, J. E. (1980). "On the Impossibility of Informationally Efficient Markets." *American Economic Review,* 70 (3): 393 – 408.

Guay, W. (2000). "Discussion of Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers." *Journal of Accounting Research,* 38: 43 – 51.

Guay, W. and Verrecchia, R. (2006). "Discussion of an Economic Framework for Conservative Accounting and Bushman and Piotroski (2006)." *Journal of Accounting and Economics,* 42 (1-2): 149 – 165.

Guba, E. G. (1990). "The Alternative Paradigm Dialog." In Guba, E. G. (Ed.), The Pardigm Dialog, Newbury Park, CA: Sage.

Guba, E. G. and Lincoln, Y. S. (2005). "Paradigmatic Controversies, Contradictions, and Emerging Confluences." in *The SAGE Handbook of Qualitative Research,* 3rd edition London: SAGE.

Haberly, D. (2013). "White Knights from the Gulf: Sovereign Wealth Fund Investment and the Evolution of German Industrial Finance." *Economic Geography,* 90 (3): 293 – 320.

Hail, L. and Leuz, C. (2006). "International Differences in the Cost of Capital: Do Legal Institutions and Securities Regulation Matter?" *Journal of Accounting Research,* 44 (3): 485 – 531.

Haller, A. and Wehrfritz, M. (2013). "The Impact of National GAAP and Accounting Traditions on IFRS Policy Selection: Evidence from Germany." *Journal of International Accounting, Auditing and Taxation,* 22 (1): 39 – 56.

Haneef, N. (2013). "Empirical Research Consolidation: A Generic Overview and a Classification Scheme for Methods." *Quality & Quantity,* 47 (1): 383 – 410.

Hasbrouck, J. (2009). "Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data." *The Journal of Finance,* 64 (3): 1445 – 1477.

Haugen, R. A. and Baker, N. L. (1996). "Commonality in the Determinants of Expected Stock Returns." *Journal of Financial Economics,* 41 (3): 401 – 439.

Heaton, J. and Lucas, D. (1997). "Market Frictions, Savings Behaviour, and Portfolio Choice." *Macroeconomic Dynamics,* 1: 76 – 101.

Heron, J. and Reason, P. (1997). "A Participatory Inquiry Paradigm." *Qualitative Inquiry,* 3 (3): 274 – 294.

Heston, S. L., Rouwenhorst, K. G. and Wessels, R. E. (1999). "The Role of Beta and Size in the Cross-Section of European Stock Returns." *European Financial Management,* 5 (1): 9 – 27.

Hirshleifer, D. (2001). "Investor Psychology and Asset Pricing." *The Journal of Finance,* 56 (4): 1533 – 1597.

Hirshleifer, D., Hou, K. W., Teoh, S. H. and Zhang, Y. L. (2004). "Do Investors Overvalue Firms with Bloated Balance Sheets?" *Journal of Accounting and Economics,* 38 (1-3): 297 – 331.

Hirshleifer, D., Hsu, P. H. and Li, D. M. (2013). "Innovative Efficiency and Stock Returns." *Journal of Financial Economics,* 107 (3): 632 – 654.

Hirshleifer, D. and Shumway, T. (2003). "Good Day Sunshine: Stock Returns and the Weather." *The Journal of Finance,* 58 (3): 1009 – 1032.

Hong, H., Lim, T. and Stein, J. C. (2000). "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies." *The Journal of Finance,* 55 (1): 265 – 295.

Hou, K. W. (2007). "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns." *Review of Financial Studies,* 20 (4): 1113 – 1138.

Hou, K. W., Karolyi, G. A. and Kho, B. C. (2011). "What Factors Drive Global Stock Returns?" *Review of Financial Studies,* 24 (8): 2527 – 2574

Hou, K. W. and Robinson, D. T. (2006). "Industry Concentration and Average Stock Returns." *The Journal of Finance,* 61 (4): 1927 – 1956.

Huber, P. J. (1981). "Robust Statistics." New York/Chichester: John Wiley.

Huberman, G. and Halka, D. (2001). "Systematic Liquidity." *Journal of Financial Research,* 24 (2): 161 – 178.

Huberman, G. and Jiang, W. (2006). "Offering versus Choice in 401(k) Plans: Equity Exposure and Number of Funds." *The Journal of Finance,* 61 (2): 763 – 801.

Hung, M. and Subramanyam, K. R. (2007). "Financial Statement Effects of Adopting International Accounting Standards: The Case of Germany." *Review of Accounting Studies,* 12 (4): 623 – 657.

Ivković, Z. and Weisbenner, S. (2007). "Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices." *The Review of Financial Studies*, 20 (4): 1327 – 1357.

Jegadeesh, N., Kim, J., Krische, S. D. and Lee, C. M. C. (2004). "Analyzing the Analysts: When Do Recommendations Add Value? *The Journal of Finance,* 59 (3): 1083 – 1124.

Jegadeesh, N. and Titman, S. (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *The Journal of Finance*, 48 (1): 65 – 91.

Jensen, M. C. (1978). "Some Anomalous Evidence Regarding Market Efficiency." *Journal of Financial Economics*, 6 (2): 95 – 101.

Jiang, H. (2010). "Institutional Investors, Intangible Information, and the Book-to-Market Effect." *Journal of Financial Economics*, 96 (1): 98 – 126.

Jordan, S. J., Vivian, A. J. and Wohar, M. E. (2014). "Forecasting Returns: New European Evidence." *Journal of Empirical Finance*, 26: 76 – 95.

Jovanovich, F. (2008). "The Construction of the Canonical History of Financial Economics." *History of Political Economy,* 40 (2): 213 – 242.

Kahneman, D. and Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica,* 47 (2): 263 – 291.

Kahneman, D. and Tversky, A. (1984). "Choice, Values, and Frames." *American Psychologist*, 39 (4): 341 – 350.

Karolyi, G. A., Lee, K.-H. and Van Dijk, M. A. (2012). "Understanding Commonality in Liquidity Around the World." *Journal of Financial Economics*, 105 (1): 82 – 112.

Kaserer, C. and Klingler, C. (2008). "The Accrual Anomaly under Different Accounting Standards – Lessons Learned from the German Experiment." *Journal of Business Finance & Accounting*, 35 (7-8): 837 – 859.

Kaustia, M. and Knüpfer, S. (2012). "Peer Performance and Stock Market Entry". *Journal of Financial Economics*, 104 (2): 321 – 338.

Ke, B. and Ramalingegowda, S. (2005). "Do Institutional Investors Exploit the Post-Earnings Announcement Drift?" *Journal of Accounting and Economics*, 39 (1): 25 – 53.

Keim, D. (1983). "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence." *Journal of Financial Economics*, 12 (1): 13 – 32.

Keynes, J. M. (1937). "The General Theory of Employment." *The Quarterly Journal of Economics*, 51 (2): 209 – 223.

Khanna, N. and Sonti, R. (2004). "Value Creating Stock Manipulation: Feedback Effect of Stock Prices on Firm Value." *Journal of Financial Markets*, 7 (3): 237 – 270.

Kim, S. (2013). "What is Behind the Magic of O-Score? An Alternative Interpretation of Dichev's (1998) Bankruptcy Risk Anomaly." *Review of Accounting Studies*, 18 (2): 291 – 32.

Kim, J. H. and Shamsuddin, A. (2008). "Are Asian Stock Markets Efficient? Evidence from New Multiple Variance Ratio Tests." *Journal of Empirical Finance*, 15: 518 – 532.

Knight, F. H. (1921). "Risk, Uncertainty and Profit." New York : Harper and Row, 1965.

Kothari, S. P. (2001). "Capital Markets Research in Accounting." *Journal of Accounting and Economics*, 31 (1-3): 105 – 231.

Kothari, S. P., Ramanna, K. and Skinner, D. J. (2010). "Implications for GAAP from an Analysis of Positive Research in Accounting." *Journal of Accounting and Economics*, 50 (2-3): 246 – 286.

Kothari, S. P., Sabino, J. S. and Zach, T. (2005). "Implications of Survival and Data Trimming for Tests of Market Efficiency." *Journal of Accounting and Economics*, 39: 129 – 161.

Kothari, S. P. and Warner, J. B. (1997). "Measuring Long-Horizon Security Price Performance." *Journal of Financial Economics*, 43 (3): 301 – 339.

Kuo, J. M., Philip, D. and Zhang, Q. (2013). "What Drives the Disappearing Dividends Phenomenon?" *Journal of Banking and Finance*, 37 (9): 3499 – 3514.

Kwok, C. C. Y. and Tadesse, S. (2006). "National Culture and Financial Systems." *Journal of International Business Studies*, 37 (2): 227 – 247.

La Porta, R., Lopez-De-Silanes, F., Shleifer, A. and Vishny, R. W. (1997). "Legal Determinants of External Finance." *The Journal of Finance*, 52 (3): 1131 – 1150.

La Porta, R., Lopez-De-Silanes, F., Shleifer, A. and Vishny, R. W. (1998). "Law and Finance." *Journal of Political Economy*, 106 (6): 1113 – 1155.

La Porta, R., Lopez-De-Silanes, F., Shleifer, A. and Vishny, R. W. (1999). "The Quality of Government." *Journal of Law, Economics, and Organization*, 15 (1): 222 – 279.

La Porta, R., Lopez-De-Silanes, F., Shleifer, A. and Vishny, R. W. (2006). "What Works in Securities Laws?" *The Journal of* Finance, 61 (1): 1 – 32.

Lakonishok, J., Shleifer, A. and Vishny, R. (1994). "Contrarian Investment, Extrapolation and Risk." *The Journal of Finance*, 44: 1541 – 1578.

Larcker, , D. F. (1989). "Discussion of Accounting Measurement, Price-Earnings Ratios, and the Information Content of Security Prices." *Journal of Accounting Research*, 27: 145 – 152.

Lee, C., Myers, J. and Swaminathan, B. (1999). "What is the Intrinsic Value of the Dow?" *The Journal of Finance* 54 (5): 1693 – 1741.

Lee, L. F. (2012). "Incentives to Inflate Reported Cash from Operations Using Classification and Timing." *The Accounting Review*, 87 (1): 1 – 33.

LeRoy, S. (1976). "Efficient Capital Markets: Comment." *The Journal of Finance*, 31 (1): 139 – 141.

Leuz, C., Nanda, D. and Wysocki, P. D. (2003). "Earnings Management and Investor Protection: An International Comparison." *Journal of Financial Economics*, 69 (3): 505 – 527.

Lev, B. and Thiagarajan, R. (1993). "Fundamental Information Analysis." *Journal of Accounting Research*, 31 (2): 190 – 215.

Levis, M. (1989). "Stock Market Anomalies. A Re-Assessment Based on the UK Evidence." *Journal of Banking and Finance*, 13 (4-5): 675 – 696.

Lesmond, D. A., Schill, M. J. and Zhou, C. (2004). "The Illusory Nature of Momentum Profits." *Journal of Financial Economics*, 71 (2): 349 – 380.

Lewellen, J. (2010). "Accounting Anomalies and Fundamental Analysis: An Alternative View." *Journal of Accounting and Economics*, 50 (2-3): 455 – 466.

Lintner, J. (1965). "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, 47 (1): 13 – 37.

Liang, S. X. and Wei, J. K. C. (2012). "Liquidity Risk and Stock Returns Around the World." *Journal of Banking and Finance,* 36 (12): 3274 – 3288.

Liu, W. (2006). "A Liquidity-Augmented Capital Asset Pricing Model." *Journal of Financial Economics*, 82 (3): 631 – 671.

Lo, A. W. (2004). "The Adaptive Market Hypothesis." *The Journal of Portfolio Management,* 30 (5): 15 – 29.

Lo, A. W. and MacKinlay, C. (1988). "Stock Market Prices Do not Follow Random Walks: Evidence from a Simple Specification Test." *Review of Financial Studies,* 1 (1): 41 – 66.

Lo, A. W. and Mueller, M. T. (2010). "Warning: Physics Envy May Be Hazardous to Your Wealth!" *Available at SSRN.*

Lo, A. W. and Repin, D. V. (2002). "The Psychophysiology of Real-Time Financial Risk Processing." *Journal of Cognitive Neuroscience,* 14 (3): 323 – 339.

Lo, A. W., Repin, D. V. and Steenbarger, B. N. (2005). "Fear and Greed in Financial Markets: A Clinical Study of Day-Traders." *American Economic Review,* 95 (2): 352 – 359.

Longstaff, F. A. (2004). "The Flight-to-Liquidity Premium in US Treasure Bond Prices." *Journal of Business,* 77 (3): 511 – 526.

MacKenzie, D. (2003). "Long-Term Capital Management and the Sociology of Arbitrage." *Economy and Society,* 32 (3): 349 – 380.

MacKenzie, D. and Millo, Y. (2003). "Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange." *American Journal of Sociology*, 109 (1): 107 – 145.

MacKinlay, A. C. (1995). "Multifactor Models do not Explain Deviations from the CAPM." *Journal of Financial Economics,* 38 (1): 3 – 28.

MacKinlay, A. C. (1997). "Event Studies in Economics and Finance." *Journal of Economic Literature,* 35 (1): 13 – 39.

Malkiel, B. G. (2003). "A Random Walk Down Wall Street." New York: Norton.

Mann, H. B. and Whitney, D. R. (1947). "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other." *The Annals of Mathematical Statistics,* 18 (1): 50 – 60.

Marginson, D. and McAulay, L. (2008). "Exploring the Debate on Short-Termism: A Theoretical and Empirical Analysis." *Strategic Management Journal,* 29 (3): 273 – 292.

Markowitz, H. (1959). "Portfolio Selection: Efficient Diversification of Investments." New York: John Wiley.

Martin, G. S. and Puthenpurackal, J. (2008). "Imitation Is the Sincerest Form of Flattery: Warren Buffett and Berkshire Hathaway." *Available at SSRN.*

Maury, B. (2006). “Family Ownership and Firm Performance: Empirical Evidence from Western European Corporations.” *Journal of Corporate Finance,* 12 (2): 321 – 341.

McLean, R. D., Zhang, T. and Zhao, M. (2012). “Why Does the Law Matter? Investor Protection and Its Effects on Investment, Finance, and Growth.” *The Journal of Finance,* 67 (1): 313 – 350.

Megginson, W. L. and Netter, J. M. (2001). “From State to Market: A Survey of Empirical Studies on Privatization.” *Journal of Economic Literature,* 39 (2): 321 – 389.

Merton, R. (1973). “An Intertemporal Capital Asset Pricing Model.” *Econometrica,* 41: 867 – 887.

Mohanram, P. S. (2005). “Separating Winners from Losers among Low Book-to-Market Stocks Using Financial Statement Analysis.” *Review of Accounting Studies,* 10 (2-3): 133 – 170.

Mood, A. M. (1950). “Introduction to the Theory of Statistics.” New York: McGraw-Hill.

Moosa, I. A. (2013). “The Capital Asset Pricing Model (CAPM): The History of a Failed Revolutionary Idea in Finance? Comments and Extensions.” *Abacus,* 49 (1): 62 – 68.

Morgan, G. and Thomas, S. (1998). “Taxes, Dividend Yields and Returns in the UK Equity Market.” *Journal of Banking and Finance,* 22 (4): 405 – 423.

Moskowitz, T. J. and Grinblatt, M. (1999). “Do Industries Explain Momentum?” *The Journal of Finance,* 54 (4): 1249 – 1290.

Mossin, J. (1966). “Equilibrium in a Capital Asset Market.” *Econometrica,* 34 (4): 768 – 783.

Moy, R. L. (2014). “Think, Act, and Invest Like Warren Buffett: The Winning Strategy to Help You Achieve Your Financial and Life Goals.” *Financial Analysts Journal,* 70 (1): 67 – 68.

Muscarella, C. J. and Vetsuypens, M. R. (1996). “Stock Splits: Signaling or Liquidity? The Case of ADR ‘Solo-Splits’.” *Journal of Financial Economics,* 42 (1): 3 – 26.

Næs, R., Skjeltorp, J. A. and Ødegaard, B. A. (2011). “Stock Market Liquidity and the Business Cycle.” *The Journal of Finance,* 66 (1): 139 – 176.

Neave, H. R. and Worthington, P. L. (1988). “Distribution-Free Tests.” London: Unwin Hyman.

Ng, L. and Wu, F. (2010). “Peer Effects in the Trading Decisions of Individual Investors.” *Financial Management,* 39 (2): 807 – 831.

Odean, T. (1998). “Volume, Volatility, Price, and Profit when All Traders are Above Average.” *The Journal of Finance,* 53 (6): 1887 – 1934.

O’Hara, M. (2008). “Bubbles: Some Perspectives (and Loose Talk) from History.” *Review of Financial Studies,* 21 (1): 11 – 17.

Ou, J. A. and Penman, S. H. (1989). “Financial Statement Analysis and the Prediction of Stock Returns.” *Journal of Accounting and Economics*, 11 (4): 295 – 329.

Pástor, Ľ. and Stambaugh, R. F. (2003). “Liquidity Risk and Expected Stock Returns.” *Journal of Political Economy,* 111 (3): 642 – 685.

Pevzner, M., Xie, F. and Xin, X. (2013). “When Firms Talk, Do Investors Listen? The Role of Trust in Stock Market Reactions to Corporate Earnings Announcements.” *Journal of Financial Economics*, forthcoming.

Pincus, M., Rajgopal, S. and Venkatachalam, M. (2007). “The Accrual Anomaly: International Evidence.” *The Accounting Review,* 82 (1): 169 – 203.

Piotroski, J. D. (2000). “Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers.” *Journal of Accounting Research,* 38: 1 – 41.

Piotroski, J. D. (2005). “Discussion of Separating Winners from Losers among Low Book-to-Market Stocks Using Financial Statement Analysis.” *Review of Accounting Studies* 10 (2-3): 171 – 184.

Piotroski, J. D. and So, E. C. (2012). “Identifying Expectation Errors in Value/Glamour Strategies: A Fundamental Analysis Approach.” *Review of Financial Studies,* 25 (9): 2841 – 2875.

Piotroski, J. D. and Srinivasan, S. (2008). “Regulation and Bonding: The Sarbanes-Oxley Act and the Flow of International Listings.” *Journal of Accounting Research,* 46 (2): 383 – 425.

Ponterotto, J. G. (2005). “Qualitative Research in Counseling Psychology: A Primer on Research Paradigms and Philosophy of Science.” *Journal of Counseling Psychology*, 52 (2): 126 – 136.

Poterba, J. and Summers, L. H. (1988). “Mean Reversion in Stock Prices: Evidence and Implications.” *Journal of Financial Economics,* 22 (1): 27 – 59.

Punch, K. F. (2006). “Developing Effective Research Proposals.” London: Sage.

Rapach, D. E., Strauss, J. K. and Zhou, G. F. (2010). "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, 23 (2): 821 – 862.

Richardson, S., Tuna, I. and Wysocki, P. (2010). “Accounting Anomalies and Fundamental Analysis: A Review of Recent Research Advances.” *Journal of Accounting and Economics,* 50 (2-3): 410 – 454.

Ritter, J. R. (2003). “Behavioral Finance”. *Pacific-Basin Journal,* 11: 429 – 437.

Roll, R. (1981). “A Possible Explanation of the Small Firm Effect.” *The Journal of Finance,* 36: 879 – 888.

Roosenboom, P. and van Dijk, M. A. (2009). “The Market Reaction to Cross-Listings: Does the Destination Market Matter?” *Journal of Banking and Finance,* 33 (10): 1898 – 1908.

Rosenberg, B., Reid, K. and Lanstein, R. (1985) “Persuasive Evidence of Market Inefficiency.” *Journal of Portfolio Management*, 11: 9 – 16.

Ross, S. A. (1976). “The Arbitrage Theory of Capital Asset Pricing.” *Journal of Economic Theory*, 13: 341 – 360.

Ryan, R. J. (1982). “Capital Market Theory – A Case Study in Methodological Conflict.” *Journal of Business Finance and Accounting*, 9 (4): 443 – 458.

Sabherwal, S., Sarkar, S. K. and Zhang, Y. (2011). “Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News.” *Journal of Business Finance & Accounting*, 38 (9-10): 1209 – 1237.

Samuelson, P. A. (1965). “Proof that Properly Anticipated Prices Fluctuate Randomly.” *Industrial Management Review*, 6: 41 – 50.

Schiereck, D. De Bondt, W. and Weber, M. (1999). “Contrarian and Momentum Strategies in Germany.” *Financial Analysts Journal*, 55 (6): 104 – 116.

Schrimpf, A., Schröder, M. and Stehle, R. (2007). “Cross-Sectional Tests of Conditional Asset Pricing Models: Evidence from the German Stock Market.” *European Financial Management*, 13 (5): 880 – 907.

Sharpe, W. (1964). “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *The Journal of Finance*, 19 (3): 425 – 443.

Shiller, R. J. (1981). “Do Stock Prices Move too Much to Be Justified by Subsequent Changes in Dividends?” *American Economic Review*, 71 (3): 421 – 436.

Shiller, R. J. (2003). “From Efficient Markets Theory to Behavioral Finance.” *Journal of Economic* Perspectives, 17 (1): 83 – 104.

Shiller, R. J. (2010). “How Should the Financial Crisis Change How We Teach Economics?” *Journal of Economic Education*, 41 (4): 403 – 409.

Shleifer, A. and Summers, L. H. (1990). “The Noise Trader Approach to Finance.” *Journal of Economic Perspectives*, 4 (2): 19 – 33.

Shleifer, A. and Vishny, R. W. (1997). “The Limits of Arbitrage.” *The Journal of Finance*, 52 (1): 35 – 55.

Siegel, J. J. (2002). “Stocks for the Long Run”. 3rd edition. New York: MacGraw-Hill.

Siganos, A. (2012). “Can Retail Investors Exploit Stock Market Anomalies?” *Applied Financial Economics*, 22 (7): 537 – 547.

Simon, H. A. (1955). “A Behavioral Model of Rational Choice.” *The Quarterly Journal of Economics*, 69 (1): 99 – 118.

Sloan, R. G. (1996). “Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings?” *The Accounting Review*, 71 (3): 289 – 315.

Smith, C. (1976). “Option Pricing; A Review.” *Journal of Financial Economics*, 3 (1-2): 3 – 51.

Smith, G. (2012). “The Changing and Relative Efficiency of European Emerging Stock Markets.” *European Journal of Finance*, 18 (8): 689 - 708.

Spiegel, M. I. and Wang, X. (2005). “Cross-Sectional Variation in Stock Returns: Liquidity and Idiosyncratic Risk.” *Yale ICF Working Paper No. 05-13. Available at SSRN.*

Statman, M. (2005). “Normal Investors, Then and Now.” *Financial Analysts* Journal, 61 (2): 31 – 37.

Stock, J. H., and Watson, M. W. (2004). “Combination Forecasts of Output Growth in a Seven-Country Data Set.” *Journal of Forecasting*, 23 (6): 405 – 430.

Stulz, R. M. (2009). “Securites Laws, Disclosure, and National Capital Markets in the Age of Financial Globalization.” *Journal of Accounting Research*, 47 (2): 349 – 390.

Subrahmanyam, A. (2010). “The Cross-Section of Expected Stock Returns: What Have We Learnt from the Past Twenty-Five Years of Research?”. *European Financial Management*, 16 (1): 27 – 42.

Subrahmanyam, A. and Titman, S. (2001). “Feedback from Stock Prices to Cash Flows.” *The Journal of Finance*, 56 (6): 2389 – 2413.

Tuschke, A. and Sanders, G. (2003). “Antecedents and Consequences of Corporate Governance Reform: The Case of Germany.” *Strategic Management Journal*, 24 (7): 631 – 649.

Tversky, A. and Kahneman, D. (1971). “Belief in the Law of Small Numbers.” *Psychological Bulletin*, 76: 105 – 110.

Tversky, A. and Kahneman, D. (1973). “Availability: A Heuristic for Judging Frequency and Probability.” *Cognitive Psychology*, 5 (2): 207 – 232.

Tversky, A. and Kahneman, D. (1974). “Judgement under Uncertainty: Heuristics and Biases.” *Sciene*, 185 (4157): 1124 – 1131.

Tversky, A. and Kahneman, D. (1991). “Loss Aversion in Riskless Choice: A Reference-Dependent Model.” *Quarterly Journal of Economics*, 106 (4): 1039 – 1061.

Van Dijk, M. A. (2011). “Is Size Dead? A Review of the Size Effect in Equity Returns.” *Journal of Banking and Finance*, 35 (12): 3263 – 3274.

Vayanos, D. (1998). “Transaction Costs and Asset Prices: A Dynamic Equilibrium Model.” *Review of Financial Studies*, 11 (1): 1 – 58.

Von Neumann, J. and Morgenstern, O. (1944). "Theory of Games and Economic Behavior" Princeton, NJ: Princeton University Press.

Wahlen, J. M. and Wieland, M. M. (2011). "Can Financial Statement Analysis Beat Consensus Analysts' Recommendations?" *Review of Accounting Studies*, 16 (1): 89 – 115.

Watanabe, A., Xu, Y., Yao, T. and Yu, T. (2013). "The Asset Growth Effect: Insights From International Equity Markets." *Journal of Financial Economics*, 108 (2): 529 – 563.

Waud, R. N. (1970). "Public Interpretation of Federal Reserve Discount Rate Changes: Evidence on the 'Announcement Effect'." *Econometrica*, 38 (2): 231 – 250.

Welch, I. (2001). "The Top Achievements, Challenges, and Failures of Finance." *Yale ICF working paper no. 00-67. Available at SSRN.*

Welch, I. and Goyal, A. (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 21 (4): 1455 – 1508.

Whitaker, J. (2003). "The Performance of Value and Momentum Investment Portfolios: Recent Experience in the Major European Markets." *Journal of Asset Management*, 4 (4): 221 – 246.

Whitaker, J. (2004). "The Performance of Value and Momentum Investment Portfolios: Recent Experience in the Major European Markets Part 2." *Journal of Asset Management*, 5 (3): 157 – 175.

Wieland, M. M. (2011). "Identifying Consensus Analysts' Earnings Forecasts that Correctly and Incorrectly Predict an Earnings Increase." *Journal of Business Finance and Accounting*, 38 (5-6): 574 – 600.

Wilcoxon, F. (1945). "Individual Comparisons by Ranking Methods." *Biometrics Bulletin*, 1 (6): 80 – 83.

Wray, K. B. (2011). "Kuhn's Evolutionary Social Epistemology." Cambridge: Cambridge University Press.

Zhang, X. F. (2006). "Information Uncertainty and Stock Returns." *The Journal of Finance,* 61 (1): 105 – 136.

Zhang, X. F. (2007). "Accruals, Investment, and the Accrual Anomaly." *The Accounting Review*, 82 (5): 1333 – 1363.