Comparing Factor Models in European Stock Returns

Michael O’Connell

Supervisors:

Professor Jonathan Fletcher

Professor Andrew Marshall

University of Strathclyde

Department of Accounting and Finance

Doctor of Philosophy in Accounting and Finance

August 2024

Declarations

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Professor Jonathan Fletcher, and Professor Andrew Marshall, for their invaluable guidance, unwavering support, and insightful feedback throughout my PhD journey. Their expertise and encouragement have been instrumental in the completion of this dissertation.

I would also like to extend my thanks to Professor Cesare Robotti and Professor Siddartha Chib for the provision of code required to complete this dissertation.

Lastly, I would like to acknowledge the support of my family and friends, whose patience and encouragement have been a constant source of motivation.

Thank you all for your contributions and support.

Abstract

The goal of asset pricing research is to find the optimal model that explains the drivers of asset returns. Historically, this field has predominantly relied on data from the United States, given the extensive and detailed records of its financial markets. Due to the growing interdependence of international markets, recent research has shifted towards leveraging large global datasets to develop universally applicable models. However, empirical evidence suggests that these global models explain less variation in domestic returns compared to country-specific models.

This thesis investigates the effectiveness of country-specific asset pricing models across a set of European markets, utilising both classical and Bayesian methods to assess model performance. The first empirical chapter begins with evaluating the relative performance of nine asset pricing models in developed European stock markets from 1991-2022. Asymptotically valid tests of model comparison, developed by Barillas, Kan, Robotti and Shanken (2020), are conducted, where the extent of model mispricing is gauged by the squared Sharpe ratio improvement measure of Barillas and Shanken (2017). The findings reveal that the Fama and French (2018) six-factor model, with both original and updated value factors, are the top-performing models in most markets. However, variation in the absolute and relative performance of models across samples suggests that a singular optimal European asset pricing model does not exist within the classical framework.

To enhance model performance, the second empirical chapter explores the use of serial correlation in factor returns as conditioning information. Adopting the methodology of Ehsani and Linnainmaa (2022), this chapter shows that multiple investment factors in the cross-country dataset are unconditionally minimum-variance inefficient: factor returns are positively autocorrelated, while risk remains constant regardless of past returns. Using Ferson and Siegel's (2001) general framework, 'time-series efficient factors' are constructed by conditioning factor weights on historical returns to enhance the Sharpe ratios of these factors across the European markets under consideration. A number of these optimised factors achieve significantly higher average Sharpe ratios compared to the original factors, while retaining all the information contained in the original factors. When the model comparison tests of Barillas et al. (2020) are repeated with these optimised factors, the absolute performance of the lower-performing models improves, while the relative performance among the models remains consistent across markets.

In the third and final empirical chapter, the Bayesian framework of Chib, Zeng, and Zhao (2020) is used to identify the optimal combination of factors from a starting collection of 12 risk factors in each European market. The results indicate that the optimal combinations of factors are similar to the top-performing models in the classical tests. The optimal model from the scan either represents a reduced form with one or two fewer factors or an extension of the top model identified in Chapter Two, with one or two additional factors. This alignment underscores the robustness of the model selection across different testing methodologies. The changes in these optimal combinations are then examined under the assumptions of both normality and multivariate-t distributions on the factor data. Employing the methodology of Chib and Zeng (2020), the analysis reveals no significant disparities in results when a Student-t distribution is assumed for the factor data. Additionally, the extent to which the efficient factor transformation impacts the model comparison tests in each market is analysed. The findings reveal that certain efficient factors are present in the optimal combination of factors across European markets.

Table of Contents

Chapter 1 Introduction 12
1.1 Asset Pricing Overview 13
1.2 Rationale for research on this topic 15
1.3 Research Aims 20
1.4 Main Findings 21
1.5 Key Contributions 23
Chapter 2 Comparing Asset Pricing Models 24
2.1 Introduction 25
2.2 Overview of Asset Pricing Theory 27
2.2.1 Stochastic Discount Factor 27
2.2.2 Asset Pricing Model Development 29
2.2.3 Model Composition 32
2.3 Model Comparison 34
2.3.1 Mean Variance Efficiency 34
2.3.2 Test Asset Irrelevance 38
2.4 Research Methods 39
2.4.1 Pairwise Model Comparison 39
2.4.2 Multiple Model Comparison 41
2.5 Data and Models 43
2.5.1 Dataset 43
2.5.2 Notable Research in Markets Of Interest 44
2.5.3 Model Selection 46
2.5.4 Summary Statistics 48
2.6 Empirical Results 49
2.6.1 Tests of Equality of Squared Sharpe Ratios 49
2.6.2 Multiple Model Comparison Tests 52
2.6.3 Simulation Evidence 54
2.7 Conclusions 56
Appendix A 59
A.I U.K. Empirical Results 59
A.II Summary Statistics Remaining Markets 64
A.III Tests of Equality of Squared Sharpe Ratio Remaining Markets 69
A.IV Multiple Model Comparison Remaining Markets 74
A.V Simulation Evidence for Full Sample 79

Chapter 3 Time Series Efficient Factors 85
3.1 Introduction 86
3.2 Momentum in Stock Returns 88
3.2.1 Time Series Momentum 88
3.2.3 Factor Momentum 89
3.3 Research Methods 94
3.3.1 General Framework 94
3.3.2 Time Series Efficient Factors 95
3.4 Data 97
3.5 Empirical Results 97
3.5.1 Time Series Efficient Factors 97
3.5.2 Improvements in Sharpe Ratio 101
3.5.3 Time-Series Efficient versus Inefficient Factors 103
3.5.4 Momentum Factor 105
3.5.5 Model Comparison with Time-Series Efficient Factors 106
3.6 Conclusions 108
Appendix B 111
B.I U.K Empirical Results 111
B.II Risk Return Characteristics for Remaining Markets 118
B.III Realised Improvements in Sharpe Ratio for Remaining Markets 123
B.IV Efficient Factor Alpha Spanning Tests for Remaining Markets 128
B.V Momentum Factor Alpha Tests for Remaining Markets 133
B.VI Model Comparison Tests with Efficient Factors Remaining Markets 138
Chapter 4 Bayesian Tests of Model Comparison 143
4.1 Introduction 144
4.2 Literature Review 146
4.3 Research Methods 148
4.3.1 Model Scan Approach 148
4.3.2 Prior Computation 150
4.3.3 Marginal Likelihood 151
4.3.4 Summary Statistics of Risk Factors 154
4.3.5 Student-t Distribution 155
4.4 Empirical Results 159
4.4.1 Model Performance in Bayesian Framework 159
4.4.2 Bayesian Model Scan 161

4.4.3 Summary Statistics of Best Model Risk Factors 165
4.4.4 Multivariate-t Assumption....166
4.4.5 Overall Performance Gaussian v Student-t 168
4.4.6 Impact of Efficient Factors....170
4.5 Conclusions....171
Appendix C 174
C.I U.K Empirical Results....174
C.II Bayesian Model Scan for Remaining Markets 180
C.III Summary Statistics of Posterior Distributions in Top Models Remaining Markets .185
C.IV Bayesian Model Scan Assuming Multivariate t for Remaining Markets....190
C.V Gaussian v Student t Overall Performance Remaining Markets 195
C.VI Bayesian Model Scan with Time Series Efficient Factors Remaining Markets 200
C.VII Prior Computation....205
Chapter 5 Conclusions 206
5.1 Summary of Findings....207
5.1 Areas of Further Research....210
Bibliography 212

List of Tables

Table 2.1. Summary Statistics for Monthly U.K. Factor Returns 59
Table 2.2. Summary Statistics for Monthly Returns Remaining Markets 60
Table 2.3. U.K. Tests of Equality of Squared Sharpe Ratios 61
Table 2.4. Top Performing Factor Models Across European Markets 62
Table 2.5. U.K. Multiple Model Comparison 63
Table A.1. Summary Statistics for Monthly French Factor Returns 64
Table A.2. Summary Statistics for Monthly German Factor Returns....65
Table A.3. Summary Statistics for Monthly Netherlands Factor Returns 66
Table A.4. Summary Statistics for Monthly Italian Factors Returns....67
Table A.5. Summary Statistics for Spanish Monthly Factor Returns....68
Table A.6. France Tests of Equality of Squared Sharpe Ratio....69
Table A.7. Germany Tests of Equality of Squared Sharpe Ratios 70
Table A.8. Netherlands Tests of Equality of Squared Sharpe Ratios....71

Table A.9. Italy Tests of Equality of Squared Sharpe Ratios....72
Table A.10. Spain Tests of Equality of Squared Sharpe Ratios....73
Table A.11. France Multiple Model Comparison....74
Table A.12. Germany Multiple Model Comparison....75
Table A.13. Netherlands Multiple Model Comparison....76
Table A.14. Italy Multiple Model Comparison....77
Table A.15. Spain Multiple Model Comparison....78
Table A.16. U.K. Multiple Model Comparison Simulation Evidence....79
Table A.17. France Multiple Model Comparison Simulation Evidence....80
Table A.18. Germany Multiple Model Comparison Simulation Evidence....81
Table A.19. Netherlands Multiple Model Comparison Simulation Evidence....82
Table A.20. Italy Multiple Model Comparison Simulation Evidence....83
Table A.21. Spain Multiple Model Comparison Simulation Evidence....84
Table 3.1. Predictive Return Beta Values....111
Table 3.2. U.K. Returns and Volatility Conditional on Month t-1 Returns....112
Table 3.3. U.K. Realised Efficient Factor Sharpe Ratio Improvements....113
Table 3.4. Realised Efficient Factor Sharpe Ratio Improvements....114
Table 3.5. U.K. Time-Series Efficient Five-Factor Model: Alphas Spanning Tests....115
Table 3.6. U.K. Momentum versus Fama and French (2015) Five-Factor Model....116
Table 3.7. U.K. Model Comparison with Efficient Factors....117
Table B.1. France Returns and Volatility Conditional on Month t-1 Returns....118
Table B.2. Germany Returns and Volatility Conditional on Month t-1 Returns....119
Table B.3. Netherlands Returns and Volatility Conditional on Month t-1 Returns....120
Table B.4. Italy Returns and Volatility Conditional on Month t-1 Returns....121
Table B.5. Spain Returns and Volatility Conditional on Month t-1 Returns....122
Table B.6. France Realised Efficient Factor Sharpe Ratio Improvements....123
Table B.7. Germany Realised Efficient Factor Sharpe Ratio Improvements....124
Table B.8. Netherlands Realised Efficient Factor Sharpe Ratio Improvements....125
Table B.9. Italy Realised Efficient Factor Sharpe Ratio Improvements....126
Table B.10. Spain Realised Efficient Factor Sharpe Ratio Improvements....127
Table B.11. France Efficient Five-Factor Model: Alphas from Spanning Tests....128
Table B.12. Germany Efficient Five-Factor Model: Alphas from Spanning Tests....129

Table B.13. Netherlands Efficient Five-Factor Model: Alphas from Spanning Tests 130
Table B.14. Italy Efficient Five-Factor Model: Alphas from Spanning Tests 131
Table B.15. Spain Efficient Five-Factor Model: Alphas from Spanning Tests 132
Table B.16. Frace Momentum Versus Fama and French (2015) Model................................133
Table B.17. Germany Momentum Versus Fama and French (2015) Model..........................134
Table B.18. Netherlands Momentum Versus Fama and French (2015) Model 135
Table B.19. Italy Momentum Versus Fama and French (2015) Model 136
Table B.20. Spain Momentum Versus Fama and French (2015) Model................................137
Table B.21. France Model Comparison Tests with Efficient Factors 138
Table B.22. Germany Model Comparison Tests with Efficient Factors 139
Table B.23. Netherlands Model Comparison Tests with Efficient Factors............................140
Table B.24. Italy Model Comparison Tests with Efficient Factors..141
Table B.25. Spain Model Comparison Tests with Efficient Factors......................................142
Table 4.1. Model Performance in a Bayesian Framework 174
Table 4.2. U.K. Model Scan 12 Factors 175
Table 4.3. U.K. Summary Statistics Posterior Distribution Best Model Factors 176
Table 4.4. U.K. Model Scan of 12 Factors Multivariate-t Factor Distribution 177
Table 4.5. U.K. Log-Marginal Likelihoods of the Top Performing Models 178
Table 4.6. U.K. Model Scan 24 Factors 179
Table C.1. France Model Scan 12 Factors 180
Table C.2. Germany Model Scan 12 Factors 181
Table C.3. Netherlands Model Scan 12 Factors..182
Table C.4. Italy Model Scan 12 Factors..183
Table C.5. Spain Model Scan 12 Factors 184
Table C.6. France Summary Posterior Distribution of the Best Model Factors 185
Table C.7. Germany Summary Posterior Distribution of the Best Model Factors 186
Table C.8. Netherlands Summary Posterior Distribution of the Best Model Factors.............187
Table C.9. Italy Summary Posterior Distribution of the Best Model Factors.........................188
Table C.10. Spain Summary Posterior Distribution of the Best Model Factors 189
Table C.11. France Model Scan of 12 Factors Multivariate-t Factor Distribution.................190
Table C.12. Germany Model Scan of 12 Factors Multivariate-t Factor Distribution.............191
Table C.13. Netherlands Model Scan of 12 Factors Multivariate-t Factor Distribution 192

Table C.14. Italy Model Scan of 12 Factors Multivariate-t Factor Distribution 193
Table C.15. Spain Model Scan of 12 Factors Multivariate-t Factor Distribution...................194
Table C.16. France Log-Marginal Likelihoods of the Top Performing Models 195
Table C.17. Germany Log-Marginal Likelihoods of the Top Performing Models 196
Table C.18. Netherlands Log-Marginal Likelihoods of the Top Performing Models.............197
Table C.19. Italy Log-Marginal Likelihoods of the Top Performing Models 198
Table C.20. Spain Log-Marginal Likelihoods of the Top Performing Models.......................199
Table C.21. France Model Scan 24 Factors 200
Table C.22. Germany Model Scan 24 Factors 201
Table C.23. Netherlands Model Scan 24 Factors...202
Table C.24. Italy Model Scan 24 Factors...203
Table C.25. Spain Model Scan 24 Factors 204

Chapter 1 Introduction

Abstract

This chapter introduces the thesis; outlining the current landscape of the asset pricing literature, rationale for this thesis, research questions to be addressed in the three empirical chapters, and the key contributions of the study.

1.1 Asset Pricing Overview

Asset pricing research aims to explain and predict financial asset returns by analysing their fundamental drivers in an uncertain world. Initially, asset valuation focused on market exposure, attributing deviations to anomalies or unique asset characteristics. However, this view has evolved to recognise that many of these anomalies actually stem from systematic factors inherent in financial markets. This shift has revealed that asset returns are influenced not only by market movements but also by various risk factors and economic indicators. Understanding these components has refined theoretical frameworks and improved empirical methods, allowing for a more detailed analysis of asset pricing.

To capture these systematic factors, researchers have developed factor models that approximate the returns of financial securities by identifying key influences. These models are termed "factor models" because the outcome depends on the factors included. They suggest that a security’s return is driven by common factors and the security's unique sensitivities to each factor (factor loadings). The goal is to identify a set of factors—such as value, momentum, size, market, quality, and low volatility—that explains the "cross-section" of returns, or the spread of returns at a specific moment. Since the introduction of the classic Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965), various factor-based models have been developed to better explain the cross-section of expected asset returns.

These factor models are tools that help investors identify and manage investment characteristics that influence the risks and returns of stocks and portfolios. There is ample evidence, both practical observations and empirical data, that portfolio managers frequently utilise the Capital Asset Pricing Model (CAPM) alongside various multifactor models to compute return expectations (see, among others, Brealey, Myers, and Allen (2016), Fischer and Wermers (2012), Gitman and Mercurio (1982), Grinold and Kahn (2000), and Jagannathan and Meier (2002)). In their 2001 study, John Graham and Campbell Harvey surveyed CFOs from approximately 4,440 firms and found that a significant majority of them rely heavily on the Capital Asset Pricing Model (CAPM) for estimating the cost of equity. Specifically, they reported that around 75% of the CFOs use CAPM as their primary tool for this purpose. This highlights the importance of reliable factor models in assisting portfolio managers in identifying key factors that influence a wide range of securities. Extending this understanding to asset management, Ang (2014) outlines how approximately 70% of the variation in active returns can be explained by exposures to systematic factors.

Considering the extensive array of factors that have emerged since the 1960s, finding the optimal model is crucial for the investment decision-making process for both professional and private investors. Model comparison tests help identify the most effective models in explaining asset return variations, enhancing our understanding of market dynamics. Traditionally, much of the research on comparing model performance has relied heavily on data from the United States, primarily due to the extensive and well-established financial market data available there. This historical preference for U.S. data is well-documented in empirical asset pricing research (Karolyi, 2016)[1]. However, drawing general conclusions solely from U.S. data can be misleading, as studies have shown that results obtained from U.S. market data do not always hold true in international markets (e.g., Goyal and Wahal, 2015; Jacobs and Muller, 2020).

In contemporary research, there has been a notable shift away from relying solely on U.S. data, as asset pricing analysis has broadened to include global markets. This trend reflects a growing recognition of the importance of adopting a global perspective in financial studies. According to Bethmann et al. (2023), this shift is driven by the evolving nature of global financial markets and the need for models that accurately represent diverse economic environments. Instead of focusing exclusively on U.S. market dynamics, researchers are now integrating regional and local data to develop global factor models. Recent studies by Hanauer (2020), Qiao, Wang, and Lam (2022), and Huber et al. (2023) illustrate this approach. These studies aggregate data from various countries to form global factors, then conduct model comparison tests on these factors to identify the optimal model that can theoretically explain the cross-sectional variation across all markets.

Nevertheless, a critical issue has arisen with this approach: global factor models have been shown to underperform country-specific models. Studies such as Griffin (2002) and Chaieb, Langlois, and Scaillet (2021) show that domestic factor models explain much more time-series variation in country-specific returns and generally have lower pricing errors relative to global models. Having a misspecified factor model can lead to inaccurate risk assessments, distorted pricing of assets, and misguided investment strategies. Such inaccuracies stem from the model's failure to capture essential market dynamics, resulting in misleading risk-return trade-offs.

[1] Karolyi (2016) finds that only 16% of all empirical studies published in the top four finance journals are non-US based papers. In the top fourteen finance journals this figure rises to 23%.

Consequently, investors might make suboptimal decisions, allocating resources based on flawed understandings of market behaviour.

In recent years, European markets have been subject to this aggregation for the purpose of testing global factor models, with the literature on model testing in a local or country-specific setting remaining quite limited. Except for the U.K., there is a lack of research focused on analysing stock returns at a local level across European markets to understand unique market drivers. Despite these research gaps, the study of linear factor models in a European setting remains a growing area. This is not surprising given that the European asset management market, valued at USD 32.20 trillion in 2023, is forecasted to grow to USD 41.61 trillion by 2028, with a compound annual growth rate of 5.26%. Nearly 55% of these assets are managed for investment funds and 45% for discretionary mandates. The market is dominated by a few countries, where the United Kingdom, France, Germany, Switzerland, Netherlands, Italy and Spain comprise about 88% of the total market[2].

The study of Pukthuanthong et al., (2023) is the most recent example of efforts to understand model performance across European markets. Their work is somewhat aligned with my broader research objective; however it lacks the extensive and comprehensive approach needed to fully understand these models in a pan-European context. This thesis makes a significant contribution to the ongoing discussion regarding the effectiveness of local models in international asset pricing. Specifically, this study aims to examine the performance of a comprehensive set of prominent factor models using both classical and Bayesian model comparison frameworks across a sample of developed European markets. A novel optimisation methodology is then applied in an effort to enhance model performance.

1.2 Rationale for research on this topic

The shift towards including non-U.S. research in asset pricing reflects a broader move towards an international perspective in financial studies. The global testing of aggregated data is based on the belief that integrated financial markets share similar forces, suggesting a single, optimal model could explain variability worldwide. This hypothesis underpins the focus on global factors in recent model comparison research, as seen in studies such as Hanauer (2020), Qiao, Wang, and Lam (2022), and Huber et al. (2023).

[2] https://www.mordorintelligence.com/industry-reports/europe-asset-management-industry/market-size

Typically, as seen in Hanauer (2020), regional factors are formed by first applying country-specific static and dynamic filters to market and accounting data. Countries are then categorised into regional groups such as Asia Pacific, Europe, Japan, and emerging markets. Within each region, factors like size (SMB) and value (HML) are constructed based on 2x3 sorts of size and other characteristics. To form global factors, the regional factors are aggregated. For example, the global size factor (SMB) is the average of the size factors calculated for each region. This standardisation ensures that the global factors are comparable across different regions and time periods

However, as mentioned previously, it has been found that these global factor models perform poorly in asset pricing tests at a local and regional level. Studies by Griffin (2002) and Hou, Karolyi, and Kho (2011) reveal that local models outperform global models within international stock markets. The metric used in these model comparison tests is typically the pricing error (alpha), which represents the unexplained variations in asset returns after accounting for the model's factors. Griffin (2002) finds that domestic factor models generally have lower pricing errors than global factor models when examining returns in the U.S., Japan, the U.K., and Canada. The author outline how country-specific three-factor models are more useful in explaining average stock returns than are international versions and concluded that for practical applications like cost of capital calculations and performance evaluations, a country-specific approach to the three-factor model is preferable. Hou, Karolyi, and Kho (2011) also tested local versus global versions of the Capital Asset Pricing Model (CAPM) and a three-factor model incorporating market, momentum, and value factors. They found that local models consistently outperformed their global counterparts, with lower pricing errors indicating better model performance.

Fama and French (2012) build on these findings by examining the performance of global and local versions of the Fama-French four-factor model. They found that local models generally surpassed global models in performance, particularly when explaining average returns on regional portfolios. For example, the local model for Japan provided a significantly tighter fit and more accurate explanation of Japanese asset returns compared to the global model. Performance was measured using various statistical metrics, including the average absolute intercepts and the Sharpe ratio for the intercepts. The results indicated that local models are better able to capture variation in regional asset returns.

This outperformance has persisted over time, with recent examples including Chaieb, Langlois, and Scaillet (2021), who examine the performance of global, regional, and local factor models in explaining individual stock returns across 47 countries from 1985 to 2018. They find that global factor models consistently underperform local factor models. The study highlights the necessity of including local market factors to accurately capture the factor structure in both developed and emerging markets. The authors demonstrate that neither global nor regional risk factors, nor currency considerations, can fully subsume the importance of local market factors. Their findings indicate that local factors carry significant risk premia and that models incorporating these local factors provide better explanations of stock returns, reaffirming the need for localised approaches in international asset pricing.

Similarly, across a comprehensive dataset covering 48 MSCI developed and emerging markets, Hollstein (2022) compares the performance of global, regional, and local versions of various factor models, including the CAPM, the Fama and French three-factor model, the Carhart four-factor model, the Fama and French five-factor model, and several others. The study focuses on the ability of these models to explain returns of anomaly portfolios across developed and emerging markets. The findings reveal that global factor models consistently underperform local factor models in explaining anomaly returns. The average absolute alphas for portfolios using global models are higher than those using local models, indicating worse performance. Specifically, global models result in annual anomaly portfolio alphas that are, on average, 1.7 percentage points higher than local models, while regional models result in alphas that are 1.1 percentage points higher than local models.

Huber et al., (2023) also note in their study of global and regional asset pricing models that while factor models which include both regional and global factor versions tend to be powerful in explaining cross sectional returns, the contribution of regional factors tends to be larger, consistent with the notion that international equity markets are partially segmented and partially integrated. A recent study by Pukthuanthong et al. in (2023) also showed local factor models generally have lower average alphas than regional and global models when explaining a range of cross-sectional anomalies across a range of global markets.

However, concentrating on a European sample, it might be intuitive to assume that the geographical closeness of European countries would lead to similar financial market behaviours. This similarity could be further supported by shared regulations, norms, and common currencies in certain regions. Consequently, one might infer that larger regional and global model, with their shared influences on financial markets, would be superior to local models.

However, academic research has demonstrated that this assumption does not hold true. Mirza and Afzal (2011) examined the performance of the Fama and French three-factor model on stocks from 15 European countries. Due to the increasing integration of global markets and the rising correlations of stocks between countries, the authors conducted their analysis at a global portfolio level. They used the MSCI EMU Index, which captures large and mid-cap representation across 10 developed countries in the European Economic and Monetary Union (EMU), as the market portfolio. Their findings revealed that the three-factor model failed to explain the variations in portfolios sorted from country returns. The findings indicate that the model performs poorly, with significant pricing errors (alphas) and an often insignificant market premium, highlighting the limitations of using a global model without accounting for local market factors. These results are consistent with Griffin (2002), who suggested that the three-factor model is domestic in nature and performs poorly for global portfolios. Some researchers, such as Fama and French (1998) and Hau (2011), suggest that stocks are globally priced.

Following Griffin's influential 2002 study, questions arose about the variation in market forces across countries and regions, despite the interconnected nature of global markets. Karolyi and Stulz (2003) reviewed the international finance literature to assess the extent to which global factors affect financial asset demands and prices. They found that theoretical asset-pricing models based on mean-variance optimising investors fail to explain the portfolio holdings of investors who exhibit a home bias. Karolyi and Stulz highlighted that this preference for local investments significantly enhances the impact of local factors on asset prices. They argued that models assuming perfectly integrated international financial markets face substantial challenges in accurately explaining both the composition and evolution of portfolio holdings over time. Their work underscores that while home bias decreases the relevance of international determinants of domestic stock prices, it does not entirely negate the influence of global factors. Instead, it suggests a complex interplay where local influences are pronounced due to investor preferences, even as global market integration and cross-country equity flows exert significant effects on asset prices. Despite the theoretical expectation that investors would hold a diversified portfolio of equities across the world if capital were fully mobile across borders, home bias in equities remains prevalent in most countries, though it is slowly decreasing and tends to be higher in emerging markets (Kilka and Webber, 2000). French and Poterba (1991) report that U.S. investors hold 93.8% of their equity portfolio in domestic stocks, Japanese investors 98.1%, and U.K. investors 82%. These figures have not fallen as expected over the course of the 2000s; for instance, in 2007, U.S. investors held more than 80% of domestic equities, which is much higher than the proportion of U.S. equities to the world market portfolio (Ardalan, 2019).

Empirically demonstrating these concerns, Chaieb, Langlois, and Scaillet (2021) explore 'aggregation bias,' where trends in aggregated data are wrongly applied to local markets. Their research shows that combining assets into portfolios can misrepresent factor exposures, emphasising the importance of considering local markets to understand asset pricing in both developed and emerging markets. They find that local market factors remain significant even when including global, regional, or currency risk factors, challenging the belief that global integration diminishes the importance of local factors. Despite this, there is still limited literature on identifying the optimal asset pricing model for individual markets, highlighting a gap in understanding local influences.

Several key features make the European financial landscape unique and valuable for financial studies. Unlike North America or Asia, where major financial markets are spread over vast areas, European markets are in closer proximity, with monetary unity and varying economic regimes. This suggests the presence of both regional and local influences. Despite this, there is a significant gap in research focused on assessing and enhancing asset pricing models specifically for European markets. Beyond the U.K., most developed European markets lack comprehensive examinations of factor models over extended periods.

The most recent study, and to the best of my knowledge the first of its kind, that aims to identify the optimal factor model for international stock markets with a focus on domestic factors, was conducted by Pukthuanthong et al. in (2023). This study evaluates the effectiveness of local, regional, and global models in explaining various cross-sectional anomalies, using a comprehensive international sample that includes 13 European markets. Using the Bayesian model scan tests of Chib et al. (2020), the author identifies the top-performing combination of investment factors in each market and compares their effectiveness in explaining 153 cross-sectional anomalies. In this analysis models with local factors perform best, while there is not much difference in the performance of regional and global models, a finding similar to Hollstein (2022). The study by Pukthuanthong et al. (2023) has limitations, beginning with its exclusive focus on the Bayesian approach for model comparison. Given the size of the sample the authors consider only 10 investment factors in their analysis, two of which are currency based factors. Their sample period runs for 21 years which is a short period of analysis relative to other recent asset pricing tests such as Chib and Zeng (2020) and Qiao, Wang, and Lam (2022).

In this thesis, I examine the performance of a comprehensive set of factor models in European stock returns within both classical and Bayesian asset pricing frameworks over an extended period of analysis. Additionally, I apply a new and innovative method to enhance these models' performance by incorporating conditioning information from past factor returns. My overarching goals are as follows: Firstly, to examine if a single model outperforms across all the markets included in the analysis. Secondly, to determine if model performance can be improved using this optimisation process involving past returns as conditioning information. Beyond these primary objectives, this thesis provides insights into the drivers of stock returns in European markets at a local level. This work marks a notable advancement in European asset pricing studies and contributes significantly to the broader field of financial research.

1.3 Research Aims

The research is presented in three further chapters. The report finishes with a chapter drawing together the main conclusions of the three papers in a unified structure.

Comparing Asset Pricing Models

The first chapter examines the performance of traditional factor models across a selection of European markets, employing the asymptotically valid tests of comparison developed by Barillas, Kan, Robotti and Shanken (2020). The aim is to identify the top performing model in each market and to examine if a single model outperforms across all markets. Furthermore, this chapter provides an overview of the asset pricing theories that underpin the empirical analysis and introduces the dataset to be used in this thesis.

The key research questions addressed in this chapter are: Which asset pricing model performs best in each European market? Is there a single asset pricing model that consistently outperforms across all European markets?

Time Series Efficient Factors

The second chapter explores the concept of 'time-series efficient factors,' where the serial correlation in factor returns is used as conditioning information to enhance the Sharpe ratios of the European factors under consideration. The chapter then examines whether a distinct momentum factor is still necessary in the models across different markets with these efficient factors. Subsequently, the model comparison tests from the previous chapter are repeated to determine the impact of these mean-variance efficient factors on both the absolute and relative performance of the models across countries.

The key research questions addressed in this chapter are: Can conditioning factor weights on historical returns enhance the mean-variance efficiency of factors? Is a distinct momentum factor still necessary when using time-series efficient factors? How do mean-variance efficient factors impact the Sharpe performance of asset pricing models across European markets?

Bayesian Tests of Model Comparison

The final empirical chapter uses the Bayesian model scanning strategy of Chib, Zeng and Zhao (2020) to identify the optimal combination of risk factors from an initial set in each country. The aim is to compare the most effective model found here with the optimal model from the classical tests. Additionally, the chapter aims to assess how the assumption of the distribution of factor data, whether Gaussian or Student-t, impacts the optimal combination of factors. Subsequently, the chapter integrates the time-series efficient factors from the previous chapter into the model scan to investigate whether these mean-variance efficient factors can enhance the performance of factor models.

The key research questions addressed in this chapter are: What is the optimal combination of risk factors in each country according to Bayesian model scanning? How does the distributional assumption (Gaussian or Student-t) impact the optimal combination of factors? Can integrating time-series efficient factors enhance the performance of factor models in a Bayesian framework?

1.4 Main Findings

Comparing Asset Pricing Models

The first empirical chapter shows that both the Fama and French (2018) and Asness et al. (2015) models perform well across the selection of markets. However, the best-performing model varies by country, indicating the presence of local influences in European stock returns. There is also consistency in the relative performance of models across markets, though it does vary by market. These findings are supported by simulation evidence.

For academics, this work significantly contributes to the literature on model comparison in a European setting, providing insights into how different models perform under varying market conditions. On a practitioner level, identifying the best-performing models at a local level allows for more accurate estimations of the cost of capital[3], enhances investment decision-making, and aids in risk assessment by tailoring strategies to specific local characteristics.

Time Series Efficient Factors

In this chapter several risk factors across the European sample are found to be mean-variance inefficient. Using the framework proposed by Ehsani and Linnainmaa (2022), conditioning factor weights on past returns significantly improves their Sharpe ratios. The findings indicate that while optimised factors reduce the importance of a distinct momentum factor, they do not eliminate its necessity. When rerunning the model comparison tests, the efficient factors enhance the absolute performance of underperforming models, with notable improvements in their squared Sharpe ratios. However, the best-performing model identified in the previous chapter and the relative performance of models remain unchanged.

For academics, this study demonstrates that certain factors in European returns are mean-variance inefficient, highlighting the importance of verifying the mean-variance efficiency of factors in future model tests. Practitioners can utilise these optimised factors to improve portfolio performance and enhance risk-adjusted returns by employing a more precise factor weighting strategy.

Bayesian Tests of Model Comparison

The model scan approach shows that the optimal model across markets is different from those identified in the classical tests. However, the optimal models do contain factors present in the top-performing models from the classical framework. This examination uses both multivariate normal and Student's t-distributions for the factor data to account for the presence of fat tails. Neither distributional assumption consistently outperforms across the sample. Additionally, the Bayesian framework shows that the time-series efficient factors enhance the performance of models across the sample.

For both academics and practitioners, this study enhances Bayesian model comparison in a European setting through a longer period of analysis, a broader range of factors, and the

[3] Using historical data can lead to inaccuracies in cost of capital estimates due to potential future market changes, regime shifts, and evolving company risk profiles.

comparison of distributional assumptions. Additionally, it provides a direct comparison between classical and Bayesian model comparison tests, offering valuable insights into the effectiveness of different approaches in identifying optimal asset pricing models.

1.5 Key Contributions

Comparing Asset Pricing Models

This work contributes to the European asset pricing literature by evaluating an extensive array of models, incorporating a broad spectrum of factors over an extended timeframe in European stock returns at a local level. It highlights the local influences on stock returns in European markets, which have been underrepresented in the literature.

Time Series Efficient Factors

This research is the first outside of the U.S. to examine the mean-variance efficiency of factor returns in this manner. We show that this transformation enhances the performance of some factors and factor models, thus contributing to the understanding of optimising factor weights and their impact on asset pricing models in a European setting.

Bayesian Tests of Model Comparison

This study contributes a direct comparison between classical and Bayesian model comparison methods to the European asset pricing literature. It provides insights into the impact of distributional assumptions of European factor data on the model scan. Additionally, it is the first to include 'time-series efficient factors' in a Bayesian model scan framework.

Chapter 2 Comparing Asset Pricing Models

Abstract

This study evaluates the relative performance of nine competing traditional asset pricing models is evaluated across developed European stock markets over the period from 1991 to 2022. I conduct asymptotically valid tests of model comparison developed by Barillas, Kan, Robotti and Shanken (2020) where the extent of model mispricing is gauged by the squared Sharpe ratio improvement measure of Barillas and Shanken (2017). This study finds that the Fama and French (2018) and Asness et al. (2015) 6-factor models emerge as the dominant models in the majority of the markets examined.

2.1 Introduction

As previously mentioned, much of asset pricing research has relied heavily on U.S. data (Karolyi, 2016). Recently, there has been a notable increase in studies conducting model comparison tests on global samples to identify return drivers in non-U.S. markets (Hanauer, 2020; Qiao Wang and Lam, 2022; Huber et al., 2023). This approach involves aggregating accounting and financial data from various countries to form global risk factors for analysis.

Despite the rationale that global market interconnectedness subjects all markets to similar forces, studies (Griffin, 2002; Hou, Karolyi, and Kho, 2011; Fama and French, 2012; Chaieb, Langlois, and Scaillet, 2021) emphasise the importance of local factors in asset pricing. These studies argue that global factor models often mis specify local influences, leading to inaccurate risk assessments and misguided investment strategies. Consequently, investors may make suboptimal decisions based on flawed understandings of market behaviour.

The aim of this chapter is to test the performance of prominent asset pricing models across six distinct European markets. Rather than evaluating them on a global or broad regional scale, this study is focused on individual European countries to account for local market characteristics and variations in financial environments. Europe's developed financial markets, close geographical proximity, and, for the most part, unified economic and monetary systems make it an ideal setting for finance studies. However, differences in currency systems, such as the UK's use of the pound and its recent exit from the European Union (Brexit), present unique challenges that warrant separate consideration. Although the UK remains a major financial hub, the structural differences between the UK and Eurozone countries suggest that models designed for the broader European market may not seamlessly apply.

Furthermore, this study is motivated by the potential local influences that underscore the necessity of customising factor models to account for the distinct economic and market environments in each country. Different markets are driven by unique factors—such as industry compositions and investor behaviour —which can significantly influence asset returns. These variations mean that a one-size-fits-all approach to asset pricing may overlook key drivers specific to each market. As a result, tailored models are needed to better capture the underlying dynamics that affect returns in each country. Despite these favourable conditions, there remains a notable gap in research aimed at assessing and enhancing asset pricing models specifically within these individual European contexts. For many of the countries in my sample, there is little to no examination of a wide range of factor models over extended periods in a classical model comparison framework. By focusing on six specific European countries, this study aims to fill that gap and provide a clearer understanding of how local influences might affect the performance of these models. An overview of some asset pricing and model comparison studies conducted in these markets is provided in Section 2.5.2.

This chapter's focus on testing asset pricing models at the local level across European markets addresses an essential gap in finance research, particularly for non-U.S. regions. While global factor models have become a popular approach in asset pricing studies, these models often overlook the unique, local economic forces that influence asset returns in individual markets. By examining six European countries individually, this study emphasizes the importance of localized risk factors, addressing findings from prior studies that global models can misrepresent or dilute local effects. Given the distinct financial structures, investor behaviours, and economic conditions that vary across European markets—even in regions sharing similar currencies or regulatory frameworks—this chapter aims to refine the understanding of asset pricing models in a way that directly benefits practitioners. Through the use of Barillas et al. (2020)'s asymptotically valid comparison tests and Sharpe ratio-based performance metrics, this research not only contributes to academic literature by assessing model accuracy across diverse financial environments but also provides actionable insights for investors, who can make better-informed decisions with models tailored to specific European markets. Ultimately, this localised approach has implications for improved asset pricing, risk assessment, and investment strategies in diverse financial contexts, highlighting where local factors play a pivotal role.

This study employs the asymptotically valid model comparison tests from Barillas et al. (2020), which facilitate the comparison of non-nested factor models, aligning with the analytical objectives. Model performance is assessed using the squared Sharpe ratio improvement, following the test asset irrelevance framework of Barillas and Shanken (2017). The key idea of this framework is that if test assets are common across regressions, the relative comparison between models is driven solely by the Sharpe ratios of the factor portfolios, providing a robust economic criterion for identifying the 'best-performing model'. My focus is twofold: first, to identify the best-performing asset pricing model in each European market within the sample. This approach is intended to provide practical insights for market participants, highlighting the cross-sectional variability. Second, to determine whether a single model can consistently outperform across the European sample.

These tests find that the Fama and French (2018) 6-factor model with both the original value factor and the updated value factor of Asness and Frazzini (2013) emerges as the dominant models in the majority of the markets examined. However, major consistency in the absolute and relative performance of models across markets is not observed. This indicates that local factors influence asset returns in a European setting, suggesting that a single model is not the optimal choice for all markets. Simulation evidence provides robustness to these results.

2.2 Overview of Asset Pricing Theory

2.2.1 Stochastic Discount Factor

Before exploring factor models, or model comparison frameworks, it is important to establish a foundational understanding of asset pricing theory that underpins all pricing models, not just factor models. Asset pricing theory aims to explain the reasons for variation in asset returns across different assets, known as cross-sectional predictability, and the fluctuation of asset returns over time, referred to as time-series predictability. Additionally, it seeks to understand why stocks typically yield higher returns than riskless bonds, a phenomenon known as the equity premium. Asset pricing theory fundamentally originates from a straightforward concept that the price of an asset equals the expected discounted payoff.

The Law of One Price (LOP) principle asserts that in an efficient market, assets with identical payoffs should sell for the same price. Hansen and Richard (1987) extend this idea to imply that there should be no discrepancies in the pricing of any assets if there are no market frictions like transaction costs or restrictions on trading. Harrison and Kreps (1979), Hansen and Richard (1987), and Ross (1978) show that if the LOP holds in financial markets, then a stochastic discount factor (m_{t+1}) exists such that we can write the price of asset i at time t as:

$$P_{it} = E_t\{m_{t+1}(P_{it+1} + D_{it+1})\}, \qquad \text{for } i=1,..,N \qquad (2.1)$$

where D_{it+1} is the amount of dividends or other payments received at $t + 1$, E_t term is the conditional expectation given the information available to investors at time t, and N is the number of test assets. To calculate the price or value of a financial asset, you multiply its expected future payoffs, which is the combination of future price plus dividends or other payments, by the Stochastic Discount Factor (SDF). This process discounts these payoffs to their present value, setting that the asset's price equals the present value of its expected future payoffs. This means that the pricing error (the difference between the asset's expected price and its market price) is zero.

The No Arbitrage (NA) principle states that market forces tend to align prices of financial assets so as to eliminate arbitrage opportunities. An arbitrage opportunity arises if assets can be combined in a portfolio with zero cost, no chance of loss and positive probability of gain. If the No Arbitrage (NA) condition holds in financial markets, the m_{t+1} will be positive in each period (Cochrane, 2005). Equation 2.1 upholds the NA principle, suggesting that portfolios with only non-negative, potentially positive payoffs should also be priced positively. This is in line with the foundational work of Harrison and Kreps (1979) and Hansen and Richard (1987), who assert that any failure in these conditions opens the door to arbitrage. For instance, a negative SDF would reverse the relationship between future payoffs and their present values, allowing investors to profit by shorting assets with positive expected payoffs since their discounted present value becomes artificially high. If the NA condition is rigorously upheld, m_{t+1} remains positive, ensuring that markets are free from arbitrage as per Cochrane (2005). This means that all potential arbitrage opportunities, which would allow an investor to secure a risk-free profit without any investment, are either non-existent or have been neutralised by market forces.

Empirical asset pricing works with returns and payoffs to standardise the scale of outright asset prices. The gross return of an asset can be defined as $R_{t+1} = (D_{t+1} + P_{t+1})P_t$. Then equation 2.1 is equivalent to

$$E_t(m_{t+1}R_{t+1}) = 1 \qquad (2.2)$$

where the expected discounted returns (considering all states of the world and their probabilities) should equate to the asset's current price, normalised to 1 in this theoretical framework. If the asset payoffs are considered as excess returns (returns above the risk-free rate), the expected value of the discounted excess returns should be zero, represented as:

$$E_t(m_{t+1}r_{it+1}) = 0 \qquad (2.3)$$

where r_{it+1} is the excess return of asset i at time *t+1*. This formulation is central in empirical asset pricing, especially when testing models across various assets.

In addition to these principles, the concepts of complete and incomplete markets are crucial in understanding the application and uniqueness of the SDF. In complete markets, every possible future state of the world can be perfectly hedged or replicated using a portfolio of existing assets. This implies that there are enough securities in the market to span all possible payoffs, resulting in a unique and well-defined SDF. The unique SDF ensures consistent pricing across all assets, adhering to the Law of One Price (LOP) and preventing arbitrage opportunities. This concept is rooted in the Arrow-Debreu model from 1954.

Conversely, in incomplete markets, not all possible future states can be hedged or replicated using the available assets, leading to the potential for multiple SDFs. This scenario can result in different SDFs pricing different subsets of assets correctly, causing inconsistencies in asset pricing. The inability to hedge all future states means that some payoffs cannot be perfectly priced, leading to mispricing and potential arbitrage opportunities. Without a unique SDF, the relationship between future payoffs and present values may not hold uniformly, and the market may not adhere strictly to the LOP. As a result, incomplete markets might allow for temporary or persistent arbitrage opportunities due to the inability to trade in certain states.

2.2.2 Asset Pricing Model Development

Early specifications of the SDF were consumption-based, linking asset prices to their covariances with consumption growth. Breeden and Litzenberger (1978) developed the Consumption Capital Asset Pricing Model (CCAPM), where asset prices reflect their potential to enhance future consumption, and utility is modelled with diminishing marginal returns (Epstein and Zin, 1991; Weil, 1989). The SDF in these models is expressed through the ratio of marginal utilities across time (Breeden, 1979). Despite theoretical appeal, consumption-based models have struggled to perform well in empirical tests, particularly in explaining cross-sectional return variations (Campbell and Cochrane, 2000; Hansen and Singleton, 1982; Hyde and Sherif, 2005).

Linear factor models have emerged as powerful tools in financial economics, primarily because they address the practical limitations of consumption-based models, providing a more robust and reliable proxy for marginal utility (Cochrane, 2005). Traditional consumption-based models often hinge on the accurate measurement of marginal utility from consumption and its relationship to asset prices—a relationship that, in practice, proves difficult to quantify directly and consistently across different market conditions. In contrast, linear factor models simplify this relationship by assuming a linear form for the stochastic discount factor (SDF), typically represented as,

$$m_{t+1} = \alpha + b_f f_{t+1} \qquad (2.6)$$

where α and b_f are parameters that adapt flexibly to empirical data.

One of the significant advantages of linear factor models is their ability to use simpler, more straightforward relationships that can still effectively capture the economic states impacting investor concerns and market fluctuations. By linearising the SDF, these models reduce the complexity involved in modelling dynamic economic interactions, making it easier to apply these models across diverse conditions without losing the nuance of economic realities. This linear specification facilitates broader applications and enhances the models' adaptability to changing economic environments.

Cochrane (2005) notes that consumption is related to broad economic indicators such as GDP growth, investment returns, and interest rates. These relationships suggest that wealth portfolio returns, which reflect the performance of a wide array of investments and economic activities, can serve as a robust proxy for the overall economic state and thus marginal utility. Cochrane (2005) demonstrates how the CAPM can be derived from a consumption-based framework in various ways, such as through assumptions of two-period quadratic utility and log utility. Each approach links the discount factor directly to the return on the wealth portfolio.

For example, with log utility, there is a proportional relationship between consumption and wealth, simplifying the substitution of wealth for consumption in asset pricing models. This means that as consumption increases, the utility or satisfaction of the investor also increases. The marginal utility of consumption under log utility is inversely proportional to consumption, allowing the discount factor m_{t+1} to be expressed as the inverse of the return on wealth: $m_{t+1} = \frac{1}{R_W}$. This means that if the return on wealth is high, the discount factor is low, and vice versa. This relationship eliminates the need to directly measure consumption data, which can be difficult, and instead uses observable market returns to estimate the discount factor. This simplification eliminates the need for direct consumption data, which can be difficult to measure accurately, and aligns the model closely with observable market returns.

In this context, rewriting equation 2.2 to reflect the above, we get $1 = E[(\alpha + bR_{W,t+1})R_{i,t+1}]$ which then when rewritten and solved for the expected return gives us the CAPM equation of

$$E(R_{i,t+1}) = R_f + \beta_i(E(R_{W,t+1}) - R_f) \quad (2.7)$$

where R_W is the return on the wealth portfolio, R_f is the risk-free rate, and β_i is the sensitivity of the i-th asset to the wealth portfolio. This setup directly links the market's overall performance to the perceived utility, providing a clearer and more direct proxy for understanding and predicting investor behaviour than traditional models.

The wealth portfolio concept is crucial for showing how linear factor models approximate marginal utility changes across economic states. By considering the portfolio's returns over two periods, these models capture how market risks affect consumption and investment decisions. Linearising models is essential in financial economics because they simplify complex relationships, making them easier to analyse and apply to real-world data. They provide robust approximations to nonlinear relationships, are less sensitive to small changes, and offer straightforward economic interpretations. Certain utility functions, such as log utility, naturally lead to linear relationships, simplifying the substitution of wealth for consumption in asset pricing models.

One common method of linearising a model is through a Taylor expansion. This technique involves expanding a nonlinear function around a point, usually the mean or equilibrium value, and using the first-order terms to approximate the function linearly. For instance, if the factor model is initially non-linear, applying a Taylor expansion around an expected value of the factor f_{t+1} can yield a linear approximation:

$$m_{t+1} \approx \alpha + b(f_{t+1} - E[f_{t+1}]) \quad (2.8)$$

This linearisation makes it feasible to empirically estimate model parameters and facilitates the application of regression techniques for testing hypotheses about asset returns.

Linear factor models simplify the relationship between economic states and asset prices by using linear representations of the SDF. This approach provides a more effective and empirically robust framework for asset pricing compared to traditional consumption-based methods. By bridging the gap between theoretical finance and practical market analysis, these models enhance our understanding of how systemic risks and economic variables influence investor behaviour and market dynamics. Consequently, the financial community gains tools that are both theoretically sound and practically applicable, improving the precision of asset pricing and investment strategy formulation.

2.2.3 Model Composition

The question is, what should one use for factors f_{t+1}? The chosen factors should serve as good proxies for aggregate marginal utility growth, meaning they should reflect changes in the economy or market that influence the utility derived from consumption. The idea is that certain economic states or events (referred to as "bad states") have a significant impact on investors' preferences regarding risk and return. These states are characterised by a heightened concern over portfolio performance. Despite some criticism of consumption-based models for their practical limitations, all factor models are essentially extensions or special cases of the consumption-based model. These models incorporate additional assumptions that allow them to use other variables as stand-ins for the growth in marginal utility, which is a central concept in understanding consumption choices and asset valuations.

Cochrane outlines that asset pricing fundamentally concerns identifying states of the world where investors are particularly wary of their portfolios underperforming. These "bad states" are moments of economic stress or downturns where the marginal utility of consumption typically rises, as each additional unit of consumption becomes more valuable. Investors are willing to sacrifice some level of average return to ensure their investments are safeguarded against these adverse conditions. The selected factors, therefore, should be indicators of such states, helping investors and economists understand when and why portfolios might underperform.

The arbitrage pricing theory (APT), introduced by Stephen A. Ross (1976), was the first major asset pricing model to include multiple risk factors. It acknowledges that numerous comprehensive risk sources, such as inflation, interest rates, and business activity, shape security returns. The specification of a multifactor model is in the form of a multivariate linear regression with N excess returns, R, and K traded factors, f . With T observations on f_t and R_t:

$$R_t = \alpha_r + \beta f_t + \varepsilon_t, t = 1, \dots, T. \quad (2.10)$$

where R_t , t , and α_r are N-vectors; β is an N × K matrix; and f is a K-vector where K is the number of factors. In this case the SDF is specified as:

$$m_{t+1} = \alpha + \sum_{k=1}^{k} \beta_k f_{kt+1} \quad (2.11)$$

Where β are multiple regression coefficients of returns R on the factors f_{kt+1}.

The APT serves as an alternative to the CAPM, both proposing a linear relationship between expected returns and their covariance with certain risk factors. While the CAPM focuses on the market portfolio return, the APT includes multiple factors. Cochrane (2005) links this to mean-variance efficiency, showing that any multiple-beta model can be expressed through a discount factor model, where the discount factor is a linear function of those risk factors.

Factors in these models are selected from stock characteristics related to cross-sectional returns, reflecting market anomalies that contradict the efficient market hypothesis (EMH). According to EMH, asset prices fully incorporate all available information, making it difficult to consistently outperform the market on a risk-adjusted basis. However, market anomalies indicate potential for abnormal returns, showing patterns or events that deviate from expected efficient behaviour. These factors represent systematic deviations from market efficiency, often supported by economic theories that explain why certain stocks or strategies may outperform others. The existence of predictable patterns in stock returns suggests that markets are not fully efficient and that investors can exploit these inefficiencies for potential gains.

Prominent stock market anomalies include the Size Effect by Banz (1981), illustrating smaller companies' tendency to outperform larger ones, and the Value Effect, where stocks with lower price-to-earnings or high book-to-market ratios excel in the long run, researched by Rosenberg, Reid, Lanstein (1985), Chan, Hamao, Lakonishok (1991), Fama and French (1993). The Momentum Effect by Jegadeesh and Titman (1993) shows that stocks with strong past performance often continue to perform well. Additionally, the Dividend Yield Anomaly, observed by Litzenberger and Ramaswamy (1979), suggests high dividend yield stocks outperform lower yield counterparts. The BAB (Betting-Against-Beta) factor, by Frazzini and Pedersen (2013), highlights the superior risk-adjusted returns of low-beta stocks, advocating for investments in lower-risk stocks.

To exploit anomalies such as those outlined above, factors are typically constructed in a long/short manner to take advantage of these discrepancies in stock performance. Long/short factors are investment strategies that involve taking positions in assets expected to increase in value (long positions) and in assets anticipated to decrease in value (short positions), aiming to profit from the relative performance of these assets. This method is foundational in developing factors that shed light on various dimensions of risk and return in financial markets. Recent advancements in factor identification have been highlighted by Chib, Lin, Pukthuanthong, and Zeng (2023). They recommend a Bayesian approach to uncovering risk factors from a vast array of potential candidates. Their approach specifically targets the derivation of slope factors from a broad set of stock characteristics through Fama and MacBeth (1973) regressions. Their follow up work utilises cubic splines, as outlined by Chib and Greenberg (2010), to generate non-parametric slope factors, employing the PAMS strategy—pruning, augmentation, and model scanning—as a comprehensive framework for factor discovery and evaluation. As more market anomalies are identified, there's a corresponding increase in the models that incorporate these anomalies, aiming to account for asset returns.

2.3 Model Comparison

2.3.1 Mean Variance Efficiency

To examine the mean-variance efficiency of a portfolio, Gibbons Ross and Shanken (1989) consider the following regression equation:

$$R_{it} = \alpha_{ip} + \beta_{ip} f_{pt} + \varepsilon_{it}, \tag{2.12}$$

where R_{it} represents the excess return on asset i in period t. f_{pt} is the excess return on the portfolio being tested. ε_{it} is the disturbance term for asset i in period t. The disturbances are assumed to be normally distributed with mean zero and a non-singular covariance matrix Σ, conditional on the excess returns for portfolio p.

Following this regression equation, the portfolio f_{pt} is considered a factor model combination, serving as a benchmark. This setup allows for an examination of how individual assets perform relative to this composite factor model. The alpha term (α), initially introduced by Jensen in 1968, plays a pivotal role in this analysis. It is used to determine whether a portfolio or asset performs beyond what this predictive factor model would anticipate, based on its assessed risk. The null hypothesis stated as $H_0: \alpha_{ip} = 0$, for $i=1,\ldots,N$ directly suggests that the intercepts α_{ip} in the multivariate linear regression model are hypothesised to be zero for each asset i. This means that each asset's excess return is assumed to be fully explained by the factors included in the regression (in this case, the excess return on a given portfolio f_{pt}), without any individual asset outperforming or underperforming systematically beyond what the model predicts.

To assess whether a portfolio is mean-variance efficient, the significance of the alpha values can be evaluated using a t-test. If these intercept values are significantly different from zero, it suggests that the portfolio may not fully capture the returns of the assets, indicating that the model might not completely encompass all relevant risks.

In this equation, the left-hand side returns, R_{it}, are often termed "Test assets" or mimicking portfolios refer to a set of financial assets, such as stocks, bonds, or portfolios, used to assess the performance of an asset pricing model. Typically, a broad and diverse range of assets is chosen as test assets to ensure that the model is robust across different types of investments. Test assets are generally bivariate portfolio analysis, as described by Fletcher (2019). Fletcher evaluates the mean-variance efficiency of linear factor models using U.K. stock returns, employing two sets of test assets: 16 portfolios sorted by size and book-to-market ratio, and another 16 sorted by size and momentum, constructed annually using value-weighted buy-and-hold monthly returns. For further insights into explanatory returns in asset pricing tests, see Fama and French (2018), who utilised 25 portfolios constructed from sorts on size and book-to-market or size and momentum. A multitude of studies have employed test assets to evaluate the effectiveness of asset pricing models in this regression framework, both within the U.S. and globally. Notable examples include the works of Fama and French (1992, 2012, 2015, 2016, 2018), Davis, Fama, and French (2000), Petkova (2006), and Ball, Gerakos, Linnainmaa, and Nikolaev (2015).

In typical tests of asset pricing models, as outlined, there can be 16 to 25 test asset portfolios acting as the left-hand side variables. Conducting separate T-tests for each asset's alpha value could lead to issues with multiple comparisons, such as increased likelihood of Type I errors (falsely rejecting the null hypothesis). Gibbons, Ross, and Shanken (1989) developed a multivariate statistical method, based on the principles of the Wald test, to examine the joint hypothesis that the alphas of multiple assets or portfolios are all equal to zero in the context of an asset pricing model. This test, often abbreviated as the GRS test, is particularly relevant in evaluating the performance of a given asset pricing model across several assets or portfolios simultaneously.

The Wald test is a statistical method used to determine whether a set of parameters in a model are jointly significant. The test statistic follows a chi-square distribution under the null hypothesis, which posits that all alpha values are zero ($H_0: \alpha_1 = \alpha_2 =..\alpha_N = 0$), and is given by:

$$T^* \left[\frac{1}{1+Sh^2(f)}\right] * \alpha' \Sigma^{-1} \alpha \tag{2.13}$$

where $\alpha' \Sigma^{-1} \alpha$ represents the weighted sum of the squared alphas, with weights given by the inverse of the residual covariance matrix, Σ. $Sh^2(f)$ is the maximum squared Sharpe (1966) performance of the K factors in the model, calculated as $Sh^2(f) = u_f' V_f^{-1} u_f$. Here, u_f is a (K,1) array of the sample mean excess returns of the K factors, V_f is the Maximum Likelihood (ML) estimate of the sample (K,K) covariance matrix of the K factor portfolio excess returns (Divide by 1/T rather than 1/(T-1)). T is the sample size.

If this statistic exceeds a critical chi-square value, the null hypothesis is rejected, indicating that the coefficients (in this case, alphas) are jointly significant. Conversely, failing to exceed the critical value implies that the coefficients are not jointly significant. The assumption of a chi-squared distribution for the Wald test, which has N degrees of freedom, holds well in large samples due to the central limit theorem. However in small samples, the actual distribution of the test statistic may not adequately approximate the chi-squared distribution.

The Gibbons et al. (1989) (GRS) test modifies the Wald test statistic to make it more applicable and interpretable in a finite sample context by transforming it into an F-statistic. This transformation adjusts the statistical framework to better align with the characteristics and constraints of finite samples, thus providing more reliable inference. The resulting test statistic follows an F-distribution, which allows for easier interpretation and comparison with critical values from the F-distribution. The statistical assumptions are that both asset and factor returns follow a multivariate normal distribution, the residuals exhibit constant variance (homoskedasticity), there is no serial correlation in the residuals, and the relationship between asset returns and factor returns is linear. The GRS test is given by:

$$\left(\frac{\mathrm{T-N-K}}{\mathrm{N}}\right) * T^* \left[\frac{1}{1+Sh^2(f)}\right] * \alpha'\Sigma^{-1}\alpha \qquad (2.14)$$

The GRS test conditional on the factors has a non-central F distribution with N and T-N-K degrees of freedom and non-centrality parameter $\lambda = \mathrm{T}\alpha'\Sigma^{-1}\alpha \,/(1 + Sh^2(f))$. Under the null hypothesis of mean-variance efficiency, $\lambda = 0$ and the GRS test has a central F distribution with N and T-N-K degrees of freedom. The factor $\left(\frac{\mathrm{T-N-K}}{\mathrm{N}}\right)$ normalises the test statistic for the number of assets (N) and the degrees of freedom remaining after accounting for the number of factors (K) and the sample size (T). This normalisation helps mitigate the impact of varying sample sizes and the number of parameters being estimated, which are critical in smaller samples. However, it is important to note that these results are only valid if the underlying statistical assumptions hold true.

Gibbons et al. (1989) also demonstrate that the GRS test links to portfolio efficiency assessments, as their test can be written as:

$$\left(\frac{\mathrm{T-N-K}}{\mathrm{N}}\right) * \left[\frac{\left(Sh^2(\mathrm{r},f)-Sh^2(f)\right)}{(1+Sh^2(f)}\right] \tag{2.15}$$

where $Sh^2(\mathrm{r}, f)$ is the maximum squared Sharpe performance of the N+K assets. $Sh^2(\mathrm{r}, f) = u'V^{-1}u$, where u is the sample mean (N+K,1) excess returns of the N+K assets, V is a (N+K,N+K) (ML estimate) of the sample covariance matrix of the N+K assets. Both formulas (2.14) and (2.15) represent different expressions of the GRS test, which essentially captures the same underlying concept. The test compares the maximum squared Sharpe performance of the N+K assets to the maximum squared Sharpe performance of the K factor portfolios. In other words, if the Sharpe ratio of the factors/test assets and factor are the same, we will not be able to reject H_0: $\alpha = 0$, and we can say the factors price the set of test assets. The further the optimal portfolio of the K factors lies from the efficient frontier of the N+K assets, the more likely it is to reject the null hypothesis of zero pricing errors. If the F-tests reject the null hypothesis for all models, it means none completely capture average returns as asset pricing models. The GRS test assumes all factors are needed; if rejected, a method of model comparison is required.

The GRS test examines the mean-variance efficiency of a single model, but to compare model performances, we need additional metrics. One key metric, as outlined previously, is Jensen's Alpha (1968), which evaluates asset performance beyond CAPM expectations. To compare the performance of different models, one can look at the alphas produced by each model. Smaller and statistically insignificant alphas indicate that the model effectively explains asset returns. The Sharpe Ratio (1966, 1994) assesses the risk-adjusted return of a portfolio by comparing its excess return over the risk-free rate to its standard deviation, providing insights into the return per unit of risk. The Treynor-Black Measure (1973) focuses on a portfolio's excess return per unit of market risk, emphasising the importance of beta in understanding investment performance. Jensen's Alpha has also been tested with additional factors over the CAPM, such as size and value factors by Fama and French (1993), and momentum by Carhart (1997). These extensions allow for a more comprehensive analysis of returns. Collectively, these measures provide robust methods for evaluating model performance.

Fama and French (2012, 2015, 2016, 2018) also examine several alpha variations to assess model performance. These include: the average absolute alpha ($A|\alpha_i|$), which measures average mispricing, with better models exhibiting lower $A|\alpha_i|$; the $A(\alpha_i^2/\sigma(r_i)^2)$ ratio, representing the average squared alphas divided by the variance of the average excess returns of the test assets, where lower values indicate better models; the $A(\alpha_i^2 - se(\alpha_i)^2)/\sigma(r_i)^2$, which adjusts for the standard errors of the alphas, reflecting real mispricing relative to the variance of the average excess returns, with lower ratios indicating better models; and the $A(se(\alpha_i)^2/A(\alpha_i)^2)$, which is the ratio of the average squared standard error to the average squared alpha, capturing the proportion of mispricing due to sampling error, with higher values suggesting better models according to Fama and French.

2.3.2 Test Asset Irrelevance

Building on the factor redundancy testing of Fama and French (2015), Barillas and Shanken (2017) address the issue of comparing models using the classic Sharpe ratio improvement metric. This metric evaluates the fit of a model by examining the improvement in the squared Sharpe ratio (expected excess return over standard deviation) when additional assets are included in the investment universe. The improvement in the squared Sharpe ratio from adding test assets R to the investment universe is a quadratic form in the test-asset alphas:

$$\alpha_R' \Sigma^{-1} \alpha_R = Sh^2(f, \mathrm{R}) - Sh^2(f) \quad (2.16)$$

where the left hand side is a measure of the model's unexplained risk-adjusted performance.

Gibbons et al. (1989), and later Barillas and Shanken (2017), highlight the importance of test assets in evaluating and comparing asset pricing models. Large alphas indicate that the model fails to capture significant portions of the return, suggesting that the Sharpe ratio of the model can be improved by including these test assets.

However, Barillas and Shanken (2017) outline how test assets are irrelevant when comparing two models using the following setup:

$$Sh^2(A, \mathrm{B}, \mathrm{r}) - Sh^2(A) = Sh^2(A, \mathrm{B}, \mathrm{r}) - Sh^2(\mathrm{B}) \quad (2.17)$$

where A and B represent different models, and r represents a set of test asset returns. This setup measures the improvement in the squared Sharpe ratio for two models, A and B, by examining the difference in their squared Sharpe ratios before and after the inclusion of factors from the other model and test assets r. If all factors and test assets are included, the expression $Sh^2(A, \mathrm{B}, \mathrm{r})$ remains the same on both sides. Therefore, the comparison simplifies to $Sh^2(A) >$

$Sh^2(B)$, indicating that the model with the higher squared Sharpe ratio is preferable, and the test assets drop out of the comparison.

Barillas and Shanken (2017) argue that what truly matters in model comparison is each model's ability to price the "excluded factors" from other models. If a model can effectively price the factors in another model (reflected in low or zero "excluded factor" alphas), it suggests that the first model captures the risk-return trade-offs inherent in the factors of the second model, and so including these factors would not improve the Sharpe ratio of that model as per Equation 2.16. This capability is more indicative of a model's adequacy than the performance improvement measured by the inclusion of test assets. For example, comparing the CAPM to the Fama French three-factor model (FF3) would involve alpha tests on the high-minus-low (HML) and small-minus-big (SMB) factors to determine if FF3 provides a better risk-return trade-off than the market factor alone.

2.4 Research Methods

2.4.1 Pairwise Model Comparison

Before continuing, it deserves emphasis that by “model comparison,” I mean here the determination of which model is superior according to a given metric. A researcher may, nonetheless, be interested in exploring how various models price particular assets, and this is certainly a form of comparison, as the term is used more generally. However, as I demonstrate, it is not the same as identifying the better model based on well-established criteria.

When comparing models, we must deal with both nested and non-nested pairs, requiring different methods for each. To compare nested models, Barillas and Shanken (2017) show that we need only focus on testing the excluded-factor restriction (test assets irrelevance as outlined previously). When models are nested, the comparison focuses on the nested structure using the heteroskedasticity-adjusted GRS test. This test evaluates whether the additional factors in the more complex model significantly improve the Sharpe ratio compared to the simpler model.

The GRS test compares the squared Sharpe ratios of the nested model to the expanded model in order to evaluate whether adding additional factors to a model significantly improves its performance. The test statistics and their corresponding p-values indicates whether the additional factors of the expanded model provide a significant improvement in performance. If the inclusion of additional factors significantly increases the squared Sharpe ratio, it suggests that the larger model offers a better risk-adjusted return.

Unlike nested models, where one model is a subset of another (and thus they share a common set of factors or variables), non-nested models do not share this hierarchy. They may include completely different sets of factors or explanatory variables. This lack of commonality makes it challenging to directly compare their explanatory power or performance since there's no baseline or shared dimension to evaluate them against each other. The improved fit of a model, as indicated by a higher squared Sharpe ratio, suggests that it better captures the dynamics of asset returns. However, quantifying this improvement requires a detailed analysis of the model's predictions against actual returns. Barillas et al. (2020)[4] develop an asymptotic test for testing non-nested models. These tests are based on the behaviour of statistical estimators as the sample size approaches infinity.

Consider two nonnested models (A and B) with factor returns f_{At} and f_{Bt} , respectively, t =1, 2,...,T. Denote the squared maximum Sharpe ratios that are attainable from the two sets of factors by $\theta_A^2 = u_A' V_A^{-1} u_A$ and $\theta_B^2 = u_B' V_B^{-1} u_B$, where u_A, u_B, V_A, and V_B are the nonzero means and invertible covariance matrices of the two sets of factors. Similarly, let the corresponding sample quantities be $\hat{\theta}_A^2 = \hat{u}_A' \hat{V}_A^{-1} \hat{u}_A$ and $\hat{\theta}_B^2 = \hat{u}_B' \hat{V}_B^{-1} \hat{u}_B$,

The asymptotic distribution of the difference in sample squared Sharpe ratios, as outlined in Proposition 1 of Barillas et al., (2020), is given by:

$$\sqrt{T}([\hat{\theta}_A^2 - \hat{\theta}_B^2] - [\theta_A^2 - \theta_B^2] \sim \mathrm{N}(0, \mathrm{E}[d_t^2]), \qquad (2.18)$$

provided that $\mathrm{E}[d_t^2] > 0$, where

$$d_t = 2(\mu_{At} - \mu_{Bt}) - (\mu^2{}_{At} - \mu^2{}_{Bt}) + (\theta_A^2 - \theta_B^2), \qquad (2.19)$$

with

$$\mu_{At} = \mu'_A V^{-1}{}_A (f_{At} - \mu_A) \text{ and } \mu_{Bt} = \mu'_B V^{-1}{}_B (f_{Bt} - \mu_B). \qquad (2.20)$$

Barillas et al. (2020) outline how the differences in Sharpe ratio as outlined in 2.18 follows a normal distribution with mean 0 and variance $\mathrm{E}[d_t^2]$. An asymptotic distribution is the probability distribution that a statistic approaches as the sample size grows indefinitely. This indicates that as the sample size T grows, the distribution of the differences in Sharpe ratios converges to a normal distribution centered around the true difference.

Barillas et al. (2020) introduced the concept of bias-adjusted squared Sharpe ratios, building on the foundational insights of Jobson and Korkie (1980), who first highlighted the challenges

[4] I am thankful to Prof Cesare Robotti for the provision of the MATLAB code on his website to perform these tests.

of estimating parameters like mean-return vectors and covariance matrices in small samples. The Sharpe ratio, a key metric for assessing investment portfolio performance, measures the excess return per unit of risk (standard deviation of returns). However, when derived from sample data, the Sharpe ratio often overestimates the true population Sharpe ratio due to biases inherent in small samples. This overestimation occurs because both the numerator (mean excess return) and the denominator (standard deviation of returns) are estimated from the same data set, leading to a biased outcome that typically inflates the sample Sharpe ratio.

To adjust for this bias, Barillas et al. (2020) multiply the squared Sharpe ratio by a correction factor $(T - K - 2)/T$, where T is the sample size (number of observations) and K is the number of factors (or parameters) estimated in the model. This adjustment reduces the squared Sharpe ratio, making it a more accurate estimate of the population squared Sharpe ratio under the assumption of joint normality of returns. Additionally, K/T, representing the proportion of the sample used to estimate the parameters, is subtracted to further adjust for the degrees of freedom consumed in parameter estimation, thus mitigating the upward bias. By adjusting for this bias, researchers and practitioners can make more accurate comparisons between different investment models or portfolios, especially when the sample sizes are relatively small.

2.4.2 Multiple Model Comparison

Suppose a researcher is considering more than two models and wants to test whether one model (the "benchmark") is at least as good as the others in terms of its squared Sharpe ratio. Consider a benchmark model that is nested in a series of alternative models. We form a single alternative model that includes all of the factors contained in the models that nests the benchmark. It is then easily demonstrated that the expanded model dominates the benchmark model if and only if one or more of the "larger" models dominate it. The null hypothesis that the benchmark model has the same squared Sharpe ratio as the alternatives is tested using pairwise nested-model comparison. Specifically, all factors excluded from the benchmark are projected onto the benchmark factors, and it is tested if these alphas are jointly zero. Rejecting this null indicates the benchmark is dominated by one or more alternative models; otherwise, the benchmark model performs as well as the other models.

The multiple-model comparison test for nonnested models is based on the multivariate inequality test of Wolak (1987, 1989). Suppose there are p models.

Let $\delta=(\delta_2 ,..., \delta_p)$ and $\hat{\delta}\hat{}=(\hat{\delta}_2 ,..., \hat{\delta}_p)$, where $\delta_i = \theta^2{}_1 - \theta^2{}_i$ and $\hat{\delta}_i = \hat{\theta}_1^2 - \hat{\theta}_i^2$ for i = 2,..., p.

Here, δ represents the differences in squared Sharpe ratios between the benchmark model and each of the alternative models. A positive value indicates the benchmark model performs at least as well as the alternative. $\hat{\delta}$ is the sample counterpart, representing the observed differences. The test is as follows:

$$H0 : \delta \geq 0 \text{ versus } H1 : \delta \in \eta_r$$

where r = $p-1$ is the number of nonnegativity restrictions. Thus, under the null hypothesis, model 1 (the benchmark) performs at least as well as models 2–p (the competing models).

The test is based on the sample counterpart of δ, $\hat{\delta}=(\hat{\delta}_2 ,..., \hat{\delta}_p)$, which has an asymptotic normal distribution with mean δ and covariance matrix $\Sigma_{\hat{\delta}}$ (the conditions for this are provided in the Online Appendix to Kan et al. (2013)). The sample counterpart $\hat{\delta}$ follows an asymptotic normal distribution as the sample size increases to infinity. This allows for the use of normal distribution theory to infer population parameters from sample data.

The test involves solving a quadratic programming problem where the goal is to minimise the weighted squared difference between $\hat{\delta}$ and δ, subject to δ being nonnegative. The likelihood-ratio (LR) measuring deviation from the null hypothesis is given by:

$$LR = T (\hat{\delta} - \tilde{\delta})' \hat{\Sigma}_{\hat{\delta}}^{-1} (\hat{\delta} - \tilde{\delta}) \qquad (2.21)$$

The LR is calculated based on the difference between the observed $\hat{\delta}$ and the optimised $\tilde{\delta}$, weighted by the inverse of the estimated covariance matrix[5]. This statistic measures the degree to which the observed data deviate from the null hypothesis scenario where the benchmark model performs at least as well as the alternatives. A large value of LR suggests that the nonnegativity restrictions do not all hold. This would lead to rejecting the null hypothesis in favour of the alternative that suggests better performance by one or more of the alternative models. To conduct statistical inference, the asymptotic distribution of LR is needed. Readers are referred to Kan et al. (2013) for its derivation and a discussion of numerical methods for calculating the p-value.

In comparing a benchmark model with a set of alternative models, those alternatives nested by the benchmark model are removed, as the null hypothesis $\delta_i \geq 0$ holds in this case. If any alternative is nested within another, the "smaller" model is removed because the larger model

5 The code uses bootstrapping (resampling) techniques to estimate the distribution of the Likelihood Ratio (LR) test statistic under the null hypothesis.

will have at least as high a squared Sharpe ratio. This ensures that comparisons are only made with models that are not inherently included in the benchmark, aligning with the null hypothesis that the benchmark's performance is not inferior to these models. The assumption of asymptotic normality of the difference in squared Sharpe ratios is crucial for the statistical test.

2.5 Data and Models

2.5.1 Dataset

I aim to strike a balance in selecting markets for my European based analysis. Choosing a wide range of markets can lead to limitations due to data availability, resulting in a shorter sample period and fewer factors than preferred, as seen in Pukthuanthong et al. (2023). Conversely, a sufficient number of samples are included to guarantee the robustness of the findings within a European context, while also highlighting the unique characteristics of each market. Additionally the robustness of factors derived from these markets is a key rationale. Larger markets typically provide more reliable data, reducing the likelihood of anomalies driven by market size or liquidity constraints. This ensures that the factors extracted are representative and robust, making them ideal for a comprehensive analysis of European market dynamics.

I have chosen six large developed markets in Europe for my study, specifically the U.K., Germany, France, Italy, Spain, and the Netherlands. As highlighted in Section 1.1, these nations are the primary hubs of asset management in Europe, collectively representing approximately 75% of the European asset management industry[6]. Furthermore, these markets are the focus of empirical model comparison tests that either encompass Europe as a whole or utilise aggregated European samples, see Hanauer et al. (2020). Apart from the U.K. market, there is a notable lack of literature on the testing and comparison of asset pricing models across the markets included in this European sample. In the next section I provide an overview of notable studies in these markets.

The set of 12 tradable investment factors is motivated by recent studies on factor model performance in both U.S. and non-U.S. markets, including works by Chib et al. (2023), Barillas et al. (2020), Hanauer (2020), Ahmed et al. (2019), Fletcher (2019), and Michou and Zhou (2016), among others. The first group of factors included stem from the Fama and French (1993,2015) and the Carhart (1997) models and include the excess returns on the Market factor and zero-cost portfolios for the size (SMB), value (HML), profitability (RMW), investment

[6] https://www.efama.org/newsroom/news/asset-managers-course-manage-eur-29-trillion-2023

(CMA), and momentum (MOM) effects in stock returns. The second group of factors include the betting against beta (BAB) factor of Frazzini and Pedersen (2014), and the timelier version (HML_M) of the value factor by Asness and Frazzini (2013). The two mispricing factors constructed from 11 Market anomalies termed Management (MMGT), and Performance (PERF) of Stambaugh and Yuan (2017) are also included. The period of analysis is between June 1991 and December 2022

I use the factor data from Hanauer and Windmueller (2021) available on globalfactorpremia.org. To mitigate survivorship bias in the stock lists for each country in their dataset, the authors employ 'dead lists' from Datastream. The authors are then left with the following number of stocks for each of the countries to form the factors which will be examined in this thesis – U.K (3,822), France (1,616), Germany (1,459), Italy (552), Spain (318), Netherlands (250). The updated value factor (HML_M) is obtained from the AQR database along with the Betting Against Beta (BAB) factor. All factors are denominated in USD.

This study only captures a subset of the factors that have been proposed in the literature. Pastor and Stambaugh (2003) propose an aggregate liquidity factor. The short-horizon behavioural factor (PEAD) and long-horizon behavioural factor (FIN) of Daniel, Hirshleifer, and Sun (2020) are also omitted due to lack of available data. All of these factors are worth exploring in a future study.

2.5.2 Notable Research in Markets Of Interest

Empirical work on UK stock market returns has attempted to identify the optimal asset pricing model using a variety of model comparison methods, see, for example, Fletcher (2001,2019,2019). Michou and Zhou (2016) note the investment and profitability influence UK stock market patterns, posing questions about performance of Fama and French (1993) three factor model in U.K stock returns. The author suggests the size and value factor should be replaced with investment and profitability factors. Nicol and Dowling (2015) also not the importance of these factors in U.K asset pricing tests and suggest that the Fama and French (2015) five factor model offers the most potential. However Foye (2018) notes that both the three- and five-factor models are unable to offer a convincing description of UK equity returns in a classical alpha based approach where the test assets are profitability and investment portfolios. Fletcher (2018) is the first to examine a wider set of potential asset pricing models in U.K. stock returns. The author adopts both a classical approach and the Bayesian approach of Barillas and Shanken (2018) to examine the mean–variance efficiency of nine UK factor pricing models and conducts multiple model comparisons. Combining the information from these two approaches the author finds that six-factor model of Fama and French (2018) emerges as the dominant model from a starting collection of models.

Recent literature tests the French market for a variety of anomalies such as, for example, herding in various sectors (Litimi, 2017), market efficiency (Boya, 2019), effects of environmental regulation (Pham and Ramiah, 2020), board gender diversity (Bruna et al,. 2019). Given the highly developed nature of the French stock market the literature on drivers of French stock returns is light. Lajili S. (2007) examine the performance of the Fama French three-factor model in the French stock market over the period June 1976 to June 2001. The author notes a positive and robust Size premium as measured by the average premium on the SMB factor, of 0.742% per month. Also noted is the lower Value premium, as measured by the HML factor of 0.597% per month. Lajili S. and. Desban, (2018), when examining find that the Size factor is redundant in the French market when examining the performance of the Fama French five factor model. The authors also note the weak Value effect and relevance of the quality-minus-junk factor in their tests model performance tests. In both studies outlined above the small number of candidate factor models are compared in a traditional alpha based regression where the test assets are different portfolio sorts on both size and value. While the studies above are very useful to both academics and investors operating in the French market the number of factors included in the analysis and the models compared is limited.

Given the highly developed nature of the German stock market the literature on drivers of German stock returns is extremely light. Artmann et al. (2012) and Hennecke et al. (2022) examine the extent to which four prominent asset pricing models can explain the cross-section of German returns, finding the Carhart four-factor model explains the cross section of average German stock returns between 2008 and 2020 best. There exists a variety of other work done of the German market which examines specific anomalies in more detail such as size, value, and momentum. Other papers have conducted a similar analysis of a proposed anomaly, see, for example, Glaser and Webber (2003); Walkshaeusl and Lobe (2014). The scope of asset pricing models in German returns is limited.

Silvestri and Veltri (2011) investigate if the Fama and French three-factor model is able to explain the variations in stock returns in Italian market with mixed results. Also testing the Fama and French (1993) three factor model, Rossi (2012) notes that beta alone cannot explain the risk-returns relationship. The results indicate that the size factor accompanied by the beta seems to have a greater explanatory power. Pirogova and Roma (2020) show that all three factors are significant in explaining Italian stock returns during the sample period 2000-2018. Unlike the previous studies mentioned above, which either found no value effect at all or no clear-cut results when testing the book-to-market variable, these authors find that the value factor is statistically significant, and the associated risk premium is of a considerable size.

In recent years much of the research done on the Spanish stock market has focused on the role of exogenous shocks such as liquidity crises or the Covid-19 pandemic, see, for example, Martinez et al., 2005; Moya-Martinez et al., 2014; Ahmar and del Val, 2020. Forner and Marhuenda (2003) find that momentum strategies generate abnormal returns in the Spanish stock market over periods of 12-months. Surprisingly, given the highly developed nature of the Spanish stock market the literature on investment factors is quite limited. The Dutch stock market demonstrates signs of inefficiency with calendar anomalies such as the January effect (Moor and Sercu, 2013) and the twist-on-the-Monday effect (Gultekin & Gultekin, 1983). The majority of work done focuses on one a singular anomaly or trading strategy. Doeswijk (1997) found that contrarian strategies yield an outperformance without a higher risk. Knopers (2014) tests the performance of value investing strategies for the Dutch stock market from 1995 to 2013. The results of this study show there is indeed a value premium on the Dutch stock market, consistent with the majority of the international evidence.

A variety of anomalies and factor models have been analysed across all markets included in this study, with the UK market receiving the most extensive attention. For the remaining markets, this study provides a pioneering analysis, evaluating the performance of a broad set of factor models and examining the behaviour of individual investment factors over an extended sample period. Where relevant, comparisons will be made throughout the document between the findings of this study and previous research.

2.5.3 Model Selection

Listed below is the set of models which will be compared across markets. The models chosen for this analysis are cited in the asset pricing literature for their ability to explain cross-sectional returns both in absolute terms and when compared to other models, as evidenced in both alpha-based frameworks like the Fama and French (2015, 2016, 2018), and in Bayesian model comparison frameworks, as seen in the work by Chib et al. (2020). Additionally, the collection of models examined aligns with sets used in recent asset pricing model comparisons tests, such as Barillas et al. (2020) for U.S. returns and Fletcher (2019) for U.K. returns, reflecting their broad applicability. Recent Bayesian approaches, as demonstrated by Chib et al. (2023) and Chib and Zeng (2020), have advocated for larger models in U.S stock returns. However, as previously noted, the choice of models is somewhat restricted by factor data availability constraints across the sample. The names of the below models serve as references to the studies that demonstrate their effectiveness.

1. Fama and French (1993) (FF3)

The FF3 model is a three-factor model. The factors are the excess return on the market index and two zero-cost portfolios that capture the size (SMB) and value/growth (HML) effects in stock returns.

2. Carhart (1997)

The Carhart model is a four-factor model. The factors are the three factors in the FF3 model and a zero-cost portfolio that captures the momentum effect (MOM) in stock returns.

3. Fama and French (2015) (FF5)

This model is a five-factor model. The factors include the factors in the FF model and two zero-cost portfolios that capture the profitability (RMW_{OP}) and investment (CMA) effects in stock returns. The SMB factor constructed using the FF5 model is used across all models.

4. Fama and French (2018) (FF6)

This model is a six-factor model, which augments the FF5 model with the momentum (MOM) factor and replaces the operating profitability factor to a cash based factor (RMW_{CB}).

5. Frazzini and Pedersen (2014) (FrazPed)

This model augments the Capital Asset Pricing Model with the ‘Betting Against Beta’ factor.

6. Stambaugh and Yuan (2017) (SY)

The SY model is a four-factor model that includes the excess market returns, and zero-cost portfolios for the size effect, and two mispricing factors termed Management (MMGT), and Performance (PERF).

7. Hou et al. (2015) (HXZ).

The HXZ model is a four-factor model, which includes the excess market returns, and zero-cost portfolios of the size, profitability (RMW_{ROE}), and investment (CMA) effects in stock returns. The SMB factor is used as the size factor in all of the models as in Chib et al. (2022).

8. Asness, Frazzini, Israel, Moskowitz (2015) (AFIM)

The ASIM model is a six-factor model, which replaces the HML factor in the FF6 model with the Asness and Frazzini (2013) timelier version (HML_M) of the value factor using more up to date price data when rebalancing.

9. Chib, Zeng, Zhao (2020) (CZZ)

This model is a five-factor model identified as the top performing model in a Bayesian model scan of 8 factors in Chib et al. (2020). It includes the excess market returns, and zero-cost portfolios of the size, return on equity (RMW_{ROE}), momentum (MOM) and investment (CMA) effects in stock returns.

2.5.4 Summary Statistics

Given the space constraints in this document and the six separate markets analysed, the results for the U.K. market are presented and interpreted first, followed by an analysis of the remaining markets. The detailed tables for these additional markets are contained in the appendices, which are referred to throughout the discussion. Table 2.1 reports summary statistics of U.K. factor excess returns between June 1991 and December 2022. The summary statistics in the table, from left to right, include the mean excess return, standard deviation of returns, Sharpe ratio, and t-statistic of monthly factor excess returns.

Table 2.1

All factors have a positive average excess return. The factor with the highest return premium is the Momentum (MOM) factor (0.94%), followed by the market factor, and Betting Against Beta. The size factor SMB (0.10%) and the updated value factor HML_M (0.07%) have the smallest average monthly return over the period. However the average monthly return for the Size and HML_M factors returns t-statistics of 0.61 and 0.36 respectively. The majority of factors return t-statics which are significant on the 10% level of significance with the highest t-statistic coming from the MOM factor (4.22).

Table 2.2

Table 2.2 present the mean return of each factor across the remaining five markets in the dataset, which reveals a trend where all factors, except for the Size factor, generally yield positive returns. The Betting Against Beta (BAB) factor consistently shows high return premiums, notably an average monthly return of 1.38% in French returns, making it a top performer along with the Market, MOM, and PERF factors. This contrasts to the results of Lajili. S (2007) who find a robust Size premium over the period 1976-2001. German market summary statistics highlight the Momentum (MOM) factor with a leading 0.92% return and the Performance (PERF) factor having the highest Sharpe ratio (Table A.2). The Dutch market shows a negative premium for Size and a leading 0.83% return for MOM, though most factors lack statistical significance. We do not see the significant value premium observed by Knopers (2014) in earlier data. The Italian market presents negative returns for three factors, with only the Investment (CMA) factor's negative premium being significant. The Momentum (MOM) factor leads with the highest monthly average return across all markets of 0.80%. Full summary statistic results for the remaining five markets in my study are detailed in Section A.II of Appendix A. The factor with the highest average Sharpe ratio across all markets is also the Performance factor (PERF) with an average of 0.166.

2.6 Empirical Results

2.6.1 Tests of Equality of Squared Sharpe Ratios

Table 2.3 presents the pairwise tests of equality of the squared Sharpe ratios for nine U.K. factor models, some nested and others nonnested, from 1991 to 2022.

Table 2.3

Panel A shows the differences between the (bias-adjusted) sample squared Sharpe ratios (column model - row model) for various pairs of models. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements of Panel A in Table 2.2 are the sample squared Sharpe ratio differences between the model in that column and the next-best model. In Panel B, p-values for the tests of equality of the squared Sharpe ratios are presented. A low p-value indicates that the difference between squared Sharpe ratio as per Panel A is significant. As previously discussed, p-values must be computed differently depending on whether the models to be compared are nested or nonnested.

For the U.K. market, the main empirical findings can be summarised as follows: The top three performing models as per Squared Sharpe ratio are the FF6, AFIM and CZZ. Panel B shows that the top three performing models have Sharpe ratios which are significantly higher than most of the remaining models in the candidate model set. This is not surprising given the fact that they contain factors which themselves have provided high returns over the period of analysis. However while the FF6 has the highest squared Sharpe ratio, the increase that it provides over the AFIM model and the CZZ model is not statistically significant, as indicated by p-values of 0.521 and 0.121 in Panel B. The FF3 three-factor model and FrazPed have the lowest relative performance from the candidate set. The FF3 and FrazPed models are outperformed by almost all other models at the 10% level of significance as seen through low p-values in Panel B. The Carhart model is significantly outperformed by both the FF6 and AFIM models but not the CZZ model.

Table 2.4 below presents the identities of the top-performing and second-best performing models in each of the remaining markets.

Table 2.4

In most cases, either the FF6 or the AFIM six-factor models return the highest or second-highest Sharpe ratio. This finding is in line with Barillas et al. (2020) who find the AFIM to be the optimal model in U.S data and also Fletcher (2019) finding the same in U.K tests of model comparison. The FrazPed model also performs well in certain markets. Neither the FF3 three-factor model nor the five-factor model are top performers in any market. Panel B reports the squared Sharpe ratio for each model across all countries. A more detailed breakdown of the results across the sample is provided below.

Full results for the remaining five markets are detailed in Section A.III of Appendix A. In the case of the French market, as shown in Table A.6 in Appendix A, the FrazPed model is the best-performing model based on the squared Sharpe ratio. The difference between this model and the FF3, FF5, Carhart, HXZ, and CZZ models is significant at the 10% level, as indicated by the low p-values in Panel B. However, the differences in the squared Sharpe ratio between the FrazPed model and the AFIM, FF6, and SY models are not significant at any level, with p-values of 0.549, 0.576, and 0.606, respectively. Therefore, we can conclude that while the FrazPed model is the best-performing model, it only significantly outperforms half of the other models under consideration.

The German market (Table A.7) has a similar profile to the U.K. market, where both the FF6 and AFIM six-factor models perform best but cannot be significantly separated, with a p-value of 0.324. The AFIM model outperforms all other models at the 5% level of significance and the Carhart model at the 10% level of significance. The FF6 model performs less well, outperforming most other models except for the SY and FF5 models, with p-values of 0.128 and 0.265, respectively.

For the Netherlands (Table A.8), the FrazPed model performs best, which is unsurprising given the performance of the BAB factor in Dutch returns. The pairwise tests cannot distinguish between the top 7 performing models, as indicated by the high p-values in Panel B. The consistently high p-values across Panel B show that the pairwise tests do not find significant differences among many of the models.

In both Italian (Table A.9) and Spanish (Table A.10) stock returns, the AFIM model has the highest squared Sharpe ratio, significantly outperforming all other models except the FF5 model in Italy (0.262) and the Carhart model (0.105) in Spain. In both cases, the FF3 model and FrazPed model are the two lowest performing model. This result contrasts with that of Pirogova and Roma (2020) who find the FF3 to be the optimal model in pricing Italian assets albeit with a smaller set of models considered.

Given the consistent performance of the momentum factor throughout the sample, it is unsurprising that the Carhart model performs moderately well in pairwise testing across various markets. Although the Carhart model is significantly outperformed by the top-performing models in certain markets, such as the U.K., this outperformance is not statistically significant in most other markets. Except for France and the Netherlands, the FrazPed model yields a moderate squared Sharpe ratio in most samples but is significantly outperformed by the leading models in most cases. Due to the strong performance of the BAB factor across markets, the FrazPed model significantly outperforms the FF3 three-factor model and, in the Netherlands, achieves the second-highest squared Sharpe performance. In contrast, the FF5 model generally returns one of the lowest squared Sharpe ratios and is significantly outperformed by other models, except in Italy. Across various markets, the FF3 model performs poorly, consistently ranking in the bottom two without any significant anomalies. The SY model has mixed performance, often ranking as the fifth or sixth best performing model but occasionally performing better, such as in France where it ranks as the second highest performing model.

The CZZ model generally performs well, frequently ranking among the top four performing models across most markets.

2.6.2 Multiple Model Comparison Tests

Up to this point, comparisons have been made between two competing models. However, when evaluating a set of models, it is useful to test whether a single model, the "benchmark," has the highest squared Sharpe ratio among all the models. To examine this, I employ the multiple-model comparison test for nonnested models based on the multivariate inequality test of Wolak (1987, 1989). The null hypothesis in this joint test is that none of the other models are superior to the benchmark, while the alternative hypothesis is that some other model has a higher (population) $\hat{\theta}^2$ than the benchmark.

The empirical results for the U.K. are presented in Table 2.5. The first column lists the benchmark models, while the second column shows their bias-adjusted sample squared Sharpe ratios, reflecting risk-adjusted returns. The third column indicates the number of alternative models each benchmark is compared against. A selective process is employed to compare only non-nested models directly, explaining the varying 'r' values for models like FF3 by excluding nested model comparisons. The likelihood ratio statistic values calculated as per equation 2.21, found in the fourth column, and their statistical significance, provided in the fifth column, test the differences between models.

Table 2.5

Table 2.5 above shows the results of the U.K. multiple model comparison tests. Naturally, because the FF6 six factor model has the highest sample squared Sharpe ratio (0.144), the p-value for this model in the joint test is very large (0.801) consistent with the conclusion that this model performs at least as well in population as the other models. The null hypothesis of equivalent performance cannot be rejected for four models in the U.K. sample, as seem through high p-values in column five.

Low squared Sharpe ratios in column two and low p-values in column 5 for models such as the FF3, FrazPed, and FF5 models indicates that they do not perform as well as the other models. A large likelihood ratio value indicates a violation of the nonnegativity constraints which implies that the benchmark model does not perform as well as the alternative models. Consequently, this prompts the rejection of the null hypothesis, opting instead for the alternative hypothesis, which posits superior performance by at least one of the alternative models.

A similar pattern is found across the remaining five markets outlined in Section A.IV in Appendix A. Naturally as the FF6 and AFIM six factor models have the highest squared Sharpe ratio across markets they return large p-values, consistent with the conclusion that these two models perform at least as well in population as the other models.

In the context of French stock returns (Table A.11), the FrazPed model emerges as the top performer, boasting a squared Sharpe ratio of 0.089. Its high p-value of 0.697 suggests that it performs as well as or better than the other models. Following closely is the SY model, which, while not the top performer, still demonstrates competitive performance with a squared Sharpe ratio of 0.068 and a p-value of 0.631, indicating comparable performance to other models. Conversely, the FF3 and Carhart models show weaker performance, with squared Sharpe ratios of 0.005 and 0.034, respectively, coupled with low p-values of 0.001 and 0.027, suggesting that the null hypothesis of equal performance can be rejected.

For the German stock returns (Table A.12), the AFIM model stands out as the top performer, boasting a squared Sharpe ratio of 0.13. Its high p-value of 0.83 indicates that it performs as well as or better than other models in the dataset. Although not the top performer, the FF6 model demonstrates competitive performance with a squared Sharpe ratio of 0.114 and a p-value of 0.533, suggesting comparable performance to other models. The Carhart model returns a squared Sharpe ratio of 0.085 and a p-value of 0.111 meaning that while this model underperforms the top two, the null hypothesis cannot be rejected. All other models reject the null hypothesis indicating inferior performance.

Within Dutch stock returns (Table A.13), the FrazPed model emerges as the top performer, with a squared Sharpe ratio of 0.054 and a p-value of 0.743, indicating comparable performance to other models. There are five other models which return large p-values which shows that even though they have Sharpe ratios lower than the top performing model they do not reject the null hypothesis of equal performance.

In the Italian stock returns (Table A.14), the AFIM model emerges as the top performers, boasting a squared Sharpe ratio of 0.109. A p-value of 0.817 shows the null hypothesis of performing as well as any other model is rejected. There are four other models which also do not reject the null hypothesis. Conversely, models such as FF3 and Carhart show relatively weaker performance, with squared Sharpe ratios of 0.005 and 0.037, respectively, coupled with p-values of 0.002 and 0.04, indicating their underperformance compared to the top models.

Within Spanish stock returns (Table A.15), the AFIM model emerges as one of the top performers, boasting a squared Sharpe ratio of 0.065. The p-value of 0.797 suggests that it does perform as well as other models in the dataset. Similarly, the FF5 model demonstrates competitive performance with a squared Sharpe ratio of 0.039 and a p-value of 0.842, indicating comparable performance to other models. However, eight of the nine models do not reject the null hypothesis of equal performance.

Across markets, the null hypothesis for the CZZ and HXZ models cannot be rejected in most cases, indicating that these models perform as well as any other model in the model space. This is not surprising, given that these two models performed well in the pairwise tests. In all cases, the hypothesis that the FF3 performs as well as any other model in the candidate set can be rejected at the 5% level of significance. The same applies to the FF5, which can be rejected at the 10% level of significance in all cases except for the tests in Italian returns (Table A.14). In each market, multiple models fail to reject the null hypothesis of equal performance.

The variations in results across countries are driven by the unique economic environments, market structures, and investment behaviours present in each region, which impact the relevance and effectiveness of different risk factors. Factors like size, value, momentum, and profitability may exhibit different risk-return profiles and levels of significance depending on local economic conditions, regulatory frameworks, and cultural attitudes toward investing. For example, a factor that captures momentum may perform well in markets characterized by higher speculative trading or where information dissemination is slower, but it may be less effective in markets with more efficient price adjustments. Additionally, economic sectors that dominate certain countries, differences in corporate governance, and varying degrees of exposure to global economic trends can also lead to discrepancies in factor performance. These market-specific nuances explain why some models consistently outperform others in particular regions but do not necessarily yield the same results across all European markets. Consequently, applying asset pricing models requires careful consideration of local market dynamics to accurately capture the drivers of stock returns.

2.6.3 Simulation Evidence

To ensure robustness in the findings, the approach of Barillas et al. (2020) is followed by examining the small-sample properties of the various test statistics via Monte Carlo simulations. The aim is to assess the reliability of the results from the empirical tests given the sample size.[7] Additionally, by simulating various scenarios, I evaluate how the statistical methods perform under different conditions. These steps are crucial to identify any potential biases or inaccuracies that could arise from applying these methods to small samples. Following the approach of Barillas et al. (2020), factor returns are drawn from a multivariate normal distribution. Actual rejection rates over 100,000 iterations are compared to the nominal 5% level of the tests. This simulation approach is used to compare the top-performing models across each sample.

In Section A.V of Appendix A, the properties of the multiple-comparison inequality test for nonnested models in small samples across the markets considered are investigated. Remember that this test's composite null hypothesis asserts that the $\hat{\theta}^2$ value for the benchmark model is equal to or higher than that of all other models being examined. To assess the size of the test, scenarios where all models share the same $\hat{\theta}^2$ value are considered, thereby maximising the probability of rejecting the null hypothesis under these conditions. In each market the top four performing models from the previous sections are simulated. If these tests are reliable the models in the simulations should not only replicate the empirical results but also behave consistently across different testing scenarios.

For example, in the case of the U.K. market (Table A.16), the four top-performing models—FF6, AFIM, Carhart, and CZZ—were simulated using the sample squared Sharpe ratios as the population $\hat{\theta}^2$. Given that FF6 has the highest $\hat{\theta}^2$, each of the other models was used as the null model in a multiple-comparison test against three alternative models. Table A.16 reports the $\hat{\theta}^2$ of each model, followed by the rejection rates at different significance levels (10%, 5%, 1%). These values represent the proportion of times the null hypothesis (that the benchmark model's $\hat{\theta}^2$ is at least as high as the other models) is rejected at different significance levels. Lower values indicate that the null hypothesis is less frequently rejected, suggesting that the model is performing well relative to the benchmark. The results for different time periods (130, 260, 390) indicate how model performance varies over time. Generally, a model that maintains low rejection rates over longer periods is considered more robust and reliable. In Table A.16, the

[7] I am thankful to Professor Cesare Robotti for the provision of this MATLAB code to perform these simulation experiments.

rejection rates for the CZZ and Carhart models are high, which is expected given their lower $\hat{\theta}^2$ values relative to FF6 and AFIM. The AFIM model shows low rejection rates, consistent with its performance being on par with the FF6 in previous tests.

These results lend robustness to the empirical findings from the model comparison tests conducted for the U.K. market. Specifically, previous tests have consistently shown that the FF6 six-factor model, incorporating both standard and updated value factors, outperforms all other models. This superiority is corroborated by the simulation evidence, which clearly demonstrates the model's effectiveness across various metrics and testing scenarios. Such consistent performance highlights the model's reliability and suitability for analysing U.K. market dynamics.

Across markets in Tables A.17 through A.21, models with strong empirical performance consistently show low rejection rates. For instance, in France (Table A.17), FrazPed excels with the highest $\hat{\theta}^2$ and maintains low rejection rates across all intervals. In Germany (Table A.18), while FF6* has the highest $\hat{\theta}^2$, FF6 also demonstrates low and stable rejection rates, suggesting robust performance. In the Netherlands (Table A.19), FF6, along with Carhart and AFIM, show low rejection rates, indicating performances on par with the leading model. In Italy (Table A.20), AFIM leads in performance, and FF6 distinguishes itself with lower rejection rates than FF5 over time. Lastly, in Spain (Table A.21), AFIM again tops with the highest $\hat{\theta}^2$, with FF6 showing greater stability in performance compared to Carhart. These results reinforce the robustness of the empirical findings, as models like AFIM and FF6 have consistently demonstrated strong performance in empirical tests across European markets, underscoring their reliability and effectiveness.

2.7 Conclusions

The focus of this chapter is twofold. First, the goal is to determine if a single model can consistently outperform others across the selected sample of European markets. Second, if no single model emerges as a consistent top performer, the aim is to identify the best-performing asset pricing model in each specific market. This approach provides insights into the different drivers of returns across markets. These objectives are achieved through comprehensive pairwise and multiple model testing, supported by simulation evidence.

From the perspective of an investor, this chapter underscores the variation in performance of risk factors across the selection of European markets throughout the analysis period. It has been observed that the momentum (MOM) and Betting Against Beta (BAB) factors have shown superior performance in the majority of countries within the sample. Conversely, underperformance has been noted in both the size (SMB) and value (HML) factor portfolios across the entirety of the sample.

Through the analysis, which includes both pairwise and multiple model comparisons, strong performance has been identified in both the Fama and French (2018) six-factor model and the Asness et al. (2015) model with the updated value factor of Asness and Frazzini (2013). These models achieved the highest or second-highest Sharpe ratio in five out of the six countries examined. Despite this, the top-performing model varies by country. For instance, the Frazzini and Pedersen (2014) model stood out as the leading model in both France and the Netherlands. The findings underscore the importance of beta timing, momentum, investment, and profitability as key drivers of stock returns during this period. Models that exclude these factors tend to underperform compared to those that incorporate them. The simulation results support these findings.

My findings align with those of Barillas et al. (2020), who found that a variant of the 6-factor model by Fama and French (2018), featuring a monthly updated version of the usual value spread, emerges as the dominant model in U.S. tests of factor models. Similarly, Fletcher (2019) observes that this six-factor model outperforms other models in U.K. model comparison tests. The results also correspond with observations by Hanauer (2020), who noted that the Asness et al. (2015) six-factor model, as found in Barillas et al. (2020), outperforms other models in aggregate samples of international markets according to the squared Sharpe ratio.

While the best-performing model in each sample has been documented, a key contribution to the European-based literature, it is evident that no single model consistently outperforms all others across the markets under consideration. This observation suggests the presence of local factors influencing stock prices in a European setting. Such a finding underscores the complexity of financial markets and highlights the importance of considering local elements in asset pricing models.

The findings of this chapter offer significant implications for both investors and academics by highlighting the nuanced performance of asset pricing models across different European markets. For investors, these results emphasise the importance of tailoring strategies to specific market conditions rather than relying on a one-size-fits-all approach. The observed outperformance of factors like momentum (MOM) and Betting Against Beta (BAB) in multiple countries suggests that incorporating these factors can yield more effective risk-adjusted returns, especially in markets where traditional factors, such as size (SMB) and value (HML), may underperform. For academics, the lack of a universally dominant model across European markets underscores the need for further research into localised drivers of returns, which can enhance the accuracy of asset pricing models beyond U.S.-centric frameworks. This chapter contributes to the European literature by identifying which factor models align more closely with market behaviours in specific countries, revealing how factors like beta timing, momentum, investment, and profitability influence asset returns in different financial environments. By documenting the variations in top-performing models across markets, this study enriches the discussion on model adaptability and underscores the importance of considering local factors in global asset pricing frameworks, informing future research directions and providing a refined lens for analysing market-specific dynamics.

Appendix A

A.I U.K. Empirical Results

Table 2.1. Summary Statistics for Monthly U.K. Factor Returns

	Mean	StDev	Sharpe Ratio	t-Mean
MKT	0.45%	4.65%	0.097	1.91**
SMB	0.10%	3.19%	0.031	0.61
HML	0.16%	3.11%	0.051	1.01
MOM	0.94%	4.36%	0.215	4.22*
RMW_{Op}	0.22%	2.22%	0.099	1.94**
CMA	0.38%	2.14%	0.175	3.44*
BAB	0.54%	4.65%	0.115	2.26*
HML_{M}	0.07%	3.61%	0.019	0.36
MGMT	0.26%	2.47%	0.103	2.03*
PERF	0.51%	2.97%	0.171	3.36*
RMW_{CB}	0.35%	2.10%	0.166	3.27*
RMW_{ROE}	0.23%	2.29%	0.102	2.01*

*The table reports summary statistics of factors between June 1991 and December 2022 in UK factors. The summary statistics include the average excess returns (Mean) (%), standard deviation (StDev) (%), Sharpe Ratio and the t-statistic of the null hypothesis that the average excess factor returns are equal to zero. * Significant at 5%, ** Significant at 10%*

Table 2.2. Summary Statistics for Monthly Returns Remaining Markets

	France	Germany	Netherlands	Italy	Spain
MKT	0.54%**	0.40%	0.64%*	0.27%	0.45%
SMB	-0.05%	-0.01%	-0.08%	-0.04%	0.11%
HML	0.20%	0.52%*	0.33%	0.15%	0.30%**
MOM	0.57%*	0.92%*	0.83%*	0.83%*	0.70%*
RMW_{Op}	0.25%*	0.42%*	0.14%	0.67%*	0.43%*
CMA	0.19%	0.44%*	0.15%	-0.24%	0.03%
BAB	1.38%*	0.49%*	0.82%*	0.58%*	0.63%*
HML_M	0.15%	0.17%*	0.09%	0.05%	0.27%
MGMT	0.29%*	0.23%	0.10%	-0.01%	0.14%
PERF	0.63%*	0.70%*	0.64%*	0.88%*	0.32%
RMW_{CB}	0.45%*	0.32%*	0.23%	0.44%*	0.00%
RMW_{ROE}	0.25%**	0.08%	0.19%	0.65%*	0.37%*

*The table reports summary statistics of factors between June 1991 and December 2022 in UK factors. The summary statistics include the average excess returns (Mean) (%), standard deviation (StDev) (%), Sharpe Ratio and the t-statistic of the null hypothesis that the average excess factor returns are equal to zero. * Significant at 5%, ** Significant at 10%*

Table 2.3. U.K. Tests of Equality of Squared Sharpe Ratios

Panel A: Difference in Squared Sharpe Performance

Model	FrazPed	FF5	SY	HXZCP	Carhart	CZZ	AFIM	FF6
FF3	-0.014	-0.059	-0.065	-0.078	-0.082	-0.106	-0.126	-0.141
FrazPed		-0.045	-0.051	-0.064	-0.068	-0.092	-0.112	-0.127
FF5			-0.006	-0.019	-0.023	-0.047	-0.067	-0.082
SY				-0.013	-0.017	-0.041	-0.061	-0.076
HXZCP					-0.004	-0.027	-0.047	-0.063
Carhart						-0.023	-0.043	-0.059
CZZ							-0.02	-0.036
AFIM								-0.016

Panel B: p-Values

Model	FrazPed	FF5	SY	HXZCP	Carhart	CZZ	AFIM	FF6
FF3	0.262	0	0.026	0.01	0	0.006	0.002	0
FrazPed		0.137	0.096	0.047	0.037	0.019	0.006	0.003
FF5			0.84	0.151	0.536	0.102	0.021	0.017
SY				0.651	0.605	0.219	0.081	0.02
HXZCP					0.92	0.346	0.155	0.07
Carhart						0.337	0.066	0
CZZ							0.007	0.121
AFIM								0.521

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for different models, some nested and others nonnested on a set of UK factor models from 1991 to 2022. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements are the sample squared Sharpe ratio differences between the model in that column and the next-best model. In Panel B, p-values for the tests of equality of the squared Sharpe ratios are reported. Low p-values indicate the difference identified in Panel A is statistically significant.

Table 2.4. Top Performing Factor Models Across European Markets

Panel A: Identity of Best Models

Country	Top Performing Model	Next Best Performing Model
United Kingdom	Fama and French (2018)	Asness et al. (2015)
France	Frazzini and Pedersen (2014)	Stambaugh and Yuan (2017)
Germany	Asness et al. (2015)	Fama and French (2018)
Netherlands	Frazzini and Pedersen (2014)	Fama and French (2018)
Italy	Asness et al. (2015)	Fama and French (2015)
Spain	Asness et al. (2015)	Chib, Zeng and Zhao (2020)

Panel B: Squared Sharpe Measures

Factor	UK	France	Germany	Netherlands	Italy	Spain
FF3	0.003	0.005	0.021	0.011	0.005	0.005
Carhart	0.085	0.034	0.085	0.049	0.037	0.037
FF5	0.062	0.028	0.069	0.015	0.087	0.029
FF6	0.144	0.067	0.114	0.049	0.062	0.033
FrazPed	0.017	0.089	0.039	0.054	0.016	0.021
SY	0.068	0.068	0.066	0.047	0.046	0.016
AFIM	0.129	0.066	0.13	0.048	0.109	0.065
HXZ	0.081	0.035	0.035	0.016	0.052	0.013
CZZ	0.109	0.04	0.104	0.045	0.068	0.039

Panel A of this table reports the identity of the top two performing models as measured by the squared Sharpe ratio. For the factor composition of these models see Section 2.3.2. Panel B reports the Squared Sharpe ratio for each model across all countries.

Table 2.5. U.K. Multiple Model Comparison

Number of Additional Rejections Using Normal Test = 0

Model	$\hat{\theta}^2$	r	LR	p-value
FF3	0.003	4	11.396	0.003
Carhart	0.085	5	3.388	0.119
FF5	0.062	5	8.6	0.018
FF6	0.144	5	0	0.801
FrazPed	0.017	5	9.134	0.006
SY	0.068	5	5.432	0.05
AFIM	0.129	5	0.412	0.567
HXZ	0.081	5	3.286	0.058
CZZ	0.109	5	2.405	0.217

This table reports the multiple model comparison tests in U.K. stock returns. $\hat{\theta}^2$ is the bias-adjusted maximum squared Sharpe performance of each model identified in the first column. LR in column four is the value of the likelihood-ratio statistic as per Wolak (1987,1989). Column five is the p value for the hypothesis that this model performs as well as any other model in the dataset.

A.II Summary Statistics Remaining Markets

Table A.1. Summary Statistics for Monthly French Factor Returns

Factor	Mean	StDev	Sharpe Ratio	t-Mean
MKT	0.54%	5.44%	0.1	1.96**
SMB	-0.05%	2.85%	-0.02	-0.36
HML	0.20%	3.68%	0.05	1.07
MOM	0.57%	4.59%	0.12	2.43*
RMW_{Op}	0.25%	2.41%	0.1	2.00*
CMA	0.19%	2.45%	0.08	1.56
BAB	1.38%	4.78%	0.29	5.65*
HML_{M}	0.15%	4.15%	0.04	0.70
MGMT	0.29%	2.75%	0.11	2.05*
PERF	0.63%	3.52%	0.18	3.48*
RMW_{CB}	0.45%	2.37%	0.19	3.72*
RMW_{ROE}	0.25%	2.67%	0.09	1.84**

The table reports summary statistics of factors between June 1991 and December 2022 in French factors. The summary statistics include the average excess returns (Mean) (%), standard deviation (StDev) (%), Sharpe Ratio and the t-statistic of the null hypothesis that the average excess factor returns are equal to zero.

* *Significant at 5%,* ** *Significant at 10%*

Table A.2. Summary Statistics for Monthly German Factor Returns

Factor	Mean	StDev	Sharpe Ratio	t-Mean
MKT	0.40%	5.73%	0.07	1.37
SMB	-0.01%	2.95%	0.00	-0.09
HML	0.52%	3.51%	0.15	2.88*
MOM	0.92%	4.66%	0.20	3.87*
RMW_{Op}	0.42%	2.27%	0.18	3.60*
CMA	0.44%	2.96%	0.15	2.91*
BAB	0.49%	4.61%	0.11	2.09*
HML_{M}	0.17%	3.75%	0.05	0.89
MGMT	0.23%	2.92%	0.08	1.52
PERF	0.70%	3.21%	0.22	4.28*
RMW_{CB}	0.32%	2.19%	0.15	2.88*
RMW_{ROE}	0.08%	2.50%	0.03	0.64

*The table reports summary statistics of factors between June 1991 and December 2022 in German factors. The summary statistics include the average excess returns (Mean) (%), standard deviation (StDev) (%), Sharpe Ratio and the t-statistic of the null hypothesis that the average excess factor returns are equal to zero. * Significant at 5%, ** Significant at 10%*

Table A.3. Summary Statistics for Monthly Netherlands Factor Returns

Factor	Mean	StDev	Sharpe Ratio	t-Mean
MKT	0.64%	5.91%	0.108	2.12*
SMB	-0.08%	3.32%	-0.025	-0.49
HML	0.33%	4.00%	0.083	1.63
MOM	0.83%	5.53%	0.15	2.94*
RMW_{Op}	0.14%	3.95%	0.036	0.71
CMA	0.15%	3.29%	0.047	0.91
BAB	0.82%	5.11%	0.161	3.16*
HML_M	0.09%	4.50%	0.02	0.40
MGMT	0.10%	3.79%	0.028	0.54
PERF	0.64%	4.46%	0.143	2.80*
RMW_{CB}	0.23%	3.15%	0.072	1.40
RMW_{ROE}	0.19%	3.58%	0.052	1.02

*The table reports summary statistics of factors between June 1991 and December 2022 in Dutch factors. The summary statistics include the average excess returns (Mean) (%), standard deviation (StDev) (%), Sharpe Ratio and the t-statistic of the null hypothesis that the average excess factor returns are equal to zero. * Significant at 5%, ** Significant at 10%*

Table A.4. Summary Statistics for Monthly Italian Factors Returns

Factor	Mean	StDev	Sharpe Ratio	t-Mean
MKT	0.27%	6.80%	0.04	0.77
SMB	-0.04%	3.18%	-0.01	-0.24
HML	0.15%	3.63%	0.04	0.83
MOM	0.83%	4.73%	0.17	3.43*
RMW_{Op}	0.67%	3.25%	0.21	4.06*
CMA	-0.24%	2.82%	-0.08	-1.67
BAB	0.58%	4.05%	0.14	2.81*
HML_{M}	0.05%	4.40%	0.01	0.24
MGMT	-0.01%	3.72%	0	-0.03
PERF	0.88%	4.48%	0.2	3.85*
RMW_{CB}	0.44%	3.65%	0.12	2.35*
RMW_{ROE}	0.65%	3.38%	0.19	3.78*

*The table reports summary statistics of factors between June 1991 and December 2022 in Italian factors. The summary statistics include the average excess returns (Mean) (%), standard deviation (StDev) (%), Sharpe Ratio and the t-statistic of the null hypothesis that the average excess factor returns are equal to zero. * Significant at 5%, ** Significant at 10%*

Table A.5. Summary Statistics for Spanish Monthly Factor Returns

Factor	Mean	StDev	Sharpe Ratio	t-Mean
MKT	0.45%	6.33%	0.072	1.38
SMB	0.11%	3.32%	0.034	0.65
HML	0.30%	3.40%	0.089	1.72**
MOM	0.70%	5.17%	0.135	2.59*
RMW_{Op}	0.43%	3.14%	0.137	2.64*
CMA	0.03%	3.11%	0.01	0.19
BAB	0.63%	5.01%	0.126	2.44*
HML_{M}	0.27%	3.96%	0.069	1.34
MGMT	0.14%	3.39%	0.04	0.77
PERF	0.32%	3.76%	0.084	1.63
RMW_{CB}	0.00%	3.70%	-0.001	-0.01
RMW_{ROE}	0.37%	3.08%	0.121	2.34*

The table reports summary statistics of factors between June 1991 and December 2022 in Spanish factors The summary statistics include the average excess returns (Mean) (%), standard deviation (StDev) (%), Sharpe Ratio and the t-statistic of the null hypothesis that the average excess factor returns are equal to zero. * *Significant at 5%,* ** *Significant at 10%*

A.III Tests of Equality of Squared Sharpe Ratio Remaining Markets

Table A.6. France Tests of Equality of Squared Sharpe Ratio

Panel A: Difference in Squared Sharpe Performance

Model	FF5	Carhart	HXZ	CZZ	AFIM	FF6	SY	FrazPed
FF3	-0.023	-0.03	-0.03	-0.035	-0.061	-0.062	-0.064	-0.084
FF5		-0.006	-0.007	-0.012	-0.038	-0.039	-0.04	-0.061
Carhart			-0.001	-0.005	-0.031	-0.033	-0.034	-0.054
HXZ				-0.005	-0.031	-0.032	-0.033	-0.054
CZZ					-0.026	-0.027	-0.029	-0.049
AFIM						-0.001	-0.003	-0.023
FF6							-0.001	-0.022
SY								-0.02

Panel B: p-Values

Model	FF5	Carhart	HXZ	CZZ	AFIM	FF6	SY	FrazPed
FF3	0.005	0	0.073	0.088	0.022	0	0.021	0.004
FF5		0.776	0.63	0.456	0.052	0.097	0.097	0.064
Carhart			0.978	0.718	0.085	0.001	0.127	0.095
HXZ				0.76	0.236	0.198	0.177	0.077
CZZ					0	0.167	0.214	0.094
AFIM						0.954	0.919	0.549
FF6							0.948	0.576
SY								0.606

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for different models, some nested and others nonnested on a set of French factor models from 1991 to 2022. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements are the sample squared Sharpe ratio differences between the model in that column and the next-best model. Panel A reports the differences in squared Sharpe ratio. In Panel B, p-values for the tests of equality of the squared Sharpe ratios are reported. Low p-values indicate the difference identified in Panel A is statistically significant.

Table A.7. Germany Tests of Equality of Squared Sharpe Ratios

Panel A: Difference in Squared Sharpe Performance

Model	FF3	HXZ	SY	FF5	Carhart	CZZ	FF6	AFIM
FrazPed	-0.01	-0.024	-0.045	-0.057	-0.073	-0.087	-0.094	-0.118
FF3		-0.014	-0.035	-0.047	-0.063	-0.077	-0.084	-0.108
HXZ			-0.021	-0.033	-0.049	-0.064	-0.07	-0.095
SY				-0.012	-0.028	-0.042	-0.049	-0.073
FF5					-0.015	-0.03	-0.037	-0.061
Carhart						-0.015	-0.021	-0.046
CZZ							-0.007	-0.031
FF6								-0.024

Panel B: p-Values

Model	FF3	HXZ	SY	FF5	Carhart	CZZ	FF6	AFIM
FrazPed	0.632	0.288	0.1	0.033	0.032	0.013	0.015	0.002
FF3		0.471	0.248	0	0	0.033	0	0.002
HXZ			0.44	0.048	0.108	0.022	0.03	0.003
SY				0.667	0.398	0.15	0.128	0.025
FF5					0.624	0.236	0.265	0.023
Carhart						0.561	0.008	0.055
CZZ							0.784	0
FF6								0.324

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for different models, some nested and others nonnested on a set of German factor models from 1991 to 2022. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements are the sample squared Sharpe ratio differences between the model in that column and the next-best model. Panel A reports the differences in squared Sharpe ratio. In Panel B, p-values for the tests of equality of the squared Sharpe ratios are reported. Low p-values indicate the difference identified in Panel A is statistically significant.

Table A.8. Netherlands Tests of Equality of Squared Sharpe Ratios

Panel A: Difference in Squared Sharpe Performance

Model	FF5	HXZ	CZZ	SY	AFIM	Carhart	FF6	FrazPed
FF3	-0.004	-0.005	-0.034	-0.036	-0.037	-0.038	-0.038	-0.043
FF5		-0.001	-0.03	-0.032	-0.033	-0.034	-0.034	-0.039
HXZ			-0.029	-0.031	-0.032	-0.033	-0.033	-0.038
CZZ				-0.002	-0.003	-0.004	-0.004	-0.009
SY					-0.001	-0.002	-0.002	-0.007
AFIM						-0.001	-0.001	-0.006
Carhart							0	-0.005
FF6								-0.005

Panel B: p-Values

Model	FF5	HXZ	CZZ	SY	AFIM	Carhart	FF6	FrazPed
FF3	0.197	0.719	0.141	0.094	0.116	0	0	0.055
FF5		0.938	0.181	0.125	0.137	0.153	0.167	0.124
HXZ			0.186	0.114	0.179	0.192	0.201	0.111
CZZ				0.928	0.142	0.732	0.751	0.754
SY					0.955	0.922	0.914	0.801
AFIM						0.932	0.934	0.846
Carhart							0.379	0.865
FF6								0.872

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for different models, some nested and others nonnested on a set of Dutch factor models from 1991 to 2022. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements are the sample squared Sharpe ratio differences between the model in that column and the next-best model. Panel A reports the differences in squared Sharpe ratio. In Panel B, p-values for the tests of equality of the squared Sharpe ratios are reported. Low p-values indicate the difference identified in Panel A is statistically significant.

Table A.9. Italy Tests of Equality of Squared Sharpe Ratios

Panel A: Difference in Squared Sharpe Performance

Model	FrazPed	Carhart	SY	HXZ	FF6	CZZ	FF5	AFIM
FF3	-0.022	-0.042	-0.052	-0.057	-0.067	-0.074	-0.092	-0.114
FrazPed		-0.02	-0.03	-0.035	-0.045	-0.052	-0.07	-0.093
Carhart			-0.01	-0.015	-0.025	-0.032	-0.05	-0.073
SY				-0.005	-0.015	-0.022	-0.04	-0.063
HXZ					-0.01	-0.017	-0.035	-0.058
FF6						-0.007	-0.025	-0.047
CZZ							-0.018	-0.041
FF5								-0.023

Panel B: p-Values

Model	FrazPed	Carhart	SY	HXZ	FF6	CZZ	FF5	AFIM
FF3	0.107	0	0.025	0.029	0	0.013	0	0.001
FrazPed		0.36	0.226	0.206	0.109	0.077	0.038	0.01
Carhart			0.686	0.616	0.006	0.17	0.135	0.013
SY				0.832	0.562	0.398	0.24	0.08
HXZ					0.741	0.419	0.259	0.052
FF6						0.792	0.417	0.066
CZZ							0.443	0
FF5								0.262

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for different models, some nested and others nonnested on a set of Italian factor models from 1991 to 2022. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements are the sample squared Sharpe ratio differences between the model in that column and the next-best model. Panel A reports the differences in squared Sharpe ratio. In Panel B, p-values for the tests of equality of the squared Sharpe ratios are reported. Low p-values indicate the difference identified in Panel A is statistically significant.

Table A.10. Spain Tests of Equality of Squared Sharpe Ratios

Panel A: Difference in Squared Sharpe Performance

Model	FrazPed	SY	HXZ	FF5	FF6	Carhart	CZZ	AFIM
FF3	-0.011	-0.012	-0.016	-0.022	-0.027	-0.032	-0.037	-0.06
FrazPed		-0.001	-0.005	-0.011	-0.016	-0.02	-0.026	-0.049
SY			-0.004	-0.01	-0.015	-0.02	-0.025	-0.048
HXZ				-0.006	-0.011	-0.016	-0.021	-0.044
FF5					-0.005	-0.01	-0.015	-0.038
FF6						-0.004	-0.01	-0.033
Carhart							-0.006	-0.028
CZZ								-0.023

Panel B: p-Values

Model	FrazPed	SY	HXZ	FF5	FF6	Carhart	CZZ	AFIM
FF3	0.446	0.45	0.364	0.007	0.003	0	0.123	0.022
FrazPed		0.966	0.811	0.634	0.523	0.415	0.322	0.106
SY			0.779	0.606	0.458	0.367	0.193	0.059
HXZ				0.721	0.648	0.531	0.247	0.096
FF5					0.818	0.685	0.413	0.061
FF6						0.794	0.547	0.056
Carhart							0.745	0.105
CZZ								0.002

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for different models, some nested and others nonnested on a set of Spanish factor models from 1991 to 2022. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements are the sample squared Sharpe ratio differences between the model in that column and the next-best model. Panel A reports the differences in squared Sharpe ratio. In Panel B, p-values for the tests of equality of the squared Sharpe ratios are reported. Low p-values indicate the difference identified in Panel A is statistically significant.

A.IV Multiple Model Comparison Remaining Markets

Table A.11. France Multiple Model Comparison

Number of Additional Rejections Using Normal Test = 0

Model	$\hat{\theta}^2$	r	LR	p-value
FF3	0.005	4	14.842	0.001
Carhart	0.034	5	7.449	0.027
FF5	0.028	5	8.078	0.024
FF6	0.067	5	0.313	0.646
FrazPed	0.089	5	0	0.697
SY	0.068	5	0.267	0.631
AFIM	0.066	5	0.359	0.606
HXZ	0.035	5	3.96	0.087
CZZ	0.04	5	4.247	0.099

This table reports the multiple model comparison tests in French stock returns. $\hat{\theta}^2$ is the bias-adjusted maximum squared Sharpe performance of each model identified in the first column. LR in column four is the value of the likelihood-ratio statistic as per Wolak (1987,1989). Column five is the p value for the hypothesis that this model performs as well as any other model in the dataset.

Table A.12. Germany Multiple Model Comparison

Number of Additional Rejections Using Normal Test = 0

Model	$\hat{\theta}^2$	r	LR	p-value
FF3	0.021	4	9.251	0.008
Carhart	0.085	5	3.601	0.111
FF5	0.069	5	5.089	0.079
FF6	0.114	5	0.392	0.533
FrazPed	0.039	5	4.327	0.06
SY	0.066	5	4.211	0.096
AFIM	0.13	5	0	0.83
HXZ	0.035	5	9.231	0.012
CZZ	0.104	5	0.154	0.716

This table reports the multiple model comparison tests in German stock returns. $\hat{\theta}^2$ is the bias-adjusted maximum squared Sharpe performance of each model identified in the first column. LR in column four is the value of the likelihood-ratio statistic as per Wolak (1987,1989). Column five is the p value for the hypothesis that this model performs as well as any other model in the dataset.

Table A.13. Netherlands Multiple Model Comparison

Number of		Additional	Rejections	Using	Normal Test = 0
Model	$\hat{\theta}^2$	r	LR	p-value	
FF3	0.011	4	6.227	0.039	
Carhart	0.049	5	0.029	0.789	
FF5	0.015	5	4.377	0.099	
FF6	0.049	5	0.026	0.779	
FrazPed	0.054	5	0	0.743	
SY	0.047	5	0.064	0.74	
AFIM	0.048	5	0.039	0.781	
HXZ	0.016	5	4.614	0.091	
CZZ	0.045	5	0.165	0.72	

This table reports the multiple model comparison tests in Dutch stock returns. $\hat{\theta}^2$ is the bias-adjusted maximum squared Sharpe performance of each model identified in the first column. LR in column four is the value of the likelihood-ratio statistic as per Wolak (1987,1989). Column five is the p value for the hypothesis that this model performs as well as any other model in the dataset.

Table A.14. Italy Multiple Model Comparison

Number of Additional Rejections Using Normal Test = 0

Model	$\hat{\theta}^2$	r	LR	p-value
FF3	0.005	4	12.557	0.002
Carhart	0.037	5	6.149	0.04
FF5	0.087	5	1.258	0.374
FF6	0.062	5	2.857	0.179
FrazPed	0.016	5	6.792	0.025
SY	0.046	5	3.068	0.164
AFIM	0.109	5	0	0.817
HXZ	0.052	5	2.797	0.176
CZZ	0.068	5	0.587	0.614

This table reports the multiple model comparison tests in Italian stock returns. $\hat{\theta}^2$ is the bias-adjusted maximum squared Sharpe performance of each model identified in the first column. LR in column four is the value of the likelihood-ratio statistic as per Wolak (1987,1989). Column five is the p value for the hypothesis that this model performs as well as any other model in the dataset.

Table A.15. Spain Multiple Model Comparison

Number of Additional Rejections Using Normal Test = 0

Model	$\hat{\theta}^2$	r	LR	p-value
FF3	0.005	4	5.894	0.048
Carhart	0.037	5	2.601	0.202
FF5	0.029	5	3.292	0.144
FF6	0.033	5	3.633	0.128
FrazPed	0.021	5	2.584	0.195
SY	0.016	5	3.714	0.129
AFIM	0.065	5	0	0.797
HXZ	0.013	5	3.008	0.144
CZZ	0.039	5	0	0.842

This table reports the multiple model comparison tests in Spanish stock returns. $\hat{\theta}^2$ is the bias-adjusted maximum squared Sharpe performance of each model identified in the first column. LR in column four is the value of the likelihood-ratio statistic as per Wolak (1987,1989). Column five is the p value for the hypothesis that this model performs as well as any other model in the dataset.

A.V Simulation Evidence for Full Sample

Table A.16. U.K. Multiple Model Comparison Simulation Evidence

			Carhart	FF6	AFIM	CZZ
$\hat{\theta}^2$						
			0.097	0.147	0.136	0.094
T		=			130	
	10%		0.401	0	0.026	0.332
	5%		0.167	0	0.005	0.168
	1%		0.014	0	0	0.023
T		=			260	
	10%		0.822	0	0.108	0.55
	5%		0.593	0	0.031	0.384
	1%		0.143	0	0.001	0.105
T		=			390	
	10%		0.969	0	0.26	0.702
	5%		0.882	0	0.104	0.548
	1%		0.451	0	0.007	0.229

This table reports the $\hat{\theta}^2$ of each model in U.K stock returns from the simulation. The rejection rates at different significance levels (10%, 5%, 1%) are then reported. These values represent the proportion of times the null hypothesis (that the benchmark model's $\hat{\theta}^2$ is at least as high as the other models) is rejected at different significance levels. "T" represents the length of the time period (or sample size) for which the statistical analysis or simulation is conducted. Each "T = 130, 260, 390" corresponds to a different time period or sample size.

Table A.17. France Multiple Model Comparison Simulation Evidence

		FrazPed	FF6	AFIM	SY
$\hat{\theta}^2$					
		0.115	0.062	0.083	0.08
T		=	130		
	10%	0.04	0.32	0.044	0.131
	5%	0.015	0.167	0.016	0.06
	1%	0.001	0.026	0.002	0.011
T		=	260		
	10%	0.023	0.753	0.096	0.21
	5%	0.007	0.58	0.047	0.106
	1%	0	0.243	0.009	0.024
T		=	390		
	10%	0.012	0.931	0.152	0.269
	5%	0.004	0.848	0.077	0.162
	1%	0	0.568	0.016	0.041

This table reports the $\hat{\theta}^2$ of each model in French stock returns from the simulation. The rejection rates at different significance levels (10%, 5%, 1%) are then reported. These values represent the proportion of times the null hypothesis (that the benchmark model's $\hat{\theta}^2$ is at least as high as the other models) is rejected at different significance levels. "T" represents the length of the time period (or sample size) for which the statistical analysis or simulation is conducted. Each "T = 130, 260, 390" corresponds to a different time period or sample size.

Table A.18. Germany Multiple Model Comparison Simulation Evidence

		Carhart	FF6	AFIM	CZZ
$\hat{\theta}^2$					
		0.097	0.137	0.149	0.083
T	=		130		
	10%	0.365	0.018	0.002	0.45
	5%	0.168	0.003	0	0.251
	1%	0.017	0	0	0.036
T	=		260		
	10%	0.726	0.053	0.002	0.734
	5%	0.504	0.014	0	0.563
	1%	0.121	0	0	0.212
T	=		390		
	10%	0.908	0.095	0.002	0.874
	5%	0.773	0.034	0	0.766
	1%	0.338	0.002	0	0.441

This table reports the $\hat{\theta}^2$ of each model in German stock returns from the simulation. The rejection rates at different significance levels (10%, 5%, 1%) are then reported. These values represent the proportion of times the null hypothesis (that the benchmark model's $\hat{\theta}^2$ is at least as high as the other models) is rejected at different significance levels. "T" represents the length of the time period (or sample size) for which the statistical analysis or simulation is conducted. Each "T = 130, 260, 390" corresponds to a different time period or sample size.

Table A.19. Netherlands Multiple Model Comparison Simulation Evidence

			Carhart	FF6	AFIM	FrazPed
$\hat{\theta}^2$						
			0.057	0.064	0.062	0.042
T		=		130		
	10%		0.039	0.002	0.005	0.254
	5%		0.008	0	0.001	0.122
	1%		0	0	0	0.013
T		=		260		
	10%		0.065	0.003	0.011	0.325
	5%		0.016	0	0.002	0.181
	1%		0.001	0	0	0.031
T		=		390		
	10%		0.105	0.002	0.018	0.379
	5%		0.027	0	0.004	0.237
	1%		0.001	0	0	0.058

This table reports the $\hat{\theta}^2$ of each model in Dutch stock returns from the simulation. The rejection rates at different significance levels (10%, 5%, 1%) are then reported. These values represent the proportion of times the null hypothesis (that the benchmark model's $\hat{\theta}^2$ is at least as high as the other models) is rejected at different significance levels. "T" represents the length of the time period (or sample size) for which the statistical analysis or simulation is conducted. Each "T = 130, 260, 390" corresponds to a different time period or sample size.

Table A.20. Italy Multiple Model Comparison Simulation Evidence

		FF5	FF6	AFIM	CZZ
$\hat{\theta}^2$					
		0.101	0.115	0.125	0.047
T = 130					
	10%	0.126	0.015	0.004	0.686
	5%	0.041	0.004	0.001	0.464
	1%	0.001	0	0	0.099
T = 260					
	10%	0.277	0.031	0.003	0.933
	5%	0.119	0.008	0.001	0.849
	1%	0.01	0	0	0.482
T = 390					
	10%	0.426	0.057	0.004	0.988
	5%	0.238	0.021	0.001	0.964
	1%	0.038	0.001	0	0.809

This table reports the $\hat{\theta}^2$ of each model in Italian stock returns from the simulation. The rejection rates at different significance levels (10%, 5%, 1%) are then reported. These values represent the proportion of times the null hypothesis (that the benchmark model's $\hat{\theta}^2$ is at least as high as the other models) is rejected at different significance levels. "T" represents the length of the time period (or sample size) for which the statistical analysis or simulation is conducted. Each "T = 130, 260, 390" corresponds to a different time period or sample size.

Table A.21. Spain Multiple Model Comparison Simulation Evidence

			Carhart	FF6	AFIM	CZZ
$\hat{\theta}^2$						
			0.049	0.071	0.083	0.038
T		=		130		
	10%		0.181	0.02	0.001	0.305
	5%		0.058	0.004	0	0.122
	1%		0.003	0	0	0.007
T		=		260		
	10%		0.481	0.077	0	0.703
	5%		0.247	0.022	0	0.456
	1%		0.027	0.001	0	0.076
T		=		390		
	10%		0.734	0.156	0	0.906
	5%		0.496	0.055	0	0.751
	1%		0.107	0.003	0	0.29

This table reports the $\hat{\theta}^2$ of each model in Spanish stock returns from the simulation. The rejection rates at different significance levels (10%, 5%, 1%) are then reported. These values represent the proportion of times the null hypothesis (that the benchmark model's $\hat{\theta}^2$ is at least as high as the other models) is rejected at different significance levels. "T" represents the length of the time period (or sample size) for which the statistical analysis or simulation is conducted. Each "T = 130, 260, 390" corresponds to a different time period or sample size.

Chapter 3 Time Series Efficient Factors

Abstract

This chapter explores the use of serial correlation in factor returns to enhance Sharpe performance of the models examined in the previous chapter. Adopting the methodology of Ehsani and Linnainmaa (2022), the analysis demonstrates that multiple investment factors in the cross-country dataset are unconditionally minimum-variance inefficient: factor returns are positively autocorrelated, while risk remains constant regardless of past returns. Using Ferson and Siegel's (2001) general framework, 'time-series efficient factors' are constructed by conditioning factor weights on historical returns to enhance the Sharpe ratios of these factors across the European markets under consideration. A number of these optimised factors achieve significantly higher average Sharpe ratios compared to the original factors, while retaining all the information contained in the original factors. When the model comparison tests from Chapter Two are repeated with these optimised factors, the absolute performance of the lower-performing models improves, while the relative performance among the models remains consistent across markets.

3.1 Introduction

Asset pricing models can be enhanced either through the addition of factors that broaden the efficient frontier or by improving the mean-variance efficiency of the existing factors. In the preceding chapter, by adopting the first strategy, I found that models incorporating the momentum factor, as suggested by Fama and French (2018) and Asness et al. (2015), consistently surpassed others in the analysis. This finding is consistent with prior observations, where the momentum factor yields some of the highest average monthly returns among the factors. In this chapter, I pursue the second strategy, aiming to enhance model performance by scrutinising the mean-variance efficiency of the underlying factors.

Ehsani and Linnainmaa (2022) challenge the traditional perspective that views momentum in individual stock returns as an independent phenomenon. Instead, they suggest that the momentum commonly observed at the stock level is not a unique or isolated effect. It is, rather, a consequence of the autocorrelations presents within the broader market factors. Autocorrelation, in this context, refers to the tendency of a factor's returns to be correlated with its past returns, creating a pattern of momentum. In an effort to leverage the predictive capabilities of autocorrelation observed in factor returns, the authors introduce a concept termed as a 'time-series efficient factor.' This innovative approach utilises autocorrelation to optimally time the original factor, aiming to reduce variance while preserving the expected return. This approach stems from the framework of Ferson and Siegel (2001) which finds an unconditionally minimum variance efficient portfolio. Such a strategy ensures that if a factor can achieve reduced variance without compromising on stable returns, an improvement in the factor's Sharpe ratio will be observed. The addition of a distinct momentum factor to a model with factors constructed in this way should not be able to enhance the model performance. This is because time-series efficient factors already capture the predictable variations in factor premiums that these momentum strategies aim to exploit.

Time-series efficient factors address Cochrane's 2011 critique of the expanding "factor zoo" in asset pricing, where he pointed out the inefficiency of many asset pricing factors in unconditional mean-variance terms. Utilising these inefficient factors as benchmarks could lead to misleading conclusions about the effectiveness of asset pricing models, mistaking minor improvements over inefficient benchmarks for genuine anomalies. Ehsani and Linnainmaa (2022) argue that employing efficient factors ensures that any observed performance improvements are due to the real contribution of new factors, rather than the baseline inefficiency of the benchmarks. This is crucial when assessing models that include a momentum factor, as seen in Chapter Two, to determine if their outperformance stems from the genuine efficacy of the momentum factor or from the inefficiencies in the model's other factors.

In this chapter, the mean-variance efficiency of factors in the European dataset is explored with two main objectives. First, assessing whether the Sharpe ratio of these factors can be improved using a time-series efficient method. Second, mitigating the influence of the momentum factor on model comparison tests by reallocating the momentum observed in these factors' returns back to the factors themselves. This adjustment could potentially identify a different top-performing model across the samples. To verify this, model comparison tests from Chapter Two are rerun.

In focusing on the time-series efficiency of asset pricing factors, this chapter aims to refine our understanding of factor-based investing by reducing reliance on an expanding array of factors, as highlighted by Cochrane (2011). This approach not only addresses theoretical concerns about the proliferation of asset pricing factors but also responds to practical challenges faced by investors who seek streamlined models with robust predictive power. By improving the mean-variance efficiency of factors using conditioning information, such as past returns, this study explores a model enhancement technique that remains largely underutilised in non-U.S. markets. This optimisation aligns with recent findings by Moreira and Muir (2017) on volatility timing, further suggesting that time-series adjusted factors can enhance model robustness in environments where standard momentum and other traditional factors have shown mixed performance. In the European context, this efficient factor approach may offer investors and academics a new lens to interpret the dynamics of returns, providing a foundation for models that respond more accurately to localized economic cycles and market conditions. Ultimately, this chapter aims to demonstrate that factor efficiency improvements can yield models that are not only statistically sound but also practically valuable, reducing the risk of overfitting and enhancing the economic interpretability of factor-based asset pricing.

The analysis shows that key investment factors, such as the size factor, exhibit mean-variance inefficiency in European markets. This finding is particularly important, as these factors have developed a reputation for generating low and often insignificant returns over the past three decades (Artmann et al., 2012; Fletcher, 2019). This chapter suggests their perceived lack of impact is more about inefficiency in the mean-variance framework rather than outright insignificance. Furthermore, the study finds that the efficient factors reduce the reliance on a distinct momentum factor across models. The model comparison tests show that the optimisation improves the Sharpe performance of some models. However, the relative Sharpe performance of models across markets remains similar to the original tests.

3.2 Momentum in Stock Returns

3.2.1 Time Series Momentum

Adjusting portfolio weights based on past returns requires a relationship between a factor's return and its historical performance. Price momentum, known for assets continuing their recent outperformance or underperformance, has been extensively studied in financial research. Early studies examined the statistical traits of price series, like return autocorrelation, to test the random walk model. Kendall and Hill (1953) noted uncorrelated initial price changes, leading to further exploration of autoregressive models. Alexander (1961) developed trading strategies using filters to identify profitable signals amid market noise, supporting the success of momentum-based strategies. This contrasted with standard risk-based theories, suggesting that momentum effects stem from behavioural biases and market frictions. Fama (1965) probed the predictive capability of historical stock prices, bolstering market efficiency theory by indicating that prices change independently, thus questioning the effectiveness of technical analysis for predicting price movements.

Expanding on Fama's insights, Roll (1984) finds evidence that bid-ask spreads contribute to serial correlation in stock returns, which impacts the variance of stock returns and biases tests of market efficiency. Lo and MacKinlay (1988) provided significant contributions to the understanding of stock return behaviours as they find evidence that stock prices do not follow a random walk, particularly noting positive autocorrelation in weekly returns, challenging the efficient market hypothesis. Other research, notably by Kahneman et al. (1982) and Shiller, further scrutinised market behaviour and questioned prevailing assumptions about stock price determinants. Basu (1977) proposed the "price-ratio" hypothesis, shedding light on valuation anomalies. Studies by Irwin and Uhrig (1984), Tomek and Querin (1984), and Sweeney (1986) delved into trading strategies and market inefficiencies.

It is important to distinguish between time-series and cross-sectional momentum. Time-series momentum evaluates an asset's future performance based on its own past returns, while cross-sectional momentum selects assets that outperform their peers. Seminal works include Jegadeesh and Titman (1993, 2001), who found that stocks with strong past performance tend to continue performing well in the medium term, and De Bondt and Thaler (1985), who examined market overreaction. Rouwenhorst (1998) extended these findings to European markets, showing consistent momentum effects across 12 countries. Carhart (1997) incorporated a momentum factor into the Fama-French three-factor model, creating a four-factor model that better explained mutual fund returns. Other relevant studies include Barberis, Shleifer, and Vishny (1998) on market reactions to news, Daniel, Hirshleifer, and Subrahmanyam (1998) on investor overconfidence, and Clare and Thomas (1995) and Dissanaike (1997), who observed momentum effects in the U.K. market.

3.2.3 Factor Momentum

3.2.3.1 Return Timing

The work of Moskowitz, Ooi, and Pedersen (2012) is considered a seminal paper in the area of time series momentum in financial markets more generally as they examine other asset classes not just individual stock prices. The authors document substantial "time series momentum" across various asset classes, including equity index, currency, commodity, and bond futures. The study covered 58 liquid instruments and observed persistence in returns spanning one to 12 months. This persistence partially reverses over longer periods, aligning with sentiment theories that suggest initial under-reaction and delayed over-reaction in markets. In their research, they showed that indices representing different asset classes, such as commodities, bonds, and currencies, could be effectively timed based on their recent performance trends. These indices, when considered as "factors" within their respective asset classes, illustrate the phenomenon of factor momentum, indicating that past performance trends can be indicative of future performance.

Gupta and Kelly (2019) note that in recent decades, academic literature and industry practice have accumulated dozens of investment factors that help explain the co-movement and average returns among individual stocks. The authors build on the work on Moskowitz et al. (2012) through the analysis of a large collection of 65 such characteristic-based factors that are widely studied in the academic literature. From this dataset, they establish factor momentum as a robust and pervasive phenomenon outlining that serial correlation in returns is the basic statistical phenomenon underlying momentum. Gupta and Kelly (2019) demonstrate that individual factors can indeed be successfully timed based on their own past performance. The authors show that a time-series momentum trading strategy that scales exposure to a given factor in proportion with its own past one-month return generates excess performance over and above the raw/original factor. Specifically, the authors buy the factor if its previous return is positive and sell if negative, then perform a regression to assess the alpha generated by this strategy. The alpha obtained from this regression represents the individual time-series momentum alpha, which is the measure of the excess performance of the time-series momentum strategy over the raw/original factor return. This individual time-series momentum alpha (i.e., after controlling for a passive investment in the factor) is positive for 61 of the 65 factors and is statistically significant for 47 of them.

The study of Gupta and Kelly (2019) also revealed that a time series "factor momentum" portfolio, which combines timing strategies of all factors, yields a notable annual Sharpe ratio of 0.84. This factor momentum significantly enhances the performance of investment strategies that use traditional momentum, industry momentum, value, and other commonly studied factors. They outline that whether the look-back window is as short as one month or as long as five years, their strategy identifies large positive momentum among factors. Gupta and Kelly (2019) discover that these phenomena are just as prevalent outside the United States, evident both in a comprehensive global sample (excluding the United States) and in a more detailed European sample encompassing the markets under their investigation in this study.

The pervasiveness of momentum over time and across asset classes has given momentum the status of an independent factor—models without momentum cannot explain it and those with momentum cannot explain anything more than just momentum (Fama and French, 2016). Ehsani and Linnainmaa (2022) show that momentum is a dynamic portfolio that times other factors. These authors first confirm the findings of Gupta and Kelly (2019) through a variety of momentum proxy tests. They show that the average factor earns 51 basis points per month following a year of gains but just 6 basis points following a year of losses. They also show that small stocks, for example, are likely to outperform big stocks when they have done so over the prior year. Using a set of 15 U.S. anomalies they also demonstrate that the past returns of factors are a significant predictor of their future returns. Time-series regressions were conducted where the dependent variable is a factor's return in a given month, and the independent variable is an indicator for the factor's performance in the preceding month. This indicator is assigned a value of one if the factor's return was positive in that period, and zero otherwise. The intercepts in from these time series regressions measure the average factor returns earned following a year of underperformance. The slope coefficient represents the average return difference between the up and down years. In these regressions all slope coefficients are positive with 10 of the 15 statistically significant at the 10% level of significance.

The main hypothesis of Ehsani and Linnainmaa (2022) is whether individual stock returns show momentum beyond what is explained by factor returns. To explore this, they compare strategies based on individual stock momentum against those based on factor momentum. Factor momentum is defined by extracting principal components from 47 factors, following Kozak, Nagel, and Santosh (2020), with higher eigenvalue components showing more momentum. Their findings suggest that factor momentum can price portfolios sorted by past one-year returns better than the traditional Carhart (1997) momentum factor. They also create momentum-neutral factors by adjusting the weights of factors to ensure they are independent of past returns. These adjusted factors yield similar returns but with reduced volatility and higher information ratios, showing that factor momentum is not incidental but an inherent part of these factors. The results indicate that factor momentum drives performance more effectively than stock-specific momentum, challenging the idea that individual stock momentum alone explains returns.

In a follow-up paper, Ehsani and Linnainmaa (2022) advance the work of Gupta and Kelly (2019) by exploring factor momentum in more depth, particularly through the lens of time-series efficiency. More specifically, they examine the autocorrelation in factor returns and show how it can be leveraged to create time-series efficient factors that maximise the Sharpe ratio by balancing return and risk more effectively. Ehsani and Linnainmaa (2022) argue that just because factor returns show autocorrelation, it does not automatically mean that investors can time the market successfully. For successful timing, the changes in expected returns must not align perfectly with changes in expected volatility. The typical factor's expected return depends on its past return, but its variance does not. This creates an opportunity for timing factors to improve their Sharpe ratios (a measure of risk-adjusted return). To exploit this, the authors create 'time-series efficient' versions of standard factors. These new versions are designed by an investor who believes in a specific statistical model (an AR(1) process) for a factor's returns. The investor uses the factor's mean, variance, and autocorrelation to form factors that are mean-variance efficient (MVE) unconditionally, which means they maximise returns for a given level of risk. These time-series efficient factors use the autocorrelation in factor returns to minimise variance while maintaining expected return, leading to higher Sharpe ratios than the original factors in U.S stock returns.

3.2.3.2 Volatility Timing

In addition to using factors' past returns for timing strategies, investors can also incorporate other predictive measures, such as valuation ratios and past volatilities, to optimise their portfolios. For instance, Haddad et al. (2020) demonstrate that book-to-market ratios, a measure of a company's valuation, can be effective in forecasting expected returns. By analysing the relationship between valuation metrics and future stock performance, investors can implement timing strategies that adjust their exposure to stocks based on their expected return potential, particularly when valuation spreads signal significant opportunities.

Another prominent timing strategy, volatility timing, is detailed in the work of Moreira and Muir (2017). Their method involves dynamically adjusting portfolio exposure based on the inverse of the previous month's realized variance. Specifically, they scale the returns of various asset pricing factors—such as the market, value, momentum, and carry trade—by increasing exposure during periods of low volatility and reducing it during times of high volatility. This approach is grounded in the principle that high volatility often corresponds with increased market risk, where maintaining the same level of exposure could lead to substantial losses. By contrast, during periods of stability, increasing exposure allows investors to capitalise on potential returns. The appeal of Moreira and Muir's approach lies in its simplicity and practicality, as it does not require complex parameter estimation and can be implemented in real-time without leverage. The strategy has proven particularly effective in improving the mean-variance trade-off, resulting in significant alphas and higher Sharpe ratios compared to traditional buy-and-hold strategies. Furthermore, volatility-managed portfolios have demonstrated robustness, outperforming during market downturns, such as recessions and financial crises, by minimising risk exposure when markets are turbulent and maximising returns during periods of calm. Importantly, the author finds that this approach remains effective even after accounting for transaction costs and potential leverage constraints, making it a practical tool for real-world investors.

Expanding on nonlinear timing approaches, Kozak et al. (2020) propose an advanced strategy that leverages a broad array of stock characteristics to forecast returns. Their nonlinear model allows for a more sophisticated understanding of the relationships between various stock attributes and future returns, using machine learning techniques to identify patterns that are not apparent in traditional linear models. By incorporating a wide range of characteristics, such as firm size, book-to-market ratio, momentum, and past volatilities, this approach can dynamically adjust portfolio exposure in a way that captures complex, nonlinear dependencies. This method stands out for its ability to adapt to changing market conditions and capture the multifaceted nature of risk and return, providing a more comprehensive framework for timing strategies. The authors find that portfolios constructed using nonlinear timing approaches often achieve superior performance compared to those built on linear timing models, as they better exploit information embedded in stock characteristics.

Further, Bollerslev, Tauchen, and Zhou (2009) contribute to the understanding of volatility timing by examining how real-time measures of volatility, such as implied volatility indices and high-frequency volatility estimates, can predict future returns. They show that periods of high implied volatility often precede lower future returns, suggesting that volatility is a key indicator of risk aversion and expected market performance. Their findings highlight the importance of integrating volatility measures into asset pricing models, as these measures provide valuable signals for adjusting risk exposure. Campbell and Cochrane (1999) also introduce the concept of consumption-based volatility timing. Their model suggests that when consumption is relatively high compared to its long-term trend, expected returns are lower, and vice versa. This relationship can be used to time the market by adjusting exposure based on the ratio of current consumption to its trend. This approach captures time-varying risk premiums and demonstrates that risk aversion fluctuates over the economic cycle, influencing expected stock returns.

Lastly, the variance risk premium, studied by Bollerslev, Todorov, and Xu (2015), provides another avenue for timing returns. The variance risk premium is the difference between implied and realised variances and serves as a predictor of future equity returns. By incorporating this premium into their models, investors can gain insights into market sentiment and adjust their portfolios accordingly. High variance risk premiums often signal increased market uncertainty, prompting a reduction in risk exposure, while low premiums suggest a more favourable risk-return environment.

This chapter focuses exclusively on return timing, rather than volatility timing, because the two strategies address different aspects of market dynamics. While volatility timing adjusts risk exposure based on changes in volatility, return timing is centred on exploiting patterns in price returns to maximise returns for a given level of risk. Given the outperformance of the momentum factor based on price returns, as demonstrated in Chapter 2, our approach emphasises return-based signals rather than volatility adjustments. Furthermore, volatility timing requires managing exposure in response to past variance, which involves a different set of assumptions and data considerations. By concentrating on return timing, we align more closely with the core principles of momentum strategies that have proven effective in our earlier analysis.

3.3 Research Methods

3.3.1 General Framework

In the previous chapter, we outlined how a significant alpha can arise from the factors included in a model not being combined in a way that forms a minimum-variance efficient portfolio. Ehsani and Linnainmaa (2022) explain that this significant alpha could result from the factors themselves not being mean-variance efficient. In other words, there is information within the factors' returns that can be utilised to improve their mean-variance profile. This chapter adopts Ehsani and Linnainmaa's methodology to enhance the model by exploiting factors' time-series properties, aiming to improve Sharpe ratios by considering the independence of returns and volatility over time.

Ehsani and Linnainmaa (2022) use the framework of Ferson and Siegal (2001) to construct time series efficient factors. Time series efficient meaning that the weight placed on a factor is a function of conditioning information, which in this case is the prior month return of the factor. Ferson and Siegal (2001) outline a framework to find an unconditionally minimum variance efficient portfolio based on some conditioning information. First, this method is described for the case of one risky asset and the risk-free asset. Then, this framework is applied to factors with past returns as the conditioning information.

Starting from a single risky asset with a return of

$$\tilde{R} = \mu(\tilde{S}) + \tilde{\varepsilon}, \tag{3.1}$$

where $\tilde{R}$ is the risky asset's return in excess of the risk-free rate, $\tilde{S}$ is the predictor (signal), $\mu(\tilde{S})$ is the expected excess return conditional on the signal, and $\tilde{\varepsilon}$ is the random noise net of the signal with a mean of zero and a variance of $\sigma^2_\varepsilon(\tilde{S})$. The efficient strategy invests $x(\tilde{S})$ in the risky asset and the remainder, $1 - x(\tilde{S})$, in the risk-free asset. The unconditional expected excess return and variance of this investment strategy are given by

$$\mu_p = E\left[x(\tilde{S})\, \mu(\tilde{S})\right] \tag{3.2}$$

$$\sigma_p^2 = E\left[x^2(\tilde{S})\left(\mu^2(\tilde{S}) + \sigma_\varepsilon^2(\tilde{S})\right)\right] - \mu_p^2 \tag{3.3}$$

Ferson and Siegel (2001) show that the portfolio that minimises σ_p^2 for a given conditional expectation $\mu_p$invests $x(\tilde{S})$ in the risky asset,

$$x(\tilde{S}) = \frac{\mu_P}{\partial}\frac{\mu(\tilde{S})}{\mu^2(\tilde{S}) + \sigma_\varepsilon^2(\tilde{S})} \tag{3.4}$$

Here μ_P denotes the unconditional expected factor returns obtained from the original factor. The conditional expected portfolio returns $\mu(\tilde{S})$, assuming an AR(1) model is used to condition the time-series Efficient factor on, and the constant ∂ are defined below.

$$\partial = \frac{\mu^2(\tilde{S})}{\mu^2(\tilde{S}) + \sigma_\varepsilon^2(\tilde{S})} \tag{3.5}$$

This optimal proportion to invest in this risky asset is a function of the portfolio's expected return and the ratio of the signal's squared expected return to the sum of its squared expected return and variance denoted in 3.5. This weighting program produces a unique mean-variance efficient portfolio. That being no other portfolio has the same unconditional return at a lower unconditional variance (Ferson and Siegel, 2001). This analysis can be applied to any sort of conditioning information.

3.3.2 Time Series Efficient Factors

Ehsani and Linnainmaa (2022) use information embedded in the factors' realised returns as the signal to create this minimum variance portfolio. In other words, the prior month return of a given factors may contain information useful in deciding the optimal weight on that factor in the following month. This case gives a closed-form solution for the MVE transformation and for the expected efficiency gain or increase in Sharpe ratio. The new factors the authors construct, using information only in factors' past returns, are weak-form efficient in the sense of Fama (1970).

Ehsani and Linnainmaa (2022) assume that past returns are related to future returns but unrelated to variance. Specifically, it is assumed that returns follow a homoscedastic autoregressive process,

$$\tilde{R}_t = \mu + \rho\tilde{R}_{t-1} + \varepsilon_t \tag{3.6}$$

$$var[\varepsilon_t|R_{t-1}] = \sigma_\varepsilon^2 \tag{3.7}$$

The factor's conditional expected return under this model is $\mu(\tilde{S}) = \mu + \rho\tilde{R}_{t-1}$. Using equations (3.6) and (3.7), the investor's optimal weight on the factor is

$$x(S_t) = \mu_p \frac{\mathrm{SR}^2+1}{\mathrm{SR}^2+\rho^2} \frac{\mu_p(1-\rho)+\rho r_{t-1}}{\left(\mu_p(1-\rho)+\rho r_{t-1}\right)^2+\sigma_\varepsilon^2} \tag{3.8}$$

In this equation, μ_p is the factor's unconditional mean, *SR* is the unconditional Sharpe ratio, ρ is the autocorrelation coefficient, and $\sigma^2_\varepsilon = (1 - p^2)\ \sigma^2$ is the constant variance of the noise term. A time-series efficient factor is defined as the portfolio that invests $x(S_t)$ on the original factor. In the empirical work, like Ehsani and Linnainmaa (2022), I use month $t - 1$ return as the conditioning information. The optimal weight on a given factor depends on the factor's mean, standard deviation, and first-order autocorrelation. The efficient constant and portfolio weights are calculated for each time point t using the inverse of the sum of the covariance matrix of residuals from equation 3.6 and the square of the conditional mean. The portfolio weight vector calculated for each time point.

The regression performed for each factor is a simple time-series regression of the factor on its lagged value. This regression helps understand how a factor's past value might influence its current value. Weights are then calculated using both the conditional mean and the conditional variance of this regression. This means that the weights are a function of both how well the factor's current value can be predicted from its past (the conditional mean) and how much uncertainty there is in this prediction (the conditional variance). Applying these weights to the original factors produces a new set of factor returns. The model uses these in-sample estimates to calculate the optimal weight an investor should place on a factor for future periods, based on the historical performance and characteristics of that factor.

As outlined by Ehsani and Linnainmaa (2022) and the weighting program is not monotone in the past return. Although the optimal weight initially increases if a past return is high, the optimal strategy begins to scale back as the past return becomes abnormally high. Similarly, the investor begins to scale back on shorting the factor when the factor's past return is very low. The investor's objective is to minimise risk while maintaining a steady average return. If a signal indicates very high potential returns, an investor might make a big bet, but to manage risk, they will invest less in risky assets, aiming for smoother returns. Essentially, a high return in the previous month allows an investor to reduce their investment in that factor, lowering risk without sacrificing returns. The intuition behind this is described by Ferson and Siegel (2001) where the primary focus is on conditional asset pricing models. These models incorporate time-varying information (like economic indicators or past asset returns) to adjust the expected returns and risks of assets or portfolios. Their approach is based on the premise that the expected returns of assets are not static but vary with the economic environment or other conditioning information.

Ferson and Siegel (2001) explain that unconditionally efficient portfolios, which maximise expected return for a given risk level without additional information, must also be conditionally efficient. However, the reverse is not true. Their model shows that when future returns are moderately predicted, portfolio weights follow traditional investment strategies. But when signals indicate extreme returns, the strategy becomes more cautious, reducing investments in risky assets despite high expected returns to manage overall risk.

3.4 Data

The analysis uses the same dataset from Chapter Two, incorporating a set of 12 tradable investment factors. These factors can be categorised into two main groups. The initial group originates from the work of Fama and French (1993, 2015) and Carhart (1997), covering the Market factor's excess returns and zero-cost portfolios for size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (MOM) effects. The second group features the betting against beta (BAB) factor introduced by Frazzini and Pedersen (2014), an updated value factor (HML_M) by Asness and Frazzini (2013), and two mispricing factors from 11 market anomalies named Management (MMGT) and Performance (PERF) by Stambaugh and Yuan (2017), as elaborated in section 2.3.1. The factor data for this analysis is retrieved from the Globalfactorpremia.org database and the AQR Capital Management database. Again, the period of analysis is between June 1991 and December 2022. All factors are denominated in USD.

3.5 Empirical Results

3.5.1 Time Series Efficient Factors

3.5.1.1 Predictive Regressions

First, the relationship between factor monthly returns is examined by employing predictive regression analyses on the set of factors in European stock returns. This method directly assesses the potential of timing investment strategies through the lens of historical data's ability to forecast future performance. By regressing future returns on past returns, the analysis aims to uncover the degree of autocorrelation, which serves as a crucial indicator of their persistence over time. A significant AR(1) coefficient suggests a robust linkage between consecutive returns, indicating the possibility of enhancing investment strategies by leveraging signals from past performance. These strategies, predicated on the historical continuity of returns, aim to achieve improved Sharpe ratios, denoting more efficient risk-adjusted returns. However, when predictive regression does not demonstrate significant autocorrelation, the viability of timing strategies based on past returns appears limited. In such cases, certain factors may lack the historical patterns necessary for adjusting investment approaches based solely on past performance.

Table 3.1

The analysis across European markets reveals varying levels of predictive power from past returns. The predictive return beta values from equation 3.7 and their t-statistics are examined in Table 3.1. The U.K. stands out with notably strong and significant β values for factors such as HML (0.337 with a t-statistic of 6.995) and HML_M (0.304 with a t-statistic of 6.225), suggesting these factors' past returns might be particularly predictive of future returns in the U.K. market. Germany shows a similar trend, with factors like BAB (0.22 with a t-statistic of 4.411) and CMA (0.206 with a t-statistic of 4.099) also indicating robust predictive relationships. In contrast, France and the Netherlands reveal lower β values and fewer instances of significant results, pointing to a less pronounced predictive power of past returns for the examined factors within these markets. Meanwhile, Italy and Spain typically exhibit lower β values and levels of significance, although exceptions exist, such as BAB in Italy (0.138 with a t-statistic of 2.713), suggesting some potential for predictive utility.

This nuanced landscape underscores the variability in the predictive power of past returns across different factors and countries. For instance, the market factor (MKT) displays positive β values across all countries, with varying degrees of significance, particularly marked in the U.K., but less so in Italy and Spain. The value and momentum factors (HML and MOM) show a general trend of positive β values across the board, with HML being notably significant in the U.K. and Germany. This indicates that value strategies may possess predictive power in these regions. Furthermore, factors related to operational efficiency and asset growth (RMW_{OP} and CMA) are highlighted as significantly predictive in specific markets like the U.K. and Germany, pointing to the potential of these investment styles in certain European contexts. The significant predictive power in certain factors and markets suggests that understanding past return patterns can benefit investment decisions, provided these insights are applied with a nuanced and context-aware approach.

3.5.1.2 Risk Return Characteristics

Despite observing significant autocorrelation in prices, evaluating potential enhancements in the key performance metric, the Sharpe ratio, also requires examining volatility co-movement. Again, due to space constraints in the main body of the document, the results for the U.K. are included in the primary tables, while results for the other five markets are detailed in various sections of Appendix B.

First, this potential gain is assessed with a more empirical approach. In Table 3.2, following the approach of Ehsani and Linnainmaa (2022), U.K. factors are assigned into two equally sized groups based on their month t-1 returns. The average annualised returns and volatilities for each group are reported. The full sample is used to create two groups of equal size: one group with returns following a period of high returns and the other group with returns following a period of low returns. The annualised returns for each of these groups for each factor are reported in the first two columns of the table. The 'H-L' column reports the difference between the two subgroups. To establish if an investor should vary factor allocations with changes in expected return, the joint dynamics of variance and returns must be analysed. The annualised volatility of each group for each factor is reported in the remaining three columns of Table 3.2.

Table 3.2

For nine of the twelve factors the difference in expected returns (H-L) based on the prior month return is significantly different from zero at a significance level of 5% and one at the 10% level. In other words, there is a statistically significant difference in the returns that follow periods of low returns and periods of high returns. The findings in differences in expected factor returns are in line with the AR(1) coefficients ρ and their t-statistics stated in the appendices. The estimates show that when the Size factor (SMB) return in month *t*-1 is low, it's average annualised return is -1.74%. If the return in month *t*-1 is high, the annualised return is 4.00%. This 5.74% difference in the returns is significant at the 10% level of significance with a t-value of 1.81. The market return is not significantly different when the prior month returns are low or high. The same can be said for the Management factor (MGMT).

The value factor (HML) and updated value factor (HML_M) see some of the largest and most significant differences between the high and low groups. As a result, the largest opportunities for increasing the Sharpe ratio when conditioning on past returns lie in factors such as HML and HML_M. For example, an investor may increase their weighting on the Value factor after a month where the factor provides a high return and decrease their weighting after a month of low return. However, the focus is on the Sharpe ratio of each factor. It is necessary to examine if the increase in return is matched with an increase in volatility, which may leave the Sharpe ratio unchanged.

The final three columns of Table 3.2 report on the relationship between past returns and future variances. The estimates show that when the market factor's return in month *t*-1 is low, its annualised volatility in month t is 18.82%. Following a high-return month, its volatility is 12.88%. As such, the variance ratio of $\left(\frac{0.1882}{0.1288}\right) = 0.68$ is statistically significantly different from one with a p-value of 0.00. This means the Market factor is significantly less, rather than more, volatile following high returns. This variance ratio is expressed as a percentage in the H/L column of Table 3.1. Unfortunately this fall in volatility is not matched with an increased return as seen through the insignificant H-L t-statistic, which indicates there may be no improvement in the Sharpe ratio for this factor when information contain in past returns is considered. The Size and Value factors exhibit statistically significant differences in returns conditioned on prior month information, however the variance ratios of 0.97 and 1.01 respectively are not statistically significant. This indicates that even though the returns provided by these factors in month *t* are greater if the return in *t*-1 is also high, this increase may be matched with an increase in volatility which may not result in an increased Sharpe ratio. The remaining factor also see no difference in volatility following periods of high and low returns.

The risk-return dynamics for each of the remaining five markets are examined in Section B.II in Appendix B. Highlighting certain examples, French returns (Table B.1) show the HML factor with the highest difference (11.97) in returns between high and low periods from the previous month. This difference is complemented by a significant reduction in volatility, as indicated by a high/low variance ratio of 0.75 and a corresponding p-value of 0.07, suggesting that periods of high returns are followed by lower volatility. In Germany (Table B.2), the BAB factor leads with the most substantial difference in returns (14.86), followed by periods of high and low returns. However, this difference in returns does not translate to a significant change in volatility, with a variance ratio of 1.10 and a p-value of 0.51. The Dutch market data (Table B.3) reveals the MGMT factor as having the largest difference in returns between high and low return periods (6.08). The volatility change associated with this difference is not significant, evidenced by a variance ratio of 1.13 and a p-value of 0.16.

Overall, in the remaining markets, at least four factors exhibit a significant difference in expected returns (H-L) based on the prior month's return, with a significance level of 10%. In four of the five remaining markets, the Market factor is significantly less, rather than more, volatile following high returns. The most extreme of these coming in the case of France (Table B.1) with a variance ratio of 0.47, however this reduction in volatility is not matched with a significant increase in return. The results from these diagnostic tests on risk-return dynamics suggest that the potential for improving the Sharpe ratio across markets primarily arises from return timing rather than volatility timing.

3.5.2 Improvements in Sharpe Ratio

Improved Sharpe ratios of individual factors can increase the maximum squared Sharpe ratios of models by enhancing the overall risk-adjusted performance. The z-statistic for the difference between the efficient and original factors' Sharpe ratios can be computed. Following Jobson and Korkie (1981), and with the correction from Memmel (2003), the test statistic for the expected difference in Sharpe ratios is

$$\mathrm{z-statistic} = \frac{\sigma_o \mu_e - \sigma_e \mu_o}{\sqrt{\theta}} \quad (3.9)$$

$$\text{where } \theta = \frac{1}{T}\left(2\sigma_e^2\sigma_o^2 - 2\,\sigma_e\,\sigma_o\,\sigma_{e,o} + \frac{1}{2}\ \mu_e^2\sigma_o^2 + \frac{1}{2}\ \mu_o^2\sigma_e^2\ -\ \frac{\mu_e\mu_o}{\sigma_o\,\sigma_e}\sigma_{e,o}^2\right). \quad (3.10)$$

σ_e and σ_u are the standard deviations of the efficient and original factors. The means of the efficient and original factors are represented by μ_o and μ_e respectively. The covariance between the two factors us represented by $\sigma_{e,o}$. θ essentially normalises the numerator of the z-statistic (the difference in risk-adjusted returns) by accounting for the volatility, covariance, and number of observations. This normalisation is crucial for making the z-statistic a valid measure of statistical significance. Letting R denote the return to the original factor and x the efficient factor's weight on the original factor. The return on the efficient factor is thus xR and the covariance that needs to be computed is:

$$cov(xR, R) = \frac{\mu_p^2(1-\rho^2)}{SR^2+\rho^2} \quad (3.11)$$

When an investment factor (like size, value, or momentum) shows a high degree of autocorrelation, it suggests that its past performance can be a reliable indicator for its future performance, at least to some extent. An efficient strategy would leverage this predictability to position the portfolio advantageously. The gains from such a strategy are contingent on the presence of autocorrelations. If the time-series predictability (i.e., autocorrelation) vanishes, the strategy no longer has a reliable pattern to exploit, and therefore, the gains from this strategy would diminish or evaporate. Factors that exhibit both high autocorrelation and lower Sharpe ratios are particularly ripe for improvement through this optimisation strategy.

Table 3.3 presents Sharpe ratio of both the original and efficient factors in U.K returns, and the z-value statistic of equation 3.10 representing the significance of the difference between the two Sharpe ratios. As noted previously, the expectation is that the largest increases in Sharpe ratios will come from factors with the most autocorrelation, as seen in Table 3.1, and from those with distinct risk-return dynamics, where returns and variances differ significantly after high and low periods. In the U.K. case, prior to examining the realised gains, it is anticipated that the Profitability (RMW) and Value factors (HML and HML_M) should deliver significant improvements in Sharpe ratios.

Table 3.3

For all factors except the momentum factor, the transformation results in an increased Sharpe ratio. The realised efficient factors confirm the expectations, as the factors which provide the largest Sharpe ratio improvements are the Value factor (HML) with an improvement of 0.23 and the Profitability factor (RMW_{OP}) with an improvement of 0.22, which have z-values of 2.89 and 3.61, respectively. Other factors which provide significant increases in Sharpe ratio include; BAB (0.09), PERF(0.07) and HML_M (0.23). The RMW_{ROE} factor also provides an increase in Sharpe ratio (0.12) at the 10% level of significance. While the Market and Size efficient factors provide an increase in Sharpe this is not statistically significant. Given the already large Sharpe ratio of the momentum factor, harvesting the information in past returns does not lead to an increase in the Sharpe ratio.

Table 3.4

Table 3.4 outlines the realised improvements in Sharpe Ratio from moving to standard factors to the time-series efficient factors. Across markets, improvements in the Sharpe ratio between the standard and efficient versions are observed for most investment factors; however, in most cases, this increase is not significant. Increases at the 5% level of significance are seen for the BAB factor (0.13) in the German sample (Table B.7), SMB (0.11) in the Dutch sample (Table B.8), and CMA (0.18) in the Italian (Table B.9).

It is noteworthy that, in the French study (Table B.6), the HML (0.12) and HML_M (0.12) factors exhibit increases in Sharpe ratios, with z-scores just marginally exceeding the 10% significance threshold, highlighting their potential relevance despite narrowly missing the conventional significance criteria, like other factors across the sample. The next step is to determine whether the inclusion of either the efficient or standard factors provides additional predictive power or information in the context of asset pricing models and investment strategies.

3.5.3 Time-Series Efficient versus Inefficient Factors

In Table 3.5, the importance of the information provided by both standard and efficient factors is assessed in an asset pricing model using the Fama and French (2015) five-factor model. Each of the five standard factors is regressed against all five efficient factors, and then the regression is reversed, with each efficient factor regressed against all five standard factors. The alphas presented in Table 3.5 can be understood in two ways. Initially, if an alpha value is not statistically significant, it suggests that incorporating the factor from the left-hand side of the regression into the asset pricing model, which includes the factors on the right-hand side, would not enhance the model's effectiveness (as per Barillas and Shanken, 2017). For instance, if the alpha of the conventional size factor is insignificant when regressed against the time-series efficient five-factor model, this implies that the addition of the standard size factor does not contribute to the improvement of the efficient model. From an investor's perspective, a non-significant alpha also implies that there would be no advantage in terms of increased Sharpe ratio for an investor who already employs the factors on the right-hand side if they were to also trade the factor on the left-hand side (according to Huberman and Kandel, 1987). This means, for example, that an investor who is already utilising the efficient factors would not see any added value from trading the standard size (SMB) factor.

Table 3.5

The above table suggests that U.K. time-series efficient factors contain all the information found in the standard U.K. factors. The first row of alpha values in Table 3.4 shows that time-series efficient factors constructed using month $t-1$ returns, apart from the size factor, are all incrementally informative about future returns at the 5% level when controlling for the five-factor model, as seen through significant alpha values. In other words, adding any of these efficient factors to the standard five-factor model would increase the Sharpe ratio available to an investor. For example when the efficient version of the value factor (HML) is regressed on the standard factors it returns an alpha value of 0.494 which is highly significant (5.83), indicating this factor would enhance the performance of the model with standard factors.

Where standard factors are regressed against efficient factors, a similar albeit less pronounced trend emerges. Specifically, only three out of the five standard factors contribute to improving the model's performance when these factors are in their efficient form. Notably, the standard value factor (HML) and size factor (SMB) do not provide significant additions, as indicated by the insignificant t-statistics of 1.16 and 0.117, respectively, for the alpha values.

Furthermore, a reduction in alpha values is observed for all factors, except for the market factor. This indicates that the model with diminished alpha values, which are statistically closer to zero, likely incorporates factors that more effectively capture the risks influencing asset returns. Such a model offers a more accurate and comprehensive depiction of those returns, suggesting a superior ability to explain the dynamics of asset pricing through the included risk factors. In other words, the standard factors contribute less to an efficient factor model, as indicated by the lower alpha values, compared to the efficient factors' contribution to the standard model.

This pattern is repeated across the remaining markets presented in Section B.IV of Appendix B. In three of the remaining five markets, namely France (Table B.11), Germany (Table B.12), and the Netherlands (Table B.13), larger and more significant alpha values are observed when the efficient factors are regressed on the standard factors, indicating that efficient factors contain more information than the standard factors. In other words, the addition of an efficient factor to a model containing standard factors would allow for an improvement in the Sharpe ratio. However, in the Italian (Table B.14) and Spanish (Table B.15) cases, this is not as clear.

The presence of positive and significant alpha values for certain standard factors across the sample suggests that, although the efficient factor transformation would enhance the performance of standard models, the extent of improvement may not be as substantial as the findings of Ehsani and Linnainmaa (2022) indicate for U.S. returns. This implies that while the transformation to efficient factors does offer benefits over traditional models in a European setting, the magnitude of these benefits can vary, potentially due to differences in market dynamics or the specific factors examined.

3.5.4 Momentum Factor

After converting factors into their efficient forms, Ehsani and Linnainmaa (2022) suggest that the distinct momentum factor, pivotal to my models as per Chapter Two, might lose its significance due to the autocorrelation that underpins this factor being addressed. This raises a critical question: Is there a need to incorporate a separate momentum factor in asset pricing models for capturing momentum in European stock returns? As the time-series efficient factors have accounted for the return predictability, the separate momentum factor, which was formed from individual stock returns to capture this predictability, becomes redundant. The efficient factors themselves should explain the momentum profits, thus making the separate momentum factor unnecessary.

To explore this, Table 3.6 presents the findings from U.K. time-series regressions that assess the momentum's influence within both the standard and efficient versions of the Fama and French (2015) five-factor model, alongside the CAPM. These analyses use the Carhart (1997) momentum factor as the dependent variable. If an insignificant alpha value is observed, it will imply that the distinct momentum factor does not contribute to model performance, challenging its essential role as highlighted in the model comparison tests in Chapter Two.

Table 3.6

Columns two and three of Table 3.6 show that Carhart's (1997) momentum factor earns CAPM and five-factor model alphas of 87 and 82 basis points per month, respectively, with t-values of 4.35 and 3.99. The final regression, detailed in column four, explores the explanation of momentum using the time-series efficient version of the five-factor model. Here, the alpha decreases to 65 basis points per month, accompanied by a t-value of 2.21. In their analysis of a U.S. sample, Ehsani and Linnainmaa (2022) deem this alpha value in the efficient factor regression insignificant, suggesting that momentum strategy profits are largely due to the time-series predictability captured by the original factors. Consequently, they infer that a separate momentum factor is unnecessary in an asset pricing model that incorporates these five factors in their efficient form. However, in this study with a U.K. sample, the findings suggest the continued need for a momentum factor, even when the other factors are presented in their efficient form. Nonetheless, the reduced alpha value does indicate less reliance on this factor.

Section B.V in Appendix B presents the results of these tests across the remaining markets with similar outcomes. The results of Ehsani and Linnainmaa (2022) could not be replicated; however, similar to the U.K., the alpha value of the distinct momentum factor decreases in some cases when regressed on the five-factor model with efficient factors. For example, in Table B.16, within the set of French factors, the alpha value falls from 0.585 to 0.061 when moving from a regression on the standard factors to the efficient factors. However, in the other five markets (Tables B.16-B.20), there is no significant change in the alpha value when moving from the model with standard factors to the model with efficient factors.

3.5.5 Model Comparison with Time-Series Efficient Factors

In this section, the impact of the efficient factor transformation on the Chapter Two model comparison tests is examined. By replacing the original factors with their efficient versions, the goal is to determine if the adjustments significantly improve the models' performance metrics and if the relative performance rankings of the models are changed. The results of the U.K. pairwise and multiple model comparison tests from Chapter Two are presented in Table 3.7 below.

Table 3.7

Panel C in Table 3.7 contains the squared Sharpe ratio of the efficient factor test in column one and the same for the original tests for comparison purposes in column five. Panel C shows a significant improvement in the squared Sharpe ratios when comparing Chapter Two's results to the new tests with efficient factors: FF3 increased from 0.003 to 0.082, Carhart from 0.085 to 0.101, and FF5 from 0.062 to 0.133. This is not surprising given that the Value (HML) factor and the Profitability (RMWOP) factor provided the highest increase in Sharpe ratios when transformed into efficient factors. As FF5 contains both of these factors, it now returns the highest squared Sharpe ratio in U.K. returns.

However, this improvement is not statistically significant compared to the Sharpe ratios of the FF6 (which decreased from 0.144 to 0.121), AFIM (which decreased from 0.129 to 0.128), and CZZ (which improved from 0.109 to 0.127) models, as shown in Panel B with p-values of the pairwise tests greater than 0.05. Apart from the rise of the FF5, the model rankings remain similar to the original tests. Improvements in Sharpe ratios are observed for most models except the highest-performing model in the original tests, the FF6, which saw a decrease, and the AFIM, which remained relatively stable with a slight decrease. Looking at the p-values in Panel B, less significant differences between all models can be observed.

The results for the remainder of the sample are presented in Section B.VI of Appendix B. In the French market (Table B.21), after incorporating efficient factors, significant improvements in the squared Sharpe ratios of several models were observed. The FF3 model's Sharpe ratio increased from 0.005 to 0.032, and the Carhart model improved from 0.034 to 0.04, demonstrating moderate gains. The most notable enhancement was seen in the FrazPed model, which jumped from 0.089 to 0.113, marking the largest increase among the models. The FF5 model also showed a significant improvement from 0.028 to 0.054, leading to a reshuffling of model rankings with the integration of efficient factors.

In the German market (Table B.22), the introduction of efficient factors yielded mixed results. The FF6 model, initially leading with a 0.114 squared Sharpe ratio, saw a reduction to 0.072, while the AFIM model decreased from 0.13 to 0.077. Conversely, the FrazPed model showed a positive movement from 0.039 to 0.054, presenting the largest proportional increase. This shift suggests a revaluation of model performance, with previously leading models like FF6 and AFIM experiencing declines, while others like FrazPed improved, altering the ranking of models based on Sharpe ratios. This is not surprising, as the FrazPed model contains the BAB factor, which, as outlined in Table B.8, shows the largest and most significant increase in Sharpe ratio when optimised.

For the Dutch market (Table B.23), after applying efficient factors, the FF3 model's squared Sharpe ratio increased significantly from 0.011 to 0.028, which is the largest improvement across models. Meanwhile, the Carhart model maintained its performance level, and the FF5 model saw an uplift from 0.015 to 0.036. Notably, the FF6 model, which had a higher initial Sharpe ratio, saw a modest increase to 0.051. These adjustments indicate a notable shift in model performances and rankings, emphasising the impact of efficient factor utilisation on enhancing model outcomes.

In the Italian market (Table B.24), the application of efficient factors resulted in the FF5 model improving from a squared Sharpe ratio of 0.067 to 0.087, and the CZZ model increasing from

0.062 to 0.068. Despite these changes, the overall landscape of model efficacy experienced minor adjustments, with the FF5 model demonstrating one of the more significant improvements. The Spanish market analysis (Table B.25) reveals an improvement in the FF3 model from 0.005 to 0.01, while the Carhart model saw a slight increase to 0.036. Despite these changes, the ranking of models based on Sharpe ratios saw a reshuffling, with some models improving and others, like the FF6 and CZZ, experiencing slight declines in their squared Sharpe ratios.

Across the European samples, the application of efficient factors tends to improve the performance metrics of several models, notably the FF3 and FF5 models, which consistently show improvements in their squared Sharpe ratios. However, top-performing models like the FF6 and AFIM do not experience significant improvement. This pattern underscores the nuanced impact of efficient factors on asset pricing models, with certain models standing out across multiple markets for their enhanced performance and adaptability to efficient factor integration. However, the inclusion of efficient factors often leads to fewer models being statistically indistinguishable in terms of performance, suggesting that the transformation allows for a convergence in model performance.

3.6 Conclusions

In asset pricing, assessing the mean-variance efficiency of factors is crucial for enhancing model performance. This process involves either adding new factors to expand the efficient frontier or improving the efficiency of existing factors. This chapter focuses on examining the mean-variance efficiency of prominent investment factors in European stock returns. The aim is to enhance the efficacy of existing factors through the use of 'time-series efficient factors'. A time-series efficient factor exploits the autocorrelation in factor returns, timing the original factor to minimise variance while maintaining the expected return. If successful, this method should enhance the Sharpe ratio of factor models without needing additional momentum factors, as it already captures the predictable variations that momentum strategies target.

The examination reveals that a significant number of factors across the European samples display serial correlation, exhibiting non-random patterns in both return and variance following periods characterised by either high or low returns. This behaviour points to potential inefficiencies in the factor set, hinting at exploitable opportunities within these factors.

Leveraging the methodology proposed by Ehsani and Linnainmaa (2022), time-series efficient factors are developed by conditioning a factor's weight on its past returns. The results show that converting traditional factors into efficient versions leads to a significant enhancement in Sharpe ratios across a wide range of factors within the European dataset. This improvement is particularly pronounced for the size (SMB) and value (HML) factors, which initially showed lower returns across markets and have generally delivered low and often insignificant returns over the past three decades.

Time-series efficient factors in each sample typically capture all the information found in the standard factors. However, across the sample, the alpha provided by the distinct momentum factor cannot be fully eliminated from a model that includes efficient factors. Time-series efficient factors allow for less dependence on a distinct momentum factor, subsequently reducing factor risk for investors. Additionally, the investability of the efficient factors and the straightforward nature of the transformation open up an expanded set of investment opportunities for both retail and institutional market participants.

Given the significant increase in the Sharpe ratio, the impact of efficient factors on the model comparison tests in Chapter Two is examined. Across European markets, incorporating efficient factors generally boosts the performance of several asset pricing models, especially the Fama and French (1993) three-factor model and Fama and French (2015) five-factor model, which consistently exhibit improved Sharpe ratios. While the absolute performance of previously underperforming models improves, the relative performance of models across samples does not experience significant change. Since not all factors are significantly improved using the efficient factor transformation, only certain factors enhance model performance. The next step is to isolate the time-series efficient factors that improve model performance across the samples.

This findings on time-series efficient factors present valuable insights for both academic research and practical investing by reshaping how we interpret traditional factors in asset pricing models. For academics, these results underscore the importance of refining factors to capture inherent patterns, such as autocorrelation, that exist within return data but are often overlooked in standard factor models. This shift from merely adding new factors toward optimizing existing ones challenges the conventional approach in asset pricing research, highlighting a more sustainable path for model development that reduces reliance on an ever-expanding factor zoo. For investors, the introduction of time-series efficient factors enhances the investability and utility of asset pricing models by potentially lowering factor risk without sacrificing return. This factor efficiency improvement provides a robust tool for portfolio construction, offering models that adapt more responsively to market conditions by using information embedded in past returns. Such a refined approach could lead to more resilient portfolios, particularly for investors in European markets, where traditional factors like size and value have shown inconsistent performance over time.

Appendix B

B.I U.K Empirical Results

Table 3.1. Predictive Return Beta Values

	U.K	France	Germany	Netherlands	Italy	Spain
MKT-RF	0.085	0.063	0.034	0.065	-0.021	0.024
	(1.674)**	(1.149)	(0.664)	(1.276)	(-0.402)	(0.455)
SMB	0.086	-0.037	-0.079	-0.088	0.005	0.052
	(1.69)**	(-0.689)	(-1.555)	(-1.728)**	(0.097)	(1.009)
HML	0.337	0.187	0.143	0.095	0.056	0.022
	(6.995)*	(3.483)*	(2.826)*	(1.875)**	(1.095)	(0.417)
MOM	0.142	0.069	0.091	0.121	0.041	0.059
	(2.801)*	(1.258)	(1.782)	(2.389)*	(0.813)	(1.139)
RMW_{OP}	0.24	-0.01	0.075	-0.027	0.011	0.047
	(4.827)*	(-0.187)	(1.475)	(-0.525)	(0.224)	(0.91)
CMA	0.147	0.161	0.206	0.029	-0.065	0.062
	(2.886)*	(2.979)*	(4.099)*	(0.574)	(-1.284)	(1.19)
BAB	0.198	0.017	0.22	0.095	0.138	0.008
	(3.958)*	(0.314)	(4.411)*	(1.865)**	(2.713)*	(0.158)
HML_M	0.304	0.19	0.111	0.095	0.045	0.071
	(6.225)*	(3.54)*	(2.195)*	(1.868)**	(0.878)	(1.371)
MGMT	0.225	0.119	0.037	0.07	0.007	-0.016
	(4.504)*	(2.196)*	(0.724)	(1.369)	(0.139)	(-0.302)
PERF	0.154	-0.06	0.021	0.011	0	0.084
	(3.026)*	(-1.106)	(0.414)	(0.209)	(-0.008)	(1.617)
RMW_{CB}	0.056	-0.048	0.087	0.048	0.115	0.023
	(1.1)	(-0.883)	(1.721)**	(0.922)	(2.267)*	(0.435)
RMW_{ROE}	0.165	0.008	0.114	-0.012	-0.009	-0.005
	(3.252)*	(0.146)	(2.242)*	(-0.244)	(-0.179)	(-0.1)

*This table reports the beta values and associated t-statistics in parentheses for the predictive regressions of equation 3.7 where the dependent variable is the factor return for time period t, and the independent variable is the factor return for time period t-1. * Significant at 5%, ** Significant at 10%*

Table 3.2. U.K. Returns and Volatility Conditional on Month t-1 Returns

	Prior Month Return			Prior Month Return		
	Low	High	H-L	Low	High	H/L
MKT-RF	2.93	7.70	4.77	18.82	12.88	68%
	(0.62)	(2.38)	(0.85)			[0]*
SMB	-1.74	4.00	5.74	11.23	10.87	97%
	(-0.62)	(1.46)	(1.81)**			[0.62]
HML	-8.14	11.86	20.00	10.34	10.43	101%
	(-3.13)	(4.52)	(5.71)*			[0.79]
MOM	-1.97	7.08	9.05	16.39	13.37	82%
	(-0.98)	(3.98)	(3.23)*			[0.16]
RMW_{OP}	1.49	7.30	5.80	7.97	7.09	89%
	(0.83)	(3.84)	(2.15)*			[0.16]
CMA	1.49	7.30	5.80	7.17	7.56	105%
	(0.83)	(3.84)	(2.15)*			[0.98]
BAB	0.12	12.44	12.32	15.65	16.43	105%
	(0.03)	(3.01)	(2.21)*			[0.55]
HML_M	0.12	11.06	10.94	11.69	12.65	108%
	(0.03)	(3.48)	(4.72)*			[0.94]
MGMT	-2.87	9.05	11.91	8.55	8.27	97%
	(-1.33)	(4.35)	(4.17)			[0.99]
PERF	1.49	10.26	8.76	10.92	9.36	86%
	(0.54)	(4.36)	(2.29)*			[0.87]
RMW_{CB}	-0.09	8.14	8.23	7.41	6.89	93%
	(-0.05)	(4.7)	(3.04)*			[0.7]
RMW_{ROE}	-0.09	8.14	8.23	7.69	7.72	1.00%
	(-0.05)	(4.7)	(4.08)*			[0.6]

*This table assigns U.K. factors into two groups based on their month t − 1 returns and reports the average annualised returns and volatilities for each group. The full sample is used to create two groups of equal size. The high-minus-low difference in average returns and the high/low ratio of variances are reported. T-statistics of the null hypothesis that the average excess factor returns are equal to zero are given in parentheses. For the variance ratio, the p-value associated with the F-test is presented in square brackets. * Significant at 5%, ** Significant at 10%*

Table 3.3. U.K. Realised Efficient Factor Sharpe Ratio Improvements

	Market	SMB	HML	MOM	RMW_{OP}	CMA
Sharpe Ratio $_{org}$	0.1	0.03	0.05	0.22	0.1	0.18
Sharpe Ratio $_{ef}$	0.12	0.06	0.28	0.19	0.32	0.2
Δ Sharpe Ratio	0.03	0.03	0.23	-0.02	0.22	0.03
z-value	0.82	0.36	2.89*	-2.05	3.61*	0.52
	BAB	HML_{M}	MGMT	PERF	RMW_{CB}	RMW_{ROE}
Sharpe Ratio $_{org}$	0.12	0.02	0.1	0.17	0.17	0.1
Sharpe Ratio $_{ef}$	0.21	0.24	0.22	0.24	0.17	0.22
Δ Sharpe Ratio	0.09	0.23	0.12	0.07	0.01	0.12
z-value	1.99*	2.85*	1.71**	2.07*	0.15	1.74**

*Table 3.2 compares the Sharpe ratios of the original U.K. factors to their time-series efficient counterparts. The table includes the original Sharpe ratio ("Sharpe Ratio org") and the Sharpe ratio of the efficient factors ("Sharpe Ratio ef"). The row "Δ Sharpe Ratio" shows the increase in Sharpe ratio from switching to the efficient version of the factor. Additionally, the z-value for the improvement, calculated from equations (3.9), is reported. * Significant at 5%, ** Significant at 10%*

Table 3.4. Realised Efficient Factor Sharpe Ratio Improvements

	Market	SMB	HML	MOM	RMW_{OP}	CMA
UK	0.03	0.03	0.23*	-0.02	0.22*	0.03
France	0.03	0.04	0.12	0.00	0.00	0.06
Germany	0.01	0.06	-0.02	0.09	0.00	0.00
Netherlands	0.02	0.11**	0.04	0.02	0.00	0.01
Italy	0.00	0.02	0.02	-0.01	0.00	0.18
Spain	0.01	0.03	0.00	0.00	-0.01	0.04

	BAB	HML_M	MGMT	PERF	RMW_{CB}	RMW_{ROE}
UK	0.09*	0.23*	0.12**	0.07*	0.01	0.12**
France	0.00	0.12	0.04	0.00	0.00	0.00
Germany	0.12*	0.03	0.00	-0.01	0.01	-0.08
Netherlands	0.00	0.09	0.08	0.00	0.03	0.00
Italy	0.01	0.02	0.01	0.00	0.00	0.00
Spain	0.00	0.02	0.00	0.02	0.02	0.00

*This table reports the realised improvements in moving from the original factor to the time series efficient factor for each factor in each country. * Significant at 5%, ** Significant at 10%*

Table 3.5. U.K. Time-Series Efficient Five-Factor Model: Alphas from Spanning Tests

Efficient Factors Conditional on:	Dependent Variable MKT	SMB	HML	RMW	CMA	GRS test F-Value
	Efficient factors regressed on standard factors					
Prior-month return	0.415	0.338	0.494	0.428	0.341	5.21
	(2.681)*	(1.448)	(5.83)*	(3.267)*	(2.929)*	0
	Standard factors regressed on efficient factors					
Prior-month return	0.554	0.199	0.019	0.242	0.271	3.12
	(2.257)*	(1.16)	(0.117)	(2.079)*	(2.453)*	0

*This table reports the alphas and t-values (in parentheses) from regressions in which the dependent variable is one of the factors of the efficient or standard five-factor model and the explanatory variables are all five factors of the other model. The Gibbons et al. (1989) test statistic is under the null hypothesis that the alphas of the five factors are jointly zero. P-values for these GRS tests are reported in square brackets. * Significant at 5%, ** Significant at 10%*

Table 3.6. U.K. Momentum versus Fama and French (2015) Five-Factor Model

	Dependant Variable		
	Momentum Factor		
Independent Variable	CAPM	FF5	$FF5_{ef}$
Alpha	0.87	0.82	0.65
	(4.35)*	(3.99)*	(2.21)*
Market	-0.31	-0.15	-0.45
	(-6.53)	(-3.21)	(-5.4)
SMB		0.05	0.02
		(1.04)	(0.29)
HML		-0.49	-0.13
		(-5.98)	(-0.9)
RMW_{op}		0.34	-0.11
		(3.15)	(-1.23)
CMA		0.33	0.24
		(3.1)	(2.37)

*This table presents estimates from time-series regressions that measure the association between the momentum factor and the CAPM, as well as both the standard and efficient versions of the Fama and French (2015) five-factor model in U.K. stock returns. The alpha values, beta coefficients, and associated t-statistics (reported in parentheses) are presented in the columns across three models. In each case the dependant variable is the momentum factor. * Significant at 5%, ** Significant at 10%.*

Table 3.7. U.K. Model Comparison with Efficient Factors

Panel A: Difference in Squared Sharpe Performance

Model	FrazPed	FF3	SY	Carhart	FF6	CZZ	AFIM	FF5
HXZ	-0.011	-0.035	-0.046	-0.054	-0.073	-0.079	-0.081	-0.085
FrazPed		-0.024	-0.035	-0.043	-0.062	-0.069	-0.07	-0.074
FF3			-0.011	-0.019	-0.039	-0.045	-0.046	-0.05
SY				-0.008	-0.027	-0.033	-0.034	-0.039
Carhart					-0.02	-0.026	-0.027	-0.031
FF6						-0.006	-0.007	-0.012
CZZ							-0.001	-0.006
AFIM								-0.004

Panel B: p-Values

Model	FrazPed	FF3	SY	Carhart	FF6	CZZ	AFIM	FF5
HXZ	0.729	0.179	0.073	0.057	0.005	0.004	0.003	0.001
FrazPed		0.382	0.298	0.16	0.093	0.095	0.087	0.053
FF3			0.675	0.003	0.001	0.193	0.152	0
SY				0.798	0.314	0.264	0.236	0.145
Carhart					0.027	0.393	0.336	0.272
FF6						0.818	0.761	0.621
CZZ							0.193	0.776
AFIM								0.788

Panel C: Multiple Model Comparison

Model	$\hat{\theta}^2$	r	LR	p-value	n	$\hat{\theta}^2_{ORG}$
FF3	0.082	4	2.049	0.249	1	0.003
Carhart	0.101	5	1.218	0.389	0	0.085
FF5	0.133	5	0	0.879	1	0.062
FF6	0.121	5	0.244	0.667	1	0.144
FrazPed	0.058	5	3.788	0.073	0	0.017
SY	0.094	5	2.144	0.244	1	0.068
AFIM	0.128	5	0.072	0.752	1	0.129
HXZ	0.048	5	12.895	0.002	0	0.081
CZZ	0.127	5	0.081	0.726	1	0.109

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for a set of U.K. time series efficient factor models from 1991 to 2022. Panel A reports the differences in squared Sharpe ratio. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. Panel B reports the associated p-values. Panel C reports the multiple model comparison tests of efficient factor models in U.K. stock returns. The $\hat{\theta}^2$ column is the bias-adjusted maximum squared Sharpe performance of these efficient models. LR is the Likelihood ratio test of Wolak (1987,1989). The p-value column is the p value for the hypothesis that this model performs as well as any other model in the dataset. $\hat{\theta}^2_{ORG}$ is the bias adjusted squared Sharpe ratio of the models with original factors.

B.II Risk Return Characteristics for Remaining Markets

Table B.1. France Returns and Volatility Conditional on Month t-1 Returns

	Prior Month Return			Prior Month Return		
	Low	High	H-L	Low	High	H/L
MKT-RF	5.09	10.78	5.69	22.06	15.07	47%
	(0.87)	(2.69)	(0.79)			[0]*
SMB	1.17	-1.76	-2.93	9.94	9.90	99%
	(0.42)	(-0.67)	(-0.73)			[0.48]
HML	-3.99	7.98	11.97	13.93	12.09	75%
	(-1.04)	(2.48)	(2.36)*			[0.07]
MOM	3.83	2.77	-1.05	8.33	8.62	107%
	(1.66)	(1.21)	(-0.33)			[0.04]*
RMW_{OP}	-3.32	7.05	10.37	8.65	9.18	113%
	(-1.39)	(2.89)	(3.13)*			[0.74]
CMA	17.58	18.90	1.32	17.34	16.73	93%
	(3.67)	(4.25)	(0.21)			[0.42]
BAB	17.58	18.90	1.32	17.34	16.73	93%
	(3.67)	(4.25)	(0.21)			[0.8]
HML_{M}	17.58	6.98	-10.60	15.87	13.41	71%
	(3.67)	(1.96)	(2.02)*			[0.04]*
MGMT	-0.12	6.28	6.40	9.94	9.39	89%
	(-0.04)	(2.52)	(2.12)*			[0.39]
PERF	10.39	5.51	-4.88	13.14	11.73	80%
	(2.86)	(1.77)	(-1.04)			[0.11]
RMW_{CB}	5.93	5.33	-0.60	8.51	8.60	102%
	(2.52)	(2.33)	(-0.18)			[0.86]
RMW_{ROE}	5.93	5.33	-0.60	8.99	10.33	132%
	(2.52)	(2.33)	(0.54)			[0.09]**

*This table assigns French factors into two groups based on their month t – 1 returns and report the average annualized returns and volatilities for each group. The full sample is used to create two groups of equal size. The high-minus-low difference in average returns and the high/low ratio of variances are reported. T-statistics of the null hypothesis that the average excess factor returns are equal to zero are given in parentheses. For the variance ratio, the p-value associated with the F-test is presented in square brackets. * Significant at 5%, ** Significant at 10%.*

Table B.2. Germany Returns and Volatility Conditional on Month t-1 Returns

	Prior Month Return			Prior Month Return		
	Low	High	H-L	Low	High	H/L
MKT-RF	1.72	7.22	5.50	22.31	16.40	54%
	(0.31)	(1.75)	(0.92)			[0]*
SMB	1.91	-2.43	-4.34	10.32	10.07	95%
	(0.74)	(-0.96)	(-1.18)			[0.73]
HML	1.50	11.08	9.58	11.22	12.88	132%
	(0.53)	(3.42)	(2.4)*			[0.06]**
MOM	4.31	17.73	13.41	15.21	16.84	123%
	(1.13)	(4.19)	(2.42)*			[0.16]
RMW_{OP}	1.47	8.35	6.88	7.50	8.03	114%
	(0.78)	(4.14)	(2.51)*			[0.35]
CMA	1.94	8.68	6.74	9.10	11.16	150%
	(0.85)	(3.1)	(1.88)**			[0.01]*
BAB	-1.46	13.40	14.86	15.45	16.21	110%
	(-0.38)	(3.29)	(2.55)*			[0.51]
HML_{M}	-0.88	5.31	6.18	12.17	13.70	127%
	(-0.29)	(1.54)	(1.34)			[0.1]
MGMT	0.20	4.94	4.74	10.52	9.56	83%
	(0.08)	(2.06)	(1.41)			[0.19]
PERF	6.46	10.20	3.75	10.53	11.62	122%
	(2.44)	(3.49)	(1.02)			[0.18]
RMW_{CB}	3.25	4.22	0.97	7.77	7.32	89%
	(1.66)	(2.29)	(0.4)			[0.41]
RMW_{ROE}	3.25	4.22	0.97	9.08	8.26	83%
	(1.66)	(2.29)	(0.2)			[0.19]

*This table assigns German factors into two groups based on their month t − 1 returns and report the average annualized returns and volatilities for each group. The full sample is used to create two groups of equal size. The high-minus-low difference in average returns and the high/low ratio of variances are reported. T-statistics of the null hypothesis that the average excess factor returns are equal to zero are given in parentheses. For the variance ratio, the p-value associated with the F-test is presented in square brackets. * Significant at 5%, ** Significant at 10%.*

Table B.3. Netherlands Returns and Volatility Conditional on Month t-1 Returns

	Prior Month Return			Prior Month Return		
	Low	High	H-L	Low	High	H/L
MKT-RF	5.60	10.51	4.91	23.27	17.10	58%
	(0.96)	(2.45)	(0.65)			[0]
SMB	1.77	-3.84	-5.61	12.06	10.84	81%
	(0.58)	(-1.41)	(-1.34)			[0.12]
HML	3.88	4.51	0.62	14.06	13.62	94%
	(1.1)	(1.32)	(0.14)			[0.75]
MOM	1.74	1.93	0.20	18.58	19.32	108%
	(0.46)	(0.64)	(0.04)			[0.53]
RMW_{OP}	1.25	2.73	1.49	15.03	12.08	65%
	(0.44)	(0.94)	(0.37)			[0.01]*
CMA	8.73	11.37	2.64	11.20	11.58	107%
	(2.14)	(2.38)	(0.41)			[0.9]
BAB	8.73	11.37	2.64	16.22	19.02	137%
	(2.14)	(2.38)	(0.41)			[0.04]*
HML_M	8.73	4.92	-3.80	16.08	14.80	85%
	(2.14)	(1.32)	(1.56)			[0.22]
MGMT	-1.45	4.63	6.08	12.62	13.40	113%
	(-0.46)	(1.38)	(1.29)			[0.16]
PERF	5.58	9.99	4.41	14.79	15.77	114%
	(1.5)	(2.52)	(0.85)			[0.22]
RMW_{CB}	1.68	4.59	2.91	10.53	10.77	105%
	(0.64)	(1.7)	(0.8)			[0.4]
RMW_{ROE}	1.68	4.59	2.91	12.58	12.20	94%
	(0.64)	(1.7)	(-0.59)			[0.85]

*This table assigns Dutch factors into two groups based on their month t − 1 returns and report the average annualized returns and volatilities for each group. The full sample is used to create two groups of equal size. The high-minus-low difference in average returns and the high/low ratio of variances are reported. T-statistics of the null hypothesis that the average excess factor returns are equal to zero are given in parentheses. For the variance ratio, the p-value associated with the F-test is presented in square brackets. * Significant at 5%, ** Significant at 10%.*

Table B.4. Italy Returns and Volatility Conditional on Month t-1 Returns

	Prior Month Return			Prior Month Return		
	Low	High	H-L	Low	High	H/L
MKT-RF	6.86	0.04	-6.83	25.35	21.29	71%
	(1.08)	(0.01)	(-0.82)			[0.02]*
SMB	-0.25	-0.77	-0.51	11.15	10.88	95%
	(-0.09)	(-0.28)	(-0.11)			[0.72]
HML	1.10	2.37	1.27	12.33	12.85	109%
	(0.35)	(0.73)	(0.34)			[0.59]
MOM	7.72	8.17	0.45	15.78	16.96	116%
	(2.56)	(3.11)	(0.11)			[0.32]
RMW_{OP}	-1.06	-4.45	-3.39	12.01	10.44	76%
	(-0.43)	(-1.85)	(-0.96)			[0.05]**
CMA	-1.06	-4.45	-3.39	9.92	9.57	93%
	(-0.43)	(-1.85)	(-0.96)			[0.63]
BAB	1.53	12.39	10.86	12.99	14.87	131%
	(0.47)	(3.32)	(2.26)*			[0.07]**
HML_{M}	0.61	0.63	0.02	16.63	13.74	68%
	(0.14)	(0.18)	(0.04)			[0.01]*
MGMT	1.40	-1.40	-2.80	13.24	12.45	88%
	(0.42)	(-0.45)	(-0.61)			[0.38]
PERF	12.70	8.22	-4.48	14.80	16.19	120%
	(3.42)	(2.02)	(-0.82)			[0.22]
RMW_{CB}	3.62	7.28	3.66	12.12	13.13	117%
	(1.19)	(2.21)	(0.8)			[0.27]
RMW_{ROE}	3.62	7.28	3.66	11.44	11.97	109%
	(1.19)	(2.21)	(0.27)			[0.53]

*This table assigns Italian factors into two groups based on their month t – 1 returns and report the average annualized returns and volatilities for each group. The full sample is used to create two groups of equal size. The high-minus-low difference in average returns and the high/low ratio of variances are reported. T-statistics of the null hypothesis that the average excess factor returns are equal to zero are given in parentheses. For the variance ratio, the p-value associated with the F-test is presented in square brackets. * Significant at 5%, ** Significant at 10%.*

Table B.5. Spain Returns and Volatility Conditional on Month t-1 Returns

	Prior Month Return			Prior Month Return		
	Low	High	H-L	Low	High	H/L
MKT-RF	5.70	5.67	-0.03	24.27	19.29	63%
	(0.92)	(1.16)	(0)			[0]*
SMB	0.48	2.58	2.09	12.42	10.57	72%
	(0.15)	(0.96)	(0.5)			[0.02]*
HML	3.35	3.98	0.64	12.22	11.41	87%
	(1.08)	(1.37)	(0.15)			[0.67]
MOM	2.99	7.29	4.30	15.55	19.97	165%
	(1.22)	(2.38)	(1.12)			[0]*
RMW_{OP}	-0.80	1.53	2.33	9.64	12.04	156%
	(-0.31)	(0.53)	(0.61)			[0]*
CMA	-0.80	1.53	2.33	10.14	11.41	127%
	(-0.31)	(0.53)	(0.61)			[0.06]
BAB	11.44	3.70	-7.74	16.69	17.90	115%
	(2.7)	(0.81)	(-1.23)			[0.22]
HML_{M}	11.44	6.53	-4.91	13.43	14.01	109%
	(2.7)	(1.84)	(1.33)			[0.4]
MGMT	3.24	-0.40	-3.64	12.18	11.28	86%
	(1.05)	(-0.14)	(-0.85)			[0.24]
PERF	2.70	4.84	2.15	11.51	14.43	157%
	(0.92)	(1.32)	(0.43)			[0.01]*
RMW_{CB}	0.31	-0.12	-0.42	12.39	13.17	113%
	(0.1)	(-0.03)	(-0.1)			[0.57]
RMW_{ROE}	0.31	-0.12	-0.42	9.97	11.29	128%
	(0.1)	(-0.03)	(1.03)			[0.07]**

*This table assigns Spanish factors into two groups based on their month t − 1 returns and report the average annualized returns and volatilities for each group. The full sample is used to create two groups of equal size. The high-minus-low difference in average returns and the high/low ratio of variances are reported. T-statistics of the null hypothesis that the average excess factor returns are equal to zero are given in parentheses. For the variance ratio, the p-value associated with the F-test is presented in square brackets. * Significant at 5%, ** Significant at 10%.*

B.III Realised Improvements in Sharpe Ratio for Remaining Markets

Table B.6. France Realised Efficient Factor Sharpe Ratio Improvements

	MKT-RF	SMB	HML	MOM	RMW_{OP}	CMA
Sharpe Ratio $_{org}$	0.12	0.00	0.04	0.12	0.11	0.06
Sharpe Ratio $_{ef}$	0.15	0.04	0.16	0.12	0.12	0.13
Δ Sharpe Ratio	0.03	0.04	0.12	0.00	0.00	0.06
z-value	0.00	0.48	1.57	0.00	0.08	0.86

	BAB	HML_M	MGMT	PERF	RMW_{CB}	RMW_{ROE}
Sharpe Ratio $_{org}$	0.31	0.02	0.09	0.18	0.19	0.10
Sharpe Ratio $_{ef}$	0.31	0.14	0.14	0.19	0.20	0.10
Δ Sharpe Ratio	0.00	0.12	0.04	0.00	0.00	0.00
z-value	0.00	1.61	0.68	0.00	0.00	0.10

*This table compares the Sharpe ratios of the original French factors to their time-series efficient counterparts. The table includes the original Sharpe ratio ("Sharpe Ratio org") and the Sharpe ratio of the efficient factors ("Sharpe Ratio ef"). The row "Δ Sharpe Ratio" shows the increase in Sharpe ratio from switching to the efficient version of the factor. Additionally, the z-value for the improvement, calculated from equations (3.9), is reported. * Significant at 5%, ** Significant at 10%*

Table B.7. Germany Realised Efficient Factor Sharpe Ratio Improvements

	MKT-RF	SMB	HML	MOM	RMW_{OP}	CMA
Sharpe Ratio $_{org}$	0.07	0.00	0.15	0.11	0.18	0.15
Sharpe Ratio $_{ef}$	0.08	0.06	0.13	0.20	0.18	0.15
Δ Sharpe Ratio	0.01	0.06	-0.02	0.09	0.00	0.00
z-value	0.38	0.76	-0.32	0.00	-0.09	0.04
	BAB	HML_M	MGMT	PERF	RMW_{CB}	RMW_{ROE}
Sharpe Ratio $_{org}$	0.11	0.05	0.08	0.22	0.15	0.18
Sharpe Ratio $_{ef}$	0.23	0.08	0.07	0.21	0.16	0.10
Δ Sharpe Ratio	0.12	0.03	0.00	-0.01	0.01	-0.08
z-value	2.10*	0.45	-0.07	0.00	0.18	-1.80

*This table compares the Sharpe ratios of the original German factors to their time-series efficient counterparts. The table includes the original Sharpe ratio ("Sharpe Ratio org") and the Sharpe ratio of the efficient factors ("Sharpe Ratio ef"). The row "Δ Sharpe Ratio" shows the increase in Sharpe ratio from switching to the efficient version of the factor. Additionally, the z-value for the improvement, calculated from equations (3.9), is reported. * Significant at 5%, ** Significant at 10%*

Table B.8. Netherlands Realised Efficient Factor Sharpe Ratio Improvements

	MKT-RF	SMB	HML	MOM	RMW_{OP}	CMA
Sharpe Ratio $_{org}$	0.11	-0.03	0.08	0.15	0.04	0.05
Sharpe Ratio $_{ef}$	0.13	0.09	0.12	0.17	0.04	0.06
Δ Sharpe Ratio	0.02	0.11	0.04	0.02	0.00	0.01
z-value	0.00	1.73**	0.70	0.00	0.03	0.14
	BAB	HML_{M}	MGMT	PERF	RMW_{CB}	RMW_{ROE}
Sharpe Ratio $_{org}$	0.18	-0.02	-0.03	0.14	0.07	0.09
Sharpe Ratio $_{ef}$	0.18	0.08	0.06	0.14	0.10	0.09
Δ Sharpe Ratio	0.00	0.09	0.08	0.00	0.03	0.00
z-value	0.00	1.21	1.10	0.00	0.48	0.11

*This table compares the Sharpe ratios of the original Dutch factors to their time-series efficient counterparts. The table includes the original Sharpe ratio ("Sharpe Ratio org") and the Sharpe ratio of the efficient factors ("Sharpe Ratio ef"). The row "Δ Sharpe Ratio" shows the increase in Sharpe ratio from switching to the efficient version of the factor. Additionally, the z-value for the improvement, calculated from equations (3.9), is reported. * Significant at 5%, ** Significant at 10%*

.

Table B.9. Italy Realised Efficient Factor Sharpe Ratio Improvements

	MKT-RF	SMB	HML	MOM	RMW_{OP}	CMA
Sharpe Ratio $_{org}$	0.04	-0.01	0.04	0.18	0.21	-0.08
Sharpe Ratio $_{ef}$	0.04	0.01	0.06	0.17	0.21	0.10
Δ Sharpe Ratio	0.00	0.02	0.02	-0.01	0.00	0.18
z-value	-0.01	0.33	0.23	0.00	0.00	3.17[1]
	BAB	HML_{M}	MGMT	PERF	RMW_{CB}	RMW_{ROE}
Sharpe Ratio $_{org}$	0.14	0.01	0.00	0.20	0.12	0.19
Sharpe Ratio $_{ef}$	0.15	0.04	0.01	0.20	0.12	0.19
Δ Sharpe Ratio	0.01	0.02	0.01	0.00	0.00	0.00
z-value	0.35	0.33	0.10	0.00	0.03	0.00

*This table compares the Sharpe ratios of the original Italian factors to their time-series efficient counterparts. The table includes the original Sharpe ratio ("Sharpe Ratio org") and the Sharpe ratio of the efficient factors ("Sharpe Ratio ef"). The row "Δ Sharpe Ratio" shows the increase in Sharpe ratio from switching to the efficient version of the factor. Additionally, the z-value for the improvement, calculated from equations (3.9), is reported. * Significant at 5%, ** Significant at 10%*

Table B.10. Spain Realised Efficient Factor Sharpe Ratio Improvements

	MKT-RF	SMB	HML	MOM	RMW_{OP}	CMA
Sharpe Ratio $_{org}$	0.07	0.03	0.09	0.13	0.14	0.01
Sharpe Ratio $_{ef}$	0.08	0.06	0.09	0.13	0.13	0.05
Δ Sharpe Ratio	0.01	0.03	0.00	0.00	-0.01	0.04
z-value	0.00	0.41	0.07	0.00	0.00	0.54
	BAB	HML_{M}	MGMT	PERF	RMW_{CB}	RMW_{ROE}
Sharpe Ratio $_{org}$	0.13	0.07	0.04	0.08	0.00	0.12
Sharpe Ratio $_{ef}$	0.12	0.09	0.04	0.10	0.02	0.13
Δ Sharpe Ratio	0.00	0.02	0.00	0.02	0.02	0.00
z-value	0.00	0.42	-0.02	0.28	0.25	0.00

*This table compares the Sharpe ratios of the original Spanish factors to their time-series efficient counterparts. The table includes the original Sharpe ratio ("Sharpe Ratio org") and the Sharpe ratio of the efficient factors ("Sharpe Ratio ef"). The row "Δ Sharpe Ratio" shows the increase in Sharpe ratio from switching to the efficient version of the factor. Additionally, the z-value for the improvement, calculated from equations (3.9), is reported. * Significant at 5%, ** Significant at 10%*

B.IV Efficient Factor Alpha Spanning Tests for Remaining Markets

Table B.11. France Efficient Five-Factor Model: Alphas from Spanning Tests

Efficient Factors Conditional on:	Dependent Variable				
	MKT	SMB	HML	RMW	CMA
		Efficient factors regressed on standard factors			
Prior-month return	0.419	0.484	0.442	0.483	0.276
	(2.538)*	(1.077)	(2.341)*	(2.42)*	(2.03)*
		Standard factors regressed on efficient factors			
Prior-month return	0.719	-0.101	0.085	0.245	0.117
	(2.581)*	(-0.673)	(0.447)	(1.966)*	(0.938)

*This table reports alphas and t-values (in parentheses) from regressions in which the dependent variable is one of the factors of the efficient or standard five-factor model and the explanatory variables are all five factors of the other model. The efficient factors are conditional on the prior-month (t − 1) return. * Significant at 5%, ** Significant at 10%*

Table B.12. Germany Efficient Five-Factor Model: Alphas from Spanning Tests

Efficient Factors Conditional on:	Dependent Variable				
	MKT	SMB	HML	RMW	CMA
	Efficient factors regressed on standard factors				
Prior-month return	0.536	0.493	0.142	0.359	0.143
	(2.033)*	(1.371)	(0.975)	(3.098)*	(1.238)
	Standard factors regressed on efficient factors				
Prior-month return	0.464	0.059	0.346	0.306	0.338
	(1.561)	(0.384)	(1.973)*	(2.654)*	(2.308)*

*This table reports alphas and t-values (in parentheses) from regressions in which the dependent variable is one of the factors of the efficient or standard five-factor model and the explanatory variables are all five factors of the other model. The efficient factors are conditional on the prior-month (t − 1) return. * Significant at 5%, ** Significant at 10%*

Table B.13. Netherlands Efficient Five-Factor Model: Alphas from Spanning Tests

Efficient Factors Conditional on:	Dependent Variable				
	MKT	SMB	HML	RMW	CMA
	Efficient factors regressed on standard factors				
Prior-month return	0.379	0.455	0.435	0.372	0.48
	(2.44)*	(1.916)**	(2.545)*	(0.698)	(1.285)
	Standard factors regressed on efficient factors				
Prior-month return	0.73	-0.101	0.386	0.136	0.194
	(2.387)*	(-0.589)	(1.847)**	(0.667)	(1.124)

*This table reports alphas and t-values (in parentheses) from regressions in which the dependent variable is one of the factors of the efficient or standard five-factor model and the explanatory variables are all five factors of the other model. The efficient factors are conditional on the prior-month (t − 1) return. * Significant at 5%, ** Significant at 10%*

Table B.14. Italy Efficient Five-Factor Model: Alphas from Spanning Tests

Efficient Factors Conditional on:	Dependent Variable				
	MKT	SMB	HML	RMW	CMA
	Efficient factors regressed on standard factors				
Prior-month return	0.239	1.345	0.176	0.406	0.454
	(0.436)	(0.779)	(0.466)	(3.914)*	(2.131)*
	Standard factors regressed on efficient factors				
Prior-month return	0.384	-0.037	0.099	0.628	-0.298
	(1.104)	(-0.218)	(0.52)	(3.681)*	(-2.026)*

*This table reports alphas and t-values (in parentheses) from regressions in which the dependent variable is one of the factors of the efficient or standard five-factor model and the explanatory variables are all five factors of the other model. The efficient factors are conditional on the prior-month (t − 1) return. * Significant at 5%, ** Significant at 10%*

Table B.15. Spain Efficient Five-Factor Model: Alphas from Spanning Tests

Efficient Factors Conditional on:	Dependent Variable				
	MKT	SMB	HML	RMW	CMA
	Efficient factors regressed on standard factors				
Prior-month return	0.348	0.309	0.348	0.411	0.341
	(1.319)	(0.972)	(1.537)	(2.599)*	(0.867)
	Standard factors regressed on efficient factors				
Prior-month return	0.509	0.082	0.306	0.39	-0.037
	(1.547)	(0.483)	(1.706)**	(2.356)	(-0.229)

*This table reports alphas and t-values (in parentheses) from regressions in which the dependent variable is one of the factors of the efficient or standard five-factor model and the explanatory variables are all five factors of the other model. The efficient factors are conditional on the prior-month (t − 1) return. * Significant at 5%, ** Significant at 10%*

B.V Momentum Factor Alpha Tests for Remaining Markets

Table B.16. Frace Momentum Versus Fama and French (2015) Five-Factor Model

	Dependant Variable		
	Momentum Factor		
Independent Variable	CAPM	FF5	$FF5_{ef}$
Alpha	0.708 (3.111)*	0.585 (2.765)*	0.061 (2.571)*
Market	-0.245 (-5.911)	-0.208 (-5.146)	-0.263 (-3.63)
SMB		0.042 (0.553)	0.048 (1.776)
HML		-0.264 (-3.543)	-0.008 (-0.101)
RMW_{op}		0.481 (4.836)	0.053 (0.872)
CMA		0.206 (2.012)	0.055 (0.58)

*This table presents estimates from time-series regressions that measure the association between the momentum factor and the CAPM, as well as both the standard and efficient versions of the Fama and French (2015) five-factor model in French stock returns. The alpha values, beta coefficients, and associated t-statistics (reported in parentheses) are presented in the columns across three models. In each case the dependant variable is the momentum factor. * Significant at 5%, ** Significant at 10%.*

Table B.17. Germany Momentum Versus Fama and French (2015) Five-Factor Model

	Dependant Variable		
	Momentum Factor		
Independent Variable	CAPM	FF5	$FF5_{ef}$
Alpha	1.032 (4.587)*	0.965 (4.3)*	1.02 (4.132)*
Market	-0.276 (-7.001)	-0.268 (-5.731)	-0.141 (-2.982)
SMB		-0.249 (-2.986)	0.031 (0.904)
HML		-0.274 (-3.843)	-0.111 (-1.227)
RMW_{op}		0.033 (0.335)	-0.058 (-0.534)
CMA		0.418 (4.683)	0.049 (-7.016)

*This table presents estimates from time-series regressions that measure the association between the momentum factor and the CAPM, as well as both the standard and efficient versions of the Fama and French (2015) five-factor model in German stock returns. The alpha values, beta coefficients, and associated t-statistics (reported in parentheses) are presented in the columns across three models. In each case the dependant variable is the momentum factor. * Significant at 5%, ** Significant at 10%.*

Table B.18. Netherlands Momentum Versus Fama and French (2015) Five-Factor Model

	Dependant Variable		
	Momentum Factor		
Independent Variable	CAPM	FF5	$FF5_{ef}$
Alpha	1.024	1.076	1.003
	(3.901)*	(2.229)*	(3.048)*
Market	-0.277	-0.276	-0.240
	(-6.015)	(-5.821)	(-2.596)
SMB		-0.141	0.018
		(-1.681)	(0.297)
HML		-0.191	-0.063
		(-2.571)	(-0.704)
RMW_{op}		0.000	-0.048
		(0.002)	(-1.621)
CMA		0.020	-0.060
		(0.231)	(-1.52)

This table presents estimates from time-series regressions that measure the association between the momentum factor and the CAPM, as well as both the standard and efficient versions of the Fama and French (2015) five-factor model in Dutch stock returns. The alpha values, beta coefficients, and associated t-statistics (reported in parentheses) are presented in the columns across three models. In each case the dependant variable is the momentum factor. Significant at 5%, ** Significant at 10%.*

Table B.19. Italy Momentum Versus Fama and French (2015) Five-Factor Model

	Dependant Variable		
	Momentum Factor		
Independent Variable	CAPM	FF5	$FF5_{ef}$
Alpha	0.892	0.491	0.755
	(3.901)*	(2.229)*	(3.048)*
Market	-0.237	-0.120	0.011
	(-6.951)	(-3.336)	(0.443)
SMB		0.014	0.012
		(0.198)	(1.542)
HML		-0.043	0.039
		(-0.568)	(1.063)
RMW_{op}		0.550	0.222
		(6.795)	(1.671)
CMA		-0.017	-0.122
		(-0.197)	(-6.952)

*This table presents estimates from time-series regressions that measure the association between the momentum factor and the CAPM, as well as both the standard and efficient versions of the Fama and French (2015) five-factor model in Italian stock returns. The alpha values, beta coefficients, and associated t-statistics (reported in parentheses) are presented in the columns across three models. In each case the dependant variable is the momentum factor. * Significant at 5%, ** Significant at 10%.*

Table B.20. Spain Momentum Versus Fama and French (2015) Five-Factor Model

	Dependant Variable		
	Momentum Factor		
Independent Variable	CAPM	FF5	$FF5_{ef}$
Alpha	0.822	0.797	0.716
	(3.23)*	(3.141)*	(2.61)*
Market	-0.267	-0.254	-0.097
	(-6.23)	(-6.017)	(-1.831)
SMB		-0.076	-0.034
		(-0.961)	(-0.751)
HML		-0.191	-0.063
		(-2.463)	(-1.003)
RMW_{op}		0.209	0.172
		(2.532)	(1.883)
CMA		-0.028	-0.021
		(-0.349)	(-0.583)

*This table presents estimates from time-series regressions that measure the association between the momentum factor and the CAPM, as well as both the standard and efficient versions of the Fama and French (2015) five-factor model in Spanish stock returns. The alpha values, beta coefficients, and associated t-statistics (reported in parentheses) are presented in the columns across three models. In each case the dependant variable is the momentum factor. * Significant at 5%, ** Significant at 10%.*

B.VI Model Comparison Tests with Efficient Factors Remaining Markets

Table B.21. France Model Comparison Tests with Efficient Factors

Panel A: Difference in Squared Sharpe Performance								
Model	Carhart	HXZ	CZZ	AFIM	FF5	FF6	SY	FrazPed
FF3	-0.008	-0.014	-0.002	-0.021	-0.021	-0.045	-0.045	-0.081
Carhart		-0.006	-0.012	-0.013	-0.013	-0.037	-0.037	-0.073
HXZ			-0.007	-0.007	-0.008	-0.031	-0.031	-0.067
CZZ				-0.001	-0.001	-0.025	-0.025	-0.061
AFIM					0	-0.024	-0.024	-0.06
FF5						-0.024	-0.024	-0.06
FF6							0	-0.036
SY								-0.036
Panel B: p-Values								
Model	Carhart	HXZ	CZZ	AFIM	FF5	FF6	SY	FrazPed
FF3	0.04	0.478	0.338	0.28	0.008	0	0.066	0.034
Carhart		0.758	0.458	0.404	0.449	0	0.117	0.053
HXZ			0.569	0.523	0.516	0.118	0.189	0.084
CZZ				0.192	0.928	0.189	0.302	0.113
AFIM					0.959	0.193	0.314	0.125
FF5						0.192	0.295	0.138
FF6							0.994	0.401
SY								0.381

Panel C: Multiple Model Comparison						
Model	$\hat{\theta}^2$	r	LR	p-value	n	$\hat{\theta}^2_{ORG}$
FF3	0.032	4	6.552	0.027	0	0.005
Carhart	0.04	5	5.389	0.057	0	0.034
FF5	0.054	5	3.667	0.169	1	0.028
FF6	0.077	5	0.704	0.491	0	0.067
FrazPed	0.113	5	0	0.646	1	0.089
SY	0.077	5	0.768	0.441	1	0.068
AFIM	0.053	5	3.961	0.149	1	0.066
HXZ	0.046	5	5.199	0.074	0	0.035
CZZ	0.053	5	4.317	0.126	1	0.04

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for a set of French time series efficient factor models from 1991 to 2022. Panel A reports the differences in squared Sharpe ratio. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. Panel B reports the associated p-values. Panel C reports the multiple model comparison tests of efficient factor models in French stock returns. The $\hat{\theta}^2$ column is the bias-adjusted maximum squared Sharpe performance of these efficient models. LR is the Likelihood ratio test of Wolak (1987,1989). The p-value column is the p value for the hypothesis that this model performs as well as any other model in the dataset. $\hat{\theta}^2_{ORG}$ is the bias adjusted squared Sharpe ratio of the models with original factors.

Table B.22. Germany Model Comparison Tests with Efficient Factors

Panel A: Difference in Squared Sharpe Performance								
Model	HXZ	SY	FrazPed	FF5	Carhart	FF6	AFIM	CZZ
FF3	-0.007	-0.032	-0.036	-0.036	-0.038	-0.053	-0.059	-0.061
HXZCP		-0.025	-0.029	-0.029	-0.031	-0.046	-0.052	-0.054
SY			-0.003	-0.003	-0.005	-0.021	-0.026	-0.029
FrazPed				0	-0.002	-0.018	-0.023	-0.025
FF5					-0.002	-0.018	-0.023	-0.026
Carhart						-0.016	-0.021	-0.023
FF6							-0.005	-0.008
AFIM								-0.003
Panel B: p-Values								
Model	HXZ	SY	FrazPed	FF5	Carhart	FF6	AFIM	CZZ
FF3	0.63	0.173	0.202	0	0	0	0.03	0.026
HXZCP		0.299	0.312	0.072	0.156	0.042	0.026	0.02
SY			0.92	0.892	0.823	0.372	0.266	0.228
FrazPed				0.996	0.95	0.626	0.508	0.464
FF5					0.93	0.462	0.196	0.156
Carhart						0.025	0.269	0.226
FF6							0.768	0.66
AFIM								0.744
Panel C: Multiple Model Comparison								
Model	$\hat{\theta}^2$	r	LR	p-value	n	$\hat{\theta}^2_{ORG}$		
FF3	0.019	4	5.399	0.052	0	0.021		
Carhart	0.057	5	1.224	0.389	0	0.085		
FF5	0.054	5	1.671	0.354	0	0.069		
FF6	0.072	5	0.087	0.727	1	0.114		
FrazPed	0.054	5	0.438	0.445	0	0.039		
SY	0.051	5	1.256	0.361	0	0.066		
AFIM	0.077	5	0	0.867	1	0.13		
HXZ	0.026	5	6.831	0.034	0	0.035		
CZZ	0.08	5	0	0.863	1	0.104		

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for a set of German time series efficient factor models from 1991 to 2022. Panel A reports the differences in squared Sharpe ratio. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. Panel B reports the associated p-values. Panel C reports the multiple model comparison tests of efficient factor models in German stock returns. The $\hat{\theta}^2$ column is the bias-adjusted maximum squared Sharpe performance of these efficient models. LR is the Likelihood ratio test of Wolak (1987,1989). The p-value column is the p value for the hypothesis that this model performs as well as any other model in the dataset. $\hat{\theta}^2_{ORG}$ is the bias adjusted squared Sharpe ratio of the models with original factors.

Table B.23. Netherlands Model Comparison Tests with Efficient Factors

Panel A: Difference in Squared Sharpe Performance

Model	FF3	FrazPed	FF5	SY	AFIM	CZZ	Carhart	FF6
HXZ	-0.004	-0.01	-0.012	-0.015	-0.017	-0.02	-0.025	-0.027
FF3		-0.006	-0.008	-0.011	-0.014	-0.016	-0.021	-0.023
FrazPed			-0.002	-0.006	-0.008	-0.01	-0.015	-0.017
FF5				-0.004	-0.006	-0.008	-0.013	-0.015
SY					-0.002	-0.004	-0.01	-0.011
AFIM						-0.002	-0.008	-0.009
CZZ							-0.005	-0.007
Carhart								-0.001

Panel B: p-Values

Model	FF3	FrazPed	FF5	SY	AFIM	CZZ	Carhart	FF6
HXZ	0.794	0.607	0.42	0.364	0.326	0.274	0.24	0.217
FF3		0.768	0.133	0.518	0.522	0.466	0.003	0.011
FrazPed			0.926	0.802	0.745	0.675	0.534	0.523
FF5				0.853	0.768	0.688	0.505	0.452
SY					0.911	0.818	0.611	0.562
AFIM						0.68	0.538	0.508
CZZ							0.678	0.637
Carhart								0.282

Panel C: Multiple Model Comparison

Model	$\hat{\theta}^2$	r	LR	p-value	n	$\hat{\theta}^2_{ORG}$
FF3	0.028	4	0.543	0.523	1	0.011
Carhart	0.049	5	0	0.828	1	0.049
FF5	0.036	5	0.567	0.55	1	0.015
FF6	0.051	5	0	0.817	1	0.049
FrazPed	0.034	5	0.409	0.5	0	0.054
SY	0.039	5	0.336	0.617	0	0.047
AFIM	0.042	5	0.437	0.611	0	0.048
HXZ	0.024	5	1.759	0.345	0	0.016
CZZ	0.044	5	0.222	0.681	0	0.045

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for a set of Dutch time series efficient factor models from 1991 to 2022. Panel A reports the differences in squared Sharpe ratio. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. Panel B reports the associated p-values. Panel C reports the multiple model comparison tests of efficient factor models in Dutch stock returns. The $\hat{\theta}^2$ column is the bias-adjusted maximum squared Sharpe performance of these efficient models. LR is the Likelihood ratio test of Wolak (1987,1989). The p-value column is the p value for the hypothesis that this model performs as well as any other model in the dataset. $\hat{\theta}^2_{ORG}$ is the bias adjusted squared Sharpe ratio of the models with original factors.

Table B.24. Italy Model Comparison Tests with Efficient Factors

Panel A: Difference in Squared Sharpe Performance

Model	FrazPed	Carhart	SY	HXZ	FF6	CZZ	AFIM	FF5
FF3	-0.023	-0.035	-0.04	-0.042	-0.058	-0.065	-0.065	-0.07
FrazPed		-0.012	-0.018	-0.019	-0.035	-0.043	-0.042	-0.047
Carhart			-0.006	-0.007	-0.023	-0.031	-0.03	-0.035
SY				-0.001	-0.018	-0.025	-0.024	-0.029
HXZ					-0.016	-0.023	-0.023	-0.028
FF6						-0.007	-0.007	-0.012
CZZ							0.001	-0.004
AFIM								-0.005

Panel B: p-Values

Model	FrazPed	Carhart	SY	HXZ	FF6	CZZ	AFIM	FF5
FF3	0.103	0	0.077	0.096	0	0.027	0.029	0
FrazPed		0.607	0.46	0.462	0.207	0.154	0.166	0.128
Carhart			0.794	0.791	0.004	0.157	0.164	0.216
SY				0.946	0.487	0.279	0.31	0.299
HXZ					0.554	0.222	0.253	0.24
FF6						0.723	0.738	0.649
CZZ							0.365	0.81
AFIM								0.765

Panel C: Multiple Model Comparison Tests

Model	$\hat{\theta}^2$	r	LR	p-value	n	$\hat{\theta}^2_{ORG}$
FF3	-0.003	4	6.467	0.027	0	0.005
Carhart	0.032	5	1.976	0.259	0	0.037
FF5	0.067	5	0	0.814	1	0.087
FF6	0.055	5	0.208	0.647	0	0.062
FrazPed	0.019	5	2.471	0.169	0	0.016
SY	0.037	5	1.171	0.4	1	0.046
AFIM	0.062	5	0.089	0.805	1	0.109
HXZ	0.039	5	1.564	0.343	1	0.052
CZZ	0.062	5	0.058	0.822	1	0.068

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for a set of Italian time series efficient factor models from 1991 to 2022. Panel A reports the differences in squared Sharpe ratio. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. Panel B reports the associated p-values. Panel C reports the multiple model comparison tests of efficient factor models in Italian stock returns. The $\hat{\theta}^2$ column is the bias-adjusted maximum squared Sharpe performance of these efficient models. LR is the Likelihood ratio test of Wolak (1987,1989). The p-value column is the p value for the hypothesis that this model performs as well as any other model in the dataset. $\hat{\theta}^2_{ORG}$ is the bias adjusted squared Sharpe ratio of the models with original factors.

Table B.25. Spain Model Comparison Tests with Efficient Factors

Panel A: Difference in Squared Sharpe Performance								
Model	SY	HXZ	FrazPed	FF5	CZZ	FF6	Carhart	AFIM
FF3	-0.006	-0.009	-0.009	-0.015	-0.021	-0.022	-0.026	-0.033
SY		-0.003	-0.003	-0.009	-0.016	-0.017	-0.02	-0.027
HXZ			-0.001	-0.006	-0.013	-0.014	-0.017	-0.024
FrazPed				-0.006	-0.012	-0.013	-0.017	-0.024
FF5					-0.006	-0.008	-0.011	-0.018
CZZ						-0.001	-0.005	-0.012
FF6							-0.004	-0.01
Carhart								-0.007
Panel B: p-Values								
Model	SY	HXZ	FrazPed	FF5	CZZ	FF6	Carhart	AFIM
FF3	0.723	0.588	0.568	0.025	0.342	0.013	0.001	0.155
SY		0.878	0.852	0.627	0.375	0.457	0.345	0.171
HXZ			0.977	0.771	0.588	0.581	0.485	0.353
FrazPed				0.795	0.569	0.551	0.443	0.342
FF5					0.719	0.724	0.597	0.32
CZZ						0.943	0.768	0.025
FF6							0.683	0.556
Carhart								0.67
Panel C: Multiple Model Comparison Tests								
Model	$\hat{\theta}^2$	r	LR	p-value	n	$\hat{\theta}^2_{ORG}$		
FF3	0.01	4	2.187	0.248	0	0.005		
Carhart	0.036	5	0.182	0.67	0	0.037		
FF5	0.025	5	0.987	0.445	1	0.029		
FF6	0.032	5	0.347	0.588	0	0.033		
FrazPed	0.019	5	0.905	0.438	1	0.021		
SY	0.016	5	1.886	0.319	1	0.016		
AFIM	0.043	5	0	0.829	1	0.065		
HXZ	0.018	5	0.862	0.427	1	0.013		
CZZ	0.031	5	0.005	0.843	1	0.039		

This table reports the asymptotically valid Barillas et al, (2020) pairwise tests of equality of the squared Sharpe ratios for a set of Spanish time series efficient factor models from 1991 to 2022. Panel A reports the differences in squared Sharpe ratio. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. Panel B reports the associated p-values. Panel C reports the multiple model comparison tests of efficient factor models in Spanish stock returns. The $\hat{\theta}^2$ column is the bias-adjusted maximum squared Sharpe performance of these efficient models. LR is the Likelihood ratio test of Wolak (1987,1989). The p-value column is the p value for the hypothesis that this model performs as well as any other model in the dataset. $\hat{\theta}^2_{ORG}$ is the bias adjusted squared Sharpe ratio of the models with original factors.

Chapter 4 Bayesian Tests of Model Comparison

Abstract

In this chapter, the Bayesian frameworks of Chib, Zeng, and Zhao (2020), and Chib and Zeng (2020) are used to identify the optimal combination of factors from a starting collection of 12 risk factors in each European market. The results indicate that the optimal combinations of factors are similar to the top-performing models in the classical tests. The changes in these optimal combinations are then examined under the assumptions of both normality and multivariate distributions on the factor data. Additionally, the extent to which the efficient factor transformation impacts the model comparison tests in each market is analysed. The findings reveal that efficient factors are present in the optimal combination of factors across European markets.

4.1 Introduction

The question of which risk factors best explain the cross-section of expected equity returns continues to draw attention due to its significant importance in theoretical and empirical finance (Cochrane 2011). According to asset pricing theory, as described by Cochrane (2005), a risk factor is defined as any variable that features in the stochastic discount factor (SDF), also known as the pricing kernel. Chapter One outlined the different ways researchers have aimed to specify this kernel, such as consumption-based models and linear factor models. Until this point, it has been noted that the Fama and French (2018) six-factor model, with both traditional and updated value factors, has outperformed other widely cited models over the period of analysis across the majority of markets. These models, however, are fixed in their construction, implying that their risk factors are unchanging. This chapter investigates whether factors in top-performing models reside in the SDF by examining all potential factor models that can be formed from the set of risk factors.

Addressing these questions in a classical alpha-based model comparison framework, such as the one used in Chapter Two, is not feasible given the large number of asset pricing models that can be formed from a collection of risk factors. Modern approaches to evaluating asset pricing models focus on identifying the optimal model from an initial set of factors. Barillas and Shanken (2018) were the first to propose an alpha-based optimisation methodology to combine various risk factors into a parsimonious asset pricing model for expected returns while eliminating redundant factors. Their Bayesian approach is useful because it can consider a large number of models, both nested and non-nested. Chib, Zeng and Zhao (2020) developed an SDF-based approach to identify the optimal combination of factors that fit into the stochastic discount factor framework. Both methodologies aim to create the most effective model by optimising factor inclusion and removing redundancies, underscoring their shared objective but distinct approaches.

In Bayesian model comparison, models are assessed by updating an initial prior belief about model performance with observed data. This process involves focusing on the interaction between the likelihood (how likely the observed data is under the proposed model) and the prior distribution (the initial belief about the model's parameters). This interaction is crucial for calculating Bayes factors and posterior probabilities, which indicate the level of support for a model from the data. Chib, Zeng, and Zhao (2020) critique the approach by Barillas and Shanken (2018), who assign improper priors to model parameters. Improper priors can excessively influence model comparison outcomes, leading to biased results. Chib et al. (2020) propose an alternative method for calculating the marginal likelihood (ML), which is the key component in Bayesian model comparison. Their method emphasises analytical solvability and reduces the influence of prior probabilities, aiming for a more balanced and accurate model assessment.

The analysis unfolds in three main parts. Initially, the focus is on the original 12 factors in the benchmark model scan. Each factor is assessed as a potential risk factor (an element of the stochastic discount factor, SDF) or as a non-risk factor. This approach allows for various combinations of risk factors, leading to different restricted factor models, which are then compared with the data. The performance of each factor combination which can be constructed from the 12 investment factors is examined for each market. The performance of the traditional models is compared against these optimal models identified in the model scan.

The results show that the model scan identifies top-performing models that align closely with those from the classical model comparison tests in Chapter Two. The optimal models either simplify or extend the previous models by adding or removing one or two factors, demonstrating the robustness of model selection. The results show that the relative performance rankings of these models remain consistent across different testing approaches, including marginal likelihood calculations. Unlike Qiao, Wang, and Lam (2022), no significant impact was found when applying a Student-t distribution to the data. Lastly, incorporating time-series efficient factors into the Bayesian model scan enhances model performance, particularly for factors that previously showed higher Sharpe ratios when conditioned on past returns.

Next, following Chib and Zeng's (2020) approach, results are compared under different distributional assumptions for factor data. Initially, a multivariate normal distribution is assumed for the model scans. To address the limitation that factor data can display fat tails, a Student-t distribution is then employed for the factors. The outcomes of this Student-t distribution approach are compared to those obtained under the normality assumption to examine the robustness of the findings. Lastly, the impact of efficient factors is examined. Chapter Three highlighted that several key factors in the analysis are mean-variance inefficient, considering their past return information. Using the model scanning framework, the aim is to isolate the efficient factors that can improve model performance, instead of replacing all factors with their efficient versions as done in the previous chapter.

This chapter addresses a fundamental challenge in finance: determining the most effective combination of risk factors to explain asset returns across diverse markets. With asset pricing models at the core of both academic research and practical investment decisions, the ability to distil an optimal set of factors from an extensive factor pool has significant implications. Traditional models, while foundational, often operate under fixed assumptions that can overlook critical local market dynamics or redundancies in factor selection. By re-evaluating model performance through a Bayesian framework, this chapter contributes to a growing body of research aimed at refining factor efficiency, minimising redundancy, and improving model adaptability. Additionally, the inclusion of time-series efficient factors responds to recent criticisms of factor proliferation, offering a streamlined approach that remains sensitive to market conditions over time. This investigation holds particular value for investors seeking reliable, data-driven tools for risk assessment, as well as for academics interested in advancing the theoretical underpinnings of asset pricing through more nuanced and responsive modelling frameworks.

4.2 Literature Review

Classical and Bayesian econometrics are two parallel approaches for statistical analyses. In conventional empirical studies, the efficiency of a portfolio, known to the researcher, is tested as a straightforward hypothesis. Using classical statistical inference methods, this hypothesis is either accepted or rejected at a set significance level. However, two significant deviations from this traditional method have emerged.

The first deviation acknowledges that the portfolio being tested is often an imperfect representation ("proxy") of a more theoretically ideal portfolio. The exact composition of this ideal portfolio is unknown, leading researchers to test a "composite hypothesis" that accounts for some level of inefficiency in the known portfolio. This shift moves away from testing for perfect efficiency towards testing for approximate efficiency, recognising the practical limitations of real-world portfolios. Researchers like Kandel and Stambaugh (1987) and Shanken (1987a) have explored ways to test these composite hypotheses of approximate efficiency using classical frequentist techniques. Both analytical studies (e.g., Shanken, 1985; Gibbons, Ross, and Shanken, 1985) and simulation studies (e.g., Stambaugh, 1981; Jobson and Korkie, 1982; MacKinlay, 1985) have examined the distribution of mean-variance efficiency tests in finite samples, addressing the challenge of accurately measuring the relevant benchmark return to ensure it appropriately reflects the theoretical ideal.

The second major change in evaluating portfolio efficiency and asset pricing models is the shift from traditional classical (frequentist) methods to Bayesian inference methods. In classical econometrics, the assumptions (priors) about the parameters are not explicitly stated, and estimators and test procedures are evaluated through repeated samples. In Bayesian inference, the process starts with prior beliefs or assumptions about the parameters. New data is then collected and analysed to calculate how likely it is to observe this data for different parameter values. This likelihood is combined with the prior beliefs using Bayes' theorem, resulting in a posterior distribution. The posterior distribution represents a revised belief about the parameters, incorporating both the prior information and the new data. It provides a range of probable values for the parameters, from which various summaries can be derived.

Barillas and Shanken (2018) developed a new Bayesian test to check if a given model accurately fits mean-variance efficiency, which has a straightforward solution[8]. This test starts with an informative assumption (prior) about the alphas, which measure how much actual returns differ from expected returns under the alternative hypothesis. By focusing on these informative priors for alphas, the goal is to determine if assets are priced correctly according to the model's assumptions. This prior is based on the expected Sharpe ratio, which measures risk-adjusted return. The model is then evaluated against this expectation. A key feature of their approach is using standard "diffuse" priors for other parameters like betas and residual covariance. This means the actual data influences beliefs about these parameters, allowing the researcher to focus on setting informative priors for alphas. The test is performed using two regressions: one with the alphas constrained to zero and one without this constraint. The residuals from these regressions are compared to calculate the Bayes Factor, which measures the relative support for the null hypothesis.

Barillas and Shanken (2018) employ this constrained regression approach to calculate the marginal likelihood (ML), which represents the support for each model under consideration. This approach allows for the simultaneous comparison of numerous models by integrating the aforementioned constraints into a regression framework, where the returns of included and excluded factors are assessed against the market and other included factors without intercepts, reinforcing the zero pricing error restriction. Barillas and Shanken show this through zero alpha

[8] See earlier work by Shanken (1987b), McCulloch and Rossi (1990, 1991), and Harvey and Zhou (1990).

restrictions on excluded factors, effectively ensuring that all assets are correctly priced by the model, aligning with the no-arbitrage principle central to SDF theory. Chib et al. (2020) criticise the approach of Barillas and Shanken (2018) and derive their own closed-form expressions of the log marginal likelihoods which accommodate advanced considerations such as model-specific priors. The constrained regression approach used by Barillas and Shanken (2018) and the SDF-based pricing condition outlined by Chib et al. (2020) share the goal of identifying the most effective combination of factors, but they differ in their methodologies.

4.3 Research Methods

4.3.1 Model Scan Approach

In the model scanning approach, researchers aim to identify the optimal combination of factors from an initial set of potential risk factors, diverging from the traditional method of using predetermined models with specific factors. Chib et al. (2020) describe this theory using the stochastic discount factor (SDF) approach, which provides a theoretical foundation for deriving linear factor models. Specifically, a linear factor model is an empirical representation of the SDF. Chib et al. (2020) outline that when seeking an optimal combination of factors within a collection, the stochastic discount factor (SDF)-based pricing condition requires a specific relationship. The joint distribution of these factors should be expressed in terms of the marginal distribution of the factors in the SDF and the conditional distribution of those not in the SDF. This conditional distribution must have zero intercepts, assuming all factors are traded. In simpler terms, if a factor is not part of the SDF, it should not show any predictable pattern or influence on asset prices when SDF factors are considered.

For a specific model M_j, Chib et al. (2020) considers risk factors $\breve{f}_{j,t}$: $k_{\breve{f},j}\ x\ 1$, and a complementary set of non-risk factors ${f^*}_{j,t}$: $k_{f^*,j}\ x\ 1$. Risk factors being factors which reside in the SDF. Following the approach of Hansen and Jagannathan (1991), Chib et al. (2020) specify the SDF as:

$$M_{j,t} = 1 - \lambda'_{\breve{f},j}\Omega^{-1}_{\breve{f},j}(\breve{f}_{j,t} - E[\breve{f}_{j,t}]), \tag{4.1}$$

where the SDF is characterised by risk factors $(\breve{f}_{j,t})$, $\lambda'_{\breve{f},j}$ are the risk factor loadings and $\Omega^{-1}_{\breve{f},j}$ is the covariance matrix of risk factors. Enforcing the pricing restrictions implied by the no-arbitrage condition as outlined in Section 1.2:

$$E[M_j, \breve{f}'_{j,t}] = 0 \text{ and } E[M_j, f^{*\prime}_{j,t}] = 0 \tag{4.2}$$

For all t, ensure that the expected returns on assets, after being adjusted by the SDF for risk, should net to zero across all assets and time periods. This means that the pricing of assets by the market, considering risk factors ($\breve{f}_{j,t}$) and non-risk factors ($f^*{}_{j,t}$) should not allow for arbitrage opportunities. The expected value of the risk factors is $\lambda_{x,j}$, a constant vector. This implies that, on average, the risk factors should align with this baseline level. Similarly, it is established that the expected value of non-risk factors is related to the risk factors through a matrix Γ_j, as $E\left[f^*{}_{j,t}\right] = \Gamma_j \lambda_{\breve{f},j}$. Here, Γ_j is a matrix that translates the loadings of the risk factors into the domain of the non-risk factors, effectively linking the two sets of factors.

Considering the decomposition of factors in this way, Chib et al. (2020) build on the Bayesian approach of Barillas and Shanken (2018), which allows for the simultaneous examination of a large number of models—something not feasible in traditional alpha-based asset pricing frameworks. In order to calculate this support for each model we require a model set-up which stems from the asset pricing theory.

Suppose that in model M_1 all K factors are risk factors which could influence asset returns.

$$\breve{f}_{1,t} = \breve{\alpha}_1 + \breve{\varepsilon}_{1,t}, \qquad \breve{\varepsilon}_{1,t} \sim N_{L,}(0, \Sigma_1), \tag{4.3}$$

Where each risk factor at time t (represented as $\breve{f}_{1,t}$) is modelled as a base value ($\breve{\alpha}_1$) plus a random deviation ($\breve{\varepsilon}_{1,t}$), which follows a normal distribution with mean 0 and a covariance matrix Σ_1.

Letting $\sigma_1 = vech(\Sigma_1)$, the nuisance parameters of M_1 are simply

$$n_1 = (\sigma_1)$$

A nuisance parameter here being a parameter that is not of direct interest but must be accounted for in the model's formulation and estimation process. For a given model M_j, lets assume we have $\breve{f}$ risk factors and f^* non-risk factors which have a joint Gaussian distribution. Then the aforementioned pricing restrictions imply that under a marginal-conditional decomposition of factors, we can represent model M_j (for j =2,3,... $\breve{J}$) in a restricted and reduced form as follows:

$$\breve{f}_{j,t} = \breve{\alpha}_j + \breve{\varepsilon}_{j,t}, \qquad \breve{\varepsilon}_{j,t} \sim N_{L,}\left(0, \Sigma_j\right), \tag{4.4}$$

$$f^*{}_{j,t} = \boldsymbol{B}^*_{j,f} \breve{f}_{j,t} + \varepsilon^*{}_{j,t}, \qquad \varepsilon^*{}_{j,t} \sim N_{K-Lj,}\left(0, \Sigma^*{}_j\right) \tag{4.5}$$

With nuisance parameters

$$n_j = (\beta^*_{j,f}, \sigma_j, \sigma^*_j)$$

where $\beta^*_{j,f} = vec(\boldsymbol{B}^*_{j,f})$, $\sigma_j = vech(\Sigma_j)$, and $\sigma^*_j = vech(\Sigma^*{}_j)$. So, each model can be represented as a combination of risk factors, which directly impact asset returns, and non-risk factors, which are influenced by the risk factors.

4.3.2 Prior Computation

The goal of the analysis is to calculate the support for each model given the data. The Bayesian method of model comparison requires setting specific priors for the parameters of each model, which are designed to be non-informative, thereby minimising any priori bias towards particular model configurations. Each model is then confronted with empirical data to assess its performance.

Traditionally in Bayesian statistics, Jeffreys' priors have been employed because they are non-informative and designed to minimally influence the outcomes. However, Barillas and Shanken (2018) apply these priors across different models in their model scanning framework without specific adjustments which has led to problems. Chib et al. (2020) argue that this practice can result in non-comparable marginal likelihoods because Jeffreys' priors can vary significantly based on how each model defines its nuisance parameters.

The main issue with this approach is that if nuisance parameters are defined differently across models or are derived solely from the data without a consistent foundational prior, the comparisons made between models can be misleading. This is because the marginal likelihood inherently reflects the influence of these parameters, and variations in their definitions can skew the comparison results. Furthermore, allowing data from the training sample to set the prior values can induce volatility and bias, which might favour some models over others unfairly.

To overcome these inconsistencies, Chib et al. (2020) propose a new method where all priors are connected through an invertible mapping to a base model. This approach ensures that each model's priors are transformations of a single, consistent prior set on a base model, maintaining uniformity and comparability. The application of invertible mappings involves Jacobian transformations, which adjust for changes in parameter space scale as the priors are translated from one model to another. This not only standardises the impact of priors across different models but also respects the individual nuances of each model by adjusting the prior scale appropriately. By using this method, Chib et al. (2020) ensure that the marginal likelihoods computed for each model are both valid and comparable, which is crucial for making fair and meaningful model evaluations and decisions based on Bayesian marginal likelihoods.

Let model M_1 in equation 4.3 have the improper prior on n_1 given by

$$c\Psi(n_1|M_1) = \mathrm{c}|\Sigma_j|^{-\boldsymbol{k}/\boldsymbol{2}} \tag{4.6}$$

where c is an arbitrary constant. Then the derived priors for other models of n_j in M_j, j = 2, 3, .., J̆, given by

$$\Psi(n_j|M_j) = \mathrm{c}|\Sigma_j|^{-(2Lj-K)/2}|\Sigma^*{}_j|^{-\boldsymbol{k}/\boldsymbol{2}} \tag{4.7}$$

For models M_j (for j =2,3,… J̆), the formula adjusts the weight of the priors based on the number of risk factors (L_j) and the total number of factors (K), thus tailoring the priors to the specific complexity and dimensionality of each model. The c constant serves to maintain consistency in scale across different models. This type of prior, which does not integrate to one, is often used in Bayesian statistics when a non-informative prior is desired. Here, it serves to impose minimal prior constraints on the model, allowing the data to primarily inform the posterior distributions. The prior $\Psi(n_j|M_j)$ is thus a combination of the base model's prior with specific adjustments that make it suitable for model M_j.

Instead of assuming a common prior across models for the parameter $\breve{\alpha}_j$, which represents the mean vector of the risk factors, a model-specific prior is introduced. This prior,

$$\breve{\alpha}_j|M_j \sim N_{Lj}(\breve{\alpha}_{j0},\ k_{\boldsymbol{j}}\,\Sigma_{\boldsymbol{j}}) \tag{4.8}$$

incorporates a mean ($\breve{\alpha}_{j0}$) and a covariance ($k_{\boldsymbol{j}}\,\Sigma_{\boldsymbol{j}}$) that are specific to each model M_j. The mean ($\breve{\alpha}_{j0}$) is derived from empirical data, making the prior sensitive to the observed characteristics of the risk factors in the training data. I provide more detail on the Chib et al. (2020) priors in Section C.7 of Appendix C.

4.3.3 Marginal Likelihood

The objective of this analysis is to determine the level of support for all models which can be formed from a set of risk factors based on the sample data concerning the factors. Chib et al. (2024) conducts a prior-posterior analysis on the model space denoted by $M = \{M_1, M_2,.., M_j\}$. Assume that each model in the model space is given an uninformative and equal prior model probability, that is, for any j, $\Pr(M_j) = 1/J$. Ensuring that these prior distributions are equivalent across models is crucial, as it guarantees that the models' comparative rankings are influenced by the evidence from the data rather than by variations in the priors.

In this model setup, the marginal likelihood for each model can be calculated, providing a quantifiable measure of support for each model from the data. This calculation integrates various elements evident in the model's framework, such as the impact of risk factors and their explanatory power over non-risk factors. The marginal likelihood evaluates how well the combinations of risk and non-risk factors, represented through their respective covariance matrices and the transformations between them, align with the observed data. By considering both the direct effects of risk factors and their influence on non-risk factors, the model captures a comprehensive picture of the underlying asset pricing dynamics. This thorough analysis helps in discerning which model best fits the empirical evidence.

Marginal likelihood is the probability of observing the given data under a specific model, integrating over all possible values of the model's parameters, thereby capturing the model's overall fit to the data. Since the model prior probabilities in the numerator and the denominator cancel out, the ranked models are indicated as follows:

$$m_{1*}(\gamma_{1:T}|M_{1*}) > m_{2*}(\gamma_{1:T}|M_{2*}) > \cdots > m_{J*}(\gamma_{1:T}|M_{J*}) \quad (4.9)$$

This ranking is the basis for determining which risk factors are best supported by the data.

This approach accounts for model uncertainty, both before and after observing the data. While the prior distribution reflects initial equal belief in each model, the posterior distribution, influenced by the data size, indicates the likelihood of each model being correct. As the sample size increases, posterior probabilities converge towards the true model or the closest approximation. The result of their analysis is a ranking of models based on these probabilities, effectively determining which models (and thereby, which risk factors) are best supported by the data.

Under the priors set out with c set equal to one, the marginal likelihood of model M_j (for j =2,3,… $\check{J}$), is given by can be split into two pieces (because of the independence of the errors and the independence of the priors) as follows:

$$\log \widehat{m}\left(\gamma_{1:T}\middle|M_j\right) = \log \widehat{m}\left(\check{f}_{1:T}\middle|M_j\right) + \log \widehat{m}\left(f^{*}_{1:T}\middle|M_j\right), \quad (4.10)$$

Where the first term on the right hand side is

$$-\frac{(K-L_j)(L_j)}{2}log2-\frac{(\breve{T})(L_j)}{2}log\pi-\frac{(L_j)}{2}\log(\breve{T}k_j+1)-\frac{(\breve{T}+L_j-K)}{2}\log|\Psi_j|+log\Gamma_{Lj-1}\frac{(\breve{T}+L_j-K)}{2}$$

(4.11)

And the second term is

$$\frac{(K-L_j)(L_j)}{2}log2-\frac{(K-L_j)(\breve{T}-L_j)}{2}log\pi-\frac{(K-L_j)}{2}\log|W_j^*|-\frac{(\breve{T})}{2}\log|\Psi_j^*|-log\Gamma_{K-Lj}\frac{(\breve{T})}{2}$$

(4.12)

Where $\breve{T}=(T-n_t)$ denotes the out-of-sample size, which is the portion of the data set that is not used during the training phase but instead is used to evaluate or test the model.

The first term on the right hand side of equation 4.10 assesses the part of the model that includes specific factors or variables of interest. It is focused on how well these included factors explain the observed data. Ψ_j, determinant of the covariance matrix is perhaps the most important term here as it reflects the spread or variability of the included factors which is defined as:

$$\Psi_j=\sum_{t=1}^{T}(\breve{f}_{j,t}-\hat{\hat{a}}_j)(\breve{f}_{j,t}-\hat{\hat{a}}_j)'+\frac{\breve{T}}{\breve{T}\,k_j+1}(\hat{\hat{a}}_j-\hat{a}_{j0})(\hat{\hat{a}}_j-\hat{a}_{j0})', \quad (4.13)$$

The second term deals with the parts of the data or additional factors that are not included in the main model. It evaluates the impact or relevance of these excluded factors. Here Ψ_j^* is the variance-covariance matrix of the residuals from a regression of the excluded factors on the included ones transformed by their respective loadings. This term measures how well the included factors explain the variation in the excluded factors which is defined as:

$$W_j^*=\sum_{t=1}^{T}(\breve{f}_{j,t}\,\breve{f}'_{j,t}),\ \Psi_j^*=\sum_{t=1}^{T}(f_{j,t}^*-\hat{B}_{j,f}^*\,\breve{f}_{j,t})(f_{j,t}^*-\hat{B}_{j,f}^*\,\breve{f}_{j,t})' \quad (4.14)$$

As above, the hat symbol denotes the least square estimates, but now calculated using the data beyond the training sample, and $(\Gamma_{(d)})$ denotes the d-dimensional multivariate gamma function.

The key terms $\log|\Psi_j|$, $\log|W_j^*|$, $\log|\Psi_j^*|$, in the log marginal likelihood are critical for evaluating model fit. They appear as negative terms in the above formulae, so larger values indicate more unexplained variability by the model, resulting in a lower log marginal likelihood and indicating a poorer fit. Conversely, smaller values for these terms suggest a better fit of the model to the data, leading to a higher log marginal likelihood. Thus, the log marginal likelihood provides a quantitative measure of how well each model explains the observed data, with higher values indicating better fit.

The closed-form expressions for both $\log m'\left(\mathrm{f}_{1:T}|M_j\right) + \log m'\left(\boldsymbol{f}^*_{\boldsymbol{1:T}}|M_j\right)$, mean that the integration typically required to compute the marginal likelihood has already been performed analytically as part of the derivation of the multivariate normal distribution's properties. In this case, the closed-form solutions make the computation of the marginal likelihood more straightforward and computationally efficient, as it avoids the need for numerical integration, which can be complex and time-consuming, especially in high-dimensional spaces.

The method for establishing such equivalent priors is outlined in Chib et al. (2020). These methods of assigning priors are outlined in Section 4.3.2. Focusing on the marginal likelihoods (ML) of the models given the data $\gamma_{1:T}$, the author calculates the posterior probability of each model using Bayes' theorem as:

$$\Pr\left(M_j|\ \gamma_{1:T}\right) = \frac{m_j(\ \gamma_{1:T}|M_j)}{\sum_{l=1}^{J} m_l(\ \gamma_{1:T}|M_j)'} \tag{4.15}$$

To ensure numerical stability during this calculation, especially when dealing with log marginal likelihoods, a normalisation constant is used. This constant is chosen within the range of log marginal likelihoods to prevent numerical overflow or underflow during exponentiation. Specifically, the smallest log marginal likelihood is adjusted by a factor to stabilise the values. The adjusted log marginal likelihoods are then exponentiated to convert them back to the original scale. Finally, these exponentiated values are normalised by dividing each by the sum of all exponentiated marginal likelihoods, ensuring they sum to one and yielding the posterior probabilities for each model. This process ensures that the comparative rankings of the models are based on the evidence from the data.

4.3.4 Summary Statistics of Risk Factors

It is possible to derive the posterior distribution of the factor premiums in a given factor model as per Chib et al. (2024). 10,000 simulation draws are used for generating the posterior distribution of the factor premiums, and the corresponding stochastic discount factor coefficients in the best factor model. This simulation involves generating a series of SDF values based on the historical variability of the model's factors. The distribution of the simulated SDF values provides insights into the risk and time-value adjustments that investors might require for different states of the world. The simulation of the stochastic discount factor (SDF) coefficients is indirectly inferred through the simulation of the factor premiums and the slope coefficients of the non-risk factors.

First, the mean and covariance of excess returns for risk factors are calculated from a training sample to establish a baseline for factor premiums. These premiums are then adjusted using a scaling factor derived from the sample period proportions. To understand the relationships between risk and non-risk factors, Ordinary Least Squares (OLS) estimates are performed. Following this, the posterior distributions of both risk factor premiums and coefficients for non-risk factors are simulated over 10,000 iterations[9].This simulation involves generating covariance matrices for risk factors and non-risk factors using inverse Wishart distributions, and factor premiums and non-risk coefficients using multivariate normal distributions. The SDF loadings are computed by inverting the covariance matrix and multiplying it by the vector of factor premiums.

4.3.5 Student-t Distribution

Barillas and Shanken (2018) as well as Chib, Zeng, and Zhao (2020) make an initial assumption that risk factors conform to a Gaussian distribution. Nevertheless, it is observed that the actual factor data frequently exhibits heavy tails, as noted by Fama (1965), Affleck-Graves and McDonald (1989), and Zhou (1993), which can pose challenges. Addressing this concern, Chib and Zeng (2020) expand the Bayesian model scan strategy where marginal likelihoods are computed based on proper priors and student-t distributions of the factors. Their research reveals that the student-t distributed factor model performs notably better than the Gaussian distributed model, particularly in the context of the US stock market. In line with this, Pukthuanthong et al. (2023) findings support the notion that factor data displays fat tails, and they advocate for the superiority of Student-t distributed models over Gaussian distributed models across a set of international markets.

If the model assumes a normal distribution but the data have heavy tails (as is often the case with stock returns), predictions and inferences made by the model could be misleading. Assuming normality for data that actually follows a student t-distribution could underestimate the probability of extreme outcomes, potentially leading to underestimating risk in financial applications. As such I plan to run the model scan assuming a student-t distribution on the factors[10]. This framework lacks a closed-form solution and as a result is computationally

[9] ‘iwishrnd’ and ‘mvnrnd’ are the MATLAB functions used for generating random numbers from inverse Wishart and multivariate normal distributions, respectively. These distributions are fundamental in Bayesian statistics, often used in the context of estimating variances and covariances in a multivariate setting.
[10] The RStudio code needed to run the Student-t model scan was obtained from Professor Siddhartha Chib’s webpage.

intensive. The central difference is in how hyperparameters are estimated. The closed-form solution computes them directly and analytically, while the Bayesian approach often requires sampling them as part of the overall parameter estimation process.

Suppose the joint distribution of the factors $f_t = (x_t, w_t)$ in each market follows the student-t distribution below:

$$f_t \sim St_d(\mu, \Omega, V_f), t > 1, \tag{4.16}$$

where μ: $d \times 1$ is the mean vector, $\Omega : d \times d$ is a positive definite dispersion matrix, and V_f is the degrees of freedom. x_t reflects the factors included in a given model while w_t represents the excluded factors. Since the Student-t distribution can be expressed as a Gamma-scale mixture of normal distributions, the following holds:

$$f_t | \tau_{f,t} \sim N_d(\mu, \tau_{f,t}^{-1}, \Omega), \tag{4.17}$$

$$\tau_{f,t} \sim G\left(\frac{V_f}{2}, \frac{V_f}{2}\right), \tag{4.18}$$

Where the scale $\tau_{f,t} > 0$ is latent.

Focusing on the distributional aspects of the factors, the marginal and conditional distributions take the restricted form of:

$$x_t = \lambda_x + \eta_{x,t,} \tag{4.19}$$

$$w_t = \Gamma x_t + \eta_{w.x,t,} \tag{4.20}$$

Where

$$\begin{pmatrix} \eta_{x,t,} \\ \eta_{w.x,t,} \end{pmatrix} | \tau_{f,t} \sim \left(0, \tau_{f,t}^{-1} \begin{pmatrix} \Omega_t & 0 \\ 0 & \Omega_{w.x} \end{pmatrix}\right), \tag{4.21}$$

And $\Omega_{w.x} = \Omega_w - \Omega'_{xw} \Omega_x^{-1} \Omega_{xw}$: $d_w x\, d_w. E[x_t] = \lambda_x, E[w_t] = \Gamma\lambda_x, \lambda_x$ are risk premia parameters and Γ is the matrix of regression coefficients in the regression of the w-factors on the x-factors. The regression coefficients and other model parameters are updated iteratively within the MCMC loop. This process involves drawing from various distributions such as the inverse-Wishart and multivariate normal and using the Gibbs sampler for parameter updates. The Gibbs sampler is a method used to sample from complex probability distributions by iteratively drawing from the conditional distributions of each variable given the others. Starting with initial parameter values, it systematically updates each parameter by sampling from its conditional distribution. This creates a sequence of samples that, over many iterations, approximates the target distribution. The Gibbs sampler is particularly useful in financial modelling for parameter estimation when dealing with complex data, such as in stock market analysis.

To estimate the marginal likelihood of each contending model, they employ the Chib (1995) method which starts with the convenient expression of the log-marginal likelihood.

$$\ln m\left(\mathrm{f}_{1:T}\middle|M_j\right) = \ln \pi\left(\theta^*\middle|M_j\right) + \ln p\left(\mathrm{f}_{1:T}\middle|M_j\theta^*\right) - \ln \pi\left(\theta^*\middle|M_j, \mathrm{f}_{1:T}\right) \qquad (4.22)$$

where $\theta^*= (\lambda_x^*\ , \gamma^*\ , \Omega_x^*\ , \Omega_{w.x}^*)$ is some chosen point, say the posterior mean. In this expression, the prior and likelihood ordinates can be found analytically.

The first term here is the prior probability of the parameters under a given model, the second term is the likelihood function, and the third term is the posterior ordinate. This third term is the tricky part which is where Chib's method offers a solution.

As previously outlined, it is essential for the priors in each model to be proper, meaning they must integrate to one over the parameter space. As per section 4.3.2, the prior distributions across different models should be, to a certain extent, consistent or comparable. This ensures that any observed variations in the marginal likelihoods are not simply a result of discrepancies in the prior distributions. MCMC is used to generate samples from the posterior distribution, and these samples are then used in a complex procedure to approximate the log marginal likelihood.

Chib and Zeng (2020) use a burn in period to locate the mean of the prior distribution. The spread of the prior distribution is largely a user-specified hyperparameter. The burn in period is only for constructing the prior and is not used in the model estimation process. The prior for λ_x^* is based on a product of Student-t distributions, chosen for their flexibility and thick tails. The priors for $\gamma^*\ , \Omega_x^*\ , \Omega_{w.x}^*$ are derived from a single inverse Wishart prior on Ω. The posterior distribution is sampled using MCMC methods, taking advantage of the scale mixture of normal representation of the student-t distribution. The algorithm involves several steps, each sampling different elements of the model from specific distributions (like Gamma and normal distributions).

As for the third term in equation 4.10, the posterior ordinate, Chib and Zeng (2020) suppress the model index and use a marginal-conditional decomposition to write

$$\ln \pi(\theta^* | \mathfrak{f}_{1:T}) = \ln \pi(\Omega^*_{w.x} | \mathfrak{f}_{1:T}) + \ln \pi\,(\lambda^*_x\,, \gamma^* | \mathfrak{f}_{1:T} \Omega^*_{w.x}) + \ln \pi(\Omega^*_x | \mathfrak{f}_{1:T}, \lambda^*_x\,, \gamma^*, \Omega^*_{w.x}) \tag{4.23}$$

Now appealing to the approach of Chib and Jeliazkov (2001) we have that

$$\ln \pi(\Omega^*_{w.x} | \mathfrak{f}_{1:T}) = \frac{\bar{E}_1\{\alpha\left(\Omega^*_{w.x}\Omega_{w.x}\middle| \mathfrak{f}_{1:T},\ \lambda_x,\ \gamma, \tau_f\right) I\omega_{dx}\left(\Omega^*_{.x}\middle|\rho_1, \Omega^*_{w.x,1}\ \right)\}}{\bar{E}_2\{\alpha\left(\Omega^*_{w.x}\Omega_{w.x}\middle| \mathfrak{f}_{1:T},\ \lambda_x,\ \gamma, \tau_f\right)\}} \tag{4.24}$$

where $\bar{E}_1$ denotes the expectation with respect to the posterior distribution $\pi\ (\theta | \mathfrak{f}_{1:T})$, and $\bar{E}_2$ denotes the expectation with respect to the distribution

$$\pi(\lambda_x, \gamma\ | \mathfrak{f}_{1:T}, \Omega^*_{w.x})\ x\ I\omega_{dx}(\Omega^*_{w.x}|\rho_1, \Omega^*_{w.x,1}\ ) \tag{4.25}$$

The former expectation can be calculated by Monte Carlo with the draws on θ from the full MCMC run. A reduced MCMC run is then performed in which $\Omega_{w.x}$ is fixed at $\Omega^*_{w.x}$ and the remaining blocks of parameters are sampled as before. In this case, certain parameters are fixed at specific values, while others continue to be sampled. This approach can simplify the sampling process and isolate the effects of specific parameters.

In a full MCMC run, all model parameters are typically allowed to vary and are sampled from their respective distributions. However, in certain cases, it might be beneficial to fix some parameters to specific values (perhaps based on prior knowledge, preliminary analysis, or other considerations) to reduce the complexity of the sampling process or to isolate the effects of certain parameters. The initial samples generated by the MCMC burn in run can be heavily influenced by the initial values of the parameters. The burn-in period helps to mitigate the impact of these initial values on the final results. By discarding these early samples, the subsequent samples used in analyses are more likely to be representative of the target distribution. This reduced MCMC run gives rise to the draws

$$\{\ \lambda^{(j)}_x, \gamma^{(j)}, \Omega^{(j)}_x, \tau^{(j)}_\lambda, \tau^{(j)}_f\ \} \tag{4.26}$$

For each of these draws, $\Omega^{(j)}_{w.x}$ is sampled from the proposal distribution.

$$I\omega_{dx}\left(\rho_1, \Omega^{(j)}_{w.x,1}\right), \tag{4.27}$$

where $\Omega^{(j)}_{w.x,1}$ are conditional posterior quantities computed at $(\gamma^{(j)}, \tau^{(j)}_f)$.

The second ordinate, $\pi(\lambda_x, \gamma \mid f_{1:T}, \Omega^*_{w.x})$ is estimated from the output of the previous reduced run as

$$\hat{\pi}(\lambda^*_x, \gamma^* \mid f_{1:T}, \Omega^*_{w.x}) = \frac{1}{J}\sum_{j=1}^{J} N_{dx}(\lambda^*_x \mid \lambda^{(j)}_x, B^{(j)}_{\lambda,T}) N_q(\gamma^* \mid \gamma^{(j)}, B^{(j)}_{\gamma,T}), \quad (4.28)$$

where $\lambda^{(j)}_x$, $B^{(j)}_{\lambda,T}$, $\gamma^{(j)}$, $B^{(j)}_{\gamma,T}$ are computed conditional on

$$(\Omega^{(j)}_x, \Omega^*_{w.}, \tau^{(j)}_f, \tau^{(j)}_\lambda). \quad (4.29)$$

Finally, $\pi(\Omega^*_x \mid f_{1:T}, \lambda^*_x, \gamma^*, \Omega^*_{w.x})$ is from the output of another reduced MCMC run. Fixing $(\lambda_x, \gamma, \Omega_{w.x})$ at $(\lambda^*_x, \gamma^*, \Omega^*_{w.x})$, the remaining blocks of parameters are sampled. Letting

$$\{\Omega^{(l)}_x, \tau^{(l)}_f\}_{l=1}^{L} \quad (4.30)$$

denote the draws in this second reduced run then the estimate of the final ordinate is given by

$$\hat{\pi}(\Omega^*_x \mid f_{1:T}, \lambda^*_x, \gamma^*, \Omega^*_{w.x}) = \frac{1}{L}\sum_{l=1}^{L} I\omega_{dx}(\Omega^*_x \mid \rho_1 - d_w, \Omega^{(l)}_{x,1}), \quad (4.31)$$

where $\Omega^{(l)}_{x,1}$ is computed conditional on $(\lambda^*_x, \tau^{(l)}_f)$.

In this framework, the marginal likelihood lacks a closed-form solution. The distinction between the first, second, and final ordinates lies in the specific components of the posterior distribution they represent, with each ordinate contributing to a piece of the overall puzzle in estimating the marginal likelihood. This methodical breakdown allows for the calculation of the marginal likelihood in models where direct calculation is infeasible.

4.4 Empirical Results

4.4.1 Model Performance in Bayesian Framework

This chapter uses the same dataset of 12 investment factors and 9 models discussed in previous chapters, as detailed in Section 2.5. The period of analysis remains 1991 to 2022.

First, the marginal likelihoods for the nine factor models will be calculated using the Bayesian approach outlined in Section 4.3.3. Assuming a multivariate normal distribution on our factor data, this method determines which model most accurately represents the data, with higher marginal likelihood (ML) values indicating a better fit. It balances fitting the data well with maintaining simplicity, ensuring models are not preferred solely for their complexity. To calculate the marginal likelihood (ML) as per equation 4.9, the model's performance is first assessed with the included factors, capturing how these factors directly account for the observed data. Then, the influence of factors left out of the model is evaluated by applying constraints to gauge their indirect impact.

Table 4.1 presents the ML and the corresponding posterior probability for each model across our set of countries. To examine if the differences in performance are significant, the difference in log marginal likelihoods (ML) between the best model and other models is provided in brackets next to the posterior probabilities. According to Chib et al. (2020), if the difference in log ML is ≤ 1.15, then the best model is indistinguishable from the alternative model, following Jeffrey's rule.

Table 4.1

According to the results in Table 4.1, the Bayesian analysis yields similar rankings and performance results for the models as the traditional methods did in Chapter Two. Specifically, in the U.K., the Chib et al. (2020) model emerged as the top performer in terms of marginal likelihood and posterior probability, outperforming the models by Fama and French (2018) and Asness et al. (2015), with minimal support for other models. The logML results indicate that the top three models have performance metrics that are statistically indistinguishable, each with logML values below 1.15. This aligns with the findings in Table 2.4, where these three models could not be significantly separated in terms of their squared Sharpe ratios. The numbers in brackets following the model names indicate the ranking as per the classical tests of Chapter Two. '1' indicates the best performing model as per squared Sharpe ratio.

For the French models, the Frazzini and Pederson (2014) model demonstrated superior performance, showing the highest marginal likelihood and posterior probability, indicating its dominance over other models. The outperformance is statistically significant with large logML values. This finding aligns with the classical tests, where this model also excelled. The Bayesian framework provides more support to this model compared with the classical analysis.

In Germany, classical tests have highlighted the superior performance of the Fama and French (2018) and Asness et al. (2015) models, a trend that persisted in the Bayesian analysis. These models showed the highest marginal likelihoods and posterior probabilities, rendering the probabilities of competing models negligible. The logML for the difference between these models is 0.83, showing they cannot be separated in terms of performance. The difference between the top model and the third highest performing model is 1.67, showing significant outperformance.

In the Netherlands, the Frazzini and Pederson (2014) model was initially indistinguishable in performance from other leading models based on the Sharpe ratio. However, within the Bayesian framework, this model clearly outperformed the others, achieving the highest marginal likelihood and posterior probability, with no significant support for alternative models. Large logML difference values show significant outperformance of the top model. This pattern was similar to that observed in Germany and France, where the Bayesian tests provide more support for top models.

In Italy, the six-factor models by Fama and French (2018) and Asness et al. (2015) led the models by securing the highest marginal likelihoods, with the rankings of the remaining models consistent with the classical analysis. Lastly, in Spain, classical tests did not strongly favour any particular model. This trend continued in the Bayesian analysis, where the top four models showed considerable posterior probabilities, indicating a more competitive landscape. The logML confirms this with no significant differences in performance for the top three models.

These results align closely with those obtained from the classical tests. Specifically, both approaches identify the same top-performing model based on the squared Sharpe ratio, yielding the highest marginal likelihood (ML) in five of six countries in the sample. Moreover, the relative performance rankings among the models remain consistent across both methodologies. Notably, the ML approach offers additional support to the top-performing models compared to the classical tests in certain samples. Moving forward, the next step is to analyse how these models perform against the optimal models that could be derived from the factors for each sample.

4.4.2 Bayesian Model Scan

The empirical analysis begins by running the model scan using all 12 original factors across each market in the sample. With 4,095 possible models, an equal prior probability is assigned to each, following the methods of Chib et al. (2020) and Chib et al. (2024). Table 4.2 presents the empirical results of the U.K. model scan.

Panel A of the table reports the results for the top six models in terms of the highest posterior probability. This includes the posterior probability of each model, the ratio of the posterior probability to the prior probability, and the difference in log marginal likelihoods (ML) between the best model (M1) and other models. According to Chib et al. (2020), if the difference in log ML is ≤ 1.15, then the best model is indistinguishable from the alternative model, following Jeffrey's rule. Panel B details the identity of the factors in the top six models from the model scan. Panel C reports the performance of the traditional models relative to the optimal combinations.

Table 4.2

Table 4.2 reveals that the best model includes the Market, MOM, HML_M, and RMW_{CB} factors, with a posterior probability of 0.03986. The second-best model, which includes three of these factors along with the size factor (SMB) and the investment factor (CMA), has a posterior probability of 0.03421. The next four best models have posterior probabilities ranging from 0.03230 to 0.02576. The ratio of posterior probability to prior probability shows a substantial increase across these six best models. The differences in log marginal likelihoods (ML) in Panel A are all below 1.15, indicating that the top six models are statistically indistinguishable in terms of performance.

The role of the market index in all of the best factor models is consistent with Harvey and Liu (2021) who outline that the market factor is the dominant factor for individual stocks because it substantially reduces pricing errors. The momentum factor (MOM) is present in each of the top 6 performing models which is not surprising given the high average monthly return it provides. Panel C of Table 4.1 shows that the traditional factor models perform poorly in the model scan. The posterior probabilities of each model are essentially zero, with the Chib et al. (2020) model returning a posterior probability of 0.00083. The poor performance of the traditional factor models is consistent with Chib et al. (2024).

Findings from the U.K. model scan suggest some similarities between the optimal model identified and the top-performing models in the classical framework. The four-factor model best supported by the data contains the market factor, momentum factor, updated value factor, and a profitability factor. While the models by Asness et al. (2015), Fama and French (2018), and Chib et al. (2020) outperform all others in the classical framework, the model scan indicates that the size factor (SMB) and investment factor (CMA) are not present in the optimal model. In effect, the model scan returns an optimal model that includes relevant factors and excludes redundant ones.

Model scan results for the remaining five markets in my study can be found in Section C.II in Appendix C. Results from the French model scan (Table C.1) indicate that the four-factor model best supported by the data contains the market factor and the Betting against Beta factor (BAB) as per the Frazzini and Pederson (2014) which performs best in the classical framework. However this four-factor model also contains the updated value factor (HML_M) of Asness and Frazzini (2013) and the performance factor (PERF) of Stambaugh and Yuan (2017). The model with the second highest posterior probability contains the same factors as the top performing model along with the size factor (SMB). These two models outperform all other 4,093 potential models as seen by the differences in log marginal likelihoods (ML) in Panel A are above 1.15 for models 4-6. This indicates to us that the optimal model in the French market does not come from the traditional set of factor models. In fact the top performing two-factor model of Frazzini and Pederson (2014) ranks in 118 out of 4,095 in models best supported by the data.

The German model scan (Table C.2) indicates that the five-factor model best supported by the data, with a posterior probability of 0.22766, contains the market factor, momentum factor (MOM), updated value factor (HML_M), profitability factor (RMW_{CB}), and the Frazzini and Pederson (2014) Betting against Beta factor (BAB). The model with the second-highest posterior probability of 0.06692 contains the same factors as the top-performing model, along with the Stambaugh and Yuan (2017) performance factor (PERF). The top performing model significantly outperforms all other potential models based on differences in ML. The Asness et al. (2015) and Fama and French (2018) models significantly outperformed all other models in the classical tests. The top-performing model includes four factors from the Asness et al. (2015) model. This suggests that the optimal model in the German market does not come from the predetermined set, but the model scan does identify an optimal model that closely aligns with the Asness et al. (2015) model.

The only case where the model scan returns an almost exact combination of factors included in the dominant model from the classical tests of Chapter Two is in the Netherlands (Table C.3). In this instance, the Frazzini and Pedersen (2014) two-factor model emerges as the dominant model in the classical tests of Chapter Two. The model scan returns the highest posterior probability of 0.15855 for a three-factor model comprising of the market factor and the Betting against Beta factor (BAB), supplemented by the momentum factor (MOM). The second-highest performing model includes the Stambaugh and Yuan (2017) performance factor (PERF) along with the same three factors from the top model. The top two models identified in this model scan significantly outperform all other models.

Table C.4 shows that the best factor model in the model scan for the Italian stock market is a five-factor model with a posterior probability of 0.16565. The factors in this model are the market factor, momentum (MOM), updated value factor (HML_M), betting against beta (BAB), and profitability factor (RMW_{ROE}). The second-highest performing model is a four-factor model with the same factors included except for the momentum factor (MOM). The top two models significantly outperform all other possible models from the model scan. In the classical framework, the Asness et al. (2015) six-factor model emerges as the dominant factor model from the candidate set. Four of the six factors are included in the optimal model from the model scan. The size factor and the investment factor are omitted, which is not surprising given the low return provided by these factors in this market over the period of analysis.

In the Spanish market (Table C.5), the model with the highest posterior probability in the model scan is a five-factor model. The factors in this model are the market factor, momentum (MOM), updated value factor (HML_M), betting against beta (BAB), and profitability factor (RMW_{OP}), with a posterior probability of 0.05516. The top six models from the model scan cannot be statistically separated. The dominant model emerging from the classical comparison is the Asness et al. (2015) six-factor model. The model best supported by the data contains three of the factors included in the top performing Asness et al. (2015) model from the classical framework.

The results from the Bayesian model scan, when compared to the classical approach, remain consistent across different markets, including the U.K. In every instance, the optimal model closely mirrors the top-performing model within each market but is fine-tuned by either incorporating an additional factor or eliminating a redundant one[11]. For example, in the Dutch sample, a single factor is added, while in the Italian sample, the size and investment factors are omitted due to their redundancy, indicating their absence in the SDF These findings highlight the robustness and consistency between the two approaches in identifying superior models and underscore the ability of the model scan approach to refine and enhance the performance of these models. Panel C of Tables C.1-C.5 shows that the traditional factor models perform poorly in the model scan across all remaining samples. The posterior probabilities of most models are essentially zero. This poor performance of the traditional factor models is consistent with findings from Chib et al. (2024).

[11] The difference in results arises from the methodologies used. The squared Sharpe measure focuses on maximising risk-adjusted returns from historical data, while the Bayesian approach incorporates model uncertainty and prior beliefs.

4.4.3 Summary Statistics of Best Model Risk Factors

Table 4.2 suggests that the best-performing model in terms of posterior probability for U.K. returns is a four-factor model. Chib et al. (2024) derive the posterior distribution of the factor premiums in a given factor model. Using 10,000 simulation draws, the posterior distribution of the factor premiums and the corresponding stochastic discount factor coefficients in the best factor model is generated. This simulation involves generating a series of SDF values based on the historical variability of the model factors. Each simulation calculates the SDF for each period based on the model's parameters and the values of the underlying factors or variables at that time. The distribution of the simulated SDF values provides insights into the risk and time-value adjustments that investors might require for different states of the world. The SDF coefficients are indirectly inferred through the simulation of the factor premiums and the slope coefficients of the non-risk factors. Table 4.3 reports the summary statistics of the posterior distribution of the factor premiums (Panel A) and stochastic discount factor coefficients (Panel B). The summary statistics include the mean, standard deviation (Std Dev), median, and 2.5% and 97.5% percentiles of the posterior distribution.

Table 4.3

Panel A of Table 4.3 shows that the MOM factor has the largest mean factor premium at 0.875%, followed by the Market factor at 0.467%. All of the factor premiums are significantly positive using the 95% percentile interval, with the exception of the Market factor. In Panel B of Table 4.3, all of the mean stochastic discount factor coefficients are negative for each factor and significantly negative using the 95% percentile intervals. The negative SDF coefficients reflect the compensatory mechanism required by investors for taking on additional risks associated with each factor. This finding suggests that all four factors play an important role in the stochastic discount factor in pricing assets, given the other factors in the model (Cochrane, 2005). The Market factor plays an important role even where the mean factor premium is not significantly positive.

The findings from Table 4.3 can be contextualised by comparing the posterior factor premiums to the traditional estimates in Table 2.1. Notably, the MOM factor consistently shows a strong performance, with a high mean return and Sharpe ratio in both the Bayesian analysis and traditional statistics, underscoring its significance in pricing U.K. assets. Similarly, factors like BAB and RMW_{OP} also exhibit robustness, as their posterior means align with their established positive returns. However, the variability in factor premiums, such as HML and HML_{M}, highlights potential differences in risk compensation, suggesting that certain factors may require further scrutiny in the Bayesian framework.

Panel C reports the summary statistics of the fitted stochastic discount factor values of the best model from the Bayesian model scan and a set of traditional factor models. The fitted values are estimated using the posterior mean of the stochastic discount factor coefficients. The summary statistics include the standard deviation (Std Dev), the minimum, maximum, and the proportion (Prop y<0) of fitted values that are below zero. The final column shows the correlations between the fitted values of the traditional factor models and the best model. The simulated SDFs of the 'Best' model, derived from the factor premiums and coefficients of the factors provided in Panel A of each table, show more volatility than any of the traditional models. This suggests that these optimal models may offer a more nuanced or accurate representation of market dynamics.

Section C.III in Appendix C presents the results for the remaining markets, where a similar pattern is observed. Large values for the simulated factor premia on the momentum (MOM) and betting against beta (BAB) factors are consistent across the samples, which aligns with the large average return of these factors over the sample period. The highest simulated factor premium appears in the French market (Table C.6), where the simulation returns a mean premium of 1.532% on the BAB factor. In Panel C of Tables C.6-C.10, the simulated SDFs of the 'Best' model, derived from the factor premia and coefficients provided in Panel A of each table, show more volatility than any of the traditional models. This suggests that these optimal models may offer a more nuanced or accurate representation of market dynamics in each sample.

4.4.4 Multivariate-t Assumption

Similar to the approach by Qiao, Wang, and Lam (2022), the results from the model scan assuming a joint Gaussian distribution on the factors will now be compared to those assuming a joint multivariate-t distribution. Qiao et al. (2022) provide strong evidence that models using a Student-t distribution for global factor pricing significantly outperform those using a Gaussian distribution. This highlights the importance of using multivariate Student-t distributions to account for the fat tails in global risk factor data. The Student-t distribution is often favoured over the Gaussian distribution in financial modelling due to its ability to better handle "fat tails"—a characteristic of many financial datasets where extreme outcomes (e.g., financial crises, sudden market shocks) occur more frequently than a normal distribution would predict. In a Gaussian distribution, the probability of extreme events decreases rapidly, which can lead to an underestimation of risk in datasets with higher volatility and more frequent large deviations from the mean. In contrast, the Student-t distribution has heavier tails, meaning it assigns a higher probability to these extreme outcomes. This feature allows the Student-t distribution to more accurately capture the behaviour of financial returns, which are prone to significant fluctuations. By accommodating the fat tails, models using the Student-t distribution offer a more realistic view of potential risks and returns, resulting in better predictions and assessments of asset pricing models, particularly in environments where extreme market movements are more common.

Additionally, they find that assuming a Student-t distribution for the factor data can lead to the identification of a different top-performing asset pricing model compared to when multivariate normality is assumed. This distinction matters because if factors exhibit fat tails, a Gaussian distribution may underestimate the likelihood of extreme outcomes, leading to less accurate model assessments. Conversely, a Student-t distribution, which better captures the heavier tails of the data, can provide a more realistic representation of the risk and return dynamics. By accurately reflecting the true distribution of factor returns, the simulation approach can identify the most effective asset pricing models.

Using the approach of Chib and Zeng (2020), the model scan of the set of 12 factors across the samples was rerun, assuming a multivariate-t distribution with three degrees of freedom. In this case, when the joint distribution of the risk factors follows a Student-t distribution, an initial portion of the data is first used as the training sample to obtain the prior distribution of the parameters of the factor model in order to calculate the marginal likelihood of each contending model. The Markov chain Monte Carlo (MCMC) method is then employed to obtain the posterior distribution of the parameters and calculate their posterior means, which are further used to calculate the marginal likelihood of the factor model. Section 4.3.5 outlines the MCMC at a high level. For full details on the calculation of the marginal likelihood under the multivariate-t assumption, see Chib and Zeng (2020).

Table 4.4

Similar to Qiao, Wang, and Lam (2022), the U.K. model scan presented in Table 4.4 shows that adopting the multivariate-t assumption leads to the identification of a different top-performing model compared to assuming a Gaussian joint distribution of factors. However, their study using global factors found that shifting from a Gaussian to a Student-t distribution resulted in the addition of four extra factors to create a seven-factor model. The results here differ. When multivariate normality is assumed, the four factors from the original model scan remain, with the addition of a single factor, SMB.

Similar to Qiao et al. (2022), increased support is found for the top-performing models when assuming a Student-t distribution. The top-performing five-factor model sees an increase in its posterior probability, rising from 0.039 to 0.065. The differences in log ML in Panel A of Table 4 remain below 1.15 for all of the top six best-performing models, indicating their performance is not significantly different. When a Gaussian distribution is assumed for the factors, these top six models are also statistically indistinguishable from each other. The support for the top models has increased slightly when assuming a Student-t distribution.

The results of the remaining model scans assuming a multivariate-t distribution are presented in Section C.IV of Appendix C. This analysis reveals diverse outcomes across the dataset. Specifically, in France and Germany (referenced in Tables C.11 & C.12), the adoption of a multivariate-t distribution for factor data results in the identification of different top-performing models compared to those identified under the assumption of multivariate normality, typically incorporating an additional factor. In France, this involves adding the momentum (MOM) factor to the four-factor model also identified from the original scan. In Germany, it includes both size (SMB) and profitability (RMW_{OP}) factors, with the BAB factor dropping out. For Germany, the distinction between the top models becomes less pronounced under the multivariate-t distribution, with the top five models showing comparable performance—a contrast to the Gaussian assumptions, where a single model clearly outperformed the others.

Conversely, in the Netherlands (Table C.13), the same three-factor model consistently ranks as the most effective regardless of the distributional assumption. The top two models' superiority is reinforced by a log-likelihood (ML) difference of more than 1.15 compared to the next four highest-performing models. In Italy (Table C.14), the Student-t assumption sees a five-factor model as the top performer with the BAB factor replaced by the CMA factor. While the change in distributional assumption does not drastically alter model composition, it diminishes their relative support, making differences between models less significant. As for Spain (Table C.15), the comparison reveals no difference in the identification and performance of the top-performing model between the two distributional assumptions.

4.4.5 Overall Performance Gaussian v Student-t

The overall performance of Gaussian versus Student-t distributed models is now examined. Higher marginal likelihoods (MLs) indicate that a model fits the data well, accounts for its complexity and prior assumptions, and captures significant data variance. This suggests strong explanatory power, especially when assessing the role of certain risk factors. Qiao, Wang, and Lam (2022) reported strong evidence in favour of global factor pricing models using the Student-t distribution, as indicated through higher MLs. Table 4.5 below presents the ML for the top six performing factor models in U.K. returns under both statistical distributions on the factor data.

Table 4.5

The Gaussian distributed factor models return a higher ML, indicating they fit the data better than the Student-t distributed model. The higher MLs signify that the Gaussian model more accurately captures the data's variance while adhering to prior assumptions.

Analysing the log-marginal likelihoods for Gaussian and Student-t models across five countries, as presented in Section C.V of Appendix C, reveals mixed results. Some findings indicate that Gaussian distributed models outperform Student-t models, while others show the opposite. In France (Table C.16), Germany (Table C.17), and the Netherlands (Table C.19), higher log-marginal likelihoods for Student-t models indicate a better fit to the factor data. Conversely, in Italy (Table C.20) and Spain (Table C.21), Gaussian models are favoured, suggesting a more accurate representation of these market data. This distinction underscores the importance of selecting the right model based on its ability to accurately explain both the observed data and the variability of factors not included. While Qiao, Wang, and Lam (2022) argue for the superiority of Student-t models in global factor models, the country-specific analysis illustrates that this generalisation does not uniformly apply in a European setting, emphasising the need for a nuanced approach in model selection.

The differences in model performance between European factors in this study and the global factors examined by Qiao et al. (2022) are likely due to two key factors specific to European markets. First, the statistical characteristic of European factor returns may differ from those of global factors; for instance, European markets might exhibit less pronounced fat tails, resulting in fewer extreme events and making the Gaussian distribution more suitable for modelling factor data in certain countries. Second, country-specific economic conditions and unique risk factors may contribute to a more stable pattern of returns in some European countries, such as Spain and Italy, where the Gaussian model appears to provide a better fit. Together, these factors suggest that the observed differences are driven by local market conditions and statistical properties, highlighting the need for a tailored approach to model selection in European contexts.

4.4.6 Impact of Efficient Factors

As previously noted, only some factors benefit from the efficient factor optimisation, making it plausible that certain efficient factors contribute to incremental enhancements in model performance. To examine the impact of the efficient factor transformation on selecting the optimal asset pricing model for different samples, the model scan is rerun to include both the original factors and their efficient counterparts calculated using the framework of Ehsani and Linnainmaa (2022). This results in a starting collection of 24 factors. With these, there are now 67,108,864 possible models, to which an equal prior probability is assigned, following the approach of Chib and Zeng (2020) and Chib et al. (2024). Given that Chapter Three shows the efficient factor transformation significantly increases the Sharpe ratio of various factors across the dataset, the expectation is that these factors will be included in an optimal asset pricing model. Due to computational limitations and the large number of models to be compared, the assumption of multivariate normality for the factor returns is adopted, as assuming a multivariate t-distribution would exceed the available computing resources. This should not be an issue as no major differences are found in the results from the model scans assuming different distributions on the factors in previous sections.

Table 4.6 reports the empirical results of a model scan in U.K. factors with the time series efficient factors also included. Panel A of the Table reports the results for the top 6 models in terms of the highest posterior probability. Efficient factors are denoted with the superscript "ef". Panel A includes the posterior probability of each model and the difference in log marginal likelihoods (ML) between the best model to that of the next best model in descending order. Panel B reports the identity of the factors in the top 6 models from the model scan.

Table 4.6

Panel B of Table 4.6 shows that the best factor model from the candidate set of 24 factors is a seven-factor model which includes some efficient factors. More specifically the efficient versions of the Value (HML) factor, profitability factor (RMW_{OP}) and the betting against beta (BAB) factor are included in the top performing model along with the following original factors {Market, BAB, HML_M, RMW_{CB}. All of the original factors have been retained in the top-performing model; however, the addition of these three efficient factors enhances the performance of the asset pricing model. The inclusion of these specific efficient factors is not surprising, given the significant increase in Sharpe performance for these factors when their weight is conditioned on previous returns in the U.K. market, as noted in Table 3.2. For example the largest and most significant increase in Sharpe ratio after the efficient factor transformation was for the RMW_{OP} factor of 0.22 (3.61). It is not surprising to see these efficient factors present in the best performing model. The posterior probability of the top model is 0.01018 indicating that there is moderate to small support from the data for this particular model from the possible set. The next six best models have a posterior probability that ranges between 0.00937 and 0.00545. The difference in log ML in Panel A of Table 4.5 is below 1.15 for the top six models and so the best model is statistically indistinguishable from the other top five models.

Section C.VI in Appendix C presents the results for the remaining five markets. In each market, factors that exhibit a notable enhancement in their Sharpe ratio, when their weight is determined by their historical return, have been identified and incorporated into the optimal models during the model scanning process. Such factors, which have shown a significant improvement in Sharpe ratio (Section 3.3.2), are included in the highest-performing models. For instance, we see the inclusion of the efficient Betting Against Beta (BAB) factor in the top-performing model for the German market (Table C.22) and the inclusion of the efficient size factor (SMB) in the Dutch context (Table C.23). Additionally, some efficient factors that have demonstrated non-significant increases, such as the HML_M factor in the top-performing Spanish model (Table C.25), are also included. However, in all cases, the top-performing models return a small posterior probability. This indicates that, while they are the best-supported models, the top models are statistically indistinguishable from other models.

4.5 Conclusions

In this chapter, the analysis unfolded in three main parts. Initially, the Bayesian model scan approach proposed by Chib et al. (2020) was employed to identify the top-performing asset pricing model from a starting collection of risk factors in each of the European samples. Next, following Chib and Zeng's (2020) approach, the impact of different distributional assumptions for factor data on the choice of the optimal model was explored. The composition and support for top-performing models were compared assuming both a Gaussian and Student-t distribution. Finally, the influence of the time series efficient factor transformation on the choice of the optimal model across markets was investigated.

The initial model scan identifies optimal models that are similar to the top-performing models obtained through the classical tests of model comparison in Chapter Two. In each instance, the optimal model from the model scan either represents a reduced form with one or two fewer factors or an extension of the model identified in Chapter Two, with one or two additional factors incorporated. This alignment underscores the robustness of the model selection across different testing methodologies. The marginal likelihood for each of the original models with fixed risk factors was also calculated, revealing that the relative performance across the sample is almost identical to the relative rankings from the classical tests, further reinforcing the consistency of these findings.

Contrary to the findings of Qiao, Wang, and Lam (2022), the analysis does not reveal significant disparities in the results when a Student-t distribution assumption is applied to the factor data. While there is a slight tendency towards stronger support for the top-performing models across the sample, the composition of these top-performing models remains relatively stable. The overall marginal likelihood for the top-performing models is not consistently higher across the sample for either distributional assumption.

The final model scan highlights the positive impact of incorporating time-series efficient factor transformations on model performance. When these efficient factors are included in the Bayesian model scan framework, specific efficient factors emerge as components of the top-performing model. Notably, these efficient factors align with the findings of Chapter Three, which show that these factors have a higher Sharpe ratio when factor weights are conditioned on past returns. Additionally, some efficient factors with an insignificant Sharpe ratio increase are also included. The model scan strategy has effectively isolated these factors within the model comparison framework.

The practical implications of this work are significant. The Bayesian model scan approach reveals that traditional models often underperform compared to combinations of factors identified through the scan. This allows investors to isolate relevant risk factors that traditional analyses might miss. A key advantage of the Bayesian approach is its flexibility to incorporate different distributional assumptions, which is crucial for accurately modelling the non-normal behaviour of asset returns, such as skewness and kurtosis. Additionally, the dynamic updating mechanism of Bayesian methods keeps the model selection adaptive to new data, enhancing predictive accuracy over time.

The contributions of this chapter extend to both academic finance and practical asset management by offering a refined approach to model selection that addresses the limitations of traditional asset pricing models. For academics, this chapter demonstrates the value of Bayesian methodologies in model optimisation, highlighting how adaptive, data-driven approaches can reduce the redundancies inherent in fixed-factor models while also considering diverse distributional assumptions to better capture market realities. This work builds on prior analyses by confirming that a tailored, efficient factor approach can yield more robust and resilient pricing models across different markets, aligning with findings from earlier chapters and further emphasising the importance of conditioning factor weights. For practitioners, the Bayesian model scan provides a flexible framework that not only adapts to new data but also allows for greater precision in isolating relevant risk factors, helping investors identify optimal factor combinations that might otherwise be overlooked. The incorporation of time-series efficient factors and the flexibility to account for non-normal return distributions make this approach particularly valuable in volatile or complex markets, offering a robust tool for constructing portfolios that are both adaptable and informed by deeper market insights.

Appendix C

C.I U.K Empirical Results

Table 4.1. Model Performance in a Bayesian Framework

UK			France		
Model	ML	Prob (logML)	Model	ML	Prob (logML)
CZZ (3)	9820.74	0.53263	FrazPed (1)	8047.55	0.99994
AFIM (2)	9820.4	0.3827 (0.14)	SY (2)	8037.79	5.8E-05(4.23)
FF6 (1)	9818.34	0.04871(1.03)	Carhart (7)	8032.89	4.3E-07(6.73)
HXZ (5)	9817.51	0.02118 (1.41)	HXZ (6)	8032.77	3.8E-07(6.42)
Carhart (4)	9817.03	0.01315 (1.61)	AFIM (4)	8032.74	3.7E-07(6.43)
SY (6)	9814.71	0.00129 (2.62)	CZZ (5)	8031.84	1.5E-07(6.82)
FF5 (7)	9813.03	0.00024 (3.35)	FF3 (9)	8031.14	7.4E-08(7.13)
FrazPed (8)	9812.08	9.3E-05 (3.76)	FF6 (3)	8030.46	3.8E-08(7.42)
FF3 (9)	9807.29	7.7E-07 (5.84)	FF5 (8)	8030.21	2.9E-08(7.54)

Germany			Netherlands		
Model	ML	Prob (logML)	Model	ML	Prob (logML)
AFIM (1)	8739.06	0.85303	FrazPed (1)	7774.21	0.97336
FF6 (2)	8737.16	0.12695(0.83)	Carhart (3)	7770.17	0.01722 (1.75)
Carhart (4)	8735.21	0.01814(1.67)	CZZ (6)	7768.92	0.00493 (2.3)
CZZ (3)	8732.86	0.00173 (2.7)	SY (5)	7768.17	0.00232 (2.62)
FF5 (5)	8729.67	7.1E-05 (4.08)	FF6 (2)	7766.96	0.00069 (3.15)
SY (6)	8729.67	7.1E-05 (4.08)	FF3 (9)	7766.85	0.00062 (3.2)
HXZ (7)	8726.22	2.3E-06 (5.57)	AFIM (4)	7766.58	0.00048 (3.31)
FF3 (8)	8724.9	6E-07 (6.15)	HXZ (7)	7766.28	0.00035 (3.44)
FrazPed (9)	8724.04	2.5E-07 (6.53)	FF5 (8)	7763.7	2.7E-05 (4.56)

Italy			Spain		
Model	ML	Prob (logML)	Model	ML	Prob (logML)
AFIM (1)	8428.84	0.76395	FrazPed (8)	7830.53	0.53797
FF6 (4)	8427.32	0.16641(0.66)	AFIM (1)	7829.25	0.1487 (0.56)
FF5 (2)	8426.42	0.06784 (1.05)	CZZ (2)	7829.1	0.12863 (0.62)
FrazPed (8)	8421.49	0.00049 (3.19)	Carhart (3)	7829.1	0.12855 (0.62)
CZZ (3)	8421.28	0.0004 (3.28)	SY (7)	7827.44	0.02437 (1.34)
Carhart (7)	8421.15	0.00035 (3.34)	FF6 (4)	7826.72	0.01188 (1.66)
HXZ (5)	8420.98	0.00029 (3.42)	HXZ (6)	7826.67	0.01132 (1.68)
SY (6)	8420.87	0.00026 (3.47)	FF3 (9)	7826.15	0.00669 (1.91)
FF3 (9)	8415.11	8.3E-07 (5.96)	FF5 (5)	7824.88	0.00188 (2.46)

The table reports the results of the Bayesian marginal likelihood computations in European stock returns. The sample period is June 1991 and December 2022. The marginal likelihood of each model in presented with the corresponding posterior *probability which indicates support for the model from the data. The logML is the difference in the log ML of the best model and the model in that row.*

Table 4.2. U.K. Model Scan 12 Factors

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
1	0.03986	163.242	
2	0.03421	140.074	0.15306
3	0.03230	132.259	0.21048
4	0.03123	127.894	0.24404
5	0.02940	120.412	0.30432
6	0.02576	105.48	0.43672

Panel B: **Factors**						
1	Market	MOM	HML_M	RMW_{CB}		
2	Market	SMB	MOM	CMA	HML_M	
3	Market	SMB	MOM	HML_M	RMW_{CB}	
4	Market	MOM	CMA	HML_M	RMW_{CB}	
5	Market	SMB	MOM	HML_M		
6	Market	SMB	MOM	CMA	HML_M	RMW_{ROE}

Panel C:	Posterior Probability	Posterior/Prior
FF3	1.2E-09	4.9E-06
Carhart	2E-05	0.08363
FF5	3.7E-07	0.00153
FF6	7.6E-05	0.30989
FrazPed	1.4E-07	0.00059
SY	2E-06	0.00821
AFIM	3.3E-05	0.13474
HXZ	0.00059	2.43464
CZZ	0.00083	3.38843

The table reports the results of the Bayesian model scan of 12 factors in U.K. stock returns. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Panel C reports the posterior probabilities of the 9 candidate models examined previously.

Table 4.3. U.K. Summary Statistics of the Posterior Distribution of the Best Model Risk Factors

Panel A: **Premiums**	Mean	Std Dev	Median	2.50%	97.50%
Market	0.467	0.251	0.470	-0.028	0.953
MOM	0.875	0.251	0.871	0.391	1.368
HML_M	0.081	0.204	0.081	-0.315	0.485
RMW_{CB}	0.330	0.115	0.330	0.103	0.552

Panel B: **SDF Coeffs**	Mean	Std Dev	Median	2.50%	97.50%
Market	-3.457	1.327	-3.454	-6.075	-0.866
MOM	-8.252	1.692	-8.228	-11.663	-5.033
HML_M	-8.994	2.275	-8.988	-13.482	-4.659
RMW_{CB}	-11.083	3.155	-11.058	-17.220	-4.986

Panel C: Summary Statistics of Candidate Stochastic Discount Factor Models

	Std Dev	Minimum	Maximum	Prop y<0	Correlation
Best	0.360	-0.051	2.137	0.580	0.000
CAPM	0.098	0.670	1.451	0.000	0.278
FF3	0.105	0.528	1.553	0.000	0.229
FF5	0.318	-0.076	2.524	0.290	0.808
FF6	0.345	-0.056	2.680	0.580	0.749
AFIM	0.363	-0.335	2.834	0.870	0.785
SY	0.261	0.301	2.326	0.000	0.582

The table reports the summary statistics of the posterior distribution of the factors in the best model from the Bayesian model scan of 12 factors in U.K. stock returns. Panel A reports the summary statistics of the posterior distribution of the factor premiums (%), and Panel B reports the summary statistics of the posterior distribution of the stochastic discount factor coefficients. The summary statistics include the mean, standard deviation (Std Dev), median, 2.5% and 97.5% percentiles using 10,000 simulation draws. Panel C reports the summary statistics of the fitted stochastic discount factor values of the best model from the Bayesian model scan, and a set of traditional factor models. The fitted values are estimated using the posterior mean of the stochastic discount factor coefficients. The summary statistics include the standard deviation (Std Dev), the minimum, maximum, and the proportion (Prop y<0) of fitted values that are below zero. The final column of Panel C is the correlations between fitted values of the traditional factor models, and the best model.

Table 4.4. U.K. Model Scan of 12 Factors assuming Multivariate-t Factor Distribution

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
1	0.065427	267.9252	
2	0.052782	216.1417	0.214773
3	0.043959	180.0123	0.397683
4	0.041313	169.1759	0.459769
5	0.031584	129.3381	0.728278
6	0.029064	119.0162	0.811448

Panel B: **Factors**						
1	Market	SMB	MOM	HML_M	RMW_{CB}	
2	Market	SMB	MOM	CMA	HML_M	
3	Market	BAB	SMB	MOM	HML_M	RMW_{CB}
4	Market	MOM	CMA	HML_M	RMW_{CB}	
5	Market	SMB	MOM	HML_M		
6	Market	SMB	MOM	CMA	HML_M	RMW_{ROE}

The table reports the results of the Bayesian model scan of 12 factors in U.K. stock returns when the joint distribution of factors is assumed to follow a multivariate t distribution. The sample period is June 1991 and December 2022. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Table 4.5. U.K. Log-Marginal Likelihoods of the Top Performing Models

<u>Panel A: Top Performing Gaussian Models</u>						
						$m\ (f_{1:T} \mid M_i)$
Market	MOM	HML_M	RMW_{CB}			9811.69
Market	SMB	MOM	CMA	HML_M		9810.26
Market	SMB	MOM	HML_M	RMW_{CB}		9809.47
Market	MOM	CMA	HML_M	RMW_{CB}		9810.34
Market	SMB	MOM	HML_M			9809.43
Market	SMB	MOM	CMA	HML_M	RMW_{ROE}	9808.37
<u>Panel A: Top Performing Student-t Models</u>						
Market	SMB	MOM	HML_M	RMW_{CB}		9111.06
Market	SMB	MOM	CMA	HML_M		9110.84
Market	BAB	SMB	MOM	HML_M	RMW_{CB}	9110.66
Market	MOM	CMA	HML_M	RMW_{CB}		9110.6
Market	SMB	MOM	HML_M			9110.33
Market	SMB	MOM	CMA	HML_M	RMW_{ROE}	9110.24

The table presents the log marginal likelihoods for the best-performing U.K. models identified through both analyses. Panel A details top models assuming a Gaussian distribution, while Panel B details top models with a student-t distribution.

Table 4.6. U.K. Model Scan 24 Factors

Panel A: **Top Models** Model	Posterior Probability	ML
1	0.01018	
2	0.00937	0.08317
3	0.00742	0.31567
4	0.00597	0.53426
5	0.00558	0.60167
6	0.00545	0.62521

Panel B: **Factors**								
1	Market	MOM	HML_M	RMW_{CB}	HML^{ef}	RMW_{OP}^{ef}	BAB^{ef}	
2	Market	MOM	MGMT	HML^{ef}	RMW_{OP}^{ef}	BAB^{ef}		
3	Market	MOM	HML_M	RMW_{CB}	HML^{ef}	RMW_{OP}^{ef}	BAB^{ef}	
4	Market	MOM	CMA	HML_M	RMW_{ROE}	HML^{ef}	RMW_{OP}^{ef}	BAB^{ef}
5	Market	MOM	HML_M	RMW_{ROE}	$Market^{ef}$	HML^{ef}	BAB_{ef}	
6	Market	MOM	HML_M	RMW_{ROE}	HML^{ef}	BAB^{ef}		

The table reports the results of the Bayesian model scan of 24 factors in U.K stock returns. This set is made up of 12 factors and their efficient counterparts. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 6. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Superscript 'ef' denotes the efficient version of the factor.

C.II Bayesian Model Scan for Remaining Markets

Table C.1. France Model Scan 12 Factors

Panel A: **Top Models**		Posterior Probability		Posterior/Prior		ML
Model						
1		0.09268		379.51		
2		0.05629		230.514		0.49857
3		0.02383		97.601		1.35799
4		0.02331		95.4401		1.38038
5		0.02256		92.3629		1.41315
6		0.02053		84.0626		1.50732
Panel B: **Factors**						
1	Market	BAB	HML_M	PERF		
2	Market	SMB	BAB	HML_M	PERF	
3	Market	RMW_{OP}	BAB	HML_M	PERF	
4	Market	SMB	BAB	HML_M	PERF	RMW_{CB}
5	Market	BAB	PERF	PERF	RMW_{CB}	
6	Market	HML	BAB	PERF		
Panel C:		Posterior Probability		Posterior/Prior		
FF3		1.1E-10		4.52E-07		
Carhart		6.35E-10		2.6E-06		
FF5		4.35E-11		1.78E-07		
FF6		5.62E-11		2.3E-07		
FrazPed		0.001484		6.077379		
SY		8.56E-08		0.000351		
AFIM		5.63E-10		2.31E-06		
HXZ		5.49E-10		2.25E-06		
CZZ		2.23E-10		9.12E-07		

The table reports the results of the Bayesian model scan of 12 factors in French stock returns. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Panel C reports the posterior probabilities of the 9 candidate models examined previously.

Table C.2. Germany Model Scan 12 Factors

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
1	0.22766	932.251	
2	0.06692	274.039	1.22433
3	0.04505	184.477	1.62008
4	0.04116	168.57	1.71025
5	0.04064	166.402	1.72319
6	0.03492	143.012	1.87467

Panel B: **Factors**						
1	Market	MOM	BAB	HML_M	RMW_{CB}	
2	Market	SMB	MOM	BAB	HML_M	PERF
3	Market	MOM	CMA	BAB	HML_M	RMW_{CB}
4	Market	MOM	RMW_{OP}	BAB	HML_M	RMW_{CB}
5	Market	HML	MOM	BAB	RMW_{CB}	
6	Market	MOM	BAB	HML_M		

Panel C:	Posterior Probability	Posterior/Prior
FF3	4.8E-11	2E-07
Carhart	3.4E-08	0.00014
FF5	1.1E-09	4.3E-06
FF6	5.9E-08	0.00024
FrazPed	1.6E-07	0.00066
SY	5.3E-09	2.2E-05
AFIM	2.3E-10	9.5E-07
HXZ	1.5E-07	0.0006
CZZ	1.9E-08	7.9E-05

The table reports the results of the Bayesian model scan of 12 factors in German stock returns. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Panel C reports the posterior probabilities of the 9 candidate models examined previously.

Table C.3. Netherlands Model Scan 12 Factors

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
Model			
1	0.15855	649.263	
2	0.05679	232.567	1.02666
3	0.04378	179.285	1.28686
4	0.029	118.748	1.69883
5	0.02622	107.365	1.79961
6	0.02573	105.358	1.81848

Panel B: **Factors**				
1	Market	MOM	BAB	
2	Market	BAB	PERF	
3	Market	MOM	BAB	PERF
4	Market	HML	MOM	BAB
5	Market	MOM	CMA	BAB
6	Market	MOM	BAB	RMW_{ROE}

Panel C:	Posterior Probability	Posterior/Prior
FF3	1.4E-05	0.05574
Carhart	0.00038	1.54794
FF5	5.8E-07	0.00238
FF6	1.5E-05	0.06219
FrazPed	0.02137	87.4901
SY	5.1E-05	0.20834
AFIM	7.7E-06	0.03155
HXZ	1E-05	0.0427
CZZ	0.00011	0.44335

The table reports the results of the Bayesian model scan of 12 factors in Dutch stock returns. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Panel C reports the posterior probabilities of the 9 candidate models examined previously.

Table C.4. Italy Model Scan 12 Factors

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
Model			
1	0.16565	678.32	
2	0.08741	357.924	0.6393
3	0.03865	158.264	1.45536
4	0.03459	141.647	1.56628
5	0.03263	133.629	1.62455
6	0.03184	130.402	1.649

Panel B: **Factors**						
1	Market	MOM	BAB	HML_{M}	RMW_{ROE}	
2	MOM	BAB	HML_{M}	RMW_{ROE}		
3	Market	MOM	HML_{M}	RMW_{ROE}		
4	Market	MOM	RMW_{OP}	BAB	HML_{M}	
5	Market	MOM	BAB	HML_{M}	PERF	RMW_{ROE}
6	Market	SMB	MOM	BAB	HML_{M}	RMW_{ROE}

Panel C:	Posterior Probability	Posterior/Prior
FF3	3.06E-11	1.3E-07
Carhart	1.28E-08	5.3E-05
FF5	2.50E-06	0.01023
FF6	6.13E-06	0.0251
FrazPed	1.81E-08	7.4E-05
SY	9.72E-09	4E-05
AFIM	1.08E-08	4.4E-05
HXZ	2.81E-05	0.11524
CZZ	1.46E-08	6E-05

The table reports the results of the Bayesian model scan of 12 factors in Italian stock returns. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Panel C reports the posterior probabilities of the 9 candidate models examined previously.

Table C.5. Spain Model Scan 12 Factors

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
Model			
1	0.05516	225.897	
2	0.03183	130.337	0.54996
3	0.03019	123.634	0.60276
4	0.02738	112.131	0.70042
5	0.02729	111.765	0.70368
6	0.01781	72.9448	1.13038

Panel B: **Factors**					
1	Market	MOM	RMW_{OP}	BAB	HML_{M}
2	Market	MOM	HML_{M}	RMW_{ROE}	
3	Market	MOM	RMW_{OP}	HML_{M}	
4	MOM	RMW_{OP}	HML_{M}	RMW_{ROE}	
5	MOM	RMW_{OP}	HML_{M}		
6	Market	MOM	HML_{M}		

Panel C:	Posterior Probability	Posterior/Prior
FF3	1.8E-05	0.07574
Carhart	0.00036	1.45477
FF5	5.2E-06	0.02131
FF6	3.3E-05	0.13441
FrazPed	0.00149	6.08788
SY	6.7E-05	0.27577
AFIM	3.1E-05	0.1281
HXZ	0.00041	1.6828
CZZ	0.00036	1.45566

The table reports the results of the Bayesian model scan of 12 factors in Spanish stock returns. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Panel C reports the posterior probabilities of the 9 candidate models examined previously.

C.III Summary Statistics of Posterior Distributions in Top Models Remaining Markets

Table C.6. France Summary of the Posterior Distribution of the Best Model Risk Factors

Panel A: **Premiums**	Mean	Std Dev	Median	2.50%	97.50%
Market	0.594	0.298	0.597	0.006	1.171
BAB	1.532	0.271	1.529	1.002	2.064
HML_M	0.133	0.236	0.134	-0.329	0.604
PERF	0.644	0.194	0.643	0.259	1.027

Panel B: **SDF Coeffs**	Mean	Std Dev	Median	2.50%	97.50%
Market	-3.518	1.128	-3.518	-5.738	-1.328
BAB	-5.837	1.244	-5.812	-8.346	-3.440
HML_M	-6.101	1.874	-6.099	-9.804	-2.507
PERF	-10.022	2.317	-9.992	-14.496	-5.569

Panel C: Summary Statistics of Candidate Stochastic Discount Factor Models

	Std Dev	Minimum	Maximum	Prop y<0	Correlation
Best	0.423	-0.785	2.322	1.156	0.000
CAPM	0.105	0.618	1.419	0.000	0.252
FF3	0.118	0.452	1.548	0.000	0.258
FF5	0.237	0.088	1.951	0.000	0.472
FF6	0.245	-0.109	1.867	0.289	0.462
AFIM	0.363	-0.335	2.834	0.870	0.475
SY	0.261	0.301	2.326	0.000	0.655

The table reports the summary statistics of the posterior distribution of the factors in the best model from the Bayesian model scan of 12 factors in French stock returns. Panel A reports the summary statistics of the posterior distribution of the factor premiums (%), and Panel B reports the summary statistics of the posterior distribution of the stochastic discount factor coefficients. The summary statistics include the mean, standard deviation (Std Dev), median, 2.5% and 97.5% percentiles using 10,000 simulation draws. Panel C reports the summary statistics of the fitted stochastic discount factor values of the best model from the Bayesian model scan, and a set of traditional factor models. The fitted values are estimated using the posterior mean of the stochastic discount factor coefficients. The summary statistics include the standard deviation (Std Dev), the minimum, maximum, and the proportion (Prop y<0) of fitted values that are below zero. The final column of Panel C is the correlations between fitted values of the traditional factor models, and the best model.

Table C.7. Germany Summary of the Posterior Distribution of the Best Model Risk Factors

Panel A: **Premiums**	Mean	Std Dev	Median	2.50%	97.50%
Market	0.463	0.315	0.467	-0.167	1.070
MOM	0.966	0.263	0.963	0.455	1.478
BAB	0.450	0.241	0.451	-0.026	0.921
HML_M	0.180	0.214	0.180	-0.234	0.604
RMW_{CB}	0.363	0.123	0.364	0.120	0.599

Panel B:	Mean	Std Dev	Median	2.50%	97.50%
SDF Coeffs					
Market	-2.643	1.097	-2.641	-4.779	-0.510
MOM	-8.863	1.646	-8.815	-12.179	-5.707
BAB	-1.930	1.319	-1.932	-4.564	0.658
HML_M	-7.600	1.936	-7.566	-11.482	-3.900
RMW_{CB}	-8.085	2.642	-8.088	-13.261	-2.947

Panel C: Summary Statistics of Candidate Stochastic Discount Factor Models

	Std Dev	Minimum	Maximum	Prop y<0	Correlation
Best	0.383	-0.566	2.287	1.445	0.000
CAPM	0.079	0.737	1.289	0.000	0.278
FF3	0.188	0.159	1.911	0.000	0.229
FF5	0.383	-0.648	2.308	1.445	0.808
FF6	0.245	-0.109	1.867	0.289	0.749
AFIM	0.363	-0.335	2.834	0.870	0.785
SY	0.261	0.301	2.326	0.000	0.582

The table reports the summary statistics of the posterior distribution of the factors in the best model from the Bayesian model scan of 12 factors in German stock returns. Panel A reports the summary statistics of the posterior distribution of the factor premiums (%), and Panel B reports the summary statistics of the posterior distribution of the stochastic discount factor coefficients. The summary statistics include the mean, standard deviation (Std Dev), median, 2.5% and 97.5% percentiles using 10,000 simulation draws. Panel C reports the summary statistics of the fitted stochastic discount factor values of the best model from the Bayesian model scan, and a set of traditional factor models. The fitted values are estimated using the posterior mean of the stochastic discount factor coefficients. The summary statistics include the standard deviation (Std Dev), the minimum, maximum, and the proportion (Prop y<0) of fitted values that are below zero. The final column of Panel C is the correlations between fitted values of the traditional factor models, and the best model.

Table C.8. Netherlands Summary of the Posterior Distribution of the Best Model Risk Factors

Panel A: **Premiums**	Mean	Std Dev	Median	2.50%	97.50%
Market	0.688	0.328	0.690	0.047	1.337
MOM	0.782	0.310	0.781	0.168	1.393
BAB	0.846	0.285	0.846	0.289	1.403

Panel B: **SDF Coeffs**	Mean	Std Dev	Median	2.50%	97.50%
Market	-2.990	0.963	-2.986	-4.889	-1.137
MOM	-2.953	1.009	-2.946	-4.961	-1.004
BAB	-3.078	1.059	-3.062	-5.207	-0.998

Panel C: Summary Statistics of Candidate Stochastic Discount Factor Models

	Std Dev	Minimum	Maximum	Prop y<0	Correlation
Best	0.260	-0.017	1.758	0.289	0.000
CAPM	0.109	0.679	1.516	0.000	0.427
FF3	0.124	0.625	1.655	0.000	0.358
FF5	0.232	0.072	1.878	0.000	0.708
FF6	0.238	-0.098	1.876	0.289	0.689
AFIM	0.233	0.003	1.940	0.000	0.697
SY	0.193	0.380	1.713	0.000	0.500

The table reports the summary statistics of the posterior distribution of the factors in the best model from the Bayesian model scan of 12 factors in Dutch stock returns. Panel A reports the summary statistics of the posterior distribution of the factor premiums (%), and Panel B reports the summary statistics of the posterior distribution of the stochastic discount factor coefficients. The summary statistics include the mean, standard deviation (Std Dev), median, 2.5% and 97.5% percentiles using 10,000 simulation draws. Panel C reports the summary statistics of the fitted stochastic discount factor values of the best model from the Bayesian model scan, and a set of traditional factor models. The fitted values are estimated using the posterior mean of the stochastic discount factor coefficients. The summary statistics include the standard deviation (Std Dev), the minimum, maximum, and the proportion (Prop y<0) of fitted values that are below zero. The final column of Panel C is the correlations between fitted values of the traditional factor models, and the best model.

Table C.9. Italy Summary of the Posterior Distribution of the Best Model Risk Factors

Panel A: **Premiums**	Mean	Std Dev	Median	2.50%	97.50%
Market	0.391	0.369	0.396	-0.347	1.103
MOM	0.839	0.259	0.837	0.334	1.344
BAB	0.653	0.223	0.654	0.215	1.087
HML_M	0.050	0.248	0.052	-0.431	0.541
RMW_{ROE}	0.648	0.191	0.647	0.273	1.020

Panel B: **SDF Coeffs**	Mean	Std Dev	Median	2.50%	97.50%
Market	-2.217	0.952	-2.217	-4.052	-0.339
MOM	-6.504	1.693	-6.477	-9.854	-3.257
BAB	-3.796	1.430	-3.798	-6.675	-1.053
HML_M	-10.636	2.258	-10.593	-15.186	-6.293
RMW_{ROE}	-12.378	2.542	-12.364	-17.372	-7.456

Panel C: Summary Statistics of Candidate Stochastic Discount Factor Models

	Std Dev	Minimum	Maximum	Prop y<0	Correlation
Best	0.412	-0.282	2.325	2.023	0.000
CAPM	0.057	0.800	1.196	0.000	0.278
FF3	0.081	0.638	1.290	0.000	0.229
FF5	0.328	-0.174	2.238	0.867	0.808
FF6	0.348	-0.249	1.936	1.156	0.749
AFIM	0.363	-0.278	2.249	0.867	0.785
SY	0.261	0.301	2.326	0.000	0.582

The table reports the summary statistics of the posterior distribution of the factors in the best model from the Bayesian model scan of 12 factors in Italian stock returns. Panel A reports the summary statistics of the posterior distribution of the factor premiums (%), and Panel B reports the summary statistics of the posterior distribution of the stochastic discount factor coefficients. The summary statistics include the mean, standard deviation (Std Dev), median, 2.5% and 97.5% percentiles using 10,000 simulation draws. Panel C reports the summary statistics of the fitted stochastic discount factor values of the best model from the Bayesian model scan, and a set of traditional factor models. The fitted values are estimated using the posterior mean of the stochastic discount factor coefficients. The summary statistics include the standard deviation (Std Dev), the minimum, maximum, and the proportion (Prop y<0) of fitted values that are below zero. The final column of Panel C is the correlations between fitted values of the traditional factor models, and the best model.

Table C.10. Spain Summary of the Posterior Distribution of the Best Model Risk Factors

Panel A: **Premiums**	Mean	Std Dev	Median	2.50%	97.50%
Market	0.473	0.354	0.478	-0.235	1.157
MOM	0.664	0.289	0.662	0.100	1.227
RMW_{OP}	0.421	0.172	0.421	0.081	0.755
BAB	0.597	0.283	0.595	0.043	1.153
HML_{M}	0.221	0.224	0.223	-0.224	0.654

Panel B:	Mean	Std Dev	Median	2.50%	97.50%
SDF Coeffs					
Market	-2.365	0.972	-2.352	-4.243	-0.427
MOM	-4.357	1.456	-4.339	-7.239	-1.522
RMW_{OP}	-4.800	1.872	-4.788	-8.553	-1.130
BAB	-2.146	1.228	-2.125	-4.565	0.213
HML_{M}	-4.922	1.746	-4.941	-8.340	-1.505

Panel C: Summary Statistics of Candidate Stochastic Discount Factor Models

	Std Dev	Minimum	Maximum	Prop y<0	Correlation
Best	0.287	-0.493	1.949	0.597	0.000
CAPM	0.074	0.695	1.274	0.000	0.427
FF3	0.117	0.621	1.470	0.000	0.358
FF5	0.232	0.072	1.878	0.000	0.708
FF6	0.238	-0.098	1.876	0.289	0.689
AFIM	0.233	0.003	1.940	0.000	0.697
SY	0.193	0.380	1.713	0.000	0.500

The table reports the summary statistics of the posterior distribution of the factors in the best model from the Bayesian model scan of 12 factors in Spain stock returns. Panel A reports the summary statistics of the posterior distribution of the factor premiums (%), and Panel B reports the summary statistics of the posterior distribution of the stochastic discount factor coefficients. The summary statistics include the mean, standard deviation (Std Dev), median, 2.5% and 97.5% percentiles using 10,000 simulation draws. Panel C reports the summary statistics of the fitted stochastic discount factor values of the best model from the Bayesian model scan, and a set of traditional factor models. The fitted values are estimated using the posterior mean of the stochastic discount factor coefficients. The summary statistics include the standard deviation (Std Dev), the minimum, maximum, and the proportion (Prop y<0) of fitted values that are below zero. The final column of Panel C is the correlations between fitted values of the traditional factor models, and the best model.

C.IV Bayesian Model Scan Assuming Multivariate t for Remaining Markets

Table C.11. France Model Scan of 12 Factors assuming Multivariate-t Factor Distribution

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
1	0.0809	331.403	
2	0.0646	264.723	0.225
3	0.0426	174.552	0.641
4	0.0379	155.295	0.758
5	0.0354	145.101	0.826
6	0.0303	124.224	0.981

Panel B: **Factors**						
1	Market	MOM	BAB	HML_M	PERF	
2	Market	BAB	HML_M	PERF		
3	Market	MOM	BAB	HML_M	PERF	RMW_{CB}
4	Market	HML	BAB	PERF		
5	Market	BAB	PERF			
6	Market	SMB	BAB	HML_M	PERF	

The table reports the results of the Bayesian model scan of 12 factors in French stock returns when the joint distribution of factors is assumed to follow a multivariate t distribution. The sample period is June 1991 and December 2022. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Table C.12. Germany Model Scan of 12 Factors assuming Multivariate-t Factor Distribution

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
1	0.0898	367.551	
2	0.0806	329.912	0.10804
3	0.0457	187.072	0.67537
4	0.0346	141.634	0.95361
5	0.0276	113.004	1.17944
6	0.0227	92.9712	1.37457

Panel B: **Factors**						
1	Market	SMB	MOM	RMW_{OP}	HML_{M}	RMW_{CB}
2	Market	SMB	HML	MOM	RMW_{OP}	RMW_{CB}
3	Market	SMB	MOM	RMW_{OP}	HML_{M}	PERF
4	Market	SMB	MOM	RMW_{OP}	HML_{M}	
5	Market	HML	MOM	RMW_{OP}	RMW_{CB}	
6	Market	SMB	HML	MOM	RMW_{OP}	

The table reports the results of the Bayesian model scan of 12 factors in German stock returns when the joint distribution of factors is assumed to follow a multivariate t distribution. The sample period is June 1991 and December 2022. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Table C.13. Netherlands Model Scan of 12 Factors assuming Multivariate-t Factor Distribution

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
1	0.239	978.838	
2	0.1055	432.062	0.8178
3	0.0483	197.687	1.59968
4	0.0469	191.972	1.62902
5	0.038	155.483	1.83983
6	0.0306	125.242	2.05612

Panel B: **Factors**				
1	Market	MOM	BAB	
2	Market	MOM	BAB	PERF
3	Market	MOM	RMW_{OP}	BAB
4	Market	MOM	CMA	BAB
5	Market	MOM	BAB	RMW_{CB}
6	Market	BAB	PERF	

The table reports the results of the Bayesian model scan of 12 factors in Dutch stock returns when the joint distribution of factors is assumed to follow a multivariate t distribution. The sample period is June 1991 and December 2022. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Table C.14. Italy Model Scan of 12 Factors assuming Multivariate-t Factor Distribution

Panel A:

Top Models	Posterior Probability	Posterior/Prior	ML
1	0.1324	542.01	
2	0.0682	279.428	0.66254
3	0.0569	232.823	0.84501
4	0.0555	227.292	0.86905
5	0.0324	132.852	1.40605
6	0.0290	118.81	1.51776

Panel B:

Factors							
1	Market	MOM	RMW_{OP}	CMA	HML_M		
2	Market	MOM	RMW_{OP}	CMA	HML_M	MGMT	
3	Market	MOM	RMW_{OP}	CMA	BAB	HML_M	
4	Market	MOM	RMW_{OP}	CMA	HML_M	PERF	
5	Market	MOM	RMW_{OP}	HML_M			
6	Market	MOM	RMW_{OP}	CMA	BAB	HML_M	MGMT

The table reports the results of the Bayesian model scan of 12 factors in Italian stock returns when the joint distribution of factors is assumed to follow a multivariate t distribution. The sample period is June 1991 and December 2022. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Table C.15. Spain Model Scan of 12 Factors assuming Multivariate-t Factor Distribution

Panel A: **Top Models**	Posterior Probability	Posterior/Prior	ML
1	0.0701	287.044	
2	0.0690	282.571	0.01571
3	0.0636	260.572	0.09676
4	0.0500	204.94	0.33692
5	0.0213	87.3219	1.19004
6	0.0211	86.5624	1.19877

Panel B: **Factors**					
1	Market	MOM	RMW_{OP}	BAB	HML_{M}
2	MOM	RMW_{OP}	HML_{M}		
3	Market	MOM	RMW_{OP}	HML_{M}	
4	MOM	RMW_{OP}	BAB	HML_{M}	
5	Market	MOM	HML_{M}	PERF	
6	Market	MOM	HML_{M}		

The table reports the results of the Bayesian model scan of 12 factors in Spanish stock returns when the joint distribution of factors is assumed to follow a multivariate t distribution. The sample period is June 1991 and December 2022. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

C.V Gaussian v Student t Overall Performance Remaining Markets

Table C.16. France Log-Marginal Likelihoods of the Top Performing Models

Panel A: Top Performing Gaussian Models						$m\ (f_{1:T} \vert M_i)$
Market	BAB	HML_M	PERF			8035.21
Market	SMB	BAB	HML_M	PERF		8032.94
Market	RMW_{OP}	BAB	HML_M	PERF		8033.03
Market	SMB	BAB	HML_M	PERF	RMW_{CB}	8033.54
Market	BAB	PERF	PERF	RMW_{CB}		8033.25
Market	HML	BAB	PERF			8031.19
Panel A: Top Performing Student-t Models						
Market	MOM	BAB	HML_M	PERF		8655.80
Market	BAB	HML_M	PERF			8655.57
Market	MOM	BAB	HML_M	PERF	RMW_{CB}	8655.16
Market	HML	BAB	PERF			8655.04
Market	BAB	PERF				8654.97
Market	SMB	BAB	HML_M	PERF		8654.82

The table presents the log marginal likelihoods for the best-performing France models identified through both analyses. Panel A details models assuming a Gaussian distribution, while Panel B focuses on models with a student-t distribution.

Table C.17. Germany Log-Marginal Likelihoods of the Top Performing Models

						$m\ (f_{1:T} \mid M_i)$
Panel A: Top Performing Gaussian Models						
Market	MOM	BAB	HML_M	RMW_{CB}		8724.50
Market	SMB	MOM	BAB	HML_M	PERF	8723.36
Market	MOM	CMA	BAB	HML_M	RMW_{CB}	8722.57
Market	MOM	RMW_{OP}	BAB	HML_M	RMW_{CB}	8727.26
Market	HML	MOM	BAB	RMW_{CB}		8726.24
Market	MOM	BAB	HML_M			8725.02
Panel A: Top Performing Student-t Models						
Market	SMB	MOM	RMW_{OP}	HML_M	RMW_{CB}	9110.87
Market	SMB	HML	MOM	RMW_{OP}	RMW_{CB}	9110.68
Market	SMB	MOM	RMW_{OP}	HML_M	PERF	9110.45
Market	SMB	MOM	RMW_{OP}	HML_M		9110.17
Market	HML	MOM	RMW_{OP}	RMW_{CB}		9109.06
Market	SMB	HML	MOM	RMW_{OP}		9109.50

The table presents the log marginal likelihoods for the best-performing Germany models identified through both analyses. Panel A details top models assuming a Gaussian distribution, while Panel B details top models with a student-t distribution.

Table C.18. Netherlands Log-Marginal Likelihoods of the Top Performing Models

				$m\ (f_{1:T} \mid M_i)$
Panel A: Top Performing Gaussian Models				
Market	MOM	BAB		7771.44
Market	BAB	PERF		7769.08
Market	MOM	BAB	PERF	7769.21
Market	HML	MOM	BAB	7769.49
Market	MOM	CMA	BAB	7769.05
Market	MOM	BAB	RMW_{ROE}	7766.97
Panel A: Top Performing Student-t Models				
Market	MOM	BAB		7290.28
Market	MOM	BAB	PERF	7289.46
Market	MOM	RMW_{OP}	BAB	7288.68
Market	MOM	CMA	BAB	7288.65
Market	MOM	BAB	RMW_{CB}	7288.44
Market	BAB	PERF		7288.22

The table presents the log marginal likelihoods for the best-performing Dutch models identified through both analyses.

Panel A details top models assuming a Gaussian distribution, while Panel B details top models with a student-t distribution.

Table C.19. Italy Log-Marginal Likelihoods of the Top Performing Models

							$m\,(f_{1:T}\,|\,M_i)$
<u>Panel A: Top Performing Gaussian Models</u>							
Market	MOM	BAB	HML_M	RMW_{ROE}			8419.54
MOM	BAB	HML_M	RMW_{ROE}				8418.87
Market	MOM	HML_M	RMW_{ROE}				8417.22
Market	MOM	RMW_{OP}	BAB	HML_M			8419.46
Market	MOM	BAB	HML_M	PERF	RMW_{ROE}		8417.38
Market	SMB	MOM	BAB	HML_M	RMW_{ROE}		8417.08
<u>Panel A: Top Performing Student-t Models</u>							
Market	MOM	RMW_{OP}	CMA	HML_M			7841.21
Market	MOM	RMW_{OP}	CMA	HML_M	MGMT		7840.54
Market	MOM	RMW_{OP}	CMA	BAB	HML_M		7840.36
Market	MOM	RMW_{OP}	CMA	HML_M	PERF		7840.34
Market	MOM	RMW_{OP}	HML_M				7839.80
Market	MOM	RMW_{OP}	CMA	BAB	HML_M	MGMT	7839.69

The table presents the log marginal likelihoods for the best-performing Italian models identified through both analyses.

Panel A details top models assuming a Gaussian distribution, while Panel B details top models with a student-t distribution.

Table C.20. Spain Log-Marginal Likelihoods of the Top Performing Models

Panel A: Top Performing Gaussian Models

					$m\ (f_{1:T} \mid M_i)$
Market	MOM	RMW_{OP}	BAB	HML_{M}	7829.71
Market	MOM	HML_{M}	RMW_{ROE}		7829.06
Market	MOM	RMW_{OP}	HML_{M}		7828.51
MOM	RMW_{OP}	HML_{M}	RMW_{ROE}		7828.94
MOM	RMW_{OP}	HML_{M}			7827.49
Market	MOM	HML_{M}			7826.92

Panel A: Top Performing Student-t Models

Market	MOM	RMW_{OP}	BAB	HML_{M}	7509.73
MOM	RMW_{OP}	HML_{M}			7509.70
Market	MOM	RMW_{OP}	HML_{M}		7509.63
MOM	RMW_{OP}	BAB	HML_{M}		7509.35
Market	MOM	HML_{M}	PERF		7508.54
Market	MOM	HML_{M}			7508.39

The table presents the log marginal likelihoods for the best-performing Spanish models identified through both analyses. Panel A details top models assuming a Gaussian distribution, while Panel B details top models with a student-t distribution.

C.VI Bayesian Model Scan with Time Series Efficient Factors Remaining Markets

Table C.21. France Model Scan 24 Factors

Panel A: **Top Models** Model	Posterior Probability	ML
1	0.01405	
2	0.01102	0.24285
3	0.01063	0.27908
4	0.00848	0.50517
5	0.00799	0.5642
6	0.00712	0.67972

Panel B: **Factors**								
1	Market	BAB	HML_M	PERF	$Market^{ef}$	HML^{ef}		
2	Market	BAB	HML_M	PERF	$Market^{ef}$	HML^{ef}	RMW_{CB}^{ef}	
3	Market	BAB	HML_M	PERF	$Market^{ef}$	CMA^{ef}	$PERF^{ef}$	
4	Market	BAB	HML_M	PERF	RMW_{CB}	$Market^{ef}$	HML_M^{ef}	$PERF^{ef}$
5	Market	BAB	HML_M	PERF	$Market^{ef}$	HML^{ef}	$PERF^{ef}$	RMW_{CB}^{ef}
6	Market	BAB	HML_M	PERF	$Market^{ef}$	SMB^{ef}	$PERF^{ef}$	

The table reports the results of the Bayesian model scan of 24 factors in French stock returns. This set is made up of 13 factors and their efficient counterparts. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability. The ML column is the difference in the log ML of the best model and the next best models 2 to 6. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Superscript 'ef' denotes the efficient version of the factor.

Table C.22. Germany Model Scan 24 Factors

Panel A: **Top Models**	Posterior Probability	ML
1	0.01455	
2	0.01158	0.22808
3	0.00938	0.43944
4	0.00916	0.46276
5	0.00915	0.46455
6	0.00883	0.49915

Panel B: **Factors**									
1	MOM	RMW_{OP}	HML_{M}	$Market^{ef}$	BAB^{ef}	$PERF^{ef}$			
2	Market	SMB	MOM	HML_{M}	RMW_{OP}	$Market^{ef}$	BAB^{ef}	$PERF^{ef}$	
3	Market	SMB	MOM	HML_{M}	RMW_{OP}	$Market^{ef}$	$PERF^{ef}$		
4	MOM	RMW_{OP}	HML_{M}	$Market^{ef}$	BAB^{ef}	$PERF^{ef}$	RMW_{CB}^{ef}		
5	Market	SMB	MOM	HML_{M}	RMW_{OP}	$Market^{ef}$	BAB^{ef}	$PERF^{ef}$	RMW_{CB}^{ef}
6	MOM	RMW_{OP}	HML_{M}	$Market^{ef}$	BAB^{ef}	RMW_{CB}^{ef}			

The table reports the results of the Bayesian model scan of 24 factors in German stock returns. This set is made up of 13 factors and their efficient counterparts. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability. The ML column is the difference in the log ML of the best model and the next best models 2 to 6. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Superscript 'ef' denotes the efficient version of the factor.

Table C.23. Netherlands Model Scan 24 Factors

Panel A: **Top Models** Model	Posterior Probability	ML
1	0.01054	
2	0.0105	0.0038
3	0.00985	0.06856
4	0.0068	0.43816
5	0.00657	0.47308
6	0.00633	0.51087

Panel B: **Factors**						
1	Market	MOM	BAB	$Market^{ef}$	SMB^{ef}	$PERF^{ef}$
2	Market	MOM	BAB	$Market^{ef}$	$PERF^{ef}$	
3	Market	MOM	BAB	$Market^{ef}$	BAB^{ef}	$PERF^{ef}$
4	Market	MOM	BAB	$Market^{ef}$		
5	Market	BAB	PERF	$Market^{ef}$	SMB^{ef}	
6	Market	BAB	PERF	$Market^{ef}$	SMB^{ef}	MOM^{ef}

The table reports the results of the Bayesian model scan of 24 factors in Dutch stock returns. This set is made up of 13 factors and their efficient counterparts. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability. The ML column is the difference in the log ML of the best model and the next best models 2 to 6. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Superscript 'ef' denotes the efficient version of the factor.

Table C.24. Italy Model Scan 24 Factors

Panel A: **Top Models** Model	Posterior Probability	ML
1	0.01537	
2	0.01489	0.03166
3	0.01258	0.19995
4	0.01134	0.30392
5	0.00658	0.84842
6	0.00621	0.90526

Panel B: **Factors**								
1	Market	MOM	HML_M	RMW_{ROE}	BAB^{ef}	RMW_{ROE}^{ef}		
2	Market	MOM	BAB	HML_M	RMW_{ROE}	BAB^{ef}		
3	Market	MOM	BAB	HML_M	RMW_{ROE}	BAB^{ef}	RMW_{ROE}^{ef}	
4	Market	MOM	BAB	HML_M	RMW_{OP}	BAB^{ef}	RMW_{ROE}^{ef}	
5	Market	MOM	BAB	HML_M	RMW_{OP}	HML^{ef}	BAB^{ef}	
6	Market	MOM	BAB	HML_M	RMW_{OP}	HML^{ef}	BAB^{ef}	RMW_{ROE}^{ef}

The table reports the results of the Bayesian model scan of 24 factors in Italian stock returns. This set is made up of 13 factors and their efficient counterparts. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability. The ML column is the difference in the log ML of the best model and the next best models 2 to 6. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Superscript 'ef' denotes the efficient version of the factor.

Table C.25. Spain Model Scan 24 Factors

Panel A: **Top Models**	Posterior Probability	ML
Model		
1	0.00502	
2	0.00442	0.12671
3	0.00334	0.4064
4	0.00311	0.47911
5	0.0025	0.69651
6	0.00235	0.76047

Panel B: **Factors**							
1	MOM	RMW_{OP}	HML_{M}	$Market^{ef}$	BAB^{ef}	HML_{M}^{ef}	
2	MOM	RMW_{OP}	HML_{M}	$Market^{ef}$	RMW_{OP}^{ef}	BAB^{ef}	
3	MOM	RMW_{OP}	HML_{M}	$Market^{ef}$	BAB^{ef}		
4	MOM	RMW_{OP}	HML_{M}	BAB^{ef}	HML^{ef}		
5	Market	MOM	RMW_{OP}	HML_{M}	$Market^{ef}$	BAB^{ef}	HML^{ef}
6	MOM	RMW_{OP}	HML_{M}	BAB^{ef}			

The table reports the results of the Bayesian model scan of 24 factors in Spanish stock returns. This set is made up of 13 factors and their efficient counterparts. The sample period is June 1991 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability. The ML column is the difference in the log ML of the best model and the next best models 2 to 6. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan. Superscript 'ef' denotes the efficient version of the factor.

C.VII Prior Computation

To derive the priors for a given model, Chib, Zeng and Zhao (2020) start by setting a prior on a model where all factors are pricing factors, referred to as model M_1. The variance parameter (covariance matrix) $\Omega_{\breve{f},1}$ in this model is given an inverse Wishart prior:

$$\pi(n_1|\mathrm{M}_1) \propto \left|\Omega_{\breve{f},1}\right|^{-\left(\frac{\mathrm{v+k+1}}{2}\right)} \exp\left(-\frac{1}{2} tr\left(Q\Omega_{\breve{f},1}^{-1}\right)\right) \qquad \text{(C.1)}$$

where v (degrees of freedom), k (number of non-risk factors) and Ω (scale matrix) are parameters.

For any other model M_j, the parameters $n_j = (\Gamma_j, \Omega_{\breve{f},\mathrm{j}}, \Omega_{\mathrm{w},\breve{f},\mathrm{j}})$ are one-to-one functions of n_1. The Jacobian of this transformation is:

$$\left|\frac{\delta n_1}{\delta n_j}\right| \propto \left|\Omega_{\mathrm{x,j}}\right|^{\mathrm{k}_{f^*,\mathrm{j}}} \qquad \text{(C.2)}$$

Based on this, Chib et al. derive the prior for n_j (for $j > 1$) from the prior of n_1 using the change of variable formula:

$$\pi\left(n_\mathrm{j}\middle|\mathrm{M_j}\right) \propto \pi(n_1|\mathrm{M}_1)\left|\frac{\delta n_1}{\delta n_j}\right| \qquad \text{(C.3)}$$

This ensures that all priors are derived from the single distribution below, making them consistent across different models.

$$\pi\left(n_\mathrm{j}\middle|\mathrm{M_j}\right) \propto \left|\Omega_{\mathrm{x,j}}\right|^{-\frac{(\mathrm{v}-k_{f^*,j}+k_{\breve{f},j}+1)}{2}}\left|\Omega_{\mathrm{w,x,j}}\right|^{-\frac{(\mathrm{v}-k_{f^*,j}+k_{\breve{f},j}+1)}{2}} \exp\left(-\frac{1}{2} tr\left(Q_j\Omega_{\breve{f},j}^{-1}\right)\right) \qquad \text{(C.4)}$$

By this method, Chib et al. ensure that all priors are derived from a single distribution, maintaining uniformity and comparability across models. This approach addresses the limitations in the Barillas and Shanken method, ensuring valid and consistent marginal likelihoods for model comparison.

Chapter 5 Conclusions

Abstract

Despite a substantial body of academic research dedicated to assessing the performance of asset pricing models, there is a notable scarcity of studies focusing on local markets. This thesis aims to fill gaps in the asset pricing literature by evaluating the performance of linear factor models across major European markets, using different model comparison frameworks. Additionally, it explores opportunities to improve these models' performance by analysing the mean-variance efficiency of the factors included in these models.

5.1 Summary of Findings

Historically, asset pricing tests have predominantly focused on U.S. data, a trend seen in leading academic journals, largely due to data availability. However, recent trends show a shift towards incorporating international data, with a growing emphasis on a global perspective in financial research. This change reflects the increasing interconnectedness of global financial markets, highlighting the need for models that can explain cross-sectional variation across worldwide markets. As a result, a notable trend in recent asset pricing studies is the aggregation of international data into large samples, rather than examining specific countries individually. This approach, while broadening the scope, has its limitations. Empirical evidence, as noted in Section 1.2, shows that global factor models, which combine data from various regions, often underperform models tailored to individual countries. This suggests that while aggregating international data is important for understanding global asset pricing, these models may be less effective if they do not consider unique local influences.

This thesis presents an exploration and enhancement of factor pricing models within developed European markets. Through a detailed analysis spanning three empirical chapters, this work has advanced the literature on model performance in European markets. This research focuses on three main questions: First, is there a single optimal model for European returns, or do optimal models vary by country? Second, does the same model, or a set of models, outperform in both the classical framework and the Bayesian framework for model comparison? Lastly, can the performance of asset pricing models in a European setting be improved through the inclusion of time-series efficient factors?

In Chapter Two, the asset pricing theories that form the foundation for the empirical analysis in this thesis are outlined. The Stochastic Discount Factor (SDF) and its evolution over time are discussed, highlighting different specifications and the role of linear factor models as proxies for the SDF. The relative performance of nine competing neo-classical asset pricing models is then evaluated in a set of developed European stock markets between June 1991 and December 2022. Asymptotically valid tests of model comparison, developed by Barillas, Kan, Robotti, and Shanken (2020), are conducted, where the extent of model mispricing is gauged by the squared Sharpe ratio improvement measure of Barillas and Shanken (2017). This study is the first to comprehensively analyse a broad range of models over an extensive time period within the majority of the selected markets.

Through the pairwise and multiple model comparisons, it was observed that both the Fama and French (2018) six-factor model and the Asness et al. (2015) model—which substitutes the HML factor in the FF6 model with a more current version of the value factor (HML_M)—demonstrate robust performance throughout the sample. These findings align with those of Barillas et al. (2020), Hanauer (2020), and Fletcher (2019), who also observe the outperformance of these models in other markets. However, the leading model does vary across samples; for instance, the Frazzini and Pederson (2014) model stands out as the superior model in two out of six scenarios. The results suggest that elements like beta timing, momentum, investment, and profitability play significant roles in determining stock returns during this period. Consequently, models that neglect these factors tend to underperform compared to those that incorporate them across the sample. These findings are supported by simulation evidence. Similar findings are noted by Pukthuanthong et al. (2023), where no single model outperforms all others across the sample.

In Chapter Three, using the framework of Ehsani and Linnainmaa (2022), it is shown that multiple investment factors in the cross-country dataset are unconditionally minimum-variance inefficient: factor returns are positively autocorrelated while risk, conditional on past returns, remains constant. Ferson and Siegel's (2001) general framework of conditioning information is used to transform an autocorrelated standard investment factor into a "time-series efficient" factor. Comparing the efficient factors to the original factors reveals that the efficient factor transformation allows for a significant increase in the Sharpe ratio for several factors across markets, such as the BAB factor in the German sample, SMB in the Dutch sample, and CMA in the Italian sample, along with a number of U.K. factors. The observation of mean-variance inefficiency among key investment factors highlights their potential in improving model performance.

The model comparison tests from Chapter Two are repeated to assess the impact of the efficient factor transformation on both the absolute and relative performance of the models. Across the sample, models with low squared Sharpe ratios in Chapter Two, such as the Fama and French (1993) and Fama and French (2015) models, experience an increase in performance. However, higher-performing models do not exhibit this performance boost, and the relative performance of models across the sample remains unaffected by the inclusion of the efficient factors.

In Chapter Four, the risk factors contained in the models under consideration are examined to determine if they truly reside in the Stochastic Discount Factor (SDF). The Bayesian frameworks of Chib, Zeng, and Zhao (2020) and Chib and Zeng (2020) are used to identify the best model from the collection of investment factors across the different markets. In all markets, the model scan returns the optimal asset pricing model, which has a similar factor composition to that identified in the classical framework. In most cases, the optimal model is more parsimonious, with redundant factors dropping out.

Similar to recent asset pricing research by Qiao, Wang, and Lam (2022), the impact of different distributional assumptions on the factor data for model testing is examined. Assuming a Student-t distribution for the factor data provides little change to the identity of the top-performing model across the markets and the general fit of these models to the data. While Qiao et al. (2022) argue for the superiority of Student-t models in global factor models, the results from the country-specific analysis show that this does not apply in a European setting. The model scanning framework is then used to examine if the time-series efficient factors from Chapter Three enhance the performance of the asset pricing models across each sample. In all cases, the top-performing asset pricing model from this model scan contains at least one time-series efficient factor while retaining all of the original factors. This analysis concludes that the model scan allows for the identification of specific factors that enhance model performance, rather than replacing all factors with their efficient counterparts when the transformation is not valuable.

This thesis significantly enhances the asset pricing literature by concentrating on a select group of key European countries, addressing a noticeable shortfall in academic exploration within European asset pricing. The limited existing work on these markets makes this in-depth analysis especially valuable, offering investors targeted insights into stock market dynamics and model enhancement opportunities specific to their markets. This focus on Europe fills a critical research gap, as these markets have traditionally received less attention in academic studies compared to those in the U.S.

Furthermore, the findings of this thesis have profound implications for practitioners, academics, and general readers interested in asset pricing and investment strategies. For practitioners, the research provides actionable insights into the performance of asset pricing models across key European markets, emphasizing the importance of model selection tailored to local conditions and the potential gains from incorporating efficient factors into investment strategies. This practical relevance is underscored by the demonstrated improvements in Sharpe ratios and risk-adjusted returns, offering valuable guidance for portfolio construction and risk management. Academically, this thesis advances the discourse on factor efficiency and model optimization, bridging theoretical advancements with empirical testing and expanding the understanding of how Bayesian and classical methodologies compare in diverse settings. It highlights the necessity of refining factor models to incorporate market-specific dynamics rather than relying on generic global factors. For the general reader, this work demystifies complex asset pricing concepts by providing a structured and coherent analysis of how different factors drive stock returns in European markets, making the research accessible and relevant to those interested in the interplay between financial theory and real-world market behaviour.

5.1 Areas of Further Research

The focus of this thesis is on a prominent set of factors; however, the number of discovered factors is constantly growing, with Harvey and Liu (2019) documenting over 400 factors published in top journals. Considering a wider set of investment factors in future work may be interesting. Chib et al. (2023) have developed a new risk factor discovery methodology that involves reducing the cross-sectional volatility of a range of factors into ‘slope factors.’ It may be possible to use this data reduction method to find a smaller set of factors that could be analysed in a local market setting; however, data availability remains a constraint, especially for less pronounced factors. Ensuring that each factor included in the model is supported by a robust theoretical framework that explains its relevance to asset pricing is crucial. This approach helps justify the inclusion of factors based on economic theory rather than solely on historical performance.

Despite a wealth of literature on large sample tests of asset pricing, there remains a distinct lack of research focused on local asset pricing model performance. This thesis sheds light on the drivers of returns at a local level within a European setting, highlighting the need for similar studies in other regions. Research in asset pricing that examines broad classifications, such as regional, developed, or developing markets, should also be tested at a local level to ensure the accuracy of aggregated results.

The concept of combining factor models also warrants exploration. Throughout this study, models with predefined factors have underperformed relative to optimal models identified through the Bayesian model scan method. Legacy models like the Fama and French (1993) three-factor model have consistently underperformed across all comparison frameworks, despite their continued recognition. Considering the possible combination of legacy models in future research endeavours is pertinent. Each factor model typically focuses on specific market behaviours or risk premiums. By amalgamating them, investors and researchers can achieve a richer comprehension and more accurate predictions of asset returns. For instance, while the two-factor model by Frazzini and Pederson (2014) excels in the comparison of French factor models within a classical framework, the model scan reveals the performance factor (PERF) from the Stambaugh and Yuan (2017) model as one of the top six performing models. This suggests that a hybrid approach, combining these models, could potentially offer a superior framework compared to each model operating in isolation. Combining factor models can reveal interactions between factors that might not be apparent when models are used separately.

Bibliography

Affleck-Graves, J. and McDonald, B., 1989. Nonnormalities and tests of asset pricing theories. *The Journal of Finance*, *44*(4), pp.889-908.

Ahmar, A.S. and Del Val, E.B., 2020. SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain. *Science of the Total Environment*, *729*, p.138883.

Ahmed, S., Bu, Z. and Tsvetanov, D., 2019. Best of the best: A comparison of factor models. *Journal of Financial and Quantitative Analysis*, 54(4), pp.1713-1758.

Alexander, S.S., 1961. Price movements in speculative markets: Trends or random walks. *Industrial Management Review (pre-1986)*, 2(2), p.7.

Andrew Karolyi, G., 2016. Home bias, an academic puzzle. *Review of Finance*, 20(6), pp.2049-2078.

Ang, A., 2014. Asset management: A systematic approach to factor investing. Oxford University Press.

Ardalan, K., 2019. Equity home bias in international finance: A place-attachment perspective. Routledge.

Arrow, K.J. and Debreu, G., 1954. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pp.265-290.

Artmann, S., Finter, P. and Kempf, A., 2012. Determinants of expected stock returns: large sample evidence from the German market. *Journal of Business Finance & Accounting*, 39(5-6), pp.758-784.

Asness, C., and A. Frazzini., 2013. "The Devil in HML's Details." *Journal of Portfolio Management*, 39, pp.49–68.

Asness, C., Frazzini, A., Israel, R. and Moskowitz, T., 2014. Fact, fiction, and momentum investing. *The Journal of Portfolio Management*, 40(5), pp.75-92.

Ball, R. and Brown, P., 2013. An empirical evaluation of accounting income numbers. In *Financial Accounting and Equity Markets* (pp. 27-46). Routledge.

Ball, R., Gerakos, J., Linnainmaa, J.T. and Nikolaev, V.V., 2015. Deflating profitability. *Journal of Financial Economics*, 117(2), pp.225-248.

Bansal, R. and Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance*, 59(4), pp.1481-1509.

Banz, R.W., 1981. The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1), pp.3-18.

Barberis, N., Shleifer, A. and Vishny, R., 1998. A model of investor sentiment. *Journal of Financial Economics*, 49(3), pp.307-343.

Barillas, F. and J. Shanken, 2017, Which alpha?, *Review of Financial Studies*, 30,1316-1388.

Barillas, F. and J. Shanken, 2018, Comparing asset pricing models, *Journal of Finance*, 73, 715-754.

Barillas, F., Kan, R., Robotti, C. and Shanken, J. 2020. Model comparison with Sharpe ratios, *Journal of Financial and Quantitative Analysis*.

Basu, S., 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, 32(3), pp.663-682.

Bethmann, D., Bransch, F., Kvasnicka, M. and Sadrieh, A., 2023. Home bias in top economics journals. Discussion Paper No. 15965, SSRN.

Boya, C.M., 2019. From efficient markets to adaptive markets: Evidence from the French stock exchange. *Research in International Business and Finance*, 49, pp.156-165.

Brealey, R.A. and Myers, S.C., 2016. Franklin Allen." Principles of Corporate Finance. *NY: McGraw-Hill/Irwin*.

Breeden, D.T. and Litzenberger, R.H., 1978. Prices of state-contingent claims implicit in option prices. *Journal of Business*, pp.621-651.

Bruna, M.G., Dang, R., Scotto, M.J. and Ammari, A., 2019. Does board gender diversity affect firm risk-taking? Evidence from the French stock market. *Journal of Management and Governance*, 23, pp.915-938.

Campbell, J.Y. and Cochrane, J.H., 2000. Explaining the poor performance of consumption-based asset pricing models. *The Journal of Finance*, 55(6), pp.2863-2878.

Campbell, J.Y., 1987. Stock returns and the term structure. *Journal of Financial Economics*, 18(2), pp.373-399.

Carhart, M.M., 1997. On persistence in mutual fund performance. *The Journal of Finance*, 52(1), pp.57-82.

Chaieb, I., Langlois, H. and Scaillet, O., 2021. Factors and risk premia in individual international stock returns. *Journal of Financial Economics*, 141(2), pp.669-692.

Chan, L.K., Hamao, Y. and Lakonishok, J., 1991. Fundamentals and stock returns in Japan. *The Journal of Finance*, 46(5), pp.1739-1764.

Chen, L., R. Novy-Marx, and L. Zhang. 2010. An alternative three-factor model, Working paper, CKGSB, University of Rochester, and Ohio State University.

Chib, S. and Greenberg, E., 2010. Additive cubic spline regression with Dirichlet process mixture errors. *Journal of Econometrics*, 156(2), pp.322-336.

Chib, S. and X. Zeng, 2020, Which factors are risk factors in asset pricing? A model scan approach, *Journal of Business and Economic Statistics,* 38, 771-783.

Chib, S., 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432), pp.1313-1321.

Chib, S., Lin, Y.C., Pukthuanthong, K. and Zeng, X., 2023. Slope factors outperform: Evidence from a large comparative study. *Available at SSRN 3966807*.

Chib, S., Zeng, X. and L. Zhao, 2020, On comparing asset pricing models, *Journal of Finance*, 75, 551-577.

Chib, S., Zhao, L. and G. Zhou, 2022, Winners from winners: A tale of risk factors, *Working Paper*, Washington University in St Louis.

Clare, A. and Thomas, S., 1995. The overreaction hypothesis and the UK stockmarket. *Journal of Business Finance and Accounting*, 22, pp.961-961.

Cochrane, J., 2005. Asset pricing: Revised edition. Princeton university press.

Cochrane, J.H., 2011. Presidential address: Discount rates. *The Journal of Finance*, 66(4), pp.1047-1108.

Cohen, R.B. and Polk, C.K., 1998. The impact of industry factors in asset-pricing tests. *Kellogg Graduate School of Management working paper*.

Cujean, J. and Hasler, M., 2017. Why does return predictability concentrate in bad times?. *The Journal of Finance*, 72(6), pp.2717-2758.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A., 1998. Investor psychology and security market under-and overreactions. *The Journal of Finance*, 53(6), pp.1839-1885.

Daniel, K., Hirshleifer, D. and Sun, L., 2020. Short-and long-horizon behavioral factors. *The review of financial studies*, 33(4), pp.1673-1736.

Davis, J.L., Fama, E.F. and French, K.R., 2000. Characteristics, covariances, and average returns: 1929 to 1997. *The Journal of Finance*, 55(1), pp.389-406.

De Bondt, W.F. and Thaler, R., 1985. Does the stock market overreact?. *The Journal of Finance*, 40(3), pp.793-805.

De Moor, L. and Sercu, P., 2013. The smallest firm effect: An international study. *Journal of International Money and Finance*, 32, pp.129-155.

Desban, M. and Lajili Jarjir, S., 2018. Corporate ownership structure, market anomalies and asset pricing. *Journal of Asset Management*, 19, pp.316-340.

Dissanaike, G., 1997. Do stock market investors overreact?. *Journal of Business Finance & Accounting*, 24(1), pp.27-50.

Doeswijk, R.Q., 1997. Contrarian investment in the Dutch stock market. *De Economist*, 145(4), pp.573-598.

Dybvig, P.H. and Ingersoll Jr, J.E., 1982. Mean-variance theory in complete markets. *Journal of Business*, pp.233-251.

Ehsani, S. and J.T. Linnainmaa, 2022, Time-series efficient factors, *Working Paper*, Tuck School of Business.

Engle, R.F. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation,” Econometrica, 50, 987-1008.

Epstein, L.G. and Zin, S.E., 1991. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99(2), pp.263-286.

Eugene, F. and French, K., 1992. The cross-section of expected stock returns. *Journal of Finance*, 47(2), pp.427-465.

Fama, E. F., and K. R. French. “A Five-Factor Asset Pricing Model.” *Journal of Financial Economics*, 116 (2015), 1–22.

Fama, E. F., and K. R. French. “Choosing Factors.” *Journal of Financial Economics*, 128 (2018), 234–252.

Fama, E. F., and K. R. French. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics*, 33(1993), 3–56.

Fama, E.F. and French, K.R., 1996. Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, 51(1), pp.55-84.

Fama, E.F. and French, K.R., 1998. Value versus growth: The international evidence. *The Journal of Finance*, 53(6), pp.1975-1999.

Fama, E.F. and French, K.R., 2012. Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105(3), pp.457-472.

Fama, E.F. and MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3), pp.607-636.

Fama, E.F., 1965. The behaviour of stock-market prices. *The Journal of Business*, 38(1), pp.34-105.

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), pp.383-417.

Fama, E.F., 2014. Two pillars of asset pricing. *American Economic Review*, 104(6), pp.1467-1485.

Fama, E.F., French, K.R., 2006. The value premium and the CAPM. *The Journal of Finance* 61, 2163–2186.

Fama, E.F., French, K.R., 2008. Dissecting anomalies. *The Journal of Finance* 63, 1653–1678.

Ferson, W.E. and Jagannathan, R., 1996. 1 Econometric evaluation of asset pricing models. *Handbook of Statistics*, *14*, pp.1-33.

Ferson, W.E. and Siegel, A.F., 2001. The efficient use of conditioning information in portfolios. *The Journal of Finance*, 56(3), pp.967-982.

Fischer, B.R. and Wermers, R., 2012. Performance evaluation and attribution of security portfolios. Academic Press.

Fletcher, J., 2001. An examination of alternative factor models in UK stock returns. *Review of Quantitative Finance and Accounting*, 16, pp.117-130.

Fletcher, J., 2018. Bayesian tests of global factor models. *Journal of Empirical Finance*, 48, pp.279-289.

Fletcher, J., 2019. Model comparison tests of linear factor models in UK stock returns. *Finance Research Letters*, 28, pp.281-291.

Forner, C. and Marhuenda, J., 2003. Contrarian and momentum strategies in the Spanish stock market. *European Financial Management*, 9(1), pp.67-88.

Foye, J., 2018. A comprehensive test of the Fama-French five-factor model in emerging markets. *Emerging Markets Review*, 37, pp.199-222.

Frazzini, A. and Pedersen, L.H., 2014. Betting against beta. *Journal of Financial Economics*, 111(1), pp.1-25.

French, K.R. and Poterba, J., 1991. Investor diversification and international equity markets.

French, K.R., Schwert, G.W. and Stambaugh, R.F., 1987. Expected stock returns and volatility. *Journal of Financial Economics*, 19(1), pp.3-29.

Gibbons, M. R.; S. A. Ross; and J. Shanken, 1989. "A Test of the Efficiency of a Given Portfolio." *Econometrica*, 57, 1121–1152.

Gitman, L.J. and Mercurio, V.A., 1982. Cost of capital techniques used by major US firms: Survey and analysis of Fortune's 1000. *Financial Management*, pp.21-29.

Glaser, M. and Weber, M., 2003. Momentum and turnover: evidence from the German stock market. *Schmalenbach Business Review*, 55, pp.108-135.

Goyal, A. and Wahal, S., 2015. Is momentum an echo?. *Journal of Financial and Quantitative Analysis*, 50(6), pp.1237-1267.

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), pp.187-243.

Griffin, J.M., 2002. Are the Fama and French factors global or country specific?. *The Review of Financial Studies*, 15(3), pp.783-803.

Grinold, R.C. and Kahn, R.N., 2000. Active portfolio management.

Gultekin, M.N. and Gultekin, N.B., 1983. Stock market seasonality: International evidence. *Journal of financial economics*, 12(4), pp.469-481.

Gupta, T. and Kelly, B., 2019. Factor momentum everywhere. *The Journal of Portfolio Management*, 45(3), pp.13-36.

Hanauer, M.X. and Windmüller, S., 2021. Enhanced momentum strategies. *Journal of Banking & Finance*, *148*, p.106712.

Hanauer, M.X., 2020. A comparison of global factor models. *Available at SSRN 3546295*.

Hansen, L.P. and Richard, S.F., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society*, pp.587-613.

Hansen, L.P. and Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, pp.1269-1286.

Harrison, J.M. and Kreps, D.M., 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3), pp.381-408.

Harry M. Markowitz, *Portfolio Selection, Efficient Diversification of Investments* (New York: John Wiley and Sons, Inc., 1959). The major elements of the theory first appeared in his article "Portfolio Selection," *The Journal of Finance*, XII (March 1952), 77–91.

Harvey, C.R. and Liu, Y., 2021. Lucky factors. *Journal of Financial Economics*, 141(2), pp.413-435.

Harvey, C.R. and Zhou, G., 1990. Bayesian inference in asset pricing tests. *Journal of Financial Economics*, 26(2), pp.221-254.

Hau, H., 2011. Global versus local asset pricing: A new test of market integration. *The Review of Financial Studies*, 24(12), pp.3891-3940.

Hennecke, P., Jessen, R. and Topalov, M., 2022. Multi-factor Models after the Financial Crisis-An empirical analysis for the German stock market.

Hodrick, R.J. and Zhang, X., 2001. Evaluating the specification errors of asset pricing models. *Journal of Financial Economics*, 62(2), pp.327-376.

Hollstein, F., 2022. Local, regional, or global asset pricing?. *Journal of Financial and Quantitative Analysis*, 57(1), pp.291-320.

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), pp.265-295.

Hou, K., Karolyi, G.A. and Kho, B.C., 2011. What factors drive global stock returns?. *The Review of Financial Studies*, 24(8), pp.2527-2574.

Hou, K., Mo, H., Xue, C. and Zhang, L., 2019. Which factors?. *Review of Finance*, 23(1), pp.1-35.

Hou, K., Xue, C. and Zhang, L., 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3), pp.650-705.

Huber, D., Jacobs, H., Müller, S. and Preissler, F., 2023. International factor models. *Journal of Banking & Finance*, *150*, p.106819.

Huberman, G. and Kandel, S., 1987. Mean-variance spanning. *The Journal of Finance*, 42(4), pp.873-888.

Huberman, G., Kandel, S. and Stambaugh, R.F., 1987. Mimicking portfolios and exact arbitrage pricing. *The Journal of Finance*, 42(1), pp.1-9.

Hyde, S. and Sherif, M., 2005. Consumption asset pricing models: Evidence from the UK. *The Manchester School*, 73(3), pp.343-363.

Irwin, S.H. and Uhrig, J.W., 1984. Do technical analysts have holes in their shoes?. *Review of Research in Futures Markets*, 3(3), pp.264-277.

Jacobs, H. and Müller, S., 2020. Anomalies across the globe: Once public, no longer existent?. *Journal of Financial Economics*, 135(1), pp.213-230.

Jagannathan, R. and Meier, I., 2002. Do we need CAPM for capital budgeting?.

Jegadeesh, N. and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), pp.65-91.

Jegadeesh, N. and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), pp.65-91.

Jegadeesh, N. and Titman, S., 2001. Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, 56(2), pp.699-720.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, 23(2), pp.389-416.

Jobson, J.D. and Korkie, B., 1982. Potential performance and tests of portfolio efficiency. *Journal of Financial Economics*, 10(4), pp.433-466.

Kahneman, D., Tversky, A. and Slovic, P., 1982. Judgments of and by representativeness.

Kan, R.; C. Robotti; and J. Shanken, 2013. "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology." *The Journal of Finance*, 68, 2617–2649.

Karolyi, G.A. and Stulz, R.M., 2003. Are financial assets priced locally or globally?. *Handbook of the Economics of Finance*, *1*, pp.975-1020.

Kendall, M.G. and Hill, A.B., 1953. The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General)*, 116(1), pp.11-34.

Kilka, M. and Weber, M., 2000. Home bias in international stock return expectations. *The Journal of Psychology and Financial Markets*, *1*(3-4), pp.176-192.

Knopers, F.E.J., 2014. *Value investing: Evidence from the Dutch stock market* (Master's thesis, University of Twente).

Kozak, S., Nagel, S. and Santosh, S., 2020. Shrinking the cross-section. *Journal of Financial Economics*, 135(2), pp.271-292.

Lajili-Jarjir, S., 2007. Explaining the cross-section of stock returns in France: characteristics or risk factors?. *European Journal of Finance*, 13(2), pp.145-158.

Lee, C.M. and Swaminathan, B., 2000. Price momentum and trading volume. *The Journal of Finance*, 55(5), pp.2017-2069.

Lettau, M. and Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, 56(3), pp.815-849.

Lintner, J., 1965. Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4), pp.587-615.

Litimi, H., 2017. Herd behavior in the French stock market. *Review of Accounting and Finance*, 16(4), pp.497-515.

Litzenberger, R.H. and Ramaswamy, K., 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, 7(2), pp.163-195.

Lo, A.W. and MacKinlay, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, *1*(1), pp.41-66.

MacKinlay, A.C., 1985. An analysis of multivariate financial tests. The University of Chicago.

MacKinlay, A.C., 1995. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, 38(1), pp.3-28.

Martınez, M.A., Nieto, B., Rubio, G. and Tapia, M., 2005. Asset pricing and systematic liquidity risk: An empirical investigation of the Spanish stock market. *International Review of Economics & Finance*, 14(1), pp.81-103.

McCulloch, R. and Rossi, P.E., 1990. Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory. *Journal of Financial Economics*, 28(1-2), pp.7-38.

McCulloch, R. and Rossi, P.E., 1991. A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics*, 49(1-2), pp.141-168.

Merton, R.C., 1973. An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pp.867-887.

Michou, M. and Zhou, H., 2016. *On the information content of new asset pricing factors in the UK*. Working Paper, University of Edinburgh.

Mirza, N. and Afzal, A., 2011. Size and Value Premium in International Portfolios: Evidence from 15 European Countries. *Finance a Uver: Czech Journal of Economics & Finance*, 61(2).

Moskowitz, T.J. and Grinblatt, M., 1999. Do industries explain momentum?. *The Journal of Finance*, 54(4), pp.1249-1290.

Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H., 2012. Time series momentum. *Journal of Financial Economics*, 104(2), pp.228-250.

Moya-Martínez, P., Ferrer-Lapeña, R. and Escribano-Sotos, F., 2014. Oil price risk in the Spanish stock market: An industry perspective. *Economic Modelling*, 37, pp.280-290.

Nichol, E., and M. Dowling. 2014. Profitability and investment factors for UK asset pricing models. *Economics Letters* 125(3): 364–366.

Novy-Marx, R., 2012. Is momentum really momentum?. *Journal of Financial Economics*, 103(3), pp.429-453.

Pástor, Ľ. and Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3), pp.642-685.

Petkova, R., 2006. Do the Fama–French factors proxy for innovations in predictive variables?. *The Journal of Finance*, 61(2), pp.581-612.

Pham, H.N.A., Ramiah, V. and Moosa, I., 2020. The effects of environmental regulation on the stock market: the French experience. *Accounting & Finance*, 60(4), pp.3279-3304.

Pirogova, A. and Roma, A., 2020. Performance of value-and size-based strategies in the Italian stock market. *Economic Notes*, 49(1), p.e12160.

Pukthuanthong, K., Qiao, Z. and Wang, Y., 2023. No One-Size-Fits-All Tale: The Diversity and Complexity in Asset Pricing across Global Markets. *Available at SSRN 4525040*.

Qiao, Z., Wang, Y. and Lam, K.S., 2022. New evidence on Bayesian tests of global factor pricing models. *Journal of Empirical Finance*, 68, pp.160-172.

Rachev, S., Jašić, T., Stoyanov, S. and Fabozzi, F.J., 2007. Momentum strategies based on reward–risk stock selection criteria. *Journal of Banking & Finance*, 31(8), pp.2325-2346.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4), pp.1127-1139.

Rosenberg, B., Reid, K. and Lanstein, R., 1985. Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, 11(3), pp.9-16.

Ross, S. A., 1976. The Arbitrage Pricing Theory. *Journal of Economic Theory*, 13(3), 341–360.

Ross, S.A., 1978. A simple approach to the valuation of risky streams. *Journal of Business*, pp.453-475.

Rossi, F., 2012. The three-factor model: Evidence from the Italian stock market. *Research Journal of Finance and Accounting*, 3(9), pp.151-160.

Rouwenhorst, K.G., 1998. International momentum strategies. *The Journal of Finance*, 53(1), pp.267-284.

Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. *Journal of Financial Economics*, 14(3), pp.327-348.

Shanken, J., 1987. A Bayesian approach to testing portfolio efficiency. *Journal of Financial Economics*, 19(2), pp.195-215.

Shanken, J., 1987. Multivariate proxies and asset pricing relations: Living with the Roll critique. *Journal of Financial Economics*, 18(1), pp.91-110.

Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), pp.425-442.

Sharpe, W.F., 1966, Mutual fund performance, *Journal of Business* 39, 119-138.

Sharpe, W.F., 1992. Asset allocation: Management style and performance measurement. *Journal of portfolio Management*, 18(2), pp.7-19.

Silvestri, A. and Veltri, S., 2011. On the robustness of Fama and French Model: evidence from Italy. *Journal of Applied Finance and Banking*, *1*(4), pp.201-221.

Solnik, B.H., 1974. Why not diversify internationally rather than domestically?. *Financial Analysts Journal*, 30(4), pp.48-54.

Stambaugh, R.F. and Yuan, Y., 2017. Mispricing factors. *The Review of Financial Studies*, 30(4), pp.1270-1315.

Stambaugh, R.F., 1981. Missing assets, measuring the market, and testing the capital asset pricing model. The University of Chicago.

Sweeney, R.J., 1986. Beating the foreign exchange market. *The Journal of Finance*, 41(1), pp.163-182.

Tomek, W.G. and Querin, S.F., 1984. Random processes in prices and technical analysis. *The Journal of Futures Markets (pre-1986)*, 4(1), p.15.

Treynor, J.L. and Black, F., 1973. How to use security analysis to improve portfolio selection. *The Journal of Business*, 46(1), pp.66-86.

Walkshäusl, C. and Lobe, S., 2014. The Alternative Three-Factor Model: An Alternative beyond US Markets?. *European Financial Management*, 20(1), pp.33-70.

Weil, P., 1989. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics*, 24(3), pp.401-421.

Wolak, F. A. "Testing Inequality Constraints in Linear Econometric Models, 1989. *Journal of Econometrics*, 41, 205–235.

Wolak, F.A., 1987. An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, 82(399), pp.782-793.

Zhou, G., 1993. Asset-pricing tests under alternative distributions. *The Journal of Finance*, 48(5), pp.1927-1942.