Multi-Asset Factor Investing Strategies and Controversy Screening using Natural Language Processing

Ananthalakshmi Pallasena Ranganathan

Department of Accounting and Finance
Lancaster University

This dissertation is submitted for the degree of
Doctor of Philosophy in Finance

Lancaster University December 2023

Supervisors: Dr. Harald Lohre and Dr. Sandra Nolte

To my parents Vijayalakshme and Ranganathan,
husband Gokul, and our families

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements.

Ananthalakshmi Pallasena Ranganathan
December 2023

Acknowledgements

The incredible journey to my Ph.D. has been very rewarding yet challenging. I am deeply grateful to both my Ph.D. supervisors, Sandra Nolte and Harald Lohre, and, Margit Steiner, whose guidance, support, and motivation has made this journey so successful. Their dedication and passion for high-quality cutting-edge research have been the guiding light of my Ph.D. journey and my researcher career.

I am forever indebted to my mother, whose optimism, motivation, and encouragement helped me get through several difficult moments. Having never dreamt of doing a Ph.D. or becoming a researcher, I am immensely thankful for her unwavering belief in my abilities throughout my educational journey. I cannot thank my father enough for instilling the discipline and dedication that I needed for the successful completion of my Ph.D. I would like to extend my heartfelt thanks to my husband Gokul and our families who have helped me handle the pressure and supported me with immense love and joy in the last few months of this journey.

A big thanks to the Invesco Quantitative Strategies team for giving me the time and all the resources needed for the successful completion of my Ph.D. I would also like to thank the Economic and Social Research Council (ESRC), Lancaster University, and the Accounting and Finance department for the generous financial support throughout my doctoral journey.

I will forever cherish the memories and friendships I have had during this journey with IQS colleagues and fellow PhDs David Happersberger, Minh Thang Ho, Joshua Kothe and all my best friends from school and college, especially, Monnisha, who continue to bring light, laughter and joy into my life.

Abstract

Factor investing strategies have revolutionized the landscape of equity investing, and continues to be heavily researched by academics and practitioners, leading to the documentation of more than 450 factors. However, from a practical investment perspective, much of the factor evidence documented by academics may be more apparent than real. The performance of many factors has found to be dependent on the inclusion of small- and micro-cap stocks in academic studies, although such stocks would likely be excluded from the real investment universe due to illiquidity and transaction costs. We take the perspective of an institutional investor and navigate this zoo of factors by focusing on the evidence relevant to the practicalities of factor-based investment strategies. Establishing a sound theoretical rationale is key to identifying "true" factors, and we emphasize the need to recognize data-mining concerns that may cast doubt on the relevance of many factors. Nevertheless, a parsimonious set of factors emerges in equities and other asset classes, including currencies, fixed income and commodities. Since these factors can serve as meaningful ingredients to factor-based portfolio construction, we build currency factor strategies using the G10 currencies. We show that parametric portfolio policies can help guide an optimal currency strategy when tilting towards cross-sectional factor characteristics. While currency carry serves as the main return generator in this tilting strategy, momentum and value are implicit diversifiers to potentially balance the downside of carry investing in flight-to-quality shifts of foreign exchange investors. Drawing insights from a currency timing strategy, according to time series predictors, we further examine the parametric portfolio policy's ability to mitigate the downside of the carry trade by incorporating an explicit currency factor timing element. This integrated approach to currency factor investing outperforms a naive equally weighted benchmark as well as univariate and multivariate parametric portfolio policies.

Whilst factor investing continues to grow in popularity, investors have expressed interest in aligning their investments with social values in order to maximize positive social impact. Hence, for any company, involvement in socially unethical practices not only leads to reputational damage but also financial consequences, anecdotally. To quantify the consequence of such controversial behaviour, we investigate the price impact of involvement in social controversies and find that the returns drop, on an average, by over 200 basis in the days around the outbreak of news on social violations. We identify companies following socially unethical practices from news headlines with the help of state-of-the-art language modelling approaches. Using a large sample of 1 million news headlines, we further train and fine-tune a DistilRoBERTa model to identify reports of controversial incidents in daily news feed. We map the price reaction of such controversial events using an event study approach and document negative price impact for companies with poor social practices measured via increased controversial behaviour, largely driven by small to medium market capitalization companies. Amongst the eight different social dimensions we examine, controversies surrounding violations of product safety standards, online scams and data privacy breaches significantly impact firm returns. Dissecting this result by geographies, the U.S, Australia, Europe and Emerging Market react very negatively to social controversies.

Table of contents

List of figures

List of tables

Introduction

The notion of factor investing has gained a lot of popularity in the last few decades. Factor investing can be understood as investing in a group of stocks that share similar characteristics. Initial works on factors can be traced back to the 1970's when the focus was mainly on the sensitivity of stock's return to the market. However, over the last 4 decades, with better data availability and advanced computing technologies, the literature on cross-sectional stock return predictability has documented over 400 candidate factors. The overarching finding from academic factor studies could be summarized as follows: not all risks are equally rewarded; higher risks do not necessarily translate to higher return. Building on this premise, the empirical asset pricing literature has spawned new candidate factors to explain the cross-section of returns. Especially, the last 2 decades saw researchers propose new factors by straining data or via data snooping, prompting Cochrane (2011) to consider this proliferation of factors as a 'zoo of factors'.

Managing this factor zoo is challenging, especially in the light of data mining concerns surrounding many of the candidate factors. Recent meta-studies that put forward ways of dealing with such issues, mostly fueled by data-driven and computationally intensive methods that would deem many of these new factors insignificant. Hence, we examine this literature, taking the perspective of an institutional investor and provide guidance in navigating the zoo of factors by focusing on the evidence relevant to practicalities of factor-based investment strategies.

From a pure investment perspective, we caution that much factor evidence documented by academics is more apparent than real, and the performance of many factors is dependent on the inclusion of small- and micro-cap stocks in academic studies when such stocks would be excluded from the real investment universe on grounds of illiquidity and transaction costs. However, the run-through of the academic research findings indicates that a vast majority of factors discovered in the academic literature are not exploitable by institutions. Rigorous testing procedures have uncovered that the factor premia of unreal factors disappear for value-weighted portfolios, after excluding small-cap stocks, with any remaining profits nibbled up by transaction costs.

Despite these concerns, a parsimonious set of factors surfaces in equities and other asset classes, including currency, fixed income and commodities, which can serve as meaningful ingredients to aid factor-based portfolio construction. Although such style factors seem to resurge periodically, the initial question still remains: what number of factors are likely to pass all real-world tests? Apart from rare attempts to measure mispricing, distinguishing between risk and mispricing remains a key challenge.

In addition, the advent of machine learning in asset pricing can be considered a blessing in administering the factor zoo, but also a potential curse because it can unduly increase the number of candidate factors. Such developments emphasize the need to recognize data mining concerns casting doubt on the relevance of many factors. We also revisit the requirements of a robust factor, stressing the need for deep-rooted explanations for the existence of any given factor premia, and strongly recommend looking beyond the empirical performance of portfolio sorting procedure. Hence, identifying real factors in the factor zoo still seem to be a tough choice and we emphasize that establishing a sound theoretical rationale will remain key in identifying "true" factors.

While the literature on cross-sectional return predictability in equities is fairly developed, the recent decade has seen an emergence of studies in other asset classes as well. We briefly touch on corresponding evidence, highlighting major similarities and differences to return drivers in equities. We discuss currency factor strategies in deeper detail as currency factor models have garnered attention over the last decade. Currency factor models aim to measure the exposure of each currency to different factors, so that currency portfolios may be created dynamically. But, in a portfolio setup, the estimation of optimal weights is non-trivial. Moreover, constructing an optimal portfolio in a time varying set up while exploiting return predictability is an arduous task. The numerical complexity of finding closed-form solutions to the optimization problems complicates the procedure and hence, a plausible alternative solution would be very handy.

Brandt and Santa-Clara (2006) model the optimal portfolio weights as a function of common macroeconomic variables using a quadratic criterion. They advance the static Markowitz model by expanding their asset space to include a mechanically managed portfolio, thereby turning a static choice of managed portfolios into a dynamic strategy. This naturally implies that Markowitz mean-variance optimization can be applied without the need to estimate the full covariance matrix. On the other hand, Brandt, Santa-Clara and Valkanov (2009) perform a similar procedure for a large cross-sectional choice problem.

Hence, we use the parametric portfolio policy (PPP) approach of Brandt *et al.* (2009) for currency factor tilting to exploit cross-sectional information by combining 6 well-known FX factors in a dynamic multivariate framework. We find that an optimal currency tilting strategy with the 3 prominent FX style factors such as carry, value and momentum emerges as a clear winner. For timing currencies, we use technical indicators and fundamental variables in the PPP framework of Brandt and Santa-Clara (2006) and document the relevance of technical indicators in timing currencies. However, since the carry trade is prone to crash risk during market downturns, we explore the potential of an integrated strategy that uses relevant indicators to time the carry trade. We find that a TED spread-based regime indicator helps navigate the downside of the carry trade, thereby improving the overall risk-adjusted performance. Not only is this framework easily implementable but also offers the flexibility to be used with or without conditioning variables, thereby becoming a valuable tool for practitioners.

Outside of risk-adjusted performance, another growing area of interest for practitioners has been that of sustainable investing, fuelled by the increasing awareness of the role played by companies in the long-term well-being of the society which has naturally led investors' to look for ways to incorporate such criteria into their investments. Since sustainable investments promote positive social changes without forsaking financial benefits, investors have shown eagerness to choose investments that are more aligned with their values and social outlook. This in turn has pushed investment managers to evaluate the sustainability impact of investments by either reviewing metrics such as ESG (Environment, Social and Governance) scores of different companies provided by external rating agencies or building in-house client-specific sustainable investment policies.

However, this has been found to be problematic due to the poor overlap between controversy scores published by major rating agencies and lack of transparency of their methodology. We propose a novel way of identifying companies that violate social standards by using deep learning and natural language processing to extract relevant information from news headlines. Using state-of-the-art language model BERT and a large dataset of 1 million news headlines, we document the advantages of domain adaptation via further-training and fine-tuning a DistilRoBERTa for identifying companies following controversial practices along eight different social dimensions. Through an event study approach, we find that returns drop, on an average, by over 200 basis points in the days around the outbreak of news on social violations, largely driven by smaller to medium sized companies. This negative price impact is more severe for violations of product safety standards, online scams and data privacy breaches. We also observe heterogeneity in price impact across geographies, with regions like U.S, Europe, Australia and Emerging Market reacting very negatively to social controversies.

This cumulative dissertation encompasses three individual research papers that I have written during my joint doctoral program at Lancaster University and Invesco. The first chapter, *Navigating the Factor Zoo Around the World: An Institutional Investor Perspective*, is a joint project with Prof. Söhnke M. Bartram, Prof. Peter Pope and my supervisor Harald Lohre and has been published in the *Journal of Business Economics* (Ranganathan, Bartram, Lohre and Pope, 2021) . The second chapter, *An Integrated Approach to Currency Factor Investing*, is co-authored with Houssem Braham and my supervisors, Harald Lohre and Sandra Nolte and has been accepted for publication in the *Journal of Systematic Investing* (Ranganathan, Lohre, Nolte and Braham, 2023). The third chapter, *Controversial News Detection with Controversy-BERT and Stock Price reaction to Social Controversies* is joint work with Margit Steiner, Carsten Rother and my supervisors, Harald Lohre and Sandra Nolte.

References

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *Journal of Finance*, **61** (5), 2187–2217.

—, — and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, **22** (9), 3411–3447.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, **66**, 1047–1108.

Ranganathan, A., Lohre, H., Nolte, S. and Braham, H. (2023). An integrated approach to currency factor management. *Forthcoming in the Journal of Systematic Investing.*

Ranganathan, A., Bartram, S. M., Lohre, H. and Pope, P. F. (2021). Navigating the factor zoo around the world: an institutional investor perspective. *Journal of Business Economics*, **91** (5), 655–703.

Chapter 1

Navigating the Factor Zoo Around the World: An Institutional Investor Perspective

This project is joint work with Prof. Söhnke M. Bartram, Prof. Peter Pope and my supervisor Harald Lohre and has been published in the *Journal of Business Economics.* Helpful comments and suggestions by Wolfgang Breuer (the editor), three anonymous referees, and Livia Amato, Daniele Bianchi, Gurvinder Brar, Marie Brière (Amundi), Ben Charoenwong (NUS), Daniel Giamouridis (Bank of America Merrill Lynch), Alex Gracian (Resolute Investments), Mark Hutchinson, Anastasios Kagkadis, Felix Kempf, Erik Kroon (BNP Paribas), Jens Kummer (Star Capital), Alberto Martin-Utrera, Hongwei Mo, Spyros Mesomeris (UBS), Gajen Selverajah, Giuliano de Rossi (Goldman Sachs), Laurens Swinkels (Robeco), Shuiquing Wang and sem-inar participants at the 2021 Frontiers of Factor Investing Conference are gratefully acknowledged. This work was supported by funding from the Economic and Social Research Council (UK).

1.1 Introduction

Equity portfolios tilted towards observed firm characteristics, or factors, have attracted considerable attention from scholars and investment practitioners. From an academic perspective, characteristic-based factors are often used to explain the cross-section of equity returns, with a parsimonious subset for priced factors in modelling equity risk. From an investment perspective, the objective is to harness associated return premia when constructing factor-based equity portfolios. Whether such premia exist as compensation for bearing undiversifiable risk or as reward for identifying mispricing, they are seen as the holy grail of factor investing strategies. Against this backdrop, it is not surprising that the factor literature has proliferated to what is now considered a 'zoo of factors' Cochrane (2011), containing more than 450 predictive factors.

The factor zoo's inhabitants are diverse. To illustrate, value factors combine information from financial statements and market prices to identify relatively cheap stocks, whilst momentum and reversal factors are constructed from past return series. Quality factors build on accounting numbers to identify firms with strong balance sheets and lower downside risk, while low volatility strategies exploit the covariance structure of stock returns to establish defensive portfolio strategies that generate higher risk-adjusted returns. As they embody different styles of investing, factor-based strategies promise tailored exposures to meet risk-return objectives at lower costs, appealing to institutional investors who seek to improve diversification and control specific risk factor exposures (Figure 1.C.1). This can also be seen from the 2019 FTSE Smart Beta Global Survey, which expects the adoption of such factor strategies by institutional investors to grow, especially those marketing exchange-traded funds. Furthermore, the survey reports that an increasing number of institutional investors plan to adopt a factor lens in search of parsimonious and holistic approaches to asset management.

Factor investing appeals to investors as it is built on solid theoretical and empirical foundations, with a rationale for why factors worked in the past and are expected to continue to work in the future. Persistent factor performance is likely if a factor captures undiversifiable, systematic risks for which investors demand compensation. However, persistent investor biases are often also invoked as plausibly contributing to systematic mispricing of securities. In the absence of systematic biases, mispricing should be transient, and the associated return predictability should be short-lived, unless the underlying biases continue to exist and there are reasons to believe that mispricing cannot be arbitraged away.

Whilst early factor research focuses on establishing and rationalizing single factors, recent literature features several important studies that replicate many published factors to analyze the cross-section of predictors (e.g., Green, Hand and Zhang, 2017; Hou, Xue and Zhang, 2020; Feng, Giglio and Xiu, 2020). For example, Harvey, Liu and Zhu (2016) use statistical techniques to account for data snooping biases to separate true factors from false in a set of 316 published factors. They also describe how certain factors were deemed "significant" by luck. Pukthuanthong, Roll and Subrahmanyam (2019) use principal component analysis to test whether a given systematic risk factor qualifies as genuine. By explaining the need to distinguish priced factors from predictor characteristics, they propose several innovative methods for evaluating factors. Such guidelines are important from an investor perspective to avoid disappointing performance of their factor portfolios.

The majority of factors documented in the literature has first been identified in the U.S. equity market. Subsequently, the predictive ability of some factors has also been replicated in international markets–including developed, emerging and even frontier markets–as well as other asset classes. Such evidence can be viewed as out-of-sample evidence, despite meaningful differences across countries and asset classes with regards to institutional features such as transaction costs, liquidity, factor crowding, and the number of investible assets. Of course, exceptions exist, such as the weaker performance of momentum factors in Japan compared to other markets.

The empirical evidence, especially in early studies, often focuses on returns of equally weighted factor portfolios, which may overstate the realizable factor returns if less investible small- and micro-cap stocks are important to factor performance. Transaction costs are higher for difficult-to-arbitrage stocks, such as microcaps, low liquidity and high idiosyncratic volatility stocks. In a related vein, (factor) investors may face short-selling constraints, which may limit the potential factor performance to the contribution of a factor's long leg. Also, the portfolio turnover implied by a strategy is an important determinant of realizable factor performance–a low-turnover value strategy will incur significantly lower transaction costs than higher-turnover factors, such as momentum or short-term reversal. Textual analysis and the application of machine learning techniques are among recent developments in factor research, for instance to develop new or identify a robust set of factors. Finally, broad factor concepts such as carry, value, momentum and quality apply in many asset classes, suggesting to approach factor investing through a multi-asset multi-factor lens.

From an investment perspective, there are several key aspects for investors to consider when adopting factors in the investment process. First, despite hundreds of factors proposed in the literature, the number of factors that contain independent and exploitable predictive information for the cross-section of asset returns is much smaller. Second, with the increasing availability and growth in computational power facilitating the exploitation of alternative data sources, controlling data snooping biases is key to avoiding false discoveries. Third, the evidence on factor performance is often sensitive to the selected investment universe, with returns depending on the ability to invest in small and micro-cap stocks. Such factors are irrelevant for institutional investors, because the amount of capital that can be deployed is limited, and because market impact and other transaction costs make it expensive to trade in such stocks. Accounting for such real-world frictions is important, and investors should focus on whether a given factor delivers significant performance in value-weighted portfolios after accounting for transaction costs and investment constraints related to institutional investors' mandates.

1.2 Notable Species in the Factor Zoo

1.2.1 Bird's Eye Perspective

Starting with Cochrane (2011), academics have attempted to address concerns about the expanding factor zoo. Harvey and Liu (2019) conduct a factor census to manage the growing number of factors. Figure 1.C.2 shows the cumulative growth in the number of published factors in the top three finance journals since 1964, with 105 papers published exclusively in the *Journal of Finance*. Since 2008, there has been nearly exponential growth in the number of published articles, and hence in what has become known as the factor zoo. This growing popularity of factors can also be seen from Figure 1.C.3 which uses citation data from Scopus to track the of year-wise citations of the original study documenting the most prominent factors in the literature. Further supporting this view, a recent publication by Hou, Xue and Zhang (2020) documents 452 factors that researchers have uncovered.

A factor is typically based on an asset characteristic (or predictor variable) that has power for explaining the cross-section of future asset returns. If the ensuing factor premium is found to compensate for risk, it is considered a risk factor. Conversely, if the factor premium is not predicted to capture risk by theory and cannot be rationalized with generally accepted asset pricing models, it is considered an anomaly or a mispricing factor. However, there is often ambiguity in the literature with respect to assigning a given predictive factor to either category, risk or mispricing, partly because theoretical understanding evolves inductively and dynamically as empirical regularities are uncovered.

Researchers have suggested several guidelines to identify 'true' factors (those that generate persistent expected returns as a result of bearing priced risk or exploiting persistent behavioral biases or structural impediments). True factors should have incremental explanatory power over previously identified factors Feng, Giglio and Xiu (2020). The returns to true factors are persistent over time, pervasive across samples (e.g. countries, asset classes) and can withstand definitional variations. To be implementable, any given factor needs to survive transaction costs and have a solid theoretical rationale for the existence of the associated premia. However, time variation may make it challenging to distinguish empirically between factor premia and mispricing. Before we look into how to delineate risk and mispricing, we will first introduce the more traditional style factors and corresponding asset pricing models that are typically used to rationalize new predictive factors.

1.2.2 Salient Factors and Asset Pricing Factor Models

Starting with the capital asset pricing model (CAPM), which introduced the relationship between average returns and market exposure (or market beta), researchers have been keen to identify a model that best explains the cross-section of asset returns (see Table 1.B.2). Such models are of interest to academics and practitioners alike, as they are expected to help detect robust patterns in asset returns, which can be used to formulate profitable investment strategies and control portfolio risk. However, the empirical evidence challenges the CAPM. For instance, the low volatility factor is a rebuttal to the CAPM as seen from Haugen and Heins (1972), who find that low-risk stocks yield higher risk-adjusted returns than high-risk stocks over the long run.

Merton (1973)'s Intertemporal Capital Asset Pricing Model (ICAPM) and Ross (1976)'s Arbitrage Pricing Theory (APT) were offered as alternatives to the CAPM, highlighting the need for realistic assumptions. Ross (1976) popularized the term 'factors', and his APT lays the foundations for multifactor models. The APT expresses the expected returns on individual assets as linear combinations of the returns on one or several common factors capturing sources of risk that are priced in a no-arbitrage economy. In further studies, empirically motivated factors such as Basu (1977)'s price-earnings-based value factor and Banz (1981)'s size factor further document the insufficiency of CAPM to fully explain asset returns, calling for a more profound factor pricing model.

Addressing such concerns around the CAPM, Fama and French (1993) propose a 3-factor model combining market, size and value factors that until recently has been the standard academic workhorse model to rationalize factor premia in equity returns. However, it does not explain the returns on price momentum factors, a strategy that buys stocks with high recent returns (looking back three to twelve months), and shorts stocks with low recent returns Jegadeesh and Titman (1993). Consequently, Carhart (1997) proposes a 4-factor model by extending the Fama and French 3-factor model to include a one-year momentum factor alongside size, value and market beta.

Subsequent studies identify further regularities in stock returns that even the 4-factor model fails to capture, including quality factors such as investment and profitability put forward by Novy-Marx (2013) and Aharoni, Grundy and Zeng (2013). Subsequently, Fama and French (2015) expand their 3-factor model to a 5-factor model by adding profitability (RMW) and investment (CMA) factors. By not including a momentum factor, they treat momentum as a 'premier anomaly', unexplained by the CAPM and their own model, even many years later. The two new factors, RMW and CMA, render the value factor redundant, suggesting the use of a more parsimonious 4-factor model with market, size, investment and profitability factors alone. However, since value is one of the most sought-after factors among institutional investors, the use of the 5-factor model is warranted, and it essentially gives rise to the same abnormal returns as the 4-factor model. In a related vein, Hou, Xue and Zhang (2015)'s q-factor model, which is based on investment theory, combines an investment factor, a profitability factor, a market factor and a size factor. The authors find that the q-factor model outperforms the Fama and French 3-factor model and Carhart's 4-factor model by capturing most of the anomalies that these two models fail to account for.

The growth in the number of mispricing-based factors have prompted the development of mispricing-based factor models. Instead of constructing a model based on single anomaly factors such as size or value, Stambaugh and Yuan (2017) suggest combining information across multiple anomalies and construct two mispricing factors by averaging across eleven well-accepted anomalies in order to obtain a less noisy measure of mispricing. They ultimately propose a 4-factor model by combining the two aggregate mispricing factors, labelled management and performance, with a size factor and a market factor. Similarly, in order to distinguish the two complementary aspects of mispricing, Daniel, Hirshleifer and Sun (2020) develop a 3-factor model that features a market factor, a long-horizon factor (to capture long-term mispricing due to investor overconfidence), and a short-horizon factor (to capture short-term mispricing stemming from investor underreaction).

The significant growth in the number of suggested factors in the literature has intensified the search for an asset pricing model that identifies a parsimonious set of key factors useful in explaining the cross-section of asset returns, which would deal with the factor zoo from an investor perspective as well. For instance, Daniel *et al.* (2020)'s 3-factor model is not only parsimonious but also outperforms the profitability-based model of Novy-Marx (2013), the Fama and French (2015) 5-factor model, the q-factor model (Hou *et al.*, 2015), and the mispricing model of Stambaugh and Yuan (2017) in capturing a wide range of anomalies. On the contrary, the q5-model of Hou, Mo, Xue and Zhang (2021), which augments the q-factor model by an expected investment growth factor, has been shown by the authors to outperform all the factor pricing models identified so far: their empirical evidence suggests that the q5-model outperforms eight competing factor models including the Daniel *et al.* (2020) 3-factor.

However, several studies emphasize the complexities in comparing different factor models. For instance, Barillas and Shanken (2017) discuss why simply comparing time series regression intercepts (or test portfolio alphas) across different factor models is insufficient as they might not be applicable for non-traded factors like consumption growth. They highlight the sensitivity of model rankings to the choice of test assets and suggest the use of the GRS (Gibbons, Ross and Shanken, 1989) F-Statistic for comparing nested models. Acknowledging the challenges in comparing non-nested models, they point out that the best model in terms of a single performance metric might not be as good as one would expect because the excluded-factor evidence from the best model might favor another model. Fama and French (2018) review different approaches used in the literature for comparing factor models and use the maximum squared Sharpe ratio to compare the CAPM, the Fama and French (1993) 3-factor model, the Fama and French (2015) 5-factor model, and their 5-factor model plus momentum. With more than 400 factors in the factor zoo, they highlight the issues surrounding the comparison of multiple combinations of factors in the factor zoo, and argue that it would be almost impossible to identify the surviving factors from the factor zoo with the current statistical tests due to 'clouded' levels of p-values from overtly torturing the same data over and over again.

We illustrate the challenge in pinning down the best model by reporting selected results from Hou *et al.* (2015) (q-model), Hou *et al.* (2021) (q5 model), Stambaugh and Yuan (2017) (3-factor model) and Daniel *et al.* (2020) (3-factor model) in Table 1. Columns (4) and (5) report the maximum Sharpe ratio (MS) from Table 11 (Panel B) of Hou *et al.* (2015) and Table 5 of Hou *et al.* (2021), respectively. Column (6) reports the GRS F-Statistic, which tests whether the alpha of each of the anomalies tested is equal to zero, for 73 anomalies from Table 5 of Stambaugh and Yuan (2017), and column (7) reports the GRS F-Statistic for 34 anomalies from Table 7 of Daniel *et al.* (2020). Looking at the maximum Sharpe ratio reported by Hou *et al.* (2015) and Hou *et al.* (2021) sees their q-factor and q5-models emerging as the best models. Whereas looking at the GRS test-statistics in column (6) from Stambaugh and Yuan (2017) sees their 3-factor model with the lowest GRS F-statistic. Likewise, Daniel *et al.* (2020) report that their 3-factor model has the least value of the GRS F-test statistic in column (7). Hence, the lack of a common test metric, set of test assets, and number of anomalies tested complicates direct comparison of metrics reported in published articles.

Some studies report the performance of several asset pricing models across a number of metrics. Ahmed, Bu and Tsvetanov (2019) compare ten prominent asset pricing models and find inconclusive results. In their time-series tests, the Stambaugh and Yuan (2017) 4-factor model emerges as the best performer followed by the q-factor model. However, all tested models struggle to explain the returns on small stocks. In cross-sectional tests, the q-factor model, the Fama and French (2015) 5-factor model, the Fama and French (2015) 4-factor model, and the Barillas and Shanken (2018) 6-factor model perform best, followed by the Stambaugh and Yuan (2017) 4-factor model. Given the change in model rankings from different testing procedures, the authors caution that model comparisons are highly sensitive to the choice of test assets and comparison techniques.

In a nutshell, different combinations of factors can help to create powerful factor-based models capturing variation in expected returns without gaining exposure to unintended sources of risk and ensuring as much diversification of other sources of risk as possible. Whether the individual factors in such models stem from rational asset pricing theory, crude empiricism, or both, multi-factor models have become the dominant approach to explaining variation in expected returns, and such models can guide investors in their choice of factors.

1.2.3 Evidence of Cross-sectional Return Predictability Across the Globe

Investors across the globe have been keen to adopt factor-based investment strategies. According to the Invesco Global Factor Investing Study (2019), 50%–60% of surveyed institutional investors in North America, Europe, the Middle East and Africa (EMEA), and Asia-Pacific (APAC) intend to increase their factor allocations over the next three years. This is despite the fact that the research evidence underpinning factor investing is largely based on the U.S. equity market, emphasizing the need to examine factor returns outside the U.S.

Calling it the "academic home bias" puzzle, Karolyi (2016) shows that only 23% of all empirical finance articles examine non-U.S. markets. Given the convincing U.S. evidence, there has been increasing academic interest in confirming the existence of factor premia in other regions. Despite limitations to data and breadth of non-U.S. equity markets, the primary finding has been heterogeneity in the significance of many return predictors across regions. Still, some important predictors appear to work reasonably consistently across regions. For example, Haugen and Baker (1996) note a commonality in the primary return determinants across five major markets (U.S., Germany, France, United Kingdom and Japan), especially prominent style factors such as value and momentum factors as noted below.

Value

There is extensive international evidence that various value factors on average generate positive equity return premia, especially in emerging markets. Chan, Hamao and Lakonishok (1991) investigate returns to earnings yield, the book to market ratio, and cash flow yield in Japanese and U.S. equity markets, documenting that these fundamentals strongly predict expected returns. Similarly, Capaul, Rowley and Sharpe (1993) find evidence for value premia in France, Germany, Switzerland, the United Kingdom, Japan and the United States. Fama and French (1998) also document value premia in twelve major developed international stock markets. They note the 'hazardous' distributional properties of security returns in emerging markets, although a value factor based on the book-to-market ratio seems to work in twelve of the sixteen emerging market countries in their sample. Rouwenhorst (1999) finds further evidence in support of a value factor premium in emerging markets. Similarly, studying 35 emerging markets between 1985 and 2000, Barry, Goldreyer, Lockwood and Rodriguez (2002) offer strong emerging market evidence confirming the existence of a value premium.

Momentum

Another style factor that works across the globe is momentum, as documented in many international studies (see Rouwenhorst (1998) for European countries and Griffin, Ji and Martin (2003) for global evidence). Rouwenhorst (1999) questions whether the same cross-sectional factors drive returns in developed and emerging markets, confirming slightly weaker momentum premia in emerging markets. However, more recent evidence of Griffin, Kelly and Nardari (2010) suggests the opposite, with annual momentum profits averaging 8.7% in developed markets and 11.4% in emerging markets.

While momentum is a significant return factor in most markets, academic research has pointed out that momentum strategies fail to work in Japan. Asness (2011) argues that such evidence is not casting data mining doubts on international momentum effects. When viewing value and momentum factors as a single system, Japanese return behavior is consistent with the international evidence. These findings resonate with the universal profitability of value and momentum documented in Asness, Moskowitz and Pedersen (2013) and Fama and French (2012). Hou, Karolyi and Kho (2011) also confirm that medium-term stock price momentum is priced in international equity markets and complements the value factor. Hence, the existence of value and momentum premia has been documented in several international markets.

Beyond Value and Momentum in International Stock Markets

Other factors have also been examined in global equity markets. Ang, Hodrick, Xing and Zhang (2009) find that a low-volatility strategy is profitable in 23 developed countries. Based on such findings for many equity markets across the world, researchers have concluded that there is no reward for bearing volatility risk, thereby strengthening the case for a low-volatility factor (Haugen and Baker, 2012). Blitz, Pang and Van Vliet (2013) report that the low-volatility effect has become stronger due to delegated portfolio managers who tend to divert attention from low-risk stocks, with this impact being stronger in emerging markets than in developed markets. In a related vein, Asness, Frazzini and Pedersen (2014) support the international equity market evidence of the betting-against-beta (BAB) factor. They dismiss the suggestion that industry bets drive BAB factor premia and document significant risk-adjusted returns to industry-neutral BAB portfolios across 49 U.S. industries and in 60 of 70 global industries. The international evidence of different factors dispels concerns surrounding country-specific performance of factor strategies.

1.2.4 Performance of Equity Factors Across the Globe

Translating paper returns into realized profits is an important concern of investors, especially outside of developed markets. Accounting for realistic constraints faced by institutional investors, Van der Hart, Slagter and Van Dijk (2003) find that value, momentum and earnings revisions are stronger predictors of returns than size and liquidity in 32 emerging markets. In further evidence focused on the dynamics of frontier markets relative to developed and emerging equity markets, De Groot, Pang and Swinkels (2012) report evidence of a size premium in frontier markets that is not explained by exposure to global size, value, market, or momentum factors. They also find that value and momentum strategy returns survive transaction costs. Overall, these findings suggest that many style factors are profitable both in the United States and across the globe, underpinning the growth of factor-based investment strategies in global institutional portfolios.

In order to compare the magnitude of international equity factor premia, we gather empirical evidence on the salient factors in different regions between December 2001 and July 2020. While the academic literature often analyzes factor returns in ways that introduce practical caveats (most prominently the inclusion of microcap stocks and the use of equally-weighted portfolios), we focus on established factor indices provided by MSCI. They are a better gauge for the practical efficacy of equity factors around the globe since they employ realistic weighting schemes and focus on investable universes. The Appendix provides a more detailed description of the construction of the MSCI factor indices.

Table 1.B.3 reports performance for the equity factors Value, Size, Momentum, Quality, and Low Volatility alongside the corresponding market index returns. Panel A uses equity factors built for the MSCI All Country World Index (ACWI), while the other panels use factors for the United States (Panel B), Europe, Australasia and the Far East (EAFE) (Panel C), and Emerging Markets (EM) (Panel D). The

evidence is highly consistent across all regions. To benchmark the factor return indices, the annualized return for the MSCI ACWI within the sample period is 6.84% p.a. with a volatility of 15.4%. A corresponding index investment gave rise to a modest risk-adjusted return, as measured by a Sharpe ratio of 0.33. The maximum drawdown of 54.5% occurred in the global financial crisis (GFC) of 2008/2009.

Active returns capture the factors' performance contribution relative to the index benchmark. Table 1.B.3 shows that all factors outperform the market index except for Value. Value has an annualized active return of -1.41% p.a. relative to the market and suffered a more severe drawdown in the GFC than the benchmark index. Despite earlier evidence suggesting that Value is a more procyclical investment style, it has continued to display weak performance in the second half of the sample period. While Momentum is similar to Value in terms of volatility and drawdown statistics, it has the highest return (11.1% p.a.) of all factors considered, thus outperforming the market by 4.26% p.a. This corresponds to a risk-adjusted active return of 0.59% p.a. (as measured by the information ratio capturing the active return per unit of risk relative to the benchmark portfolio).

The Quality factor is associated with similar risk-adjusted active returns (information ratio of 0.62), but represents a more defensive absolute risk-return characteristic: the volatility of Quality factor returns is 13.6%, and the maximum drawdown is almost ten percentage points lower than that of the market (45.4% versus 54.5%). The maximum drawdown is even lower for Low Volatility (38.8%), consistent with this investment style having a considerably lower market beta. Indeed, the Low Volatility factor displays the lowest volatility across all regions–its ex post volatility of 10.5% is around two thirds of global market volatility. Low Volatility nevertheless outperformed the market by 1.94% p.a. over the sample period and shows the highest Sharpe ratio among all global factors (0.67).

Empirical research has also studied the performance of factor strategies with lower implementation costs such as ETFs and mutual funds. In particular, Van Gelderen and Huij (2014) investigate the performance of prominent style factors such as low volatility, size and value in U.S. equity mutual funds. They not only evidence significant excess returns for these factor portfolios, but also find that the performance is persistent over time. In a related vein, Elton, Gruber and De Souza (2019) document that combinations of factor ETFs outperform active U.S. equity mutual funds most of the time. Still, real-world frictions may impact investors' profitability especially when switching between factors or changing asset managers frequently. To wit, Van Gelderen, Huij and Kyosev (2019) find that despite style factors having a significant premium with a buy-and-hold strategy, rebalancing costs erode a significant portion of the factor profits. Hence, despite the convincing performance of factors, the final profit earned by investors is limited by real-world frictions.

1.2.5 The Advent of Machine Learning

Promises and Pitfalls

Machine learning (ML) is a collective term that refers to using computer algorithms to infer meaningful patterns from a dataset. Depending on the selected hyperparameters, ML can be used to cater to both low- and high-dimensional setups, that is when one is facing only a few predictors or a lot of predictors. Increased data availability and computational capabilities have opened doors for ML algorithms in the investment management industry, and this class of techniques is increasingly used for return prediction and clustering of candidate factors. Approaching the factor zoo as a high-dimensional problem, ML appears to be a natural solution.

The attractiveness of ML techniques stems from their flexibility, distribution-free specification and data-driven perspective. ML techniques have been used to construct portfolios with more accurate risk and return forecasts and under more complex constraints, to devise novel trading signals and execute trades with lower transaction costs, and to improve risk modelling and forecasting by generating insights from new sources of data (Bartram, Branke and Motahari, 2020a). Other advantages of ML methods stem from their estimation procedure that allows joint testing of a large number of cross-sectional stock characteristics, focusing more on predictive accuracy and offering a framework to deeply exploit potential non-linear relationships (Freyberger, Neuhierl and Weber, 2020). Conversely, traditional econometric techniques focus more on causal inferences.

Given the required technical skills, few researchers have attempted to apply ML techniques to testing the significance of different return predictors. Gu, Kelly and Xiu (2020) exploit the ability of ML techniques to accommodate large numbers of predictors and capture potential non-linearities and predictor interactions. Based on 94 stock characteristics, they document high out-of-sample predictive R-squared for ML return forecasts, with liquidity, volatility and price trends being the most significant predictors. They trace the predictive gains of the best performing models to their ability to capture non-linear predictor interactions missed by other classical statistical methods. Similarly, Bianchi, Büchner and Tamoni (2020) use machine learning methods for predicting bond excess returns. Based on more than 100 macroeconomic and financial variables, yields included, the authors document higher out-of-sample R-squareds compared to more traditional econometric methods.

As the ultimate goal of factor investing is to cater to the investor's risk-return objectives, newer ML techniques have been explored to automate portfolio construction. To this end, Feng, Polson and Xu (2019) utilize 62 firm characteristics as inputs to train a deep learning model for U.S. equities. Augmenting the Fama-French 3-factor model with factors identified by the deep learning model, they document marginal improvements in the R-squared in the time series analysis of portfolio returns, but considerable out-of-sample performance in cross-sectional returns prediction. Such encouraging results validate the scope of artificial intelligence and ML techniques in factor investing.

ML methods have also helped uncovering weaknesses of existing factor models in dealing with the factor zoo (Freyberger *et al.*, 2020). Its multi-dimensionality calls for models that can identify incremental information in each characteristic to eliminate the factors that are subsumed in joint tests and ultimately identify the surviving ones. Furthermore, existing models do not consider nonlinear relationships between characteristics and returns (Fama and French, 2008), prompting Cochrane (2011) to suggest the usage of different techniques to overcome such limitations. Kozak, Nagel and Santosh (2020) use ML techniques to investigate 120 return predictors and find that traditional 3-factor (or even 5-factor) models are insufficient in a high dimensional setup.

Using 36 well-known return-predictive characteristics, Freyberger *et al.* (2020) find that a linear model selects 21 characteristics, whilst non-linear models select only 8, but increase Sharpe ratios by 50% out-of-sample. Their results are robust to the choice of tuning parameters, addressing data mining and overfitting concerns. Hence, at a minimum, ML techniques could help identifying surviving factors in the factor zoo. Feng, Giglio and Xiu (2020) stress the importance of choosing the correct benchmark for navigating the factor zoo and propose a model framework to select factors from a list of candidates. The improved framework aims to identify fewer significant factors that add value after controlling for the three Fama-French factors. Thus, it appears that ML techniques may help reduce the dimensionality of the factor zoo, albeit while introducing new complications and challenges.

Sceptics consider ML in asset pricing a hard bargain, however. Because ML techniques are purely driven by the specific data used for analyses, they are susceptible to data mining and overfitting. Overfitting occurs when the ML model learns the training data too well and thus may fail to work with a new dataset. Researchers have suggested that the ratio of the degrees of freedom to the number of observations in the dataset could reflect the extent of overfitting in the model. As examining the factor zoo would require joint testing of hundreds of characteristics, the large number of independent variables would imply very high degrees of freedom, potentially leading to overfitting the training dataset.

The underlying ML mechanisms are often perceived as a "black box" with questionable theoretical underpinnings. From an institutional investor perspective, the inability to attribute investment performance can render client communication a challenge. Avramov, Cheng and Metzker (2021) question the interpretability of signals derived from ML techniques and critically evaluate the contributions of ML techniques in return prediction. They find a steep decline in return predictability of ML techniques after excluding microcaps or distressed firms and adjusting for market states. ML strategies are particularly successful in specific market states, such as periods of high investor sentiment or high market volatility. ML strategies tend to have higher turnover, and hence higher implementation costs, further emphasizing the need to approach such complex techniques with caution. Borghi and de Rossi (2020) estimate a series of models along the lines of Gu *et al.* (2020) and apply trading constraints when optimizing the portfolio, i.e. they limit turnover and the amount traded in each stock based on its average dollar volume. Whilst performance is deteriorating, the conclusion that ML is superior to traditional alternatives at combining factors is unchanged.

In a related vein, Leung, Lohre, Mischlich, Shea and Stroh (2021) investigate the potential of ML techniques for predicting the cross-section of stock returns. Using a set of twenty stock characteristics in an investible global stock universe, they confirm that ML forecasts are statistically superior to those based on standard linear models. Yet, this advantage is driven by exposure to hard-to-arbitrage factors such as short-term reversal, raising doubts about the economic relevance of ML models for practical institutional investment. Indeed, the added value in real-world portfolio simulations is less pronounced and depends heavily on the ability of an investor to take risk and implement trades efficiently.

Textual Factors

Novel sources of information have been exploited by researchers and practitioners to identify newer sources of return predictability by using state of the art techniques. Natural language processing (NLP) has become an important methodology for extracting information from unstructured textual data sources. NLP has found its way into factor investing studies to extract return predictors from published financial disclosures and related materials such as 10-K filings or earnings call transcripts. NLP techniques search for patterns in financial narratives to infer properties such as sentiment or obfuscation in the words that corporate executives use in their disclosures and communications with the market. For example, in inferring executives' sentiment, financial narratives might be classified into broad groups such as positive, negative or neutral sentiment. As investor sentiment can be used as a short-term return predictor, such information could be useful during portfolio rebalancing.

Surveys of text mining in the broader field of accounting and finance highlight the information content hidden in corporate disclosures that can help predict future firm performance (Li, 2008; Kearney and Liu, 2014; El-Haj, Rayson, Walker, Young and Simaki, 2019). Quantitative data carry more easily interpretable information than qualitative data, whilst the complex and ambiguous nature of oral and verbal communications could limit the efficiency of even the most advanced text mining tools. To this end, custom dictionary techniques and topic modelling are emerging as potentially more powerful approaches. Dictionary methods use the frequency of occurrence of a list (or bag) of words as a measure (see e.g. Bartram, Brown and Conrad (2011). The limited range of commonly used dictionaries and the equal weighting of all occurrences of a word in different contexts, however, raises concerns about the reliability of such methods (Hansen, McMahon and Prat, 2017).

In contrast to dictionary-based techniques, topic modelling techniques focus on uncovering the underlying semantic structures by recognizing topics that occur in a collection of documents. The most prevalent topic modelling technique is Latent Dirichlet Allocation (LDA) proposed by Blei, Ng and Jordan (2002). LDA approaches a document as a set of different topics and then measures the dominance of each topic. To this end, Israelsen (2014) uses risk factors extracted from 10-K filings for style analysis and offers risk-based explanations for the existence of market, size, value and momentum premia, illustrating the importance of qualitative information in firm disclosures. Topic modelling can also be useful in the development of text-based multi-factor models. Using LDA to uncover the risks disclosed in a firm's 10-K filings, Lopez-Lira (2019) identifies four systematic factors (technology, production, international and demand) that help explain the cross-section of returns. This text-based 4-factor model has the smallest GRS F-statistic compared to the Fama and French (2015) 5-factor model, Stambaugh and Yuan (2017)'s mispricing-based factor model, and the Hou *et al.* (2015) q-factor model. However, a possible downside with topic modelling is that different researchers may end up identifying different topics that are inherently subjective, rendering the findings non-replicable. Hence, investors and portfolio managers need to be cautious when using such techniques for factor selection.

1.3 Factor Investing beyond Equities

A majority of the factor investing literature focuses on equities, perhaps reflecting the absence of a clear theoretical consensus on how best to identify and model drivers of equity risk and return, but also greater investor interest and the relatively rich and diverse data available for equities compared to other financial asset classes (Bartram and Dufey, 2001). Nevertheless, in recent years researchers have increasingly attempted to apply insights from equity factor research to other asset classes such as currencies, fixed income, and commodities. For instance, Asness *et al.* (2013) find evidence for the existence of value and momentum premia in currencies, government bonds, and commodities, as well as equities. Similarly, Koijen, Moskowitz, Pedersen and Vrugt (2018) provide evidence for carry trade return predictability in global equities, global bonds, currencies, commodities, U.S. Treasuries, credit, and equity index options. This section summarizes some of the recent evidence for the main non-equity asset classes, highlighting similarities and differences relative to findings for equities.

1.3.1 Currencies

Currency factor strategies are used by institutional investors both for hedging unwanted currency exposures in internationally diversified portfolios and as a stand-alone investment asset class. Researchers have noted a tendency for currency fund managers to load on standard currency style factors, such as carry, value, momentum and volatility (Pojarliev and Levich, 2008). In addition to carry, value and momentum, recent research has identified several other related factors including dollar exposure, dollar carry, factors based on macro-economic fundamentals such as output gap and the Taylor rule (Bartram, Djuranovik and Garratt, 2020b), global external imbalance (Corte, Riddiough and Sarno, 2016) and business cycle factors that identify strong and weak economies (Colacito, Riddiough and Sarno, 2020).

FX Carry

Based on early research establishing that uncovered interest rate parity (UIP) does not hold (Bilson, 1981; Fama, 1984), the FX carry factor seeks to exploit the interest rate differentials of high- and low-yielding currencies. Hansen and Hodrick (1980), Bilson (1981) and Fama (1984) address the failure of UIP in the context of the forward premium puzzle and hence can be thought of as academic precursors for carry trading in foreign currency markets. The carry trade as a cross-sectional investment strategy involves borrowing in low-interest rate currencies and investing in high-interest rate currencies. While currency movements according to UIP should negate the resulting profits, this is empirically not the case, rendering FX carry investments profitable.

Carry trades appear to be sensitive to market movements and have experienced severe crashes with extreme drawdowns of about 30% (Doskov and Swinkels, 2015). Consequently, it has been suggested that positive returns from carry could provide compensation for crash risk (e.g., Brunnermeier, Nagel and Pedersen, 2008; Farhi, Fraiberger, Gabaix, Ranciere and Verdelhan, 2009). Brunnermeier *et al.* (2008) refer to the carry trade as "going up the stairs and down the escalator" and find that carry trade unwinding happens during liquidity squeezes and periods of heightened FX volatility (see also, Menkhoff, Sarno, Schmeling and Schrimpf, 2012a). Bhansali (2007) documents a positive relation between carry trade payoffs and currency volatility and concludes that currency carry trades perform best in low-volatility regimes. Carry trade strategies in other major asset classes also seem to perform poorly during recessionary periods Koijen *et al.* (2018).

To further rationalize the existence of carry premia, Lustig and Verdelhan (2007) and Lustig, Roussanov and Verdelhan (2011) show that U.S. consumption growth explains a significant portion of carry trade returns, arguing that carry trades reflect compensation for bearing the risk of a large depreciation during global downturns Hoffmann and Studer-Suter (2017). Relatedly, the peso problem is a commonly offered explanation for carry trade performance, i.e. it is argued that investors are compensated for exposure to relatively rare events that have extreme negative outcomes such as currency devaluations (Burnside, Eichenbaum, Kleshchelski and Rebelo, 2011). However, Jurek (2014) rejects peso problem explanations for the outperformance of carry trades and finds that negative skewness is priced in the cross-section of carry trades.

FX Momentum and FX Value

Momentum strategies have also been found to be profitable in currency markets (Menkhoff, Sarno, Schmeling and Schrimpf, 2012b). Although FX momentum strategies require frequent rebalancing and hence higher transaction costs, they are an important diversifier to counteract the downside risk of carry trades (Barroso and Santa-Clara, 2015). With carry trade positions being unwound in times of crises, currency momentum-based trading will quickly anticipate such currency movements and effectively establish positions offsetting carry currency allocations.

Potentially complementing currency momentum as a short-term hedge, currency value strategies identify overvalued and undervalued currencies in order to gain exposure to expected long-term reversal of currencies to their fundamental values (Menkhoff, Sarno, Schmeling and Schrimpf, 2017). Asness *et al.* (2013) use the 5-year change in purchasing power parity (PPP) to compute currencies' value and document significant value and momentum premia for a sample of G10 currencies between 1979 and 2011. In computing value factors for currencies, Baku, Fortes, Hervé, Lezmi, Malongo, Roncalli and Xu (2020) employ alternative proxies for fundamental value including PPP, REER (real effective exchange rate), and FEER/BEER (fundamental/behavioral equilibrium exchange rates). They find that emerging market FX value factors based on PPP and FEER/BEER have higher Sharpe ratios than their G10 counterparts. Similarly, FX carry and momentum factor returns are stronger in emerging market currencies than those in developed market currencies.

Bartram *et al.* (2020b) combine eleven predictors of currency excess returns into a combined mispricing measure documenting higher signal to noise ratios compared to individual factors. Fast decay of predictor ranks and performance as well as evidence of significant returns after comprehensive risk-adjustments challenge risk explanations to currency trading strategies. The mutual diversification benefit of combining carry, value and momentum factors has been repeatedly confirmed in the literature (Ranganathan, Lohre, Nolte and Braham, 2023); in addition, such factors meaningfully expand the investment opportunity set of international multi-asset investors (Kroencke, Schindler and Schrimpf, 2014).

1.3.2 Fixed Income

Fixed income markets are characterized by comparatively lower liquidity, and tradable securities are more heterogeneous with different coupons, maturities and covenant structures. Perhaps as a consequence, there is less corresponding factor investing research. From a top-down perspective, fixed-income investors are exposed to credit and interest rate risks, yet style factors based on duration, carry, quality, and low-volatility based definitions have been proposed to manage rates. However, despite difficulties in controlling for the pricing implications of contract design features, factors pertaining to corporate bonds, also referred to as credit factors, have recently become popular among practitioners, especially given the large cross-sectional universe of corporate bonds available for analysis.

Corporate Bond Factor Investing (Credit Factors)

An early study by Hottinga, van Leeuwen and van Ijserloo (2001) explores the vastness of the fixed income market and the promising scope of factor investing strategies in corporate bonds. Various style factors based on corporate bond characteristics, often paralleling those in equity markets, have been documented as significant predictors of bond returns. Specifically, Bai, Bali and Wen (2019) show that downside risk, credit risk, and liquidity risk are priced in the cross-section of corporate bond returns and confirm that these three risk factors are not subsumed by other bond market factors. Moreover, momentum, low-volatility, and quality have been documented as significant predictors of bond returns (Israel, Palhares and Richardson, 2018). Bali, Subrahmanyam and Wen (2020b) document long-term reversals in the corporate bond market. Kelly, Palhares and Pruitt (2020) propose a new conditional factor model for individual corporate bond returns based on instrumented principal components analysis. Brooks, Gould and Richardson (2020) find that exposure to traditional risk factors largely explains the active returns of fixed income managers.

In addition to bond factors, a number of equity factors such as size, profitability, and asset growth also have predictive power for bond returns (Chordia, Goyal, Nozawa, Subrahmanyam and Tong, 2017; Jostova, Nikolova, Philipov and Stahel, 2013; Gebhardt, Hvidkjaer and Swaminathan, 2005). Bektić, Wenzler, Wegener, Schiereck and Spielmann (2019) find that the Fama and French (2015) size, value, profitability and investment factors have explanatory power for returns of U.S. high yield corporate bonds, but the relations are less pronounced for U.S. and European investment grade bonds. Avramov, Chordia, Jostova and Philipov (2019) show that investor sentiment and financial distress jointly drive bond and equity overpricing underlying market anomalies. However, according to Choi and Kim (2018) some variables (e.g., profitability and net issuance) fail to explain bond returns, and for others (e.g., investment and momentum) bond return premia are too large compared with their loadings, or hedge ratios, on equity returns of the same firms. Moreover, Bali, Goyal, Huang, Jiang and Wen (2020a) show that while equity characteristics produce significant explanatory power for bond returns, their incremental predictive power relative to bond characteristics is economically and statistically insignificant when using machine learning.

Government Bond Factor Investing (Rates Factors)

Style factors such as carry, value, momentum, and defensive also manifest in the cross-section of global government bonds, albeit deriving from a relatively small sample of international rates. Brooks and Moskowitz (2018) analyze yield curve premia and conclude that carry, value, and momentum factors better explain their cross-sectional and time-series variation than the underlying principal components. Beekhuizen, Duyvesteyn, Martens and Zomerdijk (2019) provide a thorough analysis of (yield curve) carry strategies that involve selecting maturity buckets with the highest units of carry. Whilst the basic premise of the carry trade is borrowing a low-yielding asset to invest in a high-yielding asset, carry trades in bonds are designed to capture the roll yield, which is the price increase when the longer-term bond rolls down the yield curve.

The curve carry factor for bonds uses the slope of the yield curve directly by going long on a longer maturity, say a 10-year bond, and short on shorter maturity, say a 2-year, bond. Beekhuizen *et al.* (2019) find that the curve carry strategy subsumes the defensive betting-against-beta strategy that invests in the shortest maturity buckets. Brooks, Palhares and Richardson (2018) find that combining styles including carry, value, momentum, and defensive deliver a Sharpe ratio close to one over 20 years of data. The authors emphasize the appeal of such rates factors, given that they have low sensitivity to macroeconomic variables and thus diversification benefits. In a related vein, Kothe, Lohre and Rother (2021) show how rates factors expand investors' opportunity set and can benefit their tail-hedging or return-seeking investment objectives.

1.3.3 Commodity Style Factors

Commodities find their way into institutional investor portfolios as they are typically thought of as an alternative asset class offering protection against rising inflation, in addition to offering diversification benefits because of their low correlation with traditional asset classes including equities and bonds. However, the 2008/2009 Global Financial Crisis and the 2020 Covid-19 Crisis saw commodity prices fall in tandem with other asset classes, raising questions about the diversification benefits of commodities. Still, there is ample evidence of predictability in the very heterogeneous cross-section of commodities, and commodity factors help broadening investors' opportunity set (Giamouridis, Sakkas and Tessaromatis, 2017; Blitz and De Groot, 2014). Similar to other asset classes, commodity returns can be explained by commodity style factors including carry and momentum.

Miffre (2016) and Sakkas and Tessaromatis (2020) both present overviews of relevant commodity factors that typically build on commodity fundamentals such as term-structure variables or on past price momentum. Miffre and Rallis (2007) find evidence for the existence of both short-term momentum and long-term momentum in commodities, while noting that momentum effects are not predicted by extant asset pricing models. With carry and momentum playing a role in predicting commodities returns, it seems important to identify the fundamental economic drivers of commodity factor premia (Erb and Harvey, 2006). Bakshi, Gao and Rossi (2017) show that a 3-factor commodity pricing model with market, carry, and momentum factors summarizes the cross-section of commodity returns better than one-factor or two-factor models. Hence, multi-factor models also appear relevant for commodities (Fernandez-Perez, Frijns, Fuertes and Miffre, 2018). In line with this conclusion, Hammerschmid and Lohre (2020) integrate time-series predictors and cross-sectional characteristics in a parametric portfolio policy context for commodity futures. Their final choice of three variables for the multivariate timing policy and six fundamental factors for the multivariate tilting portfolio outperforms an equally weighted benchmark.

1.3.4 Multi-Asset Multi-Factor Investing

Institutional investors do not necessarily consider different asset classes in isolation but may combine factors in multi-asset or cross-asset investment frameworks. To this end, Figure 1.C.4 compares statistics on style factor performance across asset classes, namely excess returns, volatility, Sharpe ratio and maximum drawdown. All of the presented performance statistics refer to excess returns of long-short factor portfolios. Except for equity value, all style factors display positive excess returns throughout the sample period, thus reinforcing the validity of this investment paradigm.

While the average volatility across factors is around 5%, the least and the most volatile factors emerge within commodities; commodity quality comes in at some 3% volatility whereas commodity momentum has over 11% volatility. Importantly, many style factors deliver attractive risk-adjusted returns as measured by Sharpe ratios as well (note that the underlying factor indices account for transaction costs). Whilst Sharpe ratios refer to the compensation of volatility risk, some style factor strategies come with considerable tail risk. Maximum drawdowns in Figure 1.C.4 are often below -15% for the non-equity factors, with commodity momentum and commodity carry showing maximum drawdowns below -30%. Style factor returns tend to have low correlations, and their tail risk events typically do not coincide (Chambers, Lohre and Rother, 2019). Consequently, embracing factor investing in and across different asset classes suggests ample diversification benefits for multi-asset multi-factor strategies. Indeed, the related literature typically suggests risk-based allocation schemes to harvest the associated premia in a balanced fashion (Dichtl, Drobetz, Lohre and Rother, 2021).

1.4 Rationalizing Factor Returns

The case for including specific factors in an investment strategy is undoubtedly strengthened if the role of fundamental economic factors in driving factor performance can be identified. However, it can often be difficult to distinguish empirically between risk and mispricing explanations of factor performance, especially when results are sensitive to the choice of asset pricing (or risk) model as the benchmark for expected returns. Beyond this challenge, recent research has emphasized the importance of excluding the possible role of data snooping biases as explanations for statistically significant factor returns, especially when theoretical support for a returns predictive relationship is weak. The following subsections outline how researchers are seeking to categorize factor premia with these considerations in mind.

1.4.1 Why Do Factor Premia Exist?

The proliferation in the number of published return predictors in the literature highlights the need for having strong underlying rationales. Classic explanations include risk, mispricing and statistical bias. Behaviorists argue that factor premia stem from persistent pricing errors, while supporters of rational pricing theories suggest risk-based explanations. In particular, behavioral explanations of factor premia posit that return predictability based on public information results from investors' collective behavioral biases. For example, Lakonishok, Shleifer and Vishny (1994) reflect a behavioralist perspective in arguing that the value premium arises because under bounded rationality, investors tend to extrapolate past performance, thereby causing pricing errors. Hence, when a reversal happens, out-of-favor value stocks outperform seemingly glamorous stocks, resulting in the value premium (see, e.g. Haugen, 1995).

In general, irrational investor behavior can result in market inefficiencies, oftentimes rationalizing mispricing by either over-reaction or under-reaction by investors to public information. However, a significant challenge to behavioral explanations is whether there are effective limits to arbitrage. Unless there are significant limits (or costs) to arbitrage, factor returns reflecting behavioral biases should disappear over time as investors with arbitrage capital take advantage of other investors' biases. For assets traded in developed, liquid markets where well-informed institutional investors are active, limits to arbitrage are unlikely to be a primary explanation of the predictability of many factors.

In contrast, if factor premia exist as compensation for risk, then they should be persistent over time. The most prevalent argument is that factor premia compensate for risk that the CAPM fails to account for (Fama and French, 1993, 1996). Berk (1995) is an early exponent of the idea that characteristic-based factors capture cross-sectional variation in discount rates due to unmodelled risk. He argues that holding future cash flows constant, smaller firms with lower market capitalizations have higher discount rates. Higher discount rates, in turn, imply higher expected returns. This argument can be extended to value factors constructed using market capitalization, or on a per share basis stock price, as a deflator. Fama (1998) interprets the global value factor as a risk premium, which is priced via discount rates when markets are efficient. However, whether discount rates and expected returns are higher due to (omitted) risk factors or investor mistakes is challenging to ascertain empirically.

Similar debates about the source of factor premia also occur in the literatures relating to other asset classes. To illustrate, Brunnermeier *et al.* (2008) suggest skewness or crash risk as an explanation for the FX carry trade. In contrast, Froot and Thaler (1990) favor a behavioral explanation and dismiss risk-based explanations for the forward discount bias, the key driver of carry trades. Before discussing means to disentangle the likely driver of a given factor, be it risk or mispricing, we first explore more broadly the various risk-based explanations that have been proposed.

Risk-based Explanations for Rationalizing Factor Premia

Investors will be keen to understand the potential risks they might be exposed to when engaging in factor investing strategies. Distress risk is often used to rationalize factor premia. For instance, Chan and Chen (1991) argue that the size effect is primarily driven by firms in distress, characterized by low profitability, high financial leverage, and low dividends. As such variation in returns is not captured by market returns, investors with exposure to size may simply be compensated for taking on distress risk. Fama and French (1996) report similar findings for the role of distress risk in explaining the value premium; they note that firms with high book-to-market ratios exhibit more uncertain future earnings. In a related vein, Kapadia (2011) finds that HML predicts firms' future failure rates, suggesting that the value premium arises from investors requiring compensation for bearing financial distress risk. Finally, distress risk has also been linked to explanations for momentum. For example, (Avramov, Chordia, Jostova and Philipov, 2007) find that the profitability of momentum strategies is driven by firms with high credit risk; similarly, Agarwal and Taffler (2008) suggest that momentum is related to bankruptcy risk.

Further arguments in the literature focus on return-predictive factors being correlated with other sources of priced risk. For example, Campbell and Vuolteenaho (2004) suggest that the value premium represents compensation for bearing cash flow risk, because value firms pay out a greater proportion of capital as dividends and hence have higher book/market ratios. Consequently, value investors face higher exposure to cash flow risk. Related arguments concern duration-based explanations of the value premium. Lettau and Wachter (2007, 2011) and Schröder and Esterer (2016) show that growth stocks have higher future cash flows and cash flow growth. This manifests as higher equity duration (Dechow, Sloan and Soliman, 2004). In turn, longer-duration equities have higher discount rate risk exposure. Gormsen and Lazarus (2020) find evidence of premia related to duration for five equity factors, including value, profitability, investment, low risk, and payout. Guo, Savickas, Wang and Yang (2009) suggest that value premia reflect intertemporal pricing due to strong countercyclical variations in expected value premia. Hence, the value premium tends to be high during recessionary phases and low during expansionary phases. In a related vein, Andronoudis, Dargenidou, Konstantinidi and Pope (2019) use an ICAPM framework to show that R&D intensity, which also appears to attract a premium, is associated with higher equity duration and higher discount rate betas.

Operating leverage has also been associated with the value premium in equities. Donangelo (2020) uses a production-based asset pricing model to explain the role of labor leverage in the value premium. Using firm-level labor shares as a proxy for operating leverage, he finds a positive relationship between labor share and firms' book-to-market ratios. Firms with high labor share are more exposed to priced risk, and thereby offer higher risk premia to investors.

Aretz and Pope (2018) also adopt a production perspective, through a real options model that values capital assets as a portfolio of production options. As a result of capital investment being costly to reverse, firms invest conservatively, but nevertheless can ex post have higher levels of assets in place than is optimal based on observed uncertain demand. The optionality elasticities of investment and production options causes the betas of equities to depend on the past history of demand and investment decisions. Aretz and Pope (2018) show that a measure of capacity overhang captures the optionality of equities and is a strong predictor of returns that helps explain momentum and profitability factors in pricing the cross-section of equities.

Disaster risk is less commonly studied, but it is another potential reason for the existence of factor premia (Rietz, 1988). Following Barro (2006), Nakamura, Steinsson, Barro and Ursúa (2009), Gabaix (2012), and Berkman, Jacobsen and Lee (2011, 2017), Siriwardane (2013) examines the links among value, size, and momentum premia and disaster risk, finding that the latter plays a role in explaining the cross-section of returns of the corresponding portfolios. The limited research in this area may reflect the inherent rarity of such events. Other risks that have been related to factor returns include illiquidity risk, inflation risk, country risk (Zaremba, 2016), economic risk, such as changing volatility (Lettau, Ludvigson and Wachter, 2008) and income inequality (Gollier, 2001; Hatchondo, 2008), and political risk due to policy uncertainty (Pástor and Veronesi, 2020) or unstable governments (Lam and Zhang, 2014).

Delineating Risk and Mispricing

To empirically distinguish between alternative rationales for factor premia, scholars study their out-of-sample predictive performance (McLean and Pontiff, 2016; Linnainmaa and Roberts, 2018; Bartram *et al.*, 2020b). This research compares factor returns in an original estimation sample period ("in-sample period"), the period between the end of the sample (in which the factor was identified) and the posting of the paper on SSRN ("out-of-sample period"), and the period after posting on SSRN ("post-publication period") in order to uncover trends in cross-sectional return predictability. The idea is that any factor performance due to statistical biases is likely to disappear outside the original sample period. Similarly, factor returns that reflect mispricing are expected to decay or disappear in the post-publication period, if sophisticated investors seek to arbitrage the revealed predictability. Conversely, publication should not affect factor payoffs that are compensation for risk if assets remain fairly priced given that risk.

In this vein, McLean and Pontiff (2016) and Bartram *et al.* (2020b) reject risk-based explanations in favor of mispricing-based explanations by documenting a significant decrease in post-publication profits of many anomalies in equity and currency markets, respectively. For equity markets, the empirical evidence shows a 58% reduction of anomaly returns after publication (McLean and Pontiff, 2016) and in recent years due to increased trading activity of hedge funds and lower trading costs (Chordia, Subrahmanyam and Tong, 2014). The return decay is larger for predictors with lower arbitrage costs. McLean and Pontiff (2016) also report significant correlations between yet-to-be published predictors, although such relatedness decreases after publication. In contrast, Jacobs and Müller (2020) suggest no decay of factor performance for stock markets outside the United States. Similarly, Jensen, Kelly and Pedersen (2021) develop a Bayesian framework to replicate 153 equity factors in 93 countries and find very little post-publication decay in factor performance. Such contradictory evidence within the US and across the globe calls for further examination of post-publication factor performance.

Since performance in out-of-sample and post-publication periods could both be affected by statistical bias, Linnainmaa and Roberts (2018) investigate pre-estimation sample period returns for 36 factors. They find that many factors, including profitability and investment, are significant for the in-sample period (1970–2004), but are insignificant in the pre-sample period (1926–1969). In a similar research setup, Wahal (2019) extends the sample back to 1926 and finds evidence for the existence of the profitability factor but not for the investment factor. This evidence is consistent with data-snooping biases for most factors, although alternative explanations such as changing macroeconomic regimes cannot be ruled out.

While research has traditionally interpreted risk-adjusted returns as evidence of mispricing, the success of this approach critically depends on the validity of the risk model. To this end, recent research questions interpretations of reductions in factor premia as evidence of mispricing, suggesting that the evidence could also be consistent with time-varying compensation for risk. For instance, Kelly, Pruitt and Su (2019) develop an instrumented principal component analysis (IPCA) allowing for latent factors and time-varying factor betas. Their method introduces observable characteristics as instruments for unobservable dynamic factor betas. In evidence based on U.S. equity market data, only 4 of 37 anomalies have IPCA alphas that are significantly different from zero, suggesting that many anomaly factors documented in the literature capture time-varying risk premia as opposed to reflecting market inefficiencies. However, Bartram and Grinblatt (2021) show that trading global stocks based on a regression-based measure of mispricing yields significant risk-adjusted returns. This holds true when controlling for traditional factor models (including all 50 factors from the Fama-French data library or their own 80-factor model). It also holds when using IPCA to control for time-varying risk premia even tied to mispricing itself.

Researchers have also employed other techniques to distinguish mispricing from risk-based explanations. Shleifer and Vishny (1997) argue that the returns of mispricing anomalies should be significantly higher for stocks with higher limits to arbitrage, such as those with smaller market capitalization and lower institutional ownership. Stambaugh, Yu and Yuan (2012) document an increase in mispricing-based anomaly returns during high-sentiment periods. This implies the existence of mispricing due to overpricing, which is exacerbated by short-selling constraints. However, consistent with Berk (1995), higher prices may also be consistent with lower discount rates, and this observation leads proponents of sentiment-based explanations for factor returns to distinguish between variation in sentiment as a behavioral phenomenon and the variation in discount rates as an economic phenomenon.

Beyond the existence of significant risk-adjusted returns, relatively fast decay of signal ranks and performance are also more consistent with a mispricing explanation for factor premia than for risk, as evidenced for the agnostic mispricing measure by Bartram and Grinblatt for U.S. Bartram and Grinblatt (2018) and global equities Bartram and Grinblatt (2021), and for currency predictors Bartram *et al.* (2020b). The study of book-to-market effects in corporate bond returns may also aid in the understanding of why it influences asset returns more generally, since the future cash flows of corporate bonds, particularly senior bonds, are far less risky than their equity counterparts (Bartram, Grinblatt and Nozawa, 2020c). Consequently, bond price movements have to arise largely from discount rate variation rather than from changes in projections of future cash flows. Nevertheless, delineating between risk and mispricing explanations of return predictors remains a challenge, and more powerful tests are required to distinguish between these competing explanations for factor returns.

1.4.2 The Dangers of Data Mining

Rationalizing a given factor's efficacy through either risk or mispricing explanations is only a meaningful exercise if factor performance is not statistically spurious to begin with. Of course, the collective efforts of a multitude of academic and practitioner researchers scanning the limited datasets available to them for significant patterns of return predictability suggests that false positive, and hence spurious, in-sample results are to be expected. Such collective data mining raises concerns about the out-of-sample success and thus the usefulness of return predictive factors. To mitigate data mining concerns requires more rigorous empirical testing, including out-of-sample analysis, controlling for the statistical effects of multiple hypothesis testing on the same data, as well as plausibility checks based on well-defined theoretical economic priors.

Early Studies Accounting for Data Snooping

As computing power has grown rapidly, the risk of data mining (or snooping) has become more pronounced. Despite being recognized as a potential problem nearly a century ago (Cowles, 1933), concerns about data snooping biases have only impacted the asset pricing literature quite recently. One widely cited exception is Lo and MacKinlay (1988) who caution against the increased likelihood of data-mined results with the increase in the number of publications in any given field. In related research, Lo and MacKinlay (1990) stress the potential importance of data-grouping techniques in determining the performance of return predictors by documenting a significant difference between tests of data-driven models and theory-driven models.

To illustrate the danger of identifying spurious factors, Ferson, Sarkissian and Simin (1999) show that a simulated alphabetically-sorted portfolio earns excess returns mimicking value premia, despite there being no obvious connection between the first letter of a company's name and its expected return. This example illustrates the importance of a sound economic foundation, not just statistical significance in quantile portfolio spreads. Patton and Timmermann (2010) further emphasize the importance of examining the expected returns of all portfolio quantiles when developing a trading strategy. Instead of examining the performance of extreme quantile portfolios which is common in empirical asset pricing research, they recommend that researchers should test for monotonicity of returns across all quantiles to provide greater support for systematic relations between future returns and the sorting attribute. As observed by Romano and Wolf (2013), this test may not work if expected returns follow a non-monotonic relation or are weakly increasing, and the authors provide an alternative test that is immune under such circumstances.

Researchers have also identified more subtle channels through which data mining may manifest. Sullivan, Timmermann and White (1999) explain the dangers of reusing the same data for inferences or model selection in an attempt to achieve satisfactory results. Using bootstrapping techniques, they perform a reality check on 7,846 trading rules and find that their best in-sample trading rule (the five-day moving average rule) is an insignificant predictor out-of-sample. In a related paper, Sullivan, Timmermann and White (2001) draw attention to the danger of using the same data for formulating and testing a hypothesis, thereby inadvertently increasing the chances of data mining. By using 100 years of daily data and bootstrapping techniques, they test calendar anomalies that do not have a strong theoretical motivation. The evidence suggests that calendar rule-based strategies such as the Monday effect are not as significant as originally suggested.

Overall, while individual researchers might limit their research to one or a small number of factors, data snooping biases will arise to the extent that an individual researcher retests multiple specifications for the same underlying construct, e.g., different definitions of value or momentum. Similarly, data snooping bias arises at the aggregate level as a result of different researchers investigating different potential factors but using the same data. This body of research suggests that it is important to control for data-snooping biases in evaluating claims that a factor successfully predicts asset returns.

Statistical Methods to Mitigate Data Snooping Concerns

To account for multiple testing, the statistical literature advocates controlling for the family wise error (FWE) or the false discovery proportion (FDP). The FWE represents the probability of making at least one false positive (type I) error in the family of tested factors. The FDP is a less demanding test focusing on the number of false positive results within the set of positive results. Bonferroni (1936)'s well-known test to control for the FWE involves a p-value adjustment that divides the significance level by the number of hypotheses to be tested. Obviously, given the typically low single digit t-statistics obtained in return predictive regressions, if one tested the significance of hundreds of factors (as researchers have done collectively), one would reject the statistical significance of most if not all factors if one applied this adjustment. Fortunately, the literature has developed more powerful tests, such as the StepM-method of Romano and Wolf (2005) that incorporates the dependence structure of test statistics. Leippold and Lohre (2012a,b) are early adopters of such methods in accounting for multiple testing when investigating market anomalies across the globe. They provide an implicit proof of concept of the method's power by documenting the robustness of momentum factors but not accruals factors using a battery of multiple testing procedures.

Harvey *et al.* (2016) implement the FDP testing framework to re-examine cross-sectional return patterns in equities. Of the 316 factors tested, they find between 80 and 158 false discoveries depending on the choice of statistical test. Taking this thinking to extremes, Chordia, Goyal and Saretto (2020) study over two million trading strategies from random combinations of accounting variables and basic market variables to test for the presence and magnitude of data mining. Using stricter tests, only 17 (0.04%) of 2.1 million strategies survive. In developing recommendations to underpin a rigorous testing protocol, Harvey *et al.* (2016) suggest applying higher test statistic thresholds for testing new factors, perhaps using Bayesian adjusted p-values to guard against p-hacking. Critical t-statistic thresholds would increase to 3.0 or even higher under such an approach. Recently, Bryzgalova, Huang and Julliard (2020) use a Bayesian framework to analyze 2.25 quadrillion models and conclude that only 3 factors ('HML' value, adjusted size and adjusted market) appear to be robust.

Focusing on the independent contributions of new factors that are often correlated with "old" factors, Green *et al.* (2017) suggest that the returns of new candidate factors may need to be orthogonalized against the returns of some but not all pre-existing predictors in order to establish the contribution of new return-predicting signals. In contrast, Bartram and Grinblatt (2018, 2021) bias against data mining by taking an agnostic approach, where modelling choices are non-discretionary and only based on data availability and statistical criteria, biasing against finding predictability.

Overall, recent literature emphasizes the importance of both rigor of statistical methodology combined with solid economic theoretical support in evaluating the performance and contribution of candidate return-predictive factors. The theoretical origins of many factors, including many surviving strategies, are still not well enough understood. A combination of better theoretical support and statistically rigorous backtesting will help researchers interested in factor investing mitigate some of the skepticism that they increasingly encounter from both the academic "gatekeepers" and investment professionals.

1.5 Reality Checks for Factor Investing

Practitioners and institutional investors also need to consider the implementation and feasibility of pursuing factor-based investment strategies. Implementation costs, which include direct costs (such as manager fees), indirect costs (including trading costs), and investment constraints (such as leverage constraints, turnover, limited capacity) can curtail investors' appetite for adopting factor-based allocation schemes. Concerns about capacity, measuring how much can be invested in a factor before the additional inflows lead to price pressure and a decline in returns, have been an issue for the past few years. But many of these real-world concerns are often neglected in academic work because academics lack the relevant data, and also because investors face different constraints and costs, rendering implementation cost assumptions somewhat subjective. We also discuss integrating the notion of sustainable investing in factor-based investment strategies.

1.5.1 Measuring Factors in Practice

While the academic literature has put forward different metrics to measure an asset's exposure to returns factors, in practice, one would often see a combination of such metrics. For instance, quality factors such as profitability and investment are often backed by several balance sheet indicators, including leverage, earnings quality, return on equity, accruals, asset turnover ratio, and even ESG (Environment, Social and Governance) based measures that could potentially reflect the safety of a long investment and its susceptibility to large negative returns outcomes. Hsu, Kalesnik and Kose (2019) express concerns about the definition of quality, commenting that different index providers (e.g., MSCI, FTSE Russell, S&P, Research Affiliates, EDHEC, and Deutsche Bank) have their own combination of signals for measuring the quality premium.

When examining the seven most prominent attributes used by index providers for constructing the quality factor, Hsu *et al.* (2019) find profitability, accounting quality, payout/dilution, and investment to be more reliable sources of the quality premium than capital structure, earnings stability, and growth in profitability. They emphasize the importance of thorough analysis of each potential signal of a company's 'quality', the robustness across international markets, and the ability of traditional and non-traditional metrics in capturing the underlying concept being estimated. Thinking of each signal as a noisy indicator of the underlying unobservable construct, composite factors can be thought of as a diversified combination of signals, the performance of which will be enhanced if signals are informative (less noise) and combined in ways that eliminate the overall noise in the measure as much as possible.

Similar diversification considerations apply in measuring value factors. For instance, the S&P 500 Enhanced Value Index combines three fundamental measures, book-to-price, earnings-to-price and sales-to-price. Yet, in the context of the flagging equity value premium, a recent debate has discussed whether such traditional ratios are still adequate to assess relative company valuation. Arnott, Harvey, Kalesnik and Linnainmaa (2021b) highlight the growing importance of off-balance sheet intangibles, especially in the age of technology-dominated firms. The book equity component of the traditional price-to-book ratio does not capture most internally generate intangible assets, and this has implications for misclassification of growth and value stocks. Therefore, Park (2019) suggests incorporating intangibles into standard value metrics by calculating the intangible-adjusted book-to-market ratio (ibook-to-market). This measure not only outperforms the traditional book-to-market ratio, but also survives in periods when value was reported to be "dead" as a return predictor. While this offers hope for the continued existence of value, it also stresses the need for revisiting traditional factor definitions as business models, accounting recognition, and measurement rules change over time.

1.5.2 Environmental, Social and Governance Investing

Sustainable investing aims to create a positive impact on the environment (E), society (S) and corporate governance (G). Unlike traditional style factors, ESG investing focuses on non-financial characteristics, e.g. climate change, waste management, energy efficiency, human capital, labor management, corporate governance, gender diversity, privacy, or data security. While environmental factors such as climate change and waste management attempt to encompass the financial and firm-level consequences of global warming and greenhouse gas emissions, social factors such as human capital and labor management measure a company's adherence to good workplace practices. With some of the world's largest institutional investors such as the Government Pension Investment Fund of Japan, Norway's Government Pension Fund Global, and the

Dutch pension fund ABP investing trillions of dollars in sustainable investing, it is important to understand the ways in which investors can integrate ESG objectives in their investment mandate without negatively impacting the risk-adjusted returns of their portfolio.

A survey by Krueger, Sautner and Starks (2020) documents that institutional investors believe climate risk to play an important role in the future performance of the firms in their portfolio. Surveying about 413 senior investment professionals who represented about 43% percent of the global institutional assets under management, Amel-Zadeh and Serafeim (2018) report that institutional investors consider ESG ratings as a proxy for management quality and strongly believe that the rating reflects a firm's reputational, legal and regulatory risk. They also report that institutional investors perceive stocks with good ESG ratings to be underpriced, and hence such stocks may offer higher returns than stocks with poor ESG ratings.

However, given the challenges involved in quantifying ESG-related information from firm disclosures, associated academic research has explored the complexities in defining and measuring ESG factors and formulating profitable strategies from the same. In their survey article, Liang and Renneboog (2020) trace the relationship between sustainable investing and firm performance and discuss the caveats of relying on the ratings of ESG rating agencies to compute ESG factor scores. The low correlation between the ESG ratings of different providers underscores the inconsistencies in such ratings. Such a divergence in ESG ratings across vendors and lack of established metrics to measure each of the sustainability topics are some of the reasons why ESG factor research is still inconclusive. Still, from a practitioner perspective, one needs to learn about the return effects of ESG objectives and be in a position to integrate such objectives in portfolio construction without harming risk-adjusted returns.

1.5.3 Real-World Frictions in Implementing Factors

The Role of Illiquid Small-Cap Stocks

Factor returns in academic studies have often been assessed using equal-weighted portfolios rather than value-weighted portfolios, causing small-capitalization stocks to play a more significant role. This can be important if factor returns depend on firm size, and because implementation costs are generally higher for small companies where stock market liquidity is lower. To assess the importance of equal- versus value-weighting, Hou *et al.* (2020) 452 cross-sectional returns results predictors that have been documented in the U.S. equity market. Overall, 64%-85% of these anomalies become statistically insignificant when using more realistic value-weighted portfolios and accounting for liquidity, market microstructure effects, and other trading frictions. The authors document an over-representation of micro-cap stocks and point out how the majority of the evidence from factor studies can be attributed to microcap stocks and hence are unlikely to be exploitable by institutional investors. In related research, Green *et al.* (2017) study 94 firm characteristics to identify meaningful factors in the cross-section of U.S. stock returns and find that only 25% of them are statistically significant and even fewer after excluding microcap stocks. Several other empirical studies have also noted that the exclusion of microcap stocks leads to similar conclusions, thereby raising questions about the translation of paper profits into real institutional investment gains.

Transaction Costs

The increased adoption of factor strategies has necessitated the need for real-world cost considerations, increasing the hurdles for factor adoption by practitioners. For instance, some might argue that the high gross returns of a given factor are simply reflective of the transaction cost necessary to arbitrage the very effect. Indeed, building on a mispricing measure (estimated as the deviation of firms' market values from intrinsic values), Bartram and Grinblatt (2021) study the relation between mispricing and transaction costs in 36 countries and find that gross alphas are positively related to transaction costs. Thus, environments with higher transactions costs also see higher risk-adjusted profits from mispricing. In this sense, limits to arbitrage could explain the Bartram and Grinblatt (2021) mispricing factor and potentially many other anomalies. However, Bartram and Grinblatt (2021) also show that trade implementation is important, and that in some parts of the world, notably Asia Pacific and emerging markets, risk-adjusted profits for their mispricing measure remain significant even after accounting for institutional investors' estimated trading costs.

To explore whether payoffs to return predictors survive implementation costs, some researchers use market wide data, such as NYSE Trade and Quote (TAQ) to estimate transaction costs, while others use proprietary trading data to assess performance net of trading costs. For instance, Lesmond, Schill and Zhou (2004) use TAQ data to estimate trading costs. They find that the strategies that generate the highest momentum returns are also those ones with the highest trading costs and conclude that momentum trading is not profitable net of costs. A related perspective is provided by Korajczyk and Sadka (2004) who estimate that the price impact of a fund with over \$5 billion under management implies trading costs that exceed the abnormal returns of momentum strategies.

Novy-Marx and Velikov (2016) use Hasbrouck (2009)'s bid-ask spread measure to estimate transaction costs for 23 return predictive factors. They report average trading costs for size, value and momentum strategies, respectively, of 6 bp, 5 bp and 48 bp per month, suggesting that size, value and momentum strategies survive transaction costs. The value-weighted annual returns of value and size strategies drop by roughly 60 bp (from 5.64% to 5.04% for value and from 3.96% to 3.36% for size), while the return on momentum drops from 16% to 8%. Frazzini, Israel and Moskowitz (2018) use live proprietary trading data across international equity markets and find that real-world costs based on live trades are very different than those estimated in the literature from daily or intra-day data. Using this data, Frazzini, Israel and Moskowitz (2014) show that size, value and momentum survive transaction costs, unlike the short-term reversal factor, which has high turnover.

Hence, we observe a disconnect between academic work, that tends to conclude that transaction costs erode most the factors' excess returns, and practitioner work which argues that academics are too conservative. This disconnect is perhaps related to many academic studies implicitly assuming a too aggressive trading style that can often lead to crossing the spread. Conversely, a passive execution style seems to curb market impact while retaining most of the factor signals' predictability. In this context, Frazzini *et al.* (2018) point out that in their live trade data, trades are often executed as limit orders that remain on the book for some time. In practice, portfolio construction naturally needs to adapt to portfolio size for not trading a significant part of a given stock's available dollar volume.

Such analyses help identify whether single factor strategies are really profitable on a standalone basis. Identifying the transaction cost drivers has been shown to also positively alter multi-factor-based portfolio allocations by enhancing the diversification potential. DeMiguel, Martin-Utrera, Nogales and Uppal (2020a) examine the impact of transaction costs using characteristics-based factor selection with a lasso penalty. Before transaction costs, 6 of 51 characteristics are significant in a multivariate parametric portfolio policy, while 15 are significant after transaction costs. This seemingly counter-intuitive result is due to the increased benefits from trading diversification due to offsetting positions attributable to different stock characteristic-based factors. The authors argue that trades in the underlying stocks required to rebalance different characteristics often cancel each other out, and thus, combining a larger number of characteristics allows one to substantially reduce the quantum of transactions and hence transaction costs. Such findings from the literature may aid in improving the profitability of factor strategies.

Capacity Constraints

The capacity constraints of many factor-based investing strategies also have important implications for realistic assessments of factor performance. While different definitions of capacity exist in the literature, the capacity of a strategy can be understood as the volume of additional capital that can be invested in a strategy before it becomes unprofitable. Thus, capacity constraints reflect limits on the size of trades intended to increase factor exposure due to market impact effects. This issue is investigated by Novy-Marx and Velikov (2016), who document a negative relation between factor turnover and capacity: low turnover strategies, such as value, size, and profitability, have capacities of about \$21 billion, \$20 billion, and \$131 billion, respectively, while mid-turnover strategies such as idiosyncratic volatility and momentum have capacities of about \$1.51 billion and \$5 billion, respectively.

Ratcliffe, Miranda and Ang (2017) use proprietary high frequency trading data and find that the capacity of factor strategies varies with the trading horizon. Considering a 1-day trading horizon, the MSCI USA Minimum Volatility index has a break-even capacity of around \$1.3 trillion. For a 5-day trading horizon, it jumps to \$6.7 trillion. Other factors such as size, value, quality, momentum, and even multi-factor strategies display similar behavior, suggesting that they have very high capacities. However, practitioners criticize this study because of highly concentrated positions associated with higher capacity. An active factor investing strategy that trades multiple times a year seems to be a better alternative (Blitz and Marchesini, 2019).

1.5.4 Factor Crowding

Crowded trades have been of concern to institutional investors as such crowding has been shown empirically to hamper factor performance. Factor crowding occurs when any given factor experiences a huge influx of investment. For example, Dimson and Marsh (1999) note the disappearance of the size premium after a surge in the popularity of small cap funds. Momentum has been similarly criticized for being vulnerable to poor performance after gaining popularity. Researchers have supported this view by identifying that momentum's performance is strongly affected by crowding (Lou and Polk, 2014). Concerns about capacity in factor investing have been exacerbated in recent years as a result in the growth of investment in exchange-traded funds tracking style factor indices. Crowded factors not only perform poorly but may also experience increased volatility and drawdowns.

As crowded trades continue to pose a concern for practitioners, the academic literature has identified different ways of measuring crowdedness in factors. Since the ideology of factors stems from exploiting common patterns, a reliable measure of crowdedness could be identifying the number of investors chasing (or wanting to exit) a particular strategy. Given significant commonality of factors in their alpha models, institutional investors are likely to hold similar stocks and affected by crowded trades, e.g. in case of fire sales (Jotikasthira, Lundblad and Ramadorai, 2012). Common ownership of international stocks is a predictor of returns that can be quantified in an institutional ownership return measure (Bartram, Griffin, Lim and Ng, 2015).

Interestingly, different asset classes are impacted differently by factor crowding. Baltas (2019) studies the impact of market crowding on equity momentum, equity low beta, equity quality, FX momentum, and commodity momentum. In line with the views of Stein (2009), Baltas (2019) finds that FX momentum and commodity momentum face lower drawdowns than equity momentum over six-month to one-year investment horizons when strategies are crowded. However, in the subsequent year, equity momentum strategies begin to outperform, indicating underperformance is short-lived. Relatedly, DeMiguel, Martin-Utrera and Uppal (2020b) show that crowding concerns can be alleviated by trading diversification and other institutions exploiting strategies that, when implemented concurrently, reduce their price impact. Overall, the available evidence suggests that factor crowding may not be a valid reason for long horizon investors to avoid factor investing.

1.5.5 Short-Selling

The long-short investment strategies commonly investigated in academic return-predictive factor studies inherently assume that both the long leg and the short leg contain relevant factor-related information. Hence, investors can enjoy factor-related returns relative to a cash benchmark through zero investment long-short hedge portfolios; or relative to a benchmark index by overweighting (underweighting) stocks in the long (short) leg relative to benchmark weights. This methodology has been subject to criticism as long and short portfolios may be subject to different return dynamics. To illustrate, the long and short portfolios of value, size, and momentum strategies exhibit differential exposures to term structure risk (Aretz, Bartram and Pope, 2010). This leads to an asymmetric behavior of the long and short side, which has been repeatedly observed in the academic literature.

Whilst the contribution of the short leg to the performance of long-short strategies is evident from several academic findings, investors might be limited in their ability to short stocks due to short selling constraints, borrowing costs, risks of short-selling in the form of short squeezes, etc. These limits to arbitrage can prevent sophisticated investors from trading profitably against anomalies (Miller, 1977). Indeed, Stambaugh *et al.* (2012) show that Fama-French 3-factor alphas are larger for the short leg than the long leg of the investment strategy for all but one of eleven anomalies. They further show that the short leg returns are lower when market sentiment is high. In fact, they suggest that short-selling could even enhance factor performance when combined with market sentiment and note an increase in anomaly returns especially during high-sentiment periods as the extant mispricing in those anomalies would be higher, translating into higher returns. Similarly, Chu, Hirshleifer and Ma (2020) document the causal effect of short-selling constraints on asset pricing anomalies, with the introduction of short-selling constraints shown to affect only the short legs of anomaly portfolios significantly reducing risk-adjusted long-short portfolio performance.

Shorting stocks also entails borrowing costs and risk associated with the liquidity of stocks in the short leg (Diether, Lee and Werner, 2009). Kim and Lee (2019) report shorting costs to be 0.10% per month which is about 40% of gross long-short returns of 14 factors such as return on equities, return on assets, momentum, etc. Limitations to the ability to liquidate short positions or 'short squeezes' expose investors to unintended sources of risk, which can be circumvented by underweighting the stocks in the portfolio. Such findings are often overlooked and are an important limitation in translating many academic findings into practice (Patton and Weller, 2020).

1.5.6 Are There Benefits to Factor Timing?

Time variation in factor performance presents a major challenge to institutional investors since factors can experience extended periods of underperformance. While academics have the luxury of being able to look at long-term averages, the need to document favorable performance to clients over relatively short reporting intervals creates risk for real-world asset managers. Consequently, it is conceptually appealing to avoid such painful episodes of underperformance by actively timing factor weights in an investment portfolio, moving into a factor when it is likely to perform well and out when it is expected to underperform.

Recent factor timing research focuses on utilizing past factor performance to predict future factor performance. To illustrate, Avramov, Cheng, Schreiber and Shemer (2017) show that short-term factor momentum strategies outperform an equal-weight factor allocation in a universe of 15 well-known factor strategies. In a similar vein, Gupta and Kelly (2019) offer international factor momentum evidence in a comprehensive set of 65 characteristics-based factors around the globe, where factor momentum is found to add significantly to investment strategies based on traditional momentum, industry momentum, value, and other commonly studied factors. Similarly, Arnott, Clements, Kalesnik and Linnainmaa (2021a) find that factor momentum is pervasive across all factors and that factor momentum fully subsumes industry momentum. Ehsani and Linnainmaa (2019) explain how momentum in factors translates to momentum in individual stocks and argue that factor momentum fully explains individual stock momentum.

Expected factor performance, and hence factor weights, are related to general economic and market conditions–for example momentum and value tend to perform better in bull markets, while quality and minimum volatility perform better in bear markets. As a case in point, factor timing became very popular during the global financial crisis, where markets were more driven by policy and macroeconomic events rather than firm fundamentals. Interest-rate regimes (Muijsson, Fishwick and Satchell, 2014) and business cycles (Grant, Ahmerkamp and Kosowski, 2012) have also been shown to affect the performance of different factors. Grant *et al.* (2012) document strong predictability for carry and momentum strategies with the business cycle (using dividend yield, short rate, term spread and default spread as instruments) and liquidity indicators have predictive power for factor returns across a range of asset classes.

While some of this evidence suggests that factor timing is possible if the economic and market determinants of factor performance can be anticipated in advance, the evidence is mixed on the possibility of timing factors profitably in practice. In this regard, Dichtl, Drobetz, Lohre, Rother and Vosskamp (2019) explore the value-added of active factor allocation strategies for investable global equity factors. They find equity factors to be related to lagged fundamental and technical time-series indicators and to characteristics such as factor momentum and crowding. Yet, such predictability is difficult to exploit after transaction costs. The consensus among many practitioners is that the higher turnover and associated costs of dynamic factor allocation strategies outweighs the gross return benefits that can be expected (Asness, Chandra, Ilmanen and Israel, 2017). Supporting the findings of Van Gelderen and Huij (2014), Van Gelderen, Huij and Kyosev (2019) argue that investors are better-off by choosing a buy-and-hold strategy compared to dynamic factor allocation strategies. Unfortunately for investors, factor timing seems ‘clear in hindsight but hazy ahead’ (Vanguard research, 2019).

1.6 Conclusion

The overarching findings from academic factor studies can be summarized as follows: the empirical asset pricing literature has spawned a multitude of factors to explain the cross-section of returns, and the past three decades have witnessed a heightened proliferation of new factors (Cochrane, 2011). However, only a small number of dominant factors survive after careful and rigorous testing of significance, controlling for data snooping and other research design biases, considering real-world constraints, and careful examination of the incremental contribution of specific factors to return predictability. Identifying such ‘true’ factors is challenging, especially in light of data-mining concerns.

Recent meta-studies have proposed new ways of dealing with these issues. Most are fueled by data-driven and computationally intensive methods that would deem many of the factors insignificant. In the emerging age of alternative (and potentially big) data, such data mining concerns are likely to be exacerbated as researchers constantly conceive new factors, e.g., by applying ML techniques to big data and NLP techniques to unstructured text data or invoking the various facets of available ESG criteria. At the same time, improved understanding of the risk-return relationship helps in uncovering the underlying common factors, the associated premia, and the competing explanations for their existence. Factors that capture these premia are expected to aid institutional investors in structuring portfolio allocations.

Although not immune to episodes of poor performance, factor investing has survived turbulent times such as the global financial crisis and the ongoing Covid crisis. Part of this relative success is due to factors’ underlying building blocks not moving in lockstep and thus offering diversification and downside protection benefits. Whereas, the longer-term performance of factor strategies arises from the reward associated with bearing risk, exploiting structural impediments, or behavioral biases of investors. A factor-based approach can cater to institutional investors' specific risk-return objectives at improved transparency. Indeed, multi-asset multi-factor-based investment approaches can help maximize portfolio diversification relative to traditional asset allocation by combining asset class and style factor.

From an institutional investor perspective, many factors discovered in the academic literature may not be exploitable under real-world conditions. Taking into account such challenges in translating paper profits to reality, this literature points to the importance of parsimonious and implementable asset pricing models. Rigorous testing procedures have found that the premia of spurious factors vanish for value-weighted portfolios or after excluding small-cap stocks. Any remaining profits are often accounted for by limits to arbitrage and related transaction costs. Apart from rare attempts to measure mispricing, distinguishing between risk and mispricing remains a key challenge, with some recent work suggesting that common anomalies capture time-varying factor risk. It is still of huge concern which factors are likely to pass all real-world tests, especially to investors, who are often bewildered by all the different possible options.

Appendix 1.A Data and Methodology

We compute and compare the performance of various style factors across different asset classes, as presented in Table 1.B.3 and Figure 1.C.4. The underlying data used for calculating the metrics in Table 1.B.3 and Figure 1.C.4 is provided in Table 1.B.1. Specifically, we utilize data from Bloomberg to compute the performance metrics for Equity style factors and Commodity Quality, while all other non-equity style factors are sourced from Goldman Sachs (GS) via Bloomberg Terminals.

Using the raw data from above mentioned sources, we first compute monthly returns by aggregating daily data. We then compute annualized geometric returns as:

$$Return = \sqrt{prod(1 + R_m)^{12}} - 1$$

and annualized volatility of returns as follows:

$$Volatility = \sigma_{R_m} * \sqrt{12}$$

where R_m is the monthly returns and σ_{R_m} is the sample standard deviation of monthly returns. Using respective 3-month LIBOR rates as benchmarks or risk-free rates R_b, we compute the Sharpe ratio as follows,

$$SharpeRatio \quad = \frac{R_m - R_b}{\sigma_{R_m}}$$

Table 1.B.3 also reports the Active returns, tracking error and information ratio with respect to the corresponding market return for each of the regions. Active returns are simply the difference between the respective factor return and corresponding market return. Tracking error measures the volatility of the difference between the portfolio return R_i and the market return R_{mkt} and is computed as,

$$\text{Tracking Error} = \sqrt{\frac{1}{\sqrt{12} * n} \sum (R_i - \bar{R}_{mkt})^2}$$

where n is the length of the period under consideration. Whereas Information Ratio measures the degree to which an investment has beaten the benchmark and is given by,

$$\text{Information Ratio} = \text{Active Returns}/\text{Tracking Error}$$

We source data from different style factor indices for different asset classes and across multiple regions (for equities) and we compute the performance metrics as outlined above. We only use long-only factor indices and although detailed construction of each of the style factor indices can be found in the handbooks of the respective index providers, we outline key variables used by index providers in the following subsections:

1. **Equity Style factor definitions**

    a) We use the MSCI ACWI Value index for the *Equity Value* factor. The value investment style characteristics for index construction are defined using the following three variables: book value-to-price, 12-month forward earnings-to-price, and dividend yield. Combining these three metrics with an equal weighting of 1/3, MSCI computes each securities' value and growth score and allocates the security into a value or a growth index. Security weights are computed by sorting securities according to their distance from the origin and other investment constraints, see MSCI Index methodology documents[1] for more specific details. The index is reviewed and rebalanced semi-annually in order to reflect changes in the underlying markets, creating only limited index turnover.

    b) We use the MSCI ACWI Quality index for the *Equity Quality* factor, which uses three fundamental variables: high return on equity (ROE), stable year-over-year earnings growth, and low financial leverage. Offering high trading liquidity and investment capacity, the quality index is constructed using a composite Z-score computed with an equally weighted average of the Z-scores of these three variables. The weights of individual securities are simply the product of this quality Z-score and market capitalization weight of the security in the parent index. The final quality score is obtaCapping issuer weights at 5%, the index has moderate index turnover and is rebalanced semi-annually.

    c) We use the MSCI ACWI Momentum index for the *Equity Momentum* factor, which is based on the MSCI momentum index uses the securities' recent 12-month and 6-month risk-adjusted price performance to compute a

[1] Index methodology documentation for all the MSCI indices can be downloaded from Index methodology - MSCI

momentum score. MSCI computes the security weights by multiplying the equal-weighted momentum score with the market capitalization weight in the parent index.

d) We use the MSCI ACWI Minimum Volatility index for the *Equity low-volatility* factor, which is based on the MSCI low volatility index which aims to minimize the return variance for a given covariance matrix of returns and hence takes a different approach with respect to other style factors. The index targets lower beta and volatility than its parent index and uses the Barra optimizer to optimize its parent index for the lowest absolute volatility with a certain set of constraints.

2. **Non-Equity Style factor definitions**

All non-equity style factors and their definitions were sourced from Goldman Sachs except for Commodity Quality.

2.1. **Rates factors**

a) The *Rates Momentum* factor capitalizes on the persistence of trends in short- and long-term interest rate movements. On a daily basis, the strategy evaluates the recent performance of a number of futures contracts for the U.S., Germany, Japan, and the U.K. It then takes either a long or short position on each, depending on whether actual performance has been positive or negative.

b) The *Rates Quality* factor capitalizes on the observation that risk-adjusted returns at the short end of the curve tend to be higher than those at the long end. A leveraged long position on the former versus the latter tends to capture positive excess returns as compensation for the risk premium that stems from investors having leverage constraints and favoring long-term rates. The interest rates curve strategy enters a long position on five-year U.S. bond futures, and a short position on thirty-year bond futures, as well as a long position on five-year German bond futures and a short position on ten-year German bond futures, rolling every quarter. The exposure to each future is adjusted to approximate a duration-neutral position.

c) The *Rates Value* strategy attempts to capture the bond risk premium as compensation beyond the expected rates path. The risk premium of

Rates Value is defined as the difference between bond yield (using 10-year futures) and inflation expectations (provided by Consensus Economics) for a number of G8 government bonds. The risk management of the strategy involves the construction of a portfolio on a daily basis to maximize the exposure to the risk premium, subject to constraints on risk, leverage and potentially beta to a benchmark. The execution is smoothed over 22 days with a view of limiting turnover and transaction costs.

d) The *Rates Carry* strategy benefits from upward-sloping yield curve as compensation for bearing duration, inflation and illiquidity risk. Higher carry tends to be compensation for being long riskier assets. The risk premium of Rates Carry is defined as the difference between bond yield (using 10y futures) and funding cost plus the roll down of a number of G8 government bonds. The risk management of the strategy involves the construction of a portfolio on a daily basis to maximize the exposure to the risk premium, subject to constraints on risk, leverage and potentially beta to a benchmark. The execution is smoothed over 22 days with a view of limiting turnover and transaction costs.

2.2. **FX factors**

a) The *FX Carry* strategy benefits from the overestimation of the actual depreciation of future FX spot by the FX forwards of high-yielding currencies. Ranked using the implied carry rate (FX forwards versus FX spot) of a number of currencies (G10 and EM) against the USD, the strategy goes long on single currency indices (which roll FX forwards) for the currencies with the highest carry, and short on single-currency indices for the currencies with the lowest carry. It is reviewed and rebalanced on a monthly basis.

b) Relying on the mean reversal of exchange rates, the *FX Value* strategy ranks the currencies according to the valuation measure (based on GS DEER, Dynamic Equilibrium Exchange Rate model) of a number of currencies (G10 and EM) against the USD. Rebalanced monthly, the strategy goes long on single-currency indices (which roll FX forwards) on the most undervalued currencies, and short on single-currency indices on the most overvalued currencies.

c) Capitalizing on the persistence of trends in forward exchange rate movements that are driven by both carry and spot movements, the *FX Momentum* strategy evaluates the recent performance of twenty-seven currencies against the USD and is rebalanced on a daily basis. Long/Short positions are determined based on whether actual performance has been positive or negative.

2.3. Commodity factors

a) *Commodity Carry* captures the tendency of commodities with tighter time spreads to outperform due to low inventories that drive both backdated futures curves and price appreciation, and to buy demand from consumer hedgers for protection against price spikes in undersupplied commodities. The strategy goes long on the top third and short on the bottom third of the twenty-four commodities from the S&P GSCI universe, ranked by annualized strength of front month time spreads. The strategy is rebalanced daily based on signals over the last ten days. The strategy is net of cost.

b) *Commodity Value* uses the weekly Commodity Futures Trading Commission (CFTC) positioning data to determine long and short positions. It will take long positions in commodities where the speculative positions are the most short, and short positions in commodities where the speculative positions are the most long.

c) *Commodity Momentum* in commodity returns reflects initial underreactions, or subsequent overreactions, to changes in demand. Increases or decreases in supply can take many years to implement and may subsequently overshoot the required changes to match demand. The strategy goes long on the top third and short on the bottom third of the twenty-four commodities from the S&P GSCI universe, ranked by rolling one-year excess returns of each commodity. The strategy is rebalanced daily based on signals over the last ten days. The strategy is net of cost.

d) *Commodity Quality* is long the Bloomberg Roll Select Commodity Index and short the Bloomberg Commodity Index. The Bloomberg Roll Select Commodity Index is a version of the Bloomberg Commodity Index (BCOM) that aims to mitigate the effects of contango on index performance. To do this, the index rolls into the futures contracts for each commodity with the most backwardation or least contango. The contract selection process is performed on the 4th business day of each month.

Appendix 1.B Tables

Table 1.B.1 Style factor indices used for Figure 4. The table above provides details on the style factor series used for computing the metrics shown in Figure 4 and the table below provides the descriptive statistics of the factors, reported in %, of monthly returns. Equity style factors and Commodity Quality are sourced from Bloomberg and all non-equity style factors are sourced from Goldman Sachs (GS).

Style Factor	**Equity**	**Fixed Income**	**Commodity**	**FX**
Carry		GS Interest Rate G Carry 03	GS Macro Carry Index RP14	GS FX Carry C0115
Value	MSCI ACWI Value	GS Interest Rate Value 03	GS Commodity COT Strategy COT3	GS FX Value C0114
Momentum	MSCI ACWI Momentum	GS Interest Rate Trend 03	GS Macro Momentum Index RP15	GS FX Trend C0038
Quality	MSCI ACWI Quality	GS Interest Rate Curve C0210	Bloomberg Roll Select Commodity Index minus Bloomberg Commodity Index	
Low Volatility	MSCI ACWI Minimum Volatility			

	Mean	**SD**	**Median**	**Min**	**Max**		**Mean**	**SD**	**Median**	**Min**	**Max**
Equity						*Commodity*					
Value	-0.03	0.86	-0.04	-3.43	2.64	Carry	0.29	2.45	0.47	-6.73	7.41
Quality	0.36	1.00	0.31	-2.16	4.82	Quality	0.45	0.83	0.34	-2.20	5.08
Momentum	0.56	2.11	0.55	-7.70	12.95	Momentum	0.04	3.31	-0.23	-19.45	16.47
Low Volatility	0.37	1.35	0.43	-4.33	4.91	Value	0.45	2.20	0.41	-5.55	6.94
Fixed Income						*FX*					
Carry	0.25	1.38	0.33	-7.02	9.71	Carry	0.37	1.89	0.36	-8.11	7.20
Quality	0.08	0.84	0.00	-2.22	3.85	Momentum	0.26	1.43	0.27	-3.26	5.42
Momentum	0.31	1.34	0.35	-5.79	4.91	Value	0.17	1.22	0.08	-6.25	3.61
Value	0.05	1.13	0.09	-3.52	3.15						

Table 1.B.2 Prominent asset pricing models The table outlines prominent asset pricing models from the factor literature, and lists the factors included in each model. Columns 1 and 2 indicate the name of the model and the scholarly article in which it was introduced. Column 3 lists the factor(s) used in that model. Column 4 reports the maximum Sharpe ratio (MS) results from Table 11 (Panel B) of Hou et al. (2015). Column 5 reports the maximum Sharpe ratio (MS) results from Table 5 of Hou et al. (2020a). Column 6 reports GRS F-Statistics and the corresponding significance levels for 73 anomalies, value-weighted, NYSE deciles from Table 5 of Stambaugh and Yuan (2017). Column 7 reports GRS F-Statistics and the corresponding significance levels for 34 anomalies from Table 7 of Daniel, Hirshleifer and Sun (2020). ***, ** and * represent significance at 1%, 5% and 10% levels respectively.

Model (1)	**Study** (2)	**Factor(s)** (3)	$\mathbf{MS}_q$ (4)	$\mathbf{MS}_{q5}$ (5)	$\mathbf{GRS}_{72}$ (6)	$\mathbf{GRS}_{34}$ (7)
CAPM	Sharpe (1964); Lintner (1965); Mossin (1966)	Market	0.1			3.95***
Fama and French 3-factor model	Fama and French (1993)	Market, size, and value	0.21		2.10***	3.70***
Carhart 4-factor model	Carhart (1997)	Market, size, value and momentum	0.3			3.10***
Fama and French 5-factor model	Fama and French (2015)	Market, size, value, profitability, and investment		0.32	1.79***	2.60***
Fama and French 4-factor model	Fama and French (2015)	Market, size, profitability, and investment				
q-factor model	Hou, Xue and Zhang (2015)	Market, size, investment, and profitability	0.43	0.42		2.42***
Stambaugh and Yuan 4-factor model	Stambaugh and Yuan (2017)	Market, size, management and performance		0.41	1.54***	1.71***
q5- model	Hou, Mo, Xue and Zhang (2018)	Market, size, investment, profitability and expected investment growth		0.63	1.78***	
Fama and French 5-factor model plus momentum	Fama and French (2018)	Market, size, value, profitability, investment and momentum		0.37		
Daniel, Hirshleifer and Sun 3-factor model	Daniel, Hirshleifer and Sun (2020)	Market, long-horizon and short-horizon mispricing factors		0.42		1.61**

Table 1.B.3 Equity factor performance around the world The table reports the performance of MSCI style factor indexes across different geographies between December 2001 and July 2020. We compute annualized returns, volatilities, active returns (relative to the market) as well as corresponding tracking errors. Maximum drawdown gives the maximum loss suffered within the sample period. Panel A covers the performance of global factor portfolios as given by MSCI's ACWI universe, representing large and mid-cap equity performance across 23 developed and 27 emerging markets. Panel B is for the U.S. that represents the performance of the large and mid-cap segments and aims to represent 85% of the U.S. market. In Panel C, the MSCI EAFE Index is designed to represent the performance of large and mid-cap securities across 21 developed markets, including countries in Europe, Australia and the Far East, excluding the U.S. and Canada. Panel D covers large and mid-cap securities across 26 Emerging Markets. For risk-free rates, Panel A, B and D uses 3-month US LIBOR rates and Panel C uses 3-month EUR LIBOR rates as cash input. Details on data sources and variable definitions are provided in Appendix A.1.

	Return p.a.(%)	**Volatility p.a.(%)**	**Sharpe ratio**	**Max DD %**	**Active return p.a.**	**Tracking Error %**	**Information ratio**
Panel A: Global							
Market	6.84	15.38	0.33	54.48			
Value	5.43	15.98	0.23	56.41	-1.41	2.90	-0.49
Size	8.18	17.49	0.37	58.55	1.34	4.73	0.28
Momentum	11.10	15.46	0.61	57.79	4.26	7.16	0.59
Quality	9.31	13.61	0.56	45.40	2.47	3.99	0.62
Low Volatility	8.78	10.54	0.67	38.79	1.94	7.41	0.26
Panel B: US							
Market	8.37	14.38	0.46	48.04			
Value	6.47	14.86	0.32	51.74	-1.90	3.81	-0.50
Size	8.95	16.98	0.43	52.15	0.58	4.52	0.13
Momentum	11.62	14.45	0.69	50.21	3.25	7.38	0.44
Quality	9.81	13.04	0.62	37.98	1.44	3.63	0.40
Low Volatility	8.36	11.26	0.59	40.64	-0.01	6.03	0.00
Panel C: EAFE							
Market	6.04	16.33	0.29	56.00			
Value	4.98	17.54	0.21	57.82	-1.06	3.09	-0.34
Size	7.46	16.88	0.36	57.40	1.42	3.40	0.42
Momentum	8.43	15.39	0.46	56.35	2.39	7.42	0.32
Quality	9.12	14.94	0.52	47.88	3.08	4.92	0.63
Low Volatility	8.28	11.68	0.60	41.70	2.24	7.27	0.31
Panel D: EM							
Market	9.33	21.03	0.36	61.91			
Value	8.57	21.26	0.32	59.85	-0.76	2.87	-0.26
Size	9.36	21.22	0.36	63.08	0.03	4.06	0.01
Momentum	13.32	22.05	0.53	68.54	3.99	6.92	0.58
Quality	10.95	19.79	0.47	58.44	1.62	4.32	0.38
Low Volatility	10.90	16.64	0.55	52.24	1.57	6.06	0.26

Appendix 1.C Figures

Fig. 1.C.1 Investment objectives of institutional investors The figure shows the most common objectives of asset owners when evaluating factor investing strategies. It compares the changes in priority of objectives of the survey participants from 2016-2019. Source: 2019 FTSE Smart Beta Global Survey.

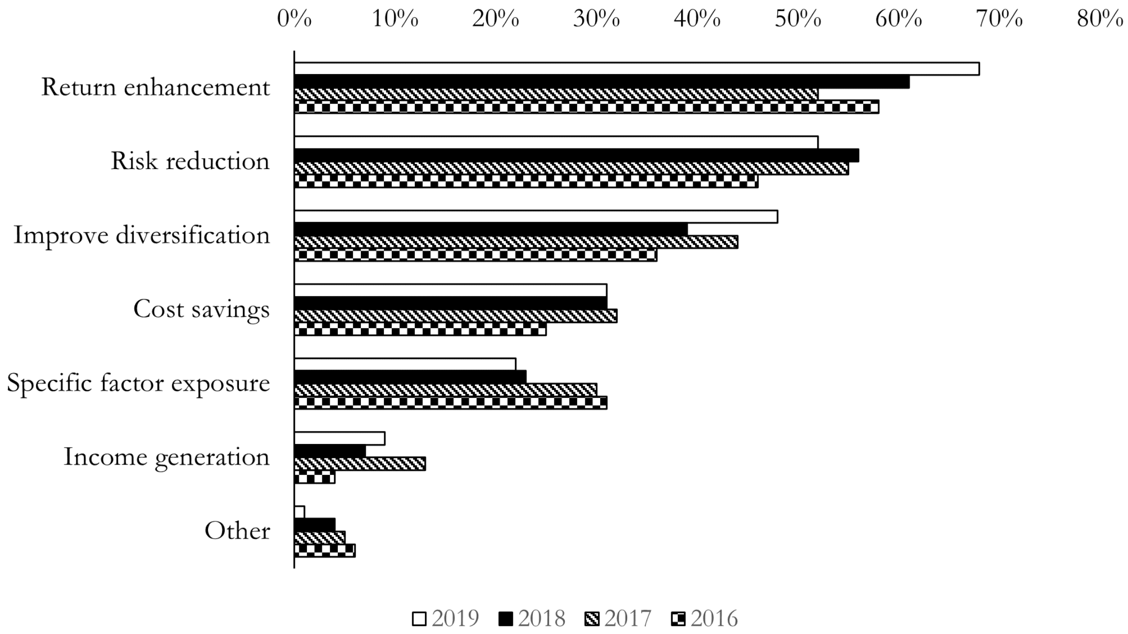
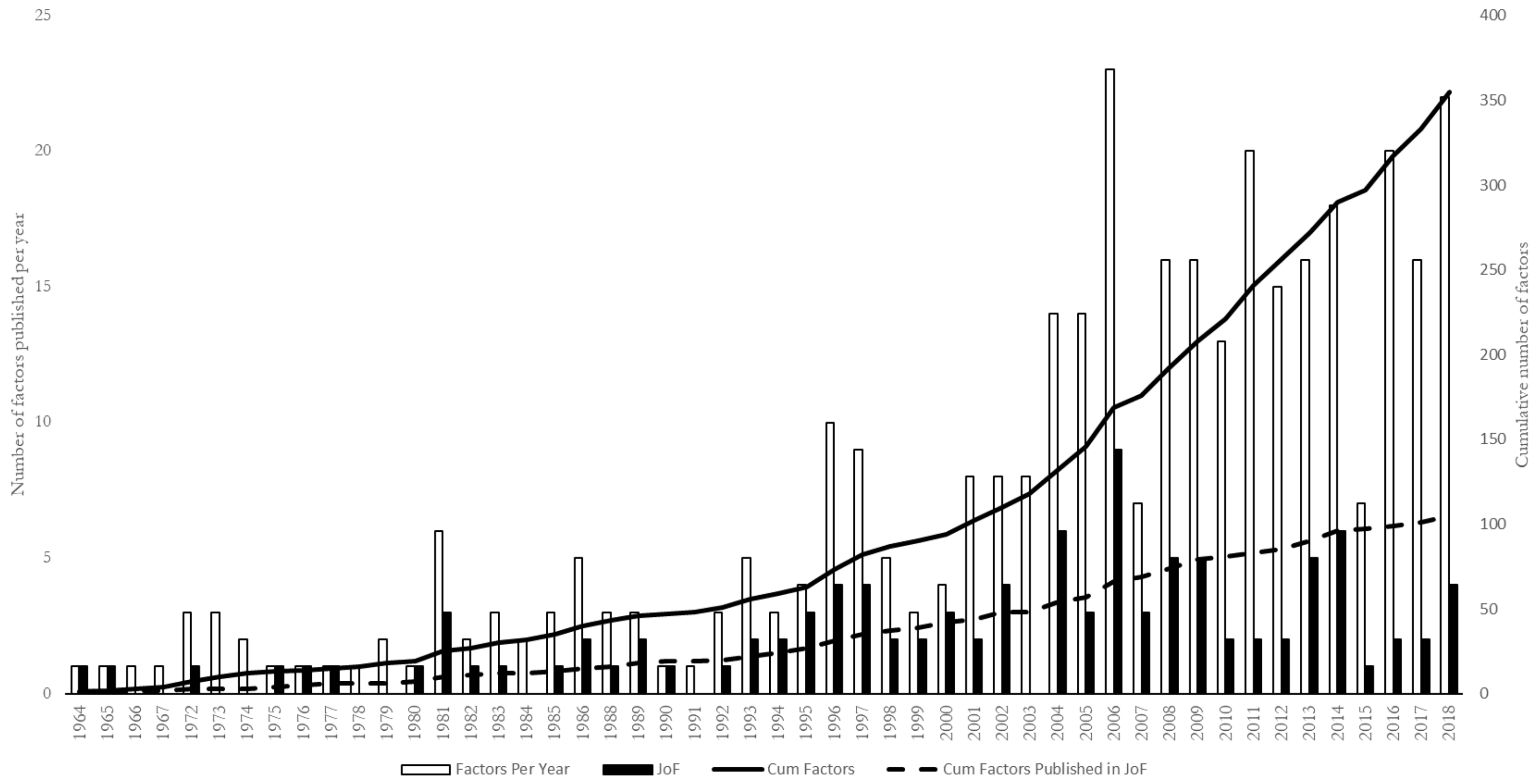


Fig. 1.C.2 Proliferation of factors since 1964 The figure shows the trend in the year-wise publication of factor studies in well-known finance journals and the number of factor studies published in the top three finance journals: Journal of Finance, Journal of Financial Economics and Review of Financial Studies since 1964. Source: Factor Census dataset of Harvey and Liu (2019).



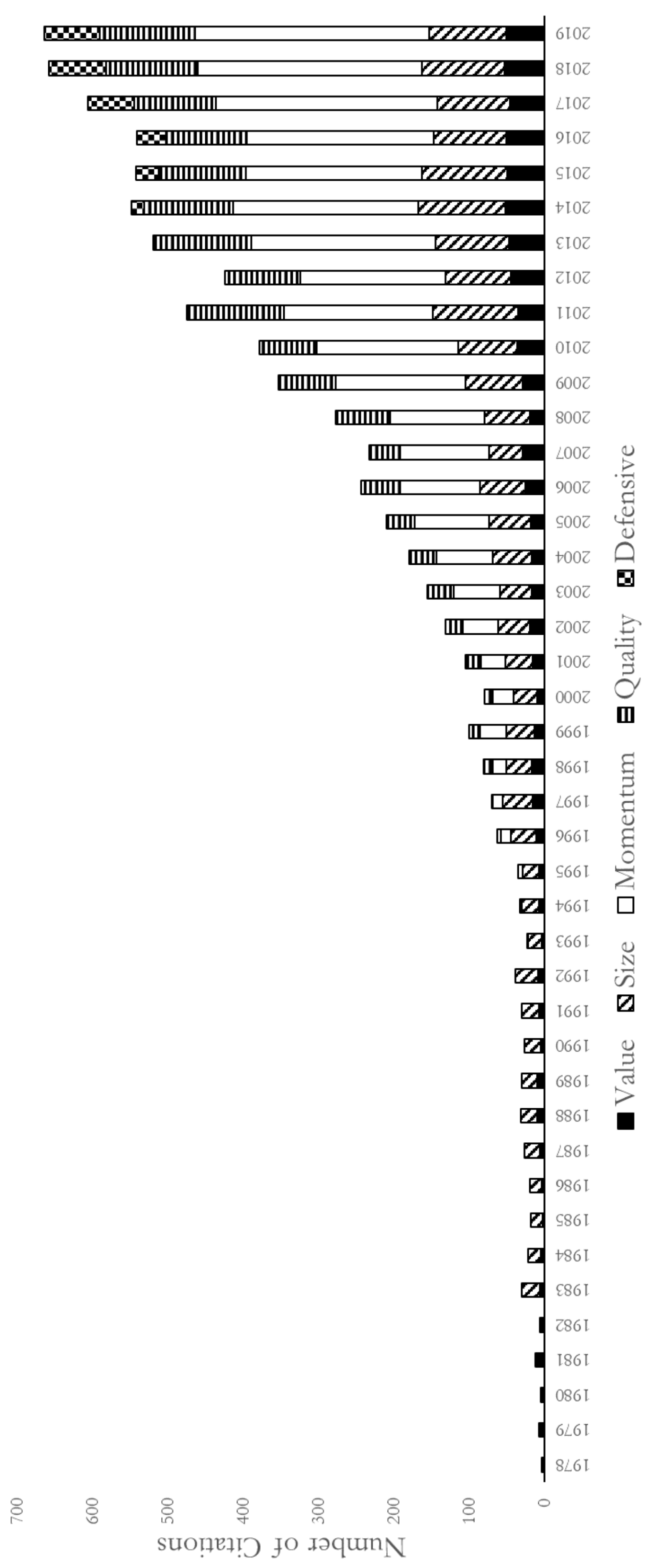
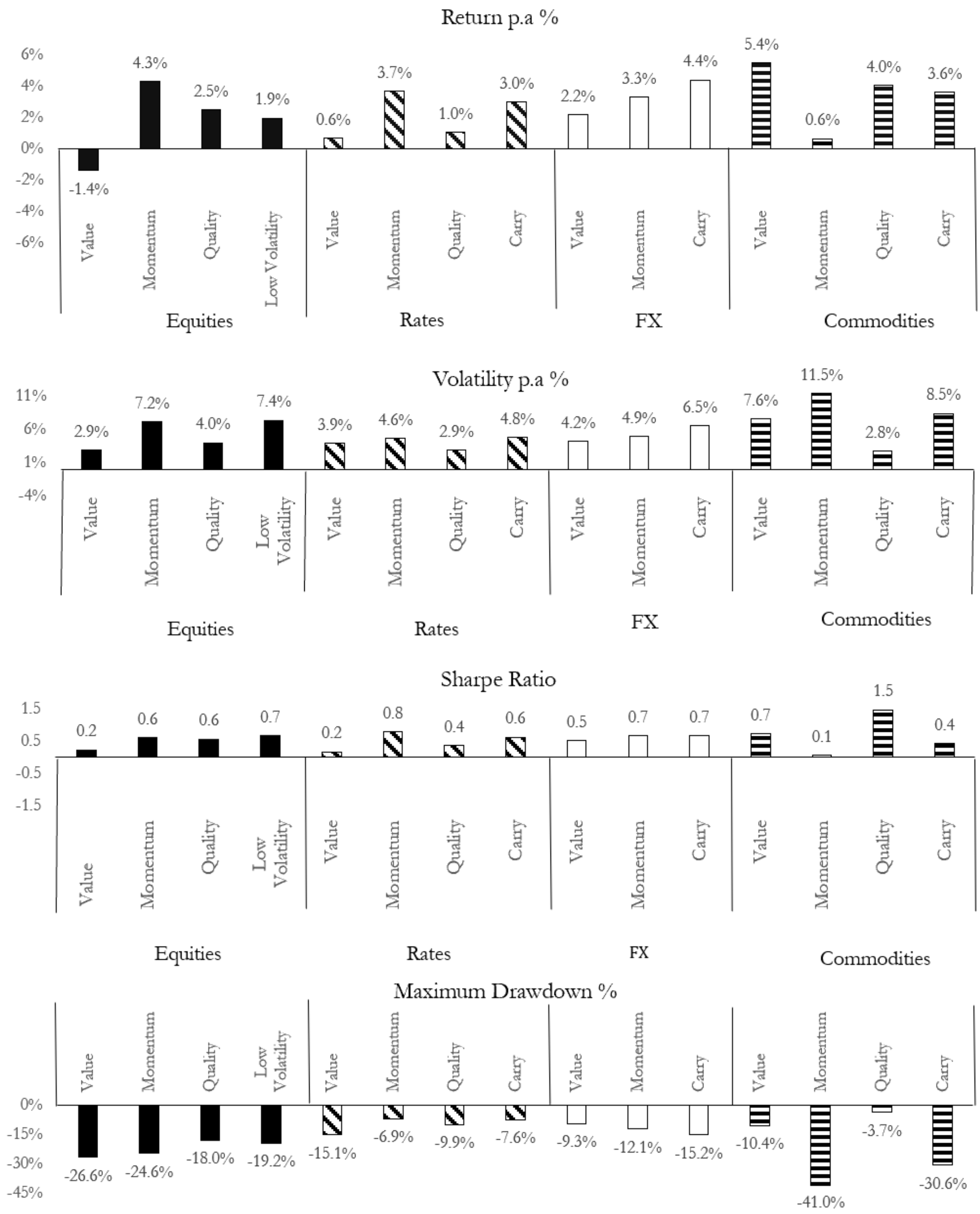


Fig. 1.C.3 Equity style factor relevance The figure shows the growth in popularity of equity style factors between 1978 and 2019. It depicts popularity of different equity factors as measured by the number of year-wise citations of the original study documenting the effect. The citation data has been sourced from Scopus. For value, the reference is Basu (1977); Banz (1981) for size; Jegadeesh and Titman (1993) for momentum; Sloan (1996) for quality; Frazzini and Pedersen (2014) for defensive. Source: Factor Census dataset of Harvey and Liu (2019).

Fig. 1.C.4 Style factor performance across asset classes The figure depicts different performance metrics of style factor performance across equity, rates, FX, and commodities. Details on data sources and variable definitions are provided in Appendices A.1. The maximum drawdowns indicate the maximum observed loss and are reported in negative to indicate the downside risk. The sample period is December 2001 and July 2020.



References

Agarwal, V. and Taffler, R. (2008). Does financial distress risk drive the momentum anomaly? *Financial Management*, **37**, 461–484.

Aharoni, G., Grundy, B. and Zeng, Q. (2013). Stock returns and the miller modigliani valuation formula: Revisiting the fama french analysis. *Journal of Financial Economics*, **110** (2), 347–357.

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

Amel-Zadeh, A. and Serafeim, G. (2018). Why and how investors use esg information: Evidence from a global survey. *Financial Analysts Journal*, **74**, 87–103.

Andronoudis, D., Dargenidou, C., Konstantinidi, E. and Pope, P. F. (2019). Conservative accounting and the pricing of risk: the case of research and development. *Available at SSRN 3006537*.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, **91**, 1–23.

Aretz, K., Bartram, S. M. and Pope, P. F. (2010). Macroeconomic risks and characteristic-based factor models. *Journal of Banking and Finance*, **34**, 1383–1399.

— and Pope, P. F. (2018). Real options models of the firm, capacity overhang, and the cross section of stock returns. *Journal of Finance*, **73**, 1363–1415.

Arnott, R. D., Clements, M., Kalesnik, V. and Linnainmaa, J. T. (2021a). *Factor momentum. Available at SSRN 3116974*.

—, Harvey, C. R., Kalesnik, V. and Linnainmaa, J. T. (2021b). Reports of value's death may be greatly exaggerated. *Financial Analysts Journal, forthcoming*.

Asness, C. (2011). Momentum in japan: The exception that proves the rule. *Journal of Portfolio Management*, **37**, 67–75.

—, Chandra, S., Ilmanen, A. and Israel, R. (2017). Contrarian factor timing is deceptively difficult. *Journal of Portfolio Management*, **43**, 72–87.

Asness, C. S., Frazzini, A. and Pedersen, L. H. (2014). Low-risk investing without industry bets. *Financial Analysts Journal*, **70**, 24–41.

—, Moskowitz, T. J. and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, **68**, 929–985.

Avramov, D., Cheng, S. and Metzker, L. (2021). *Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability. Available at SSRN 3450322*.

—, —, Schreiber, A. and Shemer, K. (2017). Scaling up market anomalies. *Journal of Investing*, **26**, 89–105.

—, Chordia, T., Jostova, G. and Philipov, A. (2007). Momentum and credit rating. *Journal of Finance*, **62**, 2503–2520.

—, —, — and — (2019). Bonds, stocks, and sources of mispricing. unpublished manuscript.

Bai, J., Bali, T. G. and Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, **131**, 619–642.

Bakshi, G., Gao, X. and Rossi, A. G. (2017). Understanding the sources of risk underlying the cross section of commodity returns. *Management Science*, **65**, 619–641.

Baku, E., Fortes, R., Hervé, K., Lezmi, E., Malongo, H., Roncalli, T. and Xu, J. (2020). Factor investing in currency markets: Does it make sense. *Journal of Portfolio Management*, **46**, 141–155.

Bali, T. G., Goyal, A., Huang, D., Jiang, F. and Wen, Q. (2020a). The cross-sectional pricing of corporate bonds using big data and machine learning.

—, Subrahmanyam, A. and Wen, Q. (2020b). Long-term reversals in the corporate bond market. *Journal of Financial Economics*.

Baltas, N. (2019). The impact of crowding in alternative risk premia investing. *Financial Analysts Journal*, **75**, 89–104.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, **9** (1), 3–18.

Barillas, F. and Shanken, J. (2017). Which alpha? *Review of Financial Studies*, **30**, 1316–1338.

— and — (2018). Comparing asset pricing models. *Journal of Finance*, **73**, 715–754.

Barro, R. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, **121**, 823–866.

Barroso, P. and Santa-Clara, P. (2015). Beyond the carry trade: Optimal currency portfolios. *Journal of Financial and Quantitative Analysis*, **50** (5), 1037–1056.

Barry, C. B., Goldreyer, E., Lockwood, L. and Rodriguez, M. (2002). Robustness of size and value effects in emerging equity markets, 1985–2000. *Emerging Markets Review*, **3**, 1–30.

Bartram, S. M., Branke, J. and Motahari, M. (2020a). *Artificial intelligence in asset management*. CFA Institute Research Foundation.

—, Brown, G. W. and Conrad, J. (2011). The effects of derivatives on firm risk and value. *Journal of Financial and Quantitative Analysis*, **46**, 967–999.

—, Djuranovik, L. and Garratt, A. (2020b). Currency anomalies. *CEPR Working Paper 15653*.

— and Dufey, G. (2001). International portfolio investment: Theory, evidence, and institutional framework. financial markets. *Institutions & Instruments*, **10**, 85–155.

—, Griffin, J. M., Lim, T.-H. and Ng, D. T. (2015). How important are foreign ownership linkages for international stock returns? *Review of Financial Studies*, **28**, 3036–3072.

— and Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, **128**, 125–147.

— and — (2021). Global market inefficiencies. *Journal of Financial Economics*, **139**, 234–259.

—, — and Nozawa, Y. (2020c). Book-to-market, mispricing, and the cross-section of corporate bond returns.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance*, **32** (3), 663–682.

Beekhuizen, P., Duyvesteyn, J., Martens, M. and Zomerdijk, C. (2019). Carry investing on the yield curve. *Financial Analysts Journal*, **75**, 51–63.

Bektić, D., Wenzler, J.-S., Wegener, M., Schiereck, D. and Spielmann, T. (2019).

Berk, J. B. (1995). A critique of size-related anomalies. *Review of Financial Studies*, **8**, 275–286.

Berkman, H., Jacobsen, B. and Lee, J. B. (2011). Time-varying rare disaster risk and stock returns. *Journal of Financial Economics*, **101**, 313–332.

—, — and — (2017). Rare disaster risk and the expected equity risk premium. *Accounting & Finance*, **57**, 351–372.

Bhansali, V. (2007). Volatility and the carry trade. *Journal of Fixed Income*, **17**, 72–84.

Bianchi, D., Büchner, M. and Tamoni, A. (2020). Bond risk premiums with machine learning. *Review of Financial Studies*, **34**, 1046–1089.

Bilson, J. F. (1981). The speculative efficiency hypothesis. *Journal of Business*, **54**, 435–451.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2002). Latent dirichlet allocation. *Advances in Neural Information Processing Systems*, pp. 601–608.

Blitz, D. and De Groot, W. (2014). Strategic allocation to commodity factor premiums. *Journal of Alternative Investments*, **17**, 103–115.

— and Marchesini, T. (2019). The capacity of factor strategies. *Journal of Portfolio Management*, **45**, 30–38.

—, Pang, J. and Van Vliet, P. (2013). The volatility effect in emerging markets. *Emerging Markets Review*, **16**, 31–45.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. **8**, 3–62.

Borghi, R. and de Rossi, G. (2020). The artificial intelligence approach to picking stocks. *Machine learning for asset management: new developments and financial applications*, pp. 115–166.

Brooks, J., Gould, T. and Richardson, S. (2020). Active fixed income illusions. *Journal of Fixed Income*, **29**, 3–16.

— and Moskowitz, T. J. (2018). Yield curve premia. available at ssrn 2956411.

—, Palhares, D. and Richardson, S. (2018). Style investing in fixed income. *Journal of Portfolio Management*, **44**, 127–139.

Brunnermeier, M. K., Nagel, S. and Pedersen, L. H. (2008). Carry trades and currency crashes. *NBER Macroeconomics Annual*, **23**, 313–348.

Bryzgalova, S., Huang, J. and Julliard, C. (2020). Bayesian solutions for the factor zoo: We just ran two quadrillion models.

Burnside, C., Eichenbaum, M., Kleshchelski, I. and Rebelo, S. (2011). Do peso problems explain the returns to the carry trade? *Review of Financial Studies*, **24**, 853–891.

Campbell, J. Y. and Vuolteenaho, T. (2004). Bad beta, good beta. *American Economic Review*, **94**, 1249–1275.

Capaul, C., Rowley, I. and Sharpe, W. F. (1993). International value and growth stock returns. *Financial Analysts Journal*, **49**, 27–36.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, **52**, 57–82.

Chambers, D., Lohre, H. and Rother, C. (2019). Tail risk management for multiasset multi-factor strategies. *Risk* & *Reward*, pp. 14–20.

Chan, K. C. and Chen, N.-F. (1991). Structural and return characteristics of small and large firms. *Journal of Finance*, **46**, 1467–1484.

Chan, L. K., Hamao, Y. and Lakonishok, J. (1991). Fundamentals and stock returns in japan. *Journal of Finance*, **46**, 1739–1764.

Choi, J. and Kim, Y. (2018). Anomalies and market (dis)integration. *Journal of Monetary Economics*, **100**, 16–34.

Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A. and Tong, Q. (2017). Are capital market anomalies common to equity and corporate bond markets? an empirical investigation. *Journal of Financial and Quantitative Analysis*, **52**, 1301–1342.

—, — and Saretto, A. (2020). Anomalies and false rejections. *Review of Financial Studies*, **33**, 2134–2179.

—, Subrahmanyam, A. and Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, **58**, 41–58.

Chu, Y., Hirshleifer, D. and Ma, L. (2020). The causal effect of limits to arbitrage on asset pricing anomalies. *Journal of Finance*, **75**, 2631–2672.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, **66**, 1047–1108.

Colacito, R., Riddiough, S. J. and Sarno, L. (2020). Business cycles and currency returns. *Journal of Financial Economics*, **137** (3), 659–678.

Corte, P. D., Riddiough, S. J. and Sarno, L. (2016). Currency premia and global imbalances. *Review of Financial Studies*, **29**, 2161–2193.

Cowles, A. (1933). Can stock market forecasters forecast? econometrica. pp. 309–324.

Daniel, K., Hirshleifer, D. and Sun, L. (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies*, **33**, 1673–1736.

De Groot, W., Pang, J. and Swinkels, L. (2012). The cross-section of stock returns in frontier emerging markets. *Journal of Empirical Finance*, **19**, 796–818.

Dechow, P. M., Sloan, R. G. and Soliman, M. T. (2004). Implied equity duration: A new measure of equity risk. *Review of Accounting Studies*, **9**, 197–228.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J. and Uppal, R. (2020a). A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies*, **33**, 2180–2222.

—, — and Uppal, R. (2020b). *What alleviates crowding in factor investing? Available at SSRN 3392875.*

Dichtl, H., Drobetz, W., Lohre, H. and Rother, C. (2021). Active factor completion strategies. *Journal of Portfolio Management*, **47**, 9–37.

—, —, —, — and Vosskamp, P. (2019). Optimal timing and tilting of equity factors. *Financial Analysts Journal*, **75**, 84–102.

Diether, K. B., Lee, K.-H. and Werner, I. M. (2009). Short-sale strategies and return predictability. *Review of Financial Studies*, **22**, 575–607.

Dimson, E. and Marsh, P. (1999). Murphy's law and market anomalies. *Journal of Portfolio Management*, **25**, 53–69.

Donangelo, A. (2020). Untangling the value premium with labor shares. *Review of Financial Studies*, **34**, 451–508.

Doskov, N. and Swinkels, L. (2015). Empirical evidence on the currency carry trade. *Journal of International Money and Finance*, **51**, 370–389.

Ehsani, S. and Linnainmaa, J. T. (2019). Factor momentum and the momentum factor. *National Bureau of Economic Research.*

El-Haj, M., Rayson, P., Walker, M., Young, S. and Simaki, V. (2019). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. *Journal of Business Finance and Accounting*, **46**, 265–306.

Elton, E. J., Gruber, M. J. and De Souza, A. (2019). Are passive funds really superior investments? an investor perspective. *Financial Analysts Journal*, **75**, 7–19.

Erb, C. B. and Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, **62**, 69–97.

Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, **14**, 319–338.

— (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, **49**, 283–306.

— and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33** (1), 3–56.

— and — (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, **51**, 55–84.

— and — (1998). Value versus growth: The international evidence. *Journal of Finance*, **53**, 1975–1999.

— and — (2008). Dissecting anomalies. *Journal of Finance*, **63** (4), 1653–1678.

— and — (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, **105** (3), 457–472.

— and — (2015). A five-factor asset pricing model. *Journal of Financial Economics*, **116**, 1–22.

— and — (2018). Choosing factors. *Journal of Financial Economics*, **128**, 234–252.

Farhi, E., Fraiberger, S. P., Gabaix, X., Ranciere, R. and Verdelhan, A. (2009). Crash risk in currency markets. *National Bureau of Economic Research*.

Feng, G., Giglio, S. and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, **75**, 1327–1370.

—, Polson, N. G. and Xu, J. (2019). Deep learning in characteristics-sorted factor models. p. 23527.

Fernandez-Perez, A., Frijns, B., Fuertes, A.-M. and Miffre, J. (2018). The skewness of commodity futures returns. *Journal of Banking and Finance*, **86**, 143–158.

Ferson, W. E., Sarkissian, S. and Simin, T. T. (1999). The alpha factor asset pricing model: A parable. *Journal of Financial Markets*, **2**, 49–68.

Frazzini, A., Israel, R. and Moskowitz, T. J. (2014). Trading costs of asset pricing anomalies. working paper.

—, — and — (2018). *Trading costs. Available at SSRN 3229719.*

Freyberger, J., Neuhierl, A. and Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, **33**, 2326–2377.

Froot, K. A. and Thaler, R. H. (1990). Anomalies: Foreign exchange. *Journal of Economic Perspectives*, **4**, 179–192.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics*, **127**, 645–700.

Gebhardt, W. R., Hvidkjaer, S. and Swaminathan, B. (2005). Stock and bond market interaction: Does momentum spill over. *Journal of Financial Economics*, **75**, 651–690.

Giamouridis, D., Sakkas, A. and Tessaromatis, N. (2017). Dynamic asset allocation with liabilities. *European Financial Management*, **23**, 254–291.

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, **57**, 1121–1152.

Gollier, C. (2001). Wealth inequality and asset pricing. *Review of Economic Studies*, **68**, 181–203.

Gormsen, N. J. and Lazarus, E. (2020). Duration-driven returns. *Available at SSRN*, **3359027**.

Grant, J., Ahmerkamp, J. D. and Kosowski, R. (2012). *Predictability in carry and momentum strategies across asset classes. Available at SSRN 2024056.*

Green, J., Hand, J. R. M. and Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies*, **30**, 4389–4436.

Griffin, J. M., Ji, X. and Martin, J. S. (2003). Momentum investing and business cycle risk: Evidence from pole to pole. *Journal of Finance*, **58**, 2515–2547.

—, Kelly, P. J. and Nardari, F. (2010). Do market efficiency measures yield correct inferences? a comparison of developed and emerging markets. *Review of Financial Studies*, **23**, 3225–3277.

Gu, S., Kelly, B. and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, **33**, 2223–2273.

Guo, H., Savickas, R., Wang, Z. and Yang, J. (2009). Is the value premium a proxy for time-varying investment opportunities? some time-series evidence. *Journal of Financial and Quantitative Analysis*, pp. 133–154.

Gupta, T. and Kelly, B. (2019). Factor momentum everywhere. *Journal of Portfolio Management*, **45**, 13–36.

Hammerschmid, R. and Lohre, H. (2020). Integrating time series and crosssectional signals for optimal commodity portfolios. *Available at SSRN*, **3504394**.

Hansen, L. P. and Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, **88**, 829–853.

Hansen, S., McMahon, M. and Prat, A. (2017). Transparency and deliberation within the fomc: A computational linguistics approach. *Quarterly Journal of Economics*, **133**, 801–870.

Harvey, C. R. and Liu, Y. (2019). A census of the factor zoo. *Available at SSRN 3341728*.

—, — and Zhu, H. (2016). … and the cross-section of expected returns. *Review of Financial Studies*, **29**, 5–68.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *Journal of Finance*, **64**, 1445–1477.

Hatchondo, J. C. (2008). A quantitative study of the role of wealth inequality on asset prices. *FRB Richmond Economic Quarterly*, **94**, 73–96.

Haugen, R. A. (1995). The new finance: The case against efficient markets.

— and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, **41**, 401–439.

— and — (2012). Low risk stocks outperform within all observable markets of the world.

— and Heins, A. J. (1972). On the evidence supporting the existence of risk premiums in the capital market.

Hoffmann, M. and Studer-Suter, R. (2017). Systematic consumption risk in currency returns. *Journal of International Money and Finance*, **74**, 187–208.

Hottinga, J., van Leeuwen, E. and van Ijserloo, J. (2001). Successful factors to select outperforming corporate bonds. *Journal of Portfolio Management*, **28**, 88–101.

Hou, K., Karolyi, G. A. and Kho, B.-C. (2011). What factors drive global stock returns? *Review of Financial Studies*, **24** (8), 2527–2574.

—, Mo, H., Xue, C. and Zhang, L. (2021). An augmented q-factor model with expected growth. *Review of Finance*, **25** (1), 1–41.

—, Xue, C. and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, **28**, 650–705.

—, — and — (2020). Replicating anomalies. *Review of Financial Studies*, **33**, 2019–2133.

Hsu, J., Kalesnik, V. and Kose, E. (2019). What is quality? *Financial Analysts Journal*, **75**, 44–61.

Israel, R., Palhares, D. and Richardson, S. A. (2018). Common factors in corporate 65 bond returns. *Journal of Investment Management*, **16**, 17–46.

Israelsen, R. (2014). Tell it like it is: Disclosed risks and factor portfolios.

Jacobs, H. and Müller, S. (2020). Anomalies across the globe: Once public, no longer existent. *Journal of Financial Economics*, **135**, 213–230.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, **48** (1), 65–91.

Jensen, T. I., Kelly, B. T. and Pedersen, L. H. (2021). Is there a replication crisis in finance?

Jostova, G., Nikolova, S., Philipov, A. and Stahel, C. W. (2013). Momentum in corporate bond returns. *Review of Financial Studies*, **26**, 1649–1693.

Jotikasthira, C., Lundblad, C. and Ramadorai, T. (2012). Asset fire sales and purchases and the international transmission of funding shocks. *Journal of Finance*, **67**, 2015–2050.

Jurek, J. W. (2014). Crash-neutral currency carry trades. *Journal of Financial Economics*, **113**, 325–347.

Kapadia, N. (2011). Tracking down distress risk. *Journal of Financial Economics*, **102**, 167–182.

Karolyi, A. (2016). Home bias, an academic puzzle. *Review of Finance*, **20**, 2049–2078.

Kearney, C. and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, **33**, 171–185.

Kelly, B. T., Palhares, D. and Pruitt, S. (2020). Modeling corporate bond returns.

—, Pruitt, S. and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, **134**, 501–524.

Kim, D. and Lee, B.-J. (2019). Shorting costs and profitability of long-short strategies.

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H. and Vrugt, E. B. (2018). **127**, 197–225.

Korajczyk, R. A. and Sadka, R. (2004). Are momentum profits robust to trading costs. *Journal of Finance*, **59**, 1039–1082.

Kothe, J., Lohre, H. and Rother, C. (2021). Rates factors and global asset allocation. *Journal of Fixed Income*, **30**, 6–25.

Kozak, S., Nagel, S. and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, **135**, 271–292.

Kroencke, T. A., Schindler, F. and Schrimpf, A. (2014). International diversification benefits with foreign exchange investment styles. *Review of Finance*, **18**, 1847–1883.

Krueger, P., Sautner, Z. and Starks, L. T. (2020). The importance of climate risks for institutional investors. *Review of Financial Studies*, **33**, 1067–1111.

Lakonishok, J., Shleifer, A. and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, **49**, 1541–1578.

Lam, S. S. and Zhang, W. (2014). Does policy uncertainty matter for international equity markets?

LEIPPOLD, M. and LOHRE, H. (2012a). Data snooping and the global accrual anomaly. *Applied Financial Economics*, **22**, 509–535.

— and — (2012b). International price and earnings momentum. *European Journal of Finance*, **18**, 535–573.

LESMOND, D. A., SCHILL, M. J. and ZHOU, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics*, **71**, 349–380.

LETTAU, M., LUDVIGSON, S. C. and WACHTER, J. A. (2008). The declining equity premium: What role does macroeconomic risk play? *Review of Financial Studies*, **21**, 1653–1687.

— and WACHTER, J. A. (2007). Why is long-horizon equity less risky? a duration-based explanation of the value premium. *Journal of Finance*, **62**, 55–92.

— and — (2011). The term structures of equity and interest rates. *Journal of Financial Economics*, **101**, 90–113.

LEUNG, E., LOHRE, H., MISCHLICH, D., SHEA, Y. and STROH, M. (2021). The promises and pitfalls of machine learning for predicting stock returns. *Journal of Financial Data Science*.

LI, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, **45**, 221–247.

LIANG, H. and RENNEBOOG, L. (2020). Corporate social responsibility and sustainable finance: A review of the literature. *Oxford Research Encyclopedia of Economics and Finance*.

LINNAINMAA, J. T. and ROBERTS, M. R. (2018). The history of the cross-section of stock returns. *Review of Financial Studies*, **31**, 2606–2649.

LO, A. W. and MACKINLAY, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, **1**, 41–66.

— and — (1990). Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, **3**, 431–467.

LOPEZ-LIRA, A. (2019). Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns.

LOU, D. and POLK, C. (2014). Comomentum: Inferring arbitrage activity from return correlations.

LUSTIG, H., ROUSSANOV, N. and VERDELHAN, A. (2011). Common risk factors in currency markets. *Review of Financial Studies*, **24**, 3731–3777.

— and VERDELHAN, A. (2007). The cross section of foreign currency risk premia and consumption growth risk. *American Economic Review*, **97**, 89–117.

MCLEAN, R. D. and PONTIFF, J. (2016). Does academic research destroy stock return predictability. *Journal of Finance*, **71**, 5–32.

Menkhoff, L., Sarno, L., Schmeling, M. and Schrimpf, A. (2012a). Carry trades and global foreign exchange volatility. *Journal of Finance*, **67**, 681–718.

—, —, — and — (2012b). Currency momentum strategies. *Journal of Financial Economics*, **106**, 660–684.

—, —, — and — (2017). Currency value. *Review of Financial Studies*, **30**, 416–441.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, **42**, 867–887.

Miffre, J. (2016). Long-short commodity investing: A review of the literature. *Journal of Commodity Markets*, **1**, 3–13.

— and Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking and Finance*, **3**, 1863–1886.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *Journal of Finance*, **32**, 1151–1168.

Muijsson, C., Fishwick, E. and Satchell, S. (2014). Taking the art out of smart beta.

Nakamura, E., Steinsson, J., Barro, R. and Ursúa, J. (2009). Crises and 58 recoveries in an empirical model of consumption disasters. working paper.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, **108** (1), 1–28.

— and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, **29**, 104–147.

Park, H. (2019). Intangible assets and the book-to-market effect. *European Financial Management*, **25**, 207–236.

Pástor, L. and Veronesi, P. (2020). Political cycles and stock returns. *Journal of Political Economy*, **128** (11), 4011–4045.

Patton, A. J. and Timmermann, A. (2010). *Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts*, **98**, 605–625.

— and Weller, B. M. (2020). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Eonomics*, **137**, 515–549.

Pojarliev, M. and Levich, R. M. (2008). Do professional currency managers beat the benchmark? *Financial Analysts Journal*, **64**, 18–32.

Pukthuanthong, K., Roll, R. and Subrahmanyam, A. (2019). A protocol for factor identification. *Review of Financial Studies*, **32**, 1573–1607.

Ranganathan, A., Lohre, H., Nolte, S. and Braham, H. (2023). An integrated approach to currency factor management. *Forthcoming in the Journal of Systematic Investing.*

Ratcliffe, R., Miranda, P. and Ang, A. (2017). Capacity of smart beta strategies from a transaction cost perspective. *Journal of Index Investing*, **8**, 39–50.

Rietz, T. A. (1988). The equity risk premium a solution. *Journal of Monetary Economics*, **22**, 117–131.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, **73**, 1237–1282.

— and — (2013). Testing for monotonicity in expected asset returns. *Journal of Empirical Finance*, **23**, 93–116.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13** (3), 341–360.

Rouwenhorst, K. G. (1998). International momentum strategies. *Journal of Finance*, **53**, 267–284.

— (1999). Local return factors and turnover in emerging stock markets. *Journal of Finance*, **54**, 1439–1464.

Sakkas, A. and Tessaromatis, N. (2020). Factor based commodity investing. *Journal of Banking & Finance*, **115**, 105807.

Schröder, D. and Esterer, F. (2016). A new measure of equity and cash flow duration: The duration-based explanation of the value premium revisited. *Journal of Money, Credit and Banking*, **48** (5), 857–900.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *Journal of Finance*, **52**, 35–55.

Siriwardane, E. (2013). The pricing of disaster risk.

Stambaugh, R. F., Yu, J. and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, **104**, 288–302.

— and Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, **30**, 1270–1315.

Stein, J. C. (2009). Presidential address: Sophisticated investors and market efficiency. *Journal of Finance*, **64**, 1517–1548.

Sullivan, R., Timmermann, A. and White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, **54**, 1647–1691.

—, — and — (2001). Dangers of data-mining: The case of calendar effects in stock returns. *Journal of Econometrics*, **105**, 249–286.

Van der Hart, J., Slagter, E. and Van Dijk, D. (2003). Stock selection strategies in emerging markets. *Journal of Empirical Finance*, **10**, 105–132.

Van Gelderen, E. and Huij, J. (2014). Academic knowledge dissemination in the mutual fund industry: can mutual funds successfully adopt factor investing strategies. *Journal of Portfolio Management*, **40**, 157–167.

—, — and Kyosev, G. (2019). Factor investing from concept to implementation. *Journal of Portfolio Management*, **45**, 125–140.

Wahal, S. (2019). The profitability and investment premium: Pre-1963 evidence. *Journal of Financial Economics*, **131**, 362–377.

Zaremba, A. (2016). Risk-based explanation for the country-level size and value effects. *Finance Research Letters*, **18**, 226–233.

Chapter 2

An Integrated Approach to Currency Factor Investing

This project is co-authored with Houssem Braham and my supervisors, Harald Lohre and Sandra Nolte and has been accepted for publication in the *Journal of Systematic Investing*. We thank the following people for their helpful comments: Guido Baltussen, Pedro Barroso, David Buckle, Philip Howard, Martin Martens, Fabio Martinetti, Mamdouh Medhat, My Nguyen, Mark Salmon and the organizers and participants of the 2018 KoLa workshop in Konstanz, the 2018 Financial Econometrics conference in Lancaster, the 2018 Global Research Meeting of Invesco Quantitative Strategies in Boston, the 2019 IFABS conference in Angers, the 2019 Mutual Funds, Hedge Funds and Factor Investing conference in Lancaster, the 2019 CEQURA conference in Munich, the 2019 DGF meeting in Essen and the 2020 Inquire UK practitioner seminar. This work was supported by funding from the Economic and Social Research Council (UK).

2.1 Introduction

Investors favor global investment portfolios because of the diversification benefits associated with investing in less correlated, international markets. These portfolios usually engage in international bond and equity investments and are therefore directly exposed to and affected by foreign exchange (FX) rate fluctuations. To manage such currency exposures, currency factor strategies such as carry, value, and momentum have become popular among institutional investors. Currency factor models aim to measure the exposure of currencies to different factors, so that currency portfolios may be created dynamically or exchange rates may be predicted.

The choice of currency factors plays a crucial role since an uninformed choice of currency factors can heavily impact portfolio performance. Verdelhan (2018) suggests a two-factor model using U.S. Dollar (USD) exchange rate return and carry exchange rate return as common factors. These two factors are chosen based on the argument that the Dollar factor represents global macro-level risk and the carry factor represents risk arising due to uncertainty. Greenaway-McGrevy, Mark, Sul and Wu (2018) also propose a two-factor model for predicting exchange rates. They conclude that the Dollar factor and the Euro factor drive exchange rates whereas the prominent carry factor does not.

The objective of this paper is to investigate an integrated approach to currency factor investing that optimally allocates currencies according to cross-sectional characteristics and time-series predictors. To this end, we build on the parametric portfolio policy (PPP) approach of Brandt and Santa-Clara (2006) and Brandt, Santa-Clara and Valkanov (2009) to assess the joint relevance of various potential predictors. Whilst the latter approach has been put forward by Barroso and Santa-Clara (2015) to construct optimal currency factor strategies based on carry, value and momentum characteristics, we contribute to this work by additionally incorporating the notion of timing. Firstly, we investigate currency timing in the parametric portfolio policy framework of Brandt and Santa-Clara (2006). Secondly, we expand the framework of Barroso and Santa-Clara (2015) to investigate the possibility of explicitly timing cross-sectional FX factors such as carry trade.

Our key results show that investing within the universe of G10 currencies using an optimal currency tilting strategy from February 1989 through December 2020 is more compelling and robust than the currency timing alternative. This finding comes mainly from stronger evidence of predictability in the cross-section of currency excess returns compared to time-series predictability with respect to the chosen fundamental variables and technical predictors. Unsurprisingly, the carry characteristic is the main driver of performance, yet the momentum and value characteristics help alleviate major carry drawdowns in volatile FX markets.

Extant literature on the carry trade, such as Menkhoff, Sarno, Schmeling and Schrimpf (2012a) and Brunnermeier, Nagel and Pedersen (2008), has examined the explicit timing of the profitability of carry trade. We investigate an extension of the tilting parametric portfolio policy to integrate the management of FX characteristics in light of meaningful conditioning information. Such currency factor investing not only offers the highest risk-adjusted performance, but also suggests that a PPP can be effectively adapted to overcome the drawbacks of models requiring forecasted expected returns.

The remainder of this paper is structured as follows: Section 2.2 reviews the FX literature and the common choice of factors used for currency tilting. Section 2.3 describes the optimal currency tilting based on the PPP framework of Brandt *et al.* (2009). Section 2.4 discusses relevant predictors for timing currencies and for timing currency factors. Section 2.5 introduces optimal currency timing and presents an integrated approach to currency factor timing. Section 2.6 concludes.

2.2 The Notion Of Currency Tilting

The idea behind characteristic-based tilting strategies is simple and straightforward. If we assume that the capital asset pricing model (CAPM) is the true market model, then tilting towards any other factor, say for example size or value in equities, should not yield superior returns. However, follow-up studies (Fama and French (1992, 1993), Carhart (1997), Ang, Hodrick, Xing and Zhang (2006)) repeatedly find strong counterevidence for the relevance of equity factors beyond the market factor such as size, momentum, value, and low-volatility.

Several factors and factor models have garnered attention since the founding of the CAPM. Still, the efficacy of style factor-based portfolio allocation has been studied primarily in equity markets. Style investing has also become popular in other asset classes such as bonds, commodities, and FX. Extending factor-based strategies to FX markets is straightforward, and common factor strategies account for a large percentage of trading volumes in FX markets. Research on currency factors has identified some prominent reasons behind the success and widespread adoption of these strategies. For example, Burnside, Eichenbaum and Rebelo (2011) examine the profitability of the two most prominent FX strategies, carry and momentum. They confirm existing evidence that payoffs to currency strategies are skewed with fat tails and that conventional risk factors alone cannot account for these returns. The authors provide possible theoretical, microstructural, and behavioral explanations for their continued profitability, and acknowledge the uncorrelated payoff to carry and momentum strategies. Such correlation patterns offer a wide scope for investors to use multiple currency strategies simultaneously.

The performance of factor strategies in the last decade indicates that even well-established factors are prone to periods of poor performance during unfavorable market conditions. Still, the low correlation among factors and the resulting diversification benefits from combining multiple factors can help weathering tough times. Hence, in a portfolio setup, the choice of currency factors, and dynamically tilting towards or away from specific factors, play a crucial role as they can shield against adverse performance effects.

Following the approach of Brandt *et al.* (2009), we build an optimal currency tilting strategy by exploiting informative currency characteristics. The literature offers a handful of proxies that have shown evidence of predictability in the cross-section of currency excess returns. We seek to integrate this information jointly in a portfolio utility context. Brandt *et al.* (2009)'s parametric portfolio policy tackles the issue of cross-sectional portfolio optimization by modeling the portfolio weights as a linear function of currency characteristics. In this study, we examine the three salient FX styles carry, value and momentum together with three additional currency factors, namely Dollar exposure, the Taylor rule and the output gap.

2.2.1 Data

The currency investment universe is comprised of the G10 currencies with USD as the base currency and the following countries' currencies: Australia (AUD), Canada (CAD), Germany (EUR), Japan (JPY), New Zealand (NZD), Norway (NOK), Sweden (SEK), Switzerland (CHF), and the United Kingdom (GBP). All the spots, forward and CPI data come from Bloomberg. We use the OECD industrial production data to compute the output gap. Our sample period spans between February 1989 through December 2020.

We start by computing daily returns for each instrument (spots, forwards and LIBOR) as follows,

$$r_t = \frac{X_t - X_{t-1}}{X_t} \tag{2.1}$$

where X refers to one of the 3 instruments (daily spots/ daily forwards or daily LIBORs). Monthly forwards, spots and cash returns for each currency are then computed by aggregating daily returns, and selecting month-end return values. Then, we define the monthly currency excess returns in USD for currency i as follows:

$$r^i_{t+1} = \frac{F^i_t}{S^i_t} - 1, \tag{2.2}$$

where F^i_t is the price of one USD expressed in foreign currency units and S^i_t is the spot price of one USD in foreign currency units. The subsequent sections discuss the theoretical underpinnings and calculation methodologies for each factor. We adopt a systematic approach by first constructing signals that capture each factor, which are subsequently used to construct an investable long-short portfolio to harness them. It is important to note that the results reported are pre-transaction costs and actual performance may vary when taking costs into account.

2.2.2 FX Carry

The FX carry trade has received a great deal of attention not only for generating high returns but also for its robustness to several other traditional factors, such as market, value, and momentum. The carry trade portfolio is constructed by buying the highest-yielding currencies and selling the lowest-yielding currencies. Researchers have posited various explanations for the carry trade performance. Farhi, Fraiberger, Gabaix, Ranciere and Verdelhan (2009) find that crash risk is responsible for 25% of carry trade returns in developed countries, while Caballero and Doyle (2012) find that carry trade returns can serve as compensation for systemic risk. They note that carry trade returns are highly correlated with equity market risk, especially during market downturns.

The carry factor seeks to exploit the failure of uncovered interest rate parity by banking on the interest rate differentials of high interest rate and low interest rate currencies. Having sparked interest in the early 1980s, Bilson (1981) and Fama (1984) attempted to solve the forward premium puzzle in order to identify the academic motivators behind carry trading in currencies. The FX carry trade strategy maintained its prominence over the years, see Galati and Melvin (2004) who attribute the surge in FX trading to the sudden rise in attractiveness of FX carry and momentum strategies. However, harvesting the FX carry trade has often been associated with collecting pennies in front of a steamroller. Researchers have identified certain cases of underperformance, for example, with liquidity squeezes (Brunnermeier, Nagel and Pedersen, 2008) or increased FX volatility (Menkhoff, Sarno, Schmeling and Schrimpf, 2012a).

To construct the FX carry factor, we take the forward discount (or premium) of a specific currency. Given the fact that covered interest rate parity empirically holds at a monthly frequency (Akram, Rime and Sarno, 2008) the forward premium is then equivalent to interest rate differentials. We compute the forward discount as:

$$fd_t^i = \frac{F_t^i}{S_t^i} - 1, \tag{2.3}$$

where F_t^i is the price of one USD expressed in foreign currency units at time t and S_t is the spot price of one USD in foreign currency units at t.

2.2.3 FX Momentum

Momentum investing has been extremely popular among equity portfolio managers and has been a subject of intense academic study for decades, beginning with Jegadeesh and Titman (1993). In the realm of currencies, exploiting short-term price momentum effects is also relevant given that they are not subsumed by any other traditional risk factor. FX momentum effects can be detected for formation periods of between one and twelve months. But three months is a common choice, because it strikes a solid balance between the goodness of signal and strategy turnover.[1]

Hence, for capturing cross-sectional FX momentum, we consider the cumulative currency return over the previous three months between the quoted and the base currency in order to capture the persistence of currency returns in the short term. We

[1] The choice of a three-month formation period is consistent with Kroencke, Schindler and Schrimpf (2013) and Barroso and Santa-Clara (2015).

thus compute the momentum signal accordingly:

$$Mom_{i,t} = \frac{S_t^i}{S_{t-3}^i}, \tag{2.4}$$

where S_t^i is the spot price of one USD expressed in foreign currency units at time t and S_{t-3}^i is the spot price of one USD in foreign currency units at $t-3$ (months).

FX momentum persists in FX markets because of impediments that restrict the deployment of arbitrage capital to exploit this phenomenon. Equity markets also seem to play a predictive role in explaining the variations in currency momentum payoffs. Okunev and White (2003) capture momentum in the cross-section of currencies and find positive evidence for existence of profits from a cross-sectional momentum-based strategy. There does not seem to be a systematic risk factor, which would explain (net) momentum returns. On the other hand, Menkhoff *et al.* (2012b) find that FX momentum returns are sensitive to transaction costs but they are less related to business cycle risk. They also confirm that FX momentum returns are much higher in currencies with high lagged idiosyncratic volatility and high-country risk ratings.

2.2.4 FX Value

To assess undervalued and overvalued currencies, FX value strategies seek to exploit long-term reversal effects in FX markets. However, there is no universally accepted rule for classifying currencies this way, therefore we need to proxy for the fundamental value of a currency. Comparing the latter with the current trading price/deviation of the exchange rate would indicate whether a currency is undervalued or overvalued.

To construct an FX value factor, we use purchasing power parity as the measure of fundamental value, assuming that goods should cost the same across countries. Currencies whose real exchange rate (RER) deviates significantly from one may be viewed as undervalued or overvalued. The FX value strategy would then seek to exploit the likely reversal of currencies that have exceeded their purchasing power parity values. To determine which measure of purchasing power parity to use, we follow Asness, Moskowitz and Pedersen (2013) and use the 60 month deviation from uncovered interest rate parity. We thus compute the cumulative real depreciation of currency i as:

$$Q_{h,t}^i = \frac{S_t^i CPI_h^i CPI_t^{US}}{S_h^i CPI_t^i}, \tag{2.5}$$

where $h = t - 60$, CPI is the consumer price index representing the price of a broad basket of goods at time period t or h in the U.S. or for the other currency i; and S_t^i and S_h^i are the spot prices of one USD in foreign currency units at t or h, respectively.

If purchasing power parity holds, then Equation (2.5) should equal one. A value below one suggests undervaluation and cumulative real depreciation for a currency, while a value above one indicates overvaluation and cumulative real appreciation. After assessing the relative strength or weakness of a currency using our value characteristic $Q_{h,t}^i$, the currency value portfolio would take long position in undervalued currencies and a short position in overvalued currencies.

The value characteristic is computed against the USD using the spot exchange rate and the consumer price index (CPI). Bloomberg provides monthly CPI data, except for Australia and New Zealand, where only quarterly data are available. For these two countries, the most recent values are carried forward to the subsequent months until the new quarterly data become available. Because we consider the EUR as the currency for Germany, we take the CPI into account only for Germany. To account for the deviation from uncovered interest rate parity, we exclude 60 months of observations, so that our sample for the value characteristic $Q_{h,t}^i$ spans between February 1994 through December 2020.

2.2.5 Macro-Based FX Factors

The currency fluctuations of a given country are, at least theoretically, linked to its economic fundamentals. Engel and West (2005) argue that exchange rates reflect the expectations of changes in macroeconomic fundamentals. They use a present value model that allows for greater emphasis on future expectations of fundamentals and find evidence of weak forecasting ability. Numerous other studies in the exchange rate predictability literature find similar weak evidence in lieu of exploiting fundamentals for predicting currency returns. Rossi (2013) reviews the literature on exchange rate forecasting and concludes that it depends on a range of elements such as predictors, forecasting horizons, sample periods, models, and forecast evaluation methods. Rossi's paper also highlights the outperformance of linear models over their more complex counterparts.

Cochrane (2017) asserts the significance of the relationship between business cycles and currency returns and emphasizes that they are necessary for the empirical validation of risk-based models. Macroeconomic conditions can hence be leveraged to understand

and explain currency excess returns. Apart from the style-based factors such as carry, value and momentum, we investigate three macro-based FX factors related to output gap, the Taylor rule and Dollar exposure, each of which is outlined below.

Output Gap

Macroeconomic theory postulates that currencies of strong economies tend to appreciate against those of weaker economies. The current strength of an economy can be deduced based on its position in the cycle, for example, whether the economy is approaching the peak of the cycle or is closer to the trough of the cycle. Specifically, a simple output gap measure can be used to capture the current state of an economy.

The output gap is defined as the difference between an economy's actual output and its maximum potential, measured as a percentage of its gross domestic product (GDP). Colacito, Riddiough and Sarno (2020) sort currencies of 27 countries on the basis of output gap in order to capture the impact of business cycles on currency returns. The resulting strategy generates a Sharpe ratio of 0.74, and a similarly high Sharpe ratio in a multivariate model with carry, value, momentum, and Dollar carry. Their output gap portfolio shows zero correlation with the carry trade portfolio, thereby suggesting diversification benefits.

We follow Colacito, Riddiough and Sarno (2020) and Bartram, Djuranovik and Garratt (2020) in constructing the output gap-based FX factor. As the output gap is not directly observable, we use detrended monthly industrial production (IP) for each country in our sample. Specifically, we use the linear projection method of Hamilton (2018) as follows,

$$IP_{i,t} = \beta_0 + \beta_1 \cdot IP_{i,t-13} + \beta_2 \cdot IP_{i,t-14} + \ldots + \beta_k \cdot IP_{i,t-24} + \epsilon_{i,t} \tag{2.6}$$

where $IP_{i,t}$ is industrial production at time t for country/currency i, β_0 is the intercept term, β_1, $\beta_2 ... \beta_k$ are the coefficients associated with each lagged industrial production terms $IP_{i,t-13}$, $IP_{i,t-14} ... IP_{i,t-24}$ respectively, and $epsilon_{i,t}$ is the error term. The inclusion of the constant term (β_0) allows for capturing the baseline level of industrial production, while β_1 accounts for the influence of the lagged industrial production terms. We use an expanding window estimation so as to include all the data available at time t.

Hence, the residuals from this regression gives the detrended output, thereby capturing the unexplained variation in industrial production caused by cyclical fluctuations which are not accounted for by 2.6. Since the goal of the output gap measure is to capture the business cycle-related deviations from potential output, using these residuals as an estimate of the output gap provides valuable insights into the current state of the economy in relation to its full potential, thereby helping identify strong and weak economies.

An output gap-based FX strategy would target exposure to stronger countries relative to the base currency USD. The strategy involves taking a long position in currencies of countries with output gaps above that of the U.S. and a short position in currencies of countries with output gaps below. In essence, this strategy entails holding a long position in a basket of currencies from strong economies and a short position in currencies from weaker economies.

The Taylor Rule

The Taylor rule evaluates changes to monetary policy based on inflation and real activity by tracking the Fed's decisions on adjusting short-term interest rates post-1987 (see Taylor (1993) for empirical application of macroeconomic policy evaluation). It provides a comprehensive framework for assessing changes in monetary policy, particularly through the lens of inflation and real economic activity. The rule, developed based on empirical evaluations of macroeconomic policy post-1987, guides decisions on adjusting short-term interest rates, with the Federal Reserve serving as a key reference point (Taylor, 1993). Nominal interest rates, according to the Taylor rule, are contingent upon the current inflation rate, the inflation gap, the output gap, and the equilibrium interest rate. In line with the linear representation proposed by Castro (2011), the Taylor rule is expressed as,

$$i_t^* = \bar{r} + \pi^* + \beta(\pi_t - \pi^*) + \gamma(y_t - y^*) \tag{2.7}$$

Thus, according to the Taylor rule, the nominal short-term interest rate (i^*) would increase if inflation (π_t) exceeds a target inflation rate (π^*) or if output (y_t) increases above its trend or potential value (y^*). Therefore, the coefficients β and γ denote the sensitivity of interest rate policy to deviations in inflation and to the output gap respectively. The coefficients β and γ in the equation signify the sensitivity of interest rate policy to deviations in inflation and the output gap, respectively.

Engel and West (2005) offer a sound reasoning for the relevance of the Taylor rule strategy for exchange rate predictability. They posit that the currencies of economies where current output is above their potential are expected to appreciate. The Taylor rule has shown consistent short-term out-of-sample forecasting ability in various follow-up studies (Mark (2009); Molodtsova and Papell (2009); Wang and Wu (2012); and Molodtsova, Nikolsko-Rzhevskyy and Papell (2008) among others). The Taylor rule-based factor is constructed from the output gap and an implicit output deflator by simply summing 1.5 times the deflator and 0.5 times the output gap.

Elevated values of Taylor rule signals, which adjust for economic conditions through the consideration of the output gap and prevailing interest rates, suggest expectations of higher (or lower) inflation compared to the target. By taking long positions in assets with high signals and short positions in currencies with low values of the signal, a Taylor Rule-based factor portfolio aims to benefit from expected outperformance in countries or currencies with favorable economic conditions.

Dollar Exposure

Building on the work of Lustig, Roussanov and Verdelhan (2014), Verdelhan (2018) develops a Dollar exposure factor that measures the average change in the USD versus other currencies, which distinguishes the local shocks from the global shocks. By sorting countries by their dollar currency betas, Verdelhan (2018) observes a highly significant spread between long and short legs. The process of building portfolios of currencies of countries sorted by their time-varying exposures to the dollar factor first involves running a regression between the exchange rate changes and interest rate differential, the carry factor, the interaction of interest rate differential and carry factor and the dollar factor given as follows,

$$\begin{aligned}\Delta\text{Exchange Rate}_{i,t} =& \alpha + \beta_1 \text{Interest Rate Differential}_{i,t} + \beta_2 \text{Carry Factor}_{i,t} \\ &+ \beta_3 (\text{Interest Rate Differential}_{i,t} \times \text{Carry Factor}_{i,t}) \\ &+ \beta_4 \text{Dollar Factor}_{i,t} + \epsilon_{i,t}\end{aligned} \tag{2.8}$$

Equation 2.8 is instrumental in evaluating how fluctuations in the Dollar factor contribute to the broader dynamics of exchange rate movements, and the resulting β_4 coefficients represents the time-varying exposures to the dollar factor. Sorting the currencies based on these coefficients reflects each country's dynamic exposure to the dollar factor. Following the methodology introduced by Verdelhan (2018), this factor takes a long position when β_4 is positive and short otherwise.

2.3 Optimal Currency Tilting

Modeling optimal portfolio weights involves evaluating a wide range of risk-reward trade-offs and investment constraints. Brandt *et al.* (2009) exploit the cross-sectional characteristics of equity returns to obtain optimal portfolio weights assuming a linear portfolio policy that models optimal portfolio weights as the sum of a benchmark weight, plus a deviation term depending on chosen characteristics. We next leverage their parametric portfolio policy framework to gauge the currency characteristics needed to generate a currency allocation and to harness the associated premia.

2.3.1 Parametric Portfolio Policy Framework

The parametric portfolio policy (PPP) framework specifically allows to model the weight of an asset as a function of its characteristics for which the coefficients are estimated by maximizing investor utility. Brandt *et al.* (2009) consider an investor seeking to maximize the conditional expected utility of her portfolio return, $r_{p,t+1}$:

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} E_t\left[u(r_{p,t+1})\right] = E_t\left[u\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right)\right], \tag{2.9}$$

where $w_{i,t}$ denotes the portfolio weight for asset i among the total number of assets N_t at time t. The authors propose to model the portfolio weight as a linear function of its characteristics $x_{i,t}$ as follows:

$$w_{i,t} = w(x_{i,t};\phi) = \overline{w_{i,t}} + \frac{1}{N_t}\phi' x_{i,t}, \tag{2.10}$$

where $\overline{w_{i,t}}$ is the weight of asset i in the benchmark portfolio, ϕ is the weight of the characteristic in the parametric portfolio that must be estimated as part of the utility maximization, and $x_{i,t}$ is the vector of cross-sectionally standardized characteristics of asset i at date t. In our case, we utilize the US 3-month LIBOR as the cash benchmark, and we compute $\overline{w_{i,t}}$ as the monthly returns associated with this as the benchmark. Parameterization (2.10) implicitly assumes that the chosen characteristics fully capture the joint distribution of asset returns. The portfolio policy is embedded in the idea of estimating the weights as a function of the characteristics, which applies to all assets over time, rather than estimating one weight for each asset.

Naturally, the cross-sectional distribution of the standardized characteristics is stationary through time and the cross-sectional mean for each standardized characteristic is zero. Thus, deviations from the benchmark are equivalent to a zero-investment portfolio. The portfolio parameterization implies that the chosen characteristics convey various aspects of the joint distribution of returns. We rewrite the optimization problem in terms of ϕ-coefficients as follows:

$$\max_{\phi} E\left[u(r_{p,t+1})\right] = E\left[u\left(\sum_{i=1}^{N_t} f(x_{i,t},\phi) r_{i,t+1}\right)\right] \tag{2.11}$$

where we parameterize the optimal portfolio weights as a function of the chosen currency characteristics. The first-order condition of the maximization problem is then given by:

$$\frac{1}{T}\sum_{t=0}^{T-1} h(r_{t+1}, x_t; \phi) \equiv \frac{1}{T}\sum_{t=0}^{T-1} u'(r_{p,t+1})\left(\frac{1}{N_t} x_t' r_{t+1}\right) = 0, \tag{2.12}$$

where $u'(r_{p,t+1})$ denotes the first derivative of the utility function. Thus, the optimization problem can be interpreted as a method of moments estimator. Based on Hansen (1982), the asymptotic covariance matrix estimator is

$$\Sigma_{\phi} \equiv AsyVar[\hat{\phi}] = \frac{1}{T}[G'V^{-1}G]^{-1}, \tag{2.13}$$

where

$$G \equiv \frac{1}{T}\sum_{t=0}^{T-1} \frac{\partial h(r_{t+1}, x_t; \phi)}{\partial \phi} = \frac{1}{T}\sum_{t=0}^{T-1} u''(r_{p,t+1})\left(\frac{1}{N_t} x_t' r_{t+1}\right)\left(\frac{1}{N_t} x_t' r_{t+1},\right)' \tag{2.14}$$

and V is a consistent estimator of the covariance matrix of $h(r,x;\phi)$ and u'' is the second derivative of the utility function. Since we construct long-short portfolios, where short selling is permitted, we ensure that the sum of the long legs and short legs is equal to zero, i.e., the long leg positions are fully financed by the short leg.

2.3.2 Naive Currency Portfolio Construction

To benchmark the performance of the PPPs, we construct naively weighted currency portfolios for each of these currency characteristics. To construct such portfolio, we rank the G10 currencies according to each characteristic. The top and bottom three currencies form the long and short legs of the portfolio, respectively, based on an equal-weighting scheme, see Kroencke, Schindler and Schrimpf (2013). Specifically, we will use the characteristics x_t defined in the previous section. The long ($\mathcal{L}_t^j$) and short ($\mathcal{S}_t^j$) legs with $N = 9$ sets of currencies are defined as follows:

$$\mathcal{L}_t^j = \begin{cases} 1 & \text{if} \quad x_t^j \geq q(x_t)_{1-p} \\ 0 & \text{if} \quad x_t^j < q(x_t)_{1-p}, \end{cases} \tag{2.15}$$

and

$$\mathcal{S}_t^j = \begin{cases} 1 & \text{if} \quad x_t^j \leq q(x_t)_p \\ 0 & \text{if} \quad x_t^j > q(x_t)_p, \end{cases} \tag{2.16}$$

where $q(x_t)$is the p-quantile of x_t and $p = \frac{3}{9}$. The naively weighted currency portfolios have a holding period of one month and are rebalanced monthly.

2.3.3 Empirical Results

We next implement the parametric portfolio policy for a mean-variance investor. We choose a conservative risk aversion coefficient of $\gamma = 10$ for the analysis to represent moderate risk aversion. We use an initial period of five years in the PPP optimization to determine the optimal coefficients. Our backtest thus begins from February 1999. The portfolios are rebalanced monthly, with an expanding window of 60 months. We also construct one-month and twelve-month cross-sectional momentum factors using the same method as for the three-month momentum.

Table 2.3.1 gives the estimation results and performance statistics for the six univariate PPPs, the multivariate PPP, and the naive models. In Panel A, we test all the chosen factors univariately. Carry and value characteristics are positively significant at the 5% level, while the Taylor rule-based factor is positively significant at the 1% level. This indicates that the PPP methodology correctly identifies the expected direction of the respective trades. In this vein, the one- and twelve-month momentum characteristics indicate price momentum effects, yet all momentum coefficients are insignificant.[2] The FX carry strategy offers the best risk-return trade-off in terms of the Sharpe ratio (0.44) followed by the Taylor rule-based strategy (0.34) and the FX

[2] Barroso and Santa-Clara (2015) observe a significant momentum coefficient (before transaction costs) when using a larger currency universe, including emerging markets, and over a longer investment horizon.

value strategy (0.31). However, it is important to note that the FX carry strategy is vulnerable to crash risk as indicated by its 27% drawdown.

In Panel B, we show FX factor performance when driven by the naive portfolio construction paradigm laid out in Section 2.2. Across the six FX characteristics, we note that risk-adjusted performance is slightly elevated relative to the optimized factor-PPPs from Panel A. In defense of the latter, we rationalize that naive factor portfolios come with the benefit of hindsight bias; while naive factor portfolios are designed to follow the prescribed rationale of a given factor throughout the sample, the PPPs have to first learn this rationale from the data before going "all-in".

Next, we investigate multivariate PPP strategies to leverage diversification benefits across FX characteristics. As a benchmark, we consider a simple 1/N-portfolio that equally weights the six naive FX factors from Panel B. This aggregate strategy gives a Sharpe ratio of 0.5. Panel C shows the results for the multivariate PPP that aims to capture the contributions and diversification coming from all six characteristics. In this context, we find carry and value to be significant (at the 1%-level). Moreover, we note that the multivariate PPP has a lower risk-adjusted performance than the naive 1/N aggregate (Sharpe ratio of 0.35 versus 0.5 for the 1/N-portfolio). To further investigate this outcome, we look into a fundamentals-only PPP in Panel D. Therein, value and Taylor-rule characteristics are significant at the 5%-level. Yet, the Sharpe ratio of this combination is similar to the performance of the respective univariate PPPs, suggesting little room for diversifying across fundamental FX factors. Against this backdrop, we lastly test a PPP that focuses on one fundamental characteristic only (namely value) together with carry and momentum. This "classic" FX factor combination produces a Sharpe ratio of 0.47 which is on par with the naive 1/N combination.

To foster intuition about how the parametric portfolio policies work, we decompose each currency weight by the six characteristics. Figure 2.3.1 illustrates the optimal weights for two currencies, CHF and NZD, over time.[3] The CHF weights are almost always negative, and are mainly driven by the carry characteristic, as expected, but also by the three fundamental characteristics value, Dollar exposure and the Taylor rule. In contrast, the NZD weights are typically positive; naturally, this outcome is mainly driven by the large carry contribution. For the remaining characteristics, contributions vary in magnitude and sign through time.

[3]The results for all other currency pairs are available upon request.

Table 2.3.1 Currency Tilting: Performance Panel A gives the estimated results of the univariate PPPs as well as the performance statistic of each investment style. Panel B gives the performance statistics for the naive portfolio construction of the six investment styles. Return, volatility and maximum drawdown figures are measured in percentage terms. Panel C groups the factors based on economic fundamentals. Panel D and E give the estimated results for the multivariate optimization with three and six factors, respectively. The sample period is February 1994 through December 2020. *, ** and *** represents significance at 10%, 5% and 1% levels, respectively.

	$\hat{\phi}$	S.E.	Return p.a. (%)	Vola p.a. (%)	Sharpe ratio	Max Draw-down (%)
Panel A: Univariate models						
Carry	1.58**	0.60	5.29	7.20	0.44	26.59
Value	1.57**	0.64	3.90	5.70	0.31	8.89
CS1Momentum	0.05	0.62	1.85	1.80	-0.17	3.93
CS3Momentum	-0.01	0.63	2.06	3.30	-0.02	7.10
CS12Momentum	0.39	0.61	2.32	3.80	0.05	10.50
DollarExposure	0.83	0.57	2.73	4.00	0.15	15.86
OutputGap	0.82	0.80	2.29	2.60	0.06	7.15
TaylorRule	1.86***	0.71	5.24	9.10	0.34	34.24
Panel B: Naive models						
Carry			5.79	7.85	0.46	23.57
Value			5.17	7.39	0.41	12.08
CS3Momentum			0.97	7.71	-0.15	39.13
DollarExposure			3.43	5.51	0.23	15.05
OutputGap			3.21	5.78	0.18	17.42
TaylorRule			5.09	7.02	0.42	26.19
Aggregate Portolio (1/N)			3.94	3.62	0.50	7.15
Panel C: 6-Factor PPP						
			6.31	12.00	0.35	35.00
Carry	2.09**	1.07				
Value	1.51**	0.67				
CS3Momentum	0.12	0.64				
DollarExposure	-1.21	0.88				
OutputGap	1.02	0.87				
TaylorRule	0.77	1.16				
Panel D: Fundamentals PPP						
			6.00	11.30	0.34	39.04
Value	1.24*	0.65				
OutputGap	0.82	0.86				
TaylorRule	1.66**	0.73				
Panel E: Carry, Value & Momentum PPP						
			6.11	8.38	0.47	18.55
Carry	1.59***	0.60				
Value	1.60**	0.62				
CS3Momentum	0.25	0.25				

Fig. 2.3.1 Decomposing optimal currency tilting weights. The figure shows the currency tilting allocation over time and the contribution of each conditioning variable. The right-hand chart is for CHF; the left-hand chart is for NZD. The sample period is February 1994 through December 2020.

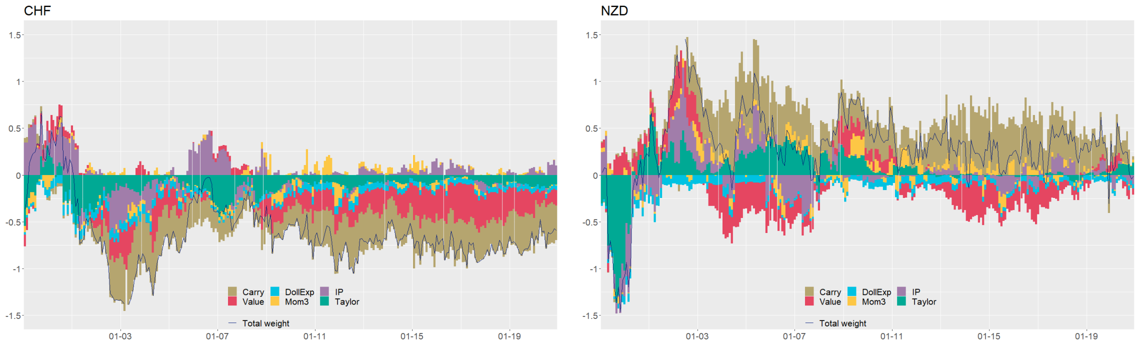
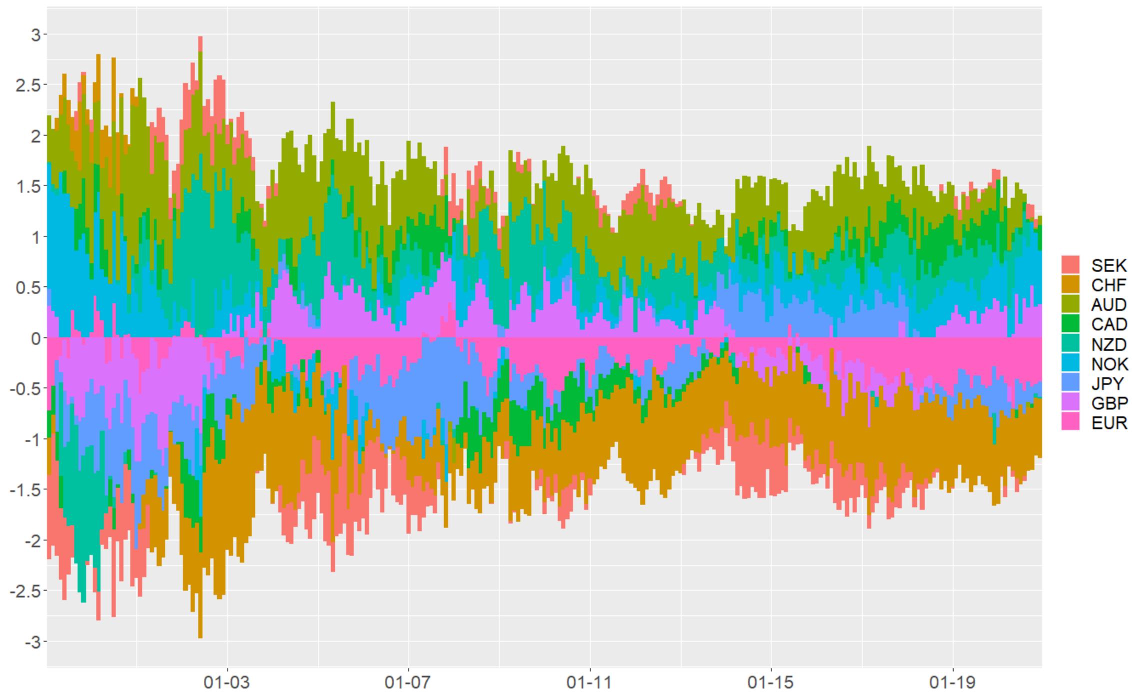


Figure 2.3.2 shows the aggregate currency allocations according to the parametric portfolio policy with no restrictions on portfolio weights. High carry currencies such as AUD and NZD have predominantly long positions; low carry currencies, such as CHF, EUR, and JPY, constitute almost the entire short leg. The SEK oscillates modestly between positive and negative weights when compared with the other currencies. CHF, followed by EUR and JPY, represent the major short positions with average weights of -60.80%, -23.69%, and -19.15% respectively. AUD, NZD, and NOK hold major long positions, with weights averaging 54.87%, 31.71%, and 24.05% respectively.

The multivariate PPP strategies combining the six currency characteristics perform better in terms of the Sharpe ratio and maximum drawdowns than any other portfolio and this outperformance can largely be attributed to the effective utilization of diversification benefits. This is evident from Figure 2.3.1, which displays a de-levering of carry trades during the 2008 period, thanks to factor diversification. This shift in allocation also reflects a desire for moderate risk aversion, and, as a result, a reduced emphasis on the carry trade. To check the robustness of these results, we vary the risk aversion parameter. In unreported results, we find that the carry and value characteristics remain significant for risk-loving and risk-averse investors alike. The other characteristics also exhibit similar behavior, in line with our findings in Panel E of Table 2.3.1. In summary, the PPPs are effective in utilizing cross-sectional currency characteristics to guide an optimal portfolio.

Fig. 2.3.2 Aggregate optimal currency tilting allocation. The sample period is February 1994 through December 2020.



2.4 The Notion Of Currency Timing and Currency Factor Timing

2.4.1 Currency Timing

In this section, we first turn to time-series information that could inform an optimal currency timing strategy in order to estimate optimal currency portfolio weights according to the PPP framework of Brandt and Santa-Clara (2006). Whether macroeconomic and financial variables can forecast FX returns is still widely debated. Nevertheless, there is evidence on the relevance of fundamental variables, interest rate-related variables (Cornell and Dietrich, 1978), and technical indicators (Cotter, Eyiah-Donkor and Potì, 2017) in forecasting FX returns. Second, rather than timing individual currencies we investigate timing the cyclicality of currency factors, focusing on the carry trade. Specifically, we illustrate how to incorporate this notion of factor timing into a PPP for optimal currency tilting as implemented in Section 2.

Fundamental Variables

We consider 14 predictor variables, as suggested by Welch and Goyal (2008)[4]: dividend price ratio (dp), dividend yield (dy), earnings price ratio (ep), dividend payout ratio (de), stock variance ($svar$), book-to-market ratio (bm), net equity expansion ($ntis$), treasury bills (tbl), long term yield (lty), long term rate of return (ltr), term spread (tms), default yield spread (dfy), default return spread (dfr) and inflation ($infl$).

It is important to ensure that the predictor variables are not correlated because lagged variables could exhibit very high first-order autocorrelations. Ferson, Sarkissian and Simin (2003) suggest "stochastic de-trending" of the lagged variable in order to avoid the bias that can result from spurious regressions. We thus standardize any predictor variable at time t by subtracting its arithmetic mean and dividing by its standard deviation. For the calculation of the mean and standard deviation we use a rolling window covering the 12 months preceding (and thus excluding) t. Furthermore, there are few standardized fundamental variables that attain extreme values, which we truncate at ± 5.

Technical Indicators

Technical indicators can time trades by recognizing the drivers of international financial markets from a behavioral perspective. Similar to Hammerschmid and Lohre (2018), we include 11 technical indicators based on two sets of trading rules related to the general concepts of momentum (MOM_k) and moving averages (MA_{s-l}).

1. *Momentum (MOM_k)*: The momentum indicator gives a buy signal if the end-of-month closing spot exchange rate indicates an upward trend, i.e., when S_t is higher than S_{t-k}, and a sell signal otherwise.

$$MOM_k = \begin{cases} 1 & \text{if} \quad S_t > S_{t-k} \\ 0 & \text{if} \quad S_t \leq S_{t-k} \end{cases} \tag{2.17}$$

   where S_t is the end-of-month closing spot exchange rate. We compute five momentum indicators for different look-back periods with $k = 1,\ 3,\ 6,\ 9$, and 12 months.

[4]The dataset is available at https://www.ivo-welch.info/professional/goyal-welch/

2. *Moving Average (MA_{s-l})*: Trading rules based on moving averages detect trends and potential breaks in such trends. The moving average at a given time t over j months is given by

$$MA_{j,t} = \frac{1}{j}\sum_{i=0}^{j-1} S_{t-i} \quad \text{for} \quad j \;=\; s,\; l, \tag{2.18}$$

where S_t is the end-of-month closing spot exchange rate of the currency; $s = 1, 2, 3$ is used for short-term moving averages, and $l = 9, 12$ is for long-term moving averages. The resulting indicator would give a buy signal when the short-term moving average crosses the long-term moving average from below, and a sell signal otherwise:

$$MA_{s-l} = \begin{cases} 1 & \text{if} \quad MA_{s,t} > MA_{l,t} \\ 0 & \text{if} \quad MA_{s,t} \leq MA_{l,t} \end{cases} \tag{2.19}$$

Hence, depending on the different long- and short-term combinations, we would have six moving average indicators for the analysis.

Predictor Variables Selection

Now that we have carefully chosen 14 fundamental variables and 11 technical predictors, it is essential to check for multicollinearity. Figure 2.4.1 shows the correlation structure (using the currency pair USD/EUR[5] as an example) for the fundamental variables and technical indicators for our entire sample from February 1989 through December 2020. As expected, and as the bottom right of the chart shows, the technical indicators are highly correlated. The fundamental variables display a heterogeneous correlation structure. While the valuation ratios *dp* and *dy* show the maximum positive correlation of 0.8, their peers, *ep* and *de*, have the highest negative correlation, which amounts to -0.7. Notably, fundamental and technical variables are fairly uncorrelated, suggesting a complementary predictive ability and suitability for our analysis.

To reduce the number of predictors, we follow Neely, Rapach, Tu and Zhou (2014) and Hammerschmid and Lohre (2018) and apply principal component analysis (PCA) separately to the fundamental and technical indicators. This procedure eliminates

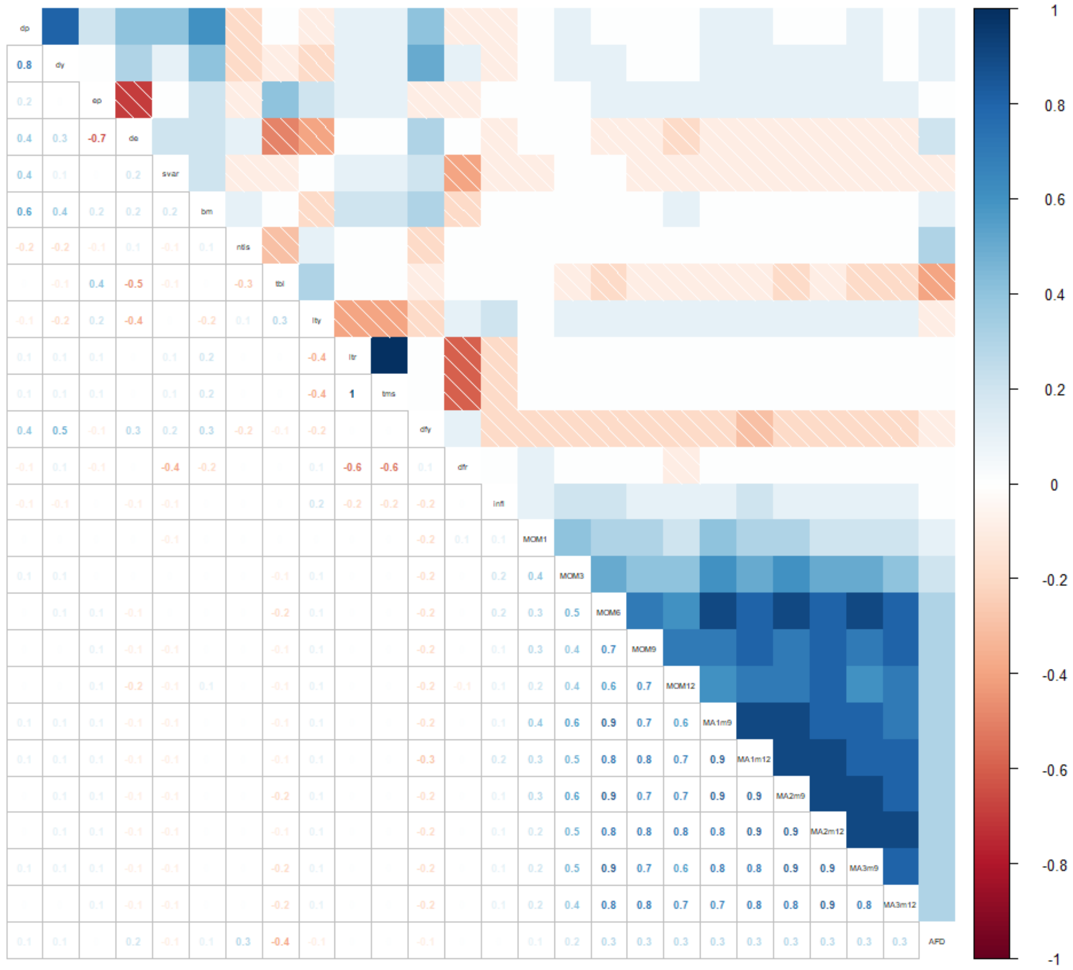
[5]We will use the USD as the benchmark currency, and we refer to each currency pair only by its matched currency, thus denoting USD/EUR as EUR.

the noise within the predictors and also provides orthogonal predictors, which helps avoid multicollinearity. The PCA results confirm our findings from the correlation map: It takes the first three principal components of the fundamental variables to jointly explain 56% of the data variation. Conversely, we only need the first principal component of the technical indicators to explain about 86% of the data variation.[6] Hence, for our analysis we use the first three principal components of the fundamental variables (denoted as F_t^{Fun1}, F_t^{Fun2}, and F_t^{Fun3}), and the first principal component of the technical indicators (denoted as F_t^{Tec}).

In addition to fundamental and technical indicators, we follow Bartram, Djuranovik and Garratt (2020) and construct a signal representing the Dollar carry trade as per Lustig, Roussanov and Verdelhan (2014). The Dollar carry trade strategy goes long on foreign currencies whenever the average foreign short-term interest rate is above the U.S. interest rate (for example during U.S. recessions). It shorts all foreign currencies otherwise. This measure captures the variations in the country-specific price of risk that the standard carry trade would fail to capture. To construct this predictor variable, we compute the average forward discount (AFD) by averaging the naive carry characteristic cross-sectionally across our currency universe.

[6]Note that these results largely hold for all currencies. For the sake of space, we do not include them here, but they are available upon request.

Fig. 2.4.1 Correlation matrix of predictor variables. The correlation between fundamental variables is shown in the top left corner. The bottom right corner shows the correlation structure of the technical indicators for EUR. The sample period is February 1989 through December 2020.



2.4.2 Currency Factor Timing

Section 2.3 documents that a PPP for currency tilting can successfully exploit cross-sectional currency characteristics. Still, we know that these factor strategies come with cyclicality that can be capitalized if anticipated in advance. Specifically, the carry trade is prone to crash risk in flight-to-quality events. Diversifying the carry signal by combining it with momentum and value signals is a viable approach in this context. However, momentum strategies can be an expensive hedge. Thus, we explore whether there are alternative ways to navigate the downside risk of the carry trade by investigating integrated parametric tilting policies conditional on carry timing signals.

Recent FX literature has explored the possibility of timing carry strategies with different indicators, such as FX volatility-based exchange rate regimes, the bid-ask spread, equity/bond returns, and the Cboe Volatility Index (VIX). Christiansen, Ranaldo and Söderlind (2011), Clarida, Davis and Pedersen (2009) offer insights into the economic consequences of high versus low distress periods, business cycles, and specific events on the carry trade. Ideally, we could condition a currency tilting policy on such information, which effectively represents an integrated approach to currency factor investing. For example, the model may anticipate the unwinding of carry trade positions during the 2008 global financial crisis. Moreover, this integrated approach could offer a deeper understanding of the variations in currency factor exposures.

Brandt *et al.* (2009) allow the coefficients that capture the joint distribution of returns to be time-variant by modifying the portfolio policy as follows:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\theta^T(z_t \otimes x_{i,t}), \tag{2.20}$$

where z_t is a vector of predictors known at time t. Hence, the effect of the characteristics on the portfolio weights will vary with the realization of the predictors z_t. To demonstrate this, Brandt *et al.* (2009) use an indicator based on the sign of the slope of the yield curve to obtain the coefficients of the parametric equity portfolio policy in order to time size, value, and momentum factors. They model the coefficients as a function of the yield curve slope, so that the effect on the joint distribution of returns can vary depending on the business cycle (as measured by the slope of the yield curve).

Liquidity- And Volatility-Based Indicators To Time The Carry Trade

We need to identify relevant predictors for timing the carry trade, which univariately has not only higher returns but also the highest risk, as seen before. Bekaert and Panayotov (2020) distinguish currency carry trades based on the Sharpe ratio and highlight the relevance of equity market risk factors for the G10 currencies between 1984 and 2014. Their results show that carry trades are driven by certain subsets of the G10 currencies. In contrast, Christiansen *et al.* (2011) use a currency volatility-based regime-dependent pricing model to capture the time-varying systematic risk of carry trades. Clarida *et al.* (2009) also explore the volatility-regime based sensitivity of carry trades. They use an Exponential Generalized Autoregressive Conditional

Heteroskedasticity (EGARCH) model during the Russian financial crisis period of 1997–1998 versus the non-crisis period and establish volatility regime sensitivity of the carry trade. Jordà and Taylor (2012) also turn to a regime-based model and conclude that such models better explain carry trades.

Carry trade returns and crash risk have also been linked consistently by numerous researchers. Brunnermeier *et al.* (2008) relate the unfavorable movements in funding liquidity and crash risk of carry trades. They explain the unwinding of carry trades when funding liquidity falls. Such evidence from the literature supports the notion of liquidity-based sensitivity of carry trades. Thus, we look for indicators that capture both liquidity and volatility, which in turn can be used for timing the carry trade.

TED Spread The TED spread is a common proxy for money market liquidity and is defined as the difference between the three-month LIBOR Market Model and three-month Treasury bill. It gauges the willingness of banks to lend money in the interbank market. The money market is said to be illiquid when the TED spread widens and vice versa. It has been observed that the TED spread naturally has a positive correlation with currency crashes.

FX Volatility We construct the FX volatility measure by using an exponential weighted moving average (EWMA)-based realized volatility, similar to Clarida *et al.* (2009). We use an exponential decay parameter λ of 0.95, which denotes a half-life in the exponential weights of 14 days in a three-month window for constructing our volatility estimates.

2.5 Optimal Currency Timing And Currency Factor Timing

2.5.1 Currency Timing

In this section, we present the Brandt and Santa-Clara (2006) framework that lends itself naturally to deriving optimal currency timing strategies. Specifically, we leverage classic timing signals as given by the fundamental variables and technical indicators introduced in Sections 3.1.1 and 3.1.2.

Methodology Of Brandt and Santa-Clara (2006)

Brandt and Santa-Clara (2006) consider a risk-averse investor who maximizes a mean-variance utility function over next period's wealth:

$$\max_{w_t} E\left[w_t' r_{t+1} - \frac{\gamma}{2} w_t' r_{t+1} r_{t+1}' w_t\right], \tag{2.21}$$

where γ is the risk-aversion parameter, w_t denotes the vector of currency factor portfolio weights and r_{t+1} is the vector of future excess return of the $N = 9$ currency pairs. The remainder is invested into the risk-free asset if the PPP is not fully invested. The Brandt and Santa-Clara (2006) methodology assumes optimal portfolio weights w_t are linear in a column vector z_t of K state variables, thereby capturing time variation in expected returns as follows:

$$w_t = \theta z_t, \tag{2.22}$$

where θ is an $(N \times K)$ matrix of parameters. Replacing the linear portfolio policy, w_t, in (2.21) yields:

$$\max_{\theta} E_t\left[(\theta z_t)' r_{t+1} - \frac{\gamma}{2}(\theta z_t)' r_{t+1} r_{t+1}' (\theta z_t)\right]. \tag{2.23}$$

with

$$(\theta z_t)' r_{t+1} = z_t' \theta' r_{t+1} = vec(\theta)'(z_t \otimes r_{t+1}), \tag{2.24}$$

where $vec(\theta)$ is a vectorization of the matrix θ into a column vector and $\otimes$ is the Kronecker product. Using $\tilde{w} = vec(\theta)$ and $\tilde{r}_{t+1} = z_t \otimes r_{t+1}$, the objective function (2.23) can be rewritten as:

$$\max_{\tilde{w}} E_t\left[\tilde{w}' \tilde{r}_{t+1} - \frac{\gamma}{2} \tilde{w}' \tilde{r}_{t+1} \tilde{r}_{t+1}' \tilde{w}\right]. \tag{2.25}$$

Hence, the original dynamic optimization problem is transformed into a static problem that can be applied to the augmented asset space represented by $\tilde{r}_{t+1}$. This represents the return vector of managed portfolios that invest in a given currency proportional to the value of given state variables. As the same $\tilde{w}$ maximizes the conditional expected utility at all t, it also maximizes the unconditional expected utility, so (2.25) is equivalent to:

$$\max_{\tilde{w}} E\left[\tilde{w}'\tilde{r}_{t+1} - \frac{\gamma}{2}\tilde{w}'\tilde{r}_{t+1}\tilde{r}'_{t+1}\tilde{w}\right]. \tag{2.26}$$

Based on the information embedded in the state variables we can determine the corresponding portfolio policy. An additional benefit of this methodology is that the PPP expresses the portfolio problem in an estimation setup. This allows for the calculation of the standard errors of portfolio weights to assess the significance of a given conditioning variable in the portfolio policy. According to Brandt and Santa-Clara (2006), we use the covariance matrix of $\tilde{w}$ to compute the standard errors as:

$$\frac{1}{\gamma^2}\frac{1}{T \;-\; N \times K}(\iota_T - \tilde{r}\tilde{w})'(\iota - \tilde{r}\tilde{w})(\tilde{r}'\tilde{r})^{-1}, \tag{2.27}$$

where ι_T denotes a $T \times K$ matrix of ones.

Empirical Results

In addition to using the AFD characteristic, we build on the PCA analysis in Section 2.4.1 and select three fundamental principal factors and one technical principal factor. Thus, we are considering five conditioning variables in total (F_1^{Fun}, F_2^{Fun}, F_3^{Fun}, F^{Tec}, AFD). The portfolio optimization will be performed out-of-sample over an expanding window. We will first use an initial window of nine years in order to compute the first optimal portfolio in February 1999 and rebalance on a monthly basis, thus aligning the dates for currency timing and tilting strategies. The risk aversion parameter γ is again fixed at ten. We implement a long-short strategy, so that long positions cancel out short positions to mimic a zero-investment strategy.

Panel A of Table 2.5.1 shows the univariate performance statistics of the AFD-only strategy, the fundamentals-based and the technicals-based strategy. The performance of a strategy using fundamental variables has a much higher drawdown than the AFD-only or the technicals-only strategy. The technicals-only strategy offers the highest return and highest Sharpe ratio with the lowest drawdown amongst the others. We constrain the weights of the PPP portfolio to 200% (in absolute terms) in order to ensure the results are comparable to those of currency tilting (Panel E of Table 2.3.1). The performance of the univariate technical strategy has a Sharpe ratio of 0.35 for the unconstrained version and 0.30 for the constrained version. Because our technical PCA is composed of indicators that capture the trend, the methodology can indeed pick up the time-series momentum phenomenon.

Panel B combines the average forward discount and the fundamental and technical PCA factors in a multivariate setup and reports the performance of the PPP strategy. We observe that the multivariate strategy offers a return of around 14% with a volatility of 49%. The drawdown of the unconstrained strategy is around 89% while that of the constrained strategy is only 21%.[7] Although currency tilting is a better alternative when comparing the constrained strategy results in Panel E of Table 2.3.1 and Panel B of Table 2.5.1, our main takeaway from Table 2.5.1 is the relevance of technical indicators in currency timing.

Table 2.5.1 Currency timing policy: θ coefficients and performance analysis
Panel A shows the univariate performance analysis (returns, volatility, Sharpe ratio and maximum drawdown) from the PPP optimization. The sample period is February 1994 through December 2020. Panel B shows the multivariate performance analysis for the parametric portfolio policy. The performance analysis includes annualized returns and volatility, Sharpe and information ratios, and maximum drawdown for the unconstrained and constrained versions. The weights in the constrained version are restricted to 200% (in absolute terms). *, **, and *** represents significance at 10%, 5% and 1% levels, respectively.

Panel A	Return p.a. (%)	Vola p.a. (%)	Sharpe Ratio	Max Drawdown (%)
AFD PPP	3.19	17.92	0.06	51.93
AFD PPP(Constrained)	1.96	5.93	-0.03	23.26
$F_{1,2,3}^{Fun}$ PPP	3.15	29.50	0.03	70.40
$F_{1,2,3}^{Fun}$ PPP(Constrained)	1.71	5.13	-0.08	13.80
F_{1}^{Tec} PPP	9.26	20.48	0.35	43.19
F_{1}^{Tec} PPP (Constrained)	4.63	8.24	0.30	15.91
Panel B	**Return p.a. (%)**	**Vola p.a. (%)**	**Sharpe Ratio**	**Max Drawdown (%)**
MV PPP	14.07	48.72	0.24	88.69
MV PPP(Constrained)	2.43	6.17	0.05	20.80

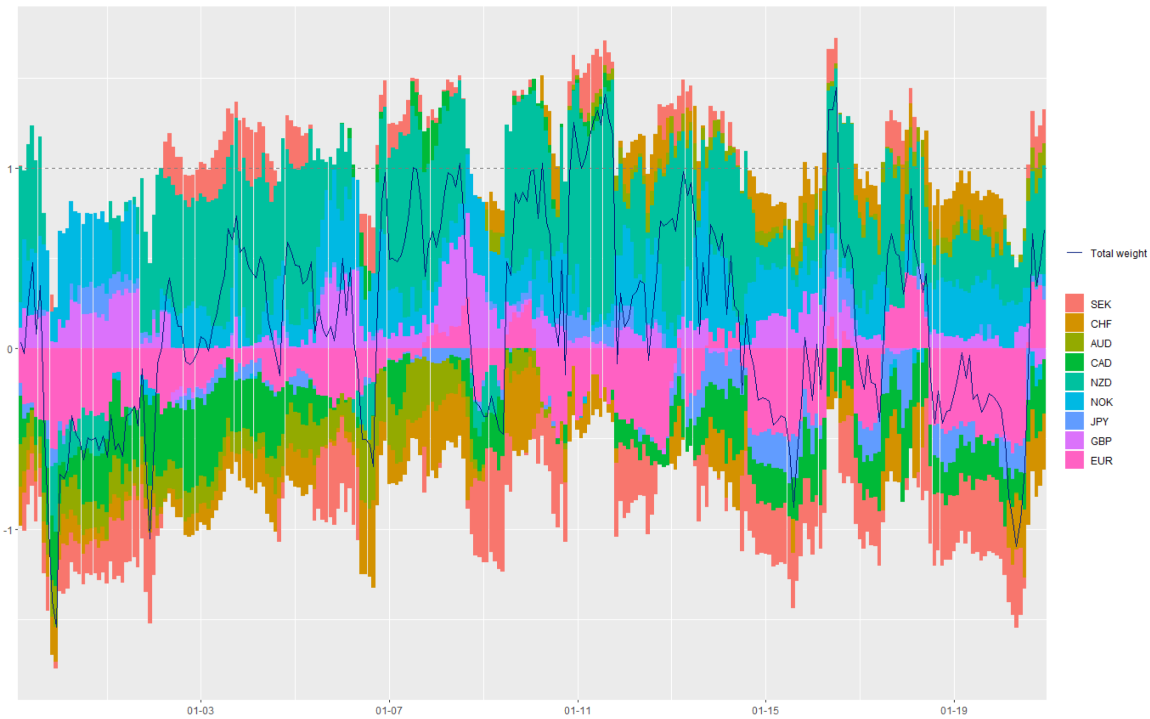
[7] In unreported results, we perform simple robustness tests for the multivariate timing portfolio for different values of the risk aversion parameter and find similar results as above.

Figure 2.5.1 shows the aggregate optimal currency timing policy allocation timed with technical indicators (F_1^{Tec}). Given that the unconstrained version is highly leveraged, we focus on a constrained strategy in which the sum of absolute weights is bound by 200%. Still, unlike the currency tilting strategy depicted in Figure 2.3.2 where some currencies take consistent long and short positions, Figure 2.5.1 lacks any such consistent patterns. Of course, this is expected because we are investigating a trend strategy for which portfolio weights are oscillating more. As visible from the constrained version, SEK is characterized by large short positions whereas NOK and NZD predominantly take long positions towards the end of the sample period. EUR oscillates throughout the sample, with large weights in both long and short legs.

We observe a marked change in the form of a reduction in both the overall allocation and the allocation across individual currencies during the 2008–09 period. For instance, the allocations for GBP drastically reduces post-2008 and even more so after the announcement of Brexit, whereas the overweight allocations in NZD and EUR revert to their pre-2008 allocations after 2009. Consistent with the fluctuations in the cumulative returns of JPY over the entire sample period, the allocations in JPY oscillate throughout the sample, with small weights in both long and short legs. Notably, we find in unreported results that all currencies contribute positively to the timing strategy's performance. The highest performance contribution is found for the NZD positioning, followed by SEK and CHF. Additionally, the first half of the sample period is characterized by higher timing returns, suggesting weaker efficacy of timing signals in the latter half of the sample period.

Although the currency timing PPP delivers higher returns, it is highly leveraged and hence, as expected, the associated risk is very high. In line with the extant literature, trend signals emerge as the strongest currency timing signals and can successfully be operationalized in the presented PPP framework. This motivates us to next investigate currency tilting in a way that operationalizes factor timing.

Fig. 2.5.1 Aggregate optimal currency timing allocation. The sample period is February 1999 through December 2020. The figure shows the allocation for the constrained version where the weights are restricted to 200%.



2.5.2 Currency Factor Timing Through Conditional Currency Tilting

Using the learnings from the previous subsection, we combine the notion of timing with the methodology followed in Section 2.3.1 to design an integrated currency factor timing strategy. This second look at Brandt et al. (2009)'s parametric portfolio policy suggests reconsidering the assumption of time-invariant coefficients.

In this subsection, we create a liquidity regime-based model indicator that is constructed as a dummy variable from the TED spread. Specifically, we build two carry characteristics: the first carry characteristic is set to zero for illiquid months (as defined by the TED spread) and equals the original characteristic otherwise. Vice versa, the second carry characteristic is equal to the original carry characteristic in illiquid months and is zero in liquid months. In the same vein, we construct FX volatility indicators where two regimes are created to differentiate between turbulent periods characterized by extreme volatility versus periods of normal volatility. We use a cut-off of 85% such that the indicator captures the 15% most volatile times. It is important to note that this cut-off was determined using information from the entire dataset, implying that the results may not be entirely out-of-sample.

We use the first five years of the sample period to initialize the parametric portfolio optimization and re-estimate the parameters on a monthly basis with an expanding window. Panel A of Table 2.5.2 includes our univariate estimation results for PPPs timing carry, along with the untimed univariate carry trade in Panel B. To first provide a proof of concept of the proposed methodology, we create a "crystal ball" indicator based on future carry trade returns. To do this, we construct a regime indicator that equals one if next month carry returns are positive, and zero otherwise. Intuitively, the regimes from this perfect foresight indicator should perfectly time the carry trade; at the very least, the conditional currency tilting policy is enabled to perfectly learn from this information as it unfolds in an expanding window estimation. Indeed, given this "crystal ball", we obtain a highly significant positive coefficient in the good regime, and likewise a significant negative coefficient in the bad regime, indicating that the PPP can distinguish between good and bad regimes. Naturally, we observe abnormal performance for the associated dynamic tilting policy, as evidenced by the Sharpe ratio of 3.77. In turn, we are confident in using this framework to test our chosen carry timing indicators.

Panel A further reports the performance statistics and the coefficients of carry portfolios timed with different liquidity and volatility indicators. Compared to the original carry portfolio, the integrated strategy timing the carry trade with the TED spread delivers a higher Sharpe ratio of 0.69 (versus 0.44 for the original carry tilting strategy). Moreover, the maximum drawdown is reduced from 26.59% to 16.06%, evidencing that the TED spread indicator helps mitigate crash risks. The estimated ϕ coefficient is significant for both low and high TED regimes but with opposite signs. This indicates the different impact of the carry characteristic on the joint distribution of returns during periods of high and low market liquidity. The positive ϕ coefficient of the low TED spread regime further stipulates the tilting of the optimal currency portfolio towards carry currencies during liquid periods. In contrast, the negative coefficient hints the tilting of the optimal portfolio towards low interest rate currencies during illiquid periods.

Next we inspect the ability of the FX volatility indicator to time the currency trade. We find a significantly positive coefficient in the low volatility regime; yet, unlike the TED spread indicator, the high volatility regime is still characterized by a positive coefficient, albeit insignificant. Hence, one is not unwinding the carry trade in these periods but though reducing the sizing of carry trade positions. In terms of strategy returns, one however experiences a reduction in risk-adjusted returns relative to the untimed version.

Panel C shows the performance of a multi-factor setup wherein the carry trade is timed using the TED spread. Comparing this multivariate result with that in Panel D of Table 2.3.1, we observe a marked improvement in the Sharpe ratio (going from 0.47 to 0.61), and a reduction in the drawdown. This highlights the improvements offered by an integrated portfolio policy approach. To rationalize the mechanics of the integrated approach, we again consider the two carry currencies, the CHF and NZD. In Figure 2.5.2, we analyse the decomposition of their optimal weights. Carry considerations dominate the short and long positions of CHF and NZD, respectively, during normal times. This confirms our results in Section 2.3. Here, the integrated parametric portfolio policy minimizes the crash risk by varying with market liquidity conditions as proxied for by the TED spread. In the same figure, we note that this timing feature was also active in 2001 during the stock market downturn. Hence, the carry trade positions get automatically adjusted whenever there is an expected drop in liquidity during such periods of financial distress. Adding further support is the expected opposite weight distribution of the high carry currency (NZD) and low carry currency (CHF).

Figure 2.5.3 shows the weights on an aggregate portfolio level. It depicts the eventual reduction of the carry trade positions, especially during the global financial crisis. It also shows the expected significant reduction in all currency weights, backed by a full investment in the risk-free rate. We also note a significant increase in the exposure to safe heaven currencies since 2000.

Table 2.5.2 Currency factor timing results. Panel A presents the estimated results of the univariate parametric portfolio policy, and the performance statistics of the carry strategy timed with different indicators. Panel B gives the estimated result for the univariate carry strategy from table 2.3.1. Panel C gives the estimated results when using different indicators to time the carry trade in a multi-factor setup including the crystal ball exercise. The sample period is February 1994 through December 2020. *, ** and *** represents significance at 10%, 5% and 1% levels, respectively.

	$\hat{\phi}$	S.E.	Return p.a. (%)	Vola p.a. (%)	Sharpe ratio	Max Drawdown (%)
Panel A: Univariate models						
Crystal ball-timed PPP			161.06	42.14	3.77	19.01
Carry x $I(low\ FutureCarry)$	-15.12***	1.48				
Carry x $II(high\ FutureCarry)$	20.48***	1.93				
TED-timed PPP			9.00	9.97	0.69	16.06
Carry x $I(low\ TED)$	2.57***	0.65				
Carry x $II(high\ TED)$	-2.22**	1.09				
FX Vol-timed PPP			4.61	7.87	0.31	28.70
Carry x $I(low\ Vol)$	1.67***	0.66				
Carry x $II(high\ Vol)$	1.27	1.37				
Panel B: Benchmark model						
Carry(Tilting)	1.58**	0.60	5.29	7.20	0.44	26.59
Panel C: Timing carry (Multivariate)						
Optimal Portfolio (Crystal-ball)			163.95	43.20	3.75	18.98
Momentum	-0.31	1.42				
Value	-1.19	1.26				
Carry x $I(low\ FutureCarry)$	-15.45***	1.51				
Carry x $II(high\ FutureCarry)$	20.84***	1.99				
Optimal Portfolio (TED-timed)			8.20	9.97	0.61	17.21
Momentum	-0.31	0.70				
Value	0.81	0.70				
Carry x $I(low\ TED)$	2.44***	0.72				
Carry x $II(high\ TED)$	-1.78	1.48				
Optimal Portfolio (FX Vol-timed)			5.26	8.49	0.37	23.21
Momentum	0.23	0.65				
Value	1.54***	0.62				
Carry x $I(low\ Vol)$	1.62**	0.67				
Carry x $II(high\ Vol)$	1.47	1.41				

Fig. 2.5.2 Decomposition of the optimal currency weights. The figure shows the currency weights decomposition in the PPP timed with FX liquidity indicator, the TED spread and the contribution of each conditioning variable. The left-hand chart is for CHF; the right-hand chart is for NZD. The sample period is February 1999 through December 2020.

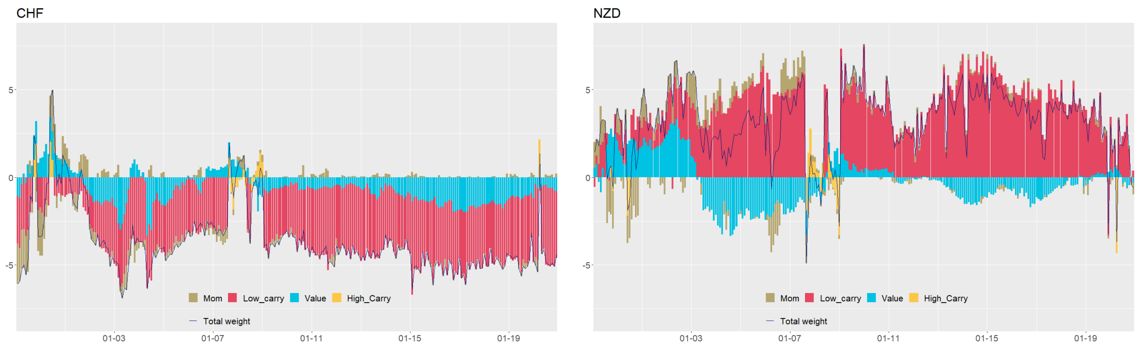
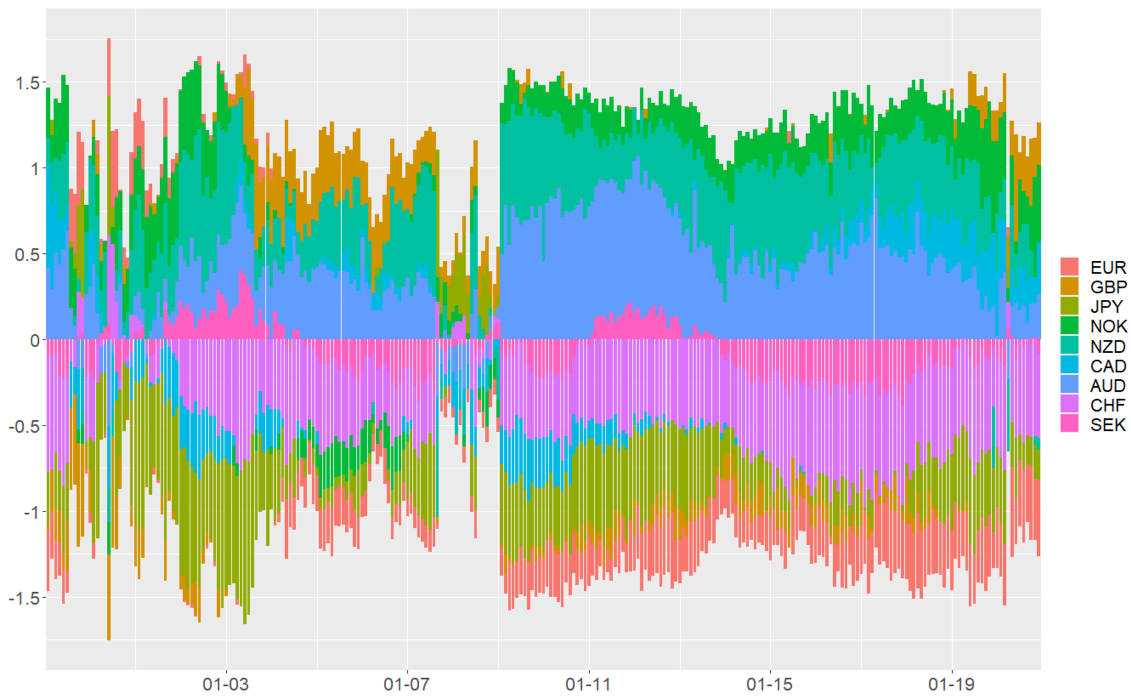


Fig. 2.5.3 Currency factor timing: aggregate allocation. The sample period is February 1999 through December 2020.



2.6 Conclusion

Extant literature on currency investing has explored the choice and relevance of style-based and macroeconomic variables mostly using univariate factor approaches or static allocations. We take a different route by focusing on a dynamic approach and a multivariate framework that combines well-known FX factors. We rely on the PPP of Brandt and Santa-Clara (2006) and Brandt *et al.* (2009), which allow for both tilting and timing currencies using salient FX characteristics and time-series indicators, respectively. As for currency tilting, we exploit cross-sectional information by using factors such as value, momentum, carry, and macro-based factors. As for currency timing, we confirm the prominent role of technical indicators whilst fundamental variables are found to add little value.

In sum, we find evidence in favor of such dynamic portfolio allocation strategies, especially for an optimal currency tilting strategy with carry, value and momentum. Yet, such optimized currency factor allocations do not outperform equal-weighted factor allocations. Such outcome prompted calibrating an integrated strategy that embeds timing carry trade positions. A TED spread-based regime indicator helps navigate the downside of the carry trade, improving the overall risk-adjusted performance of the currency allocation. From a practitioner perspective, this framework is straightforward to implement in real-time. It offers the flexibility to be used with or without conditioning variables, univariate or multivariate, and can help capture diversification benefits in a multi-factor setup.

References

Akram, Q. F., Rime, D. and Sarno, L. (2008). Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics*, **76** (2), 237–253.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, **61** (1), 259–299.

Asness, C. S., Moskowitz, T. J. and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, **68** (3), 929–985.

Barroso, P. and Santa-Clara, P. (2015). Beyond the carry trade: Optimal currency portfolios. *Journal of Financial and Quantitative Analysis*, **50** (5), 1037–1056.

Bartram, S. M., Djuranovik, L. and Garratt, A. (2020). *Currency Anomalies*. Working paper, University of Warwick.

Bekaert, G. and Panayotov, G. (2020). Good carry, bad carry. *Journal of Financial and Quantitative Analysis*, **55** (4), 1063–1094.

Bilson, J. F. (1981). The speculative efficiency hypothesis. *Journal of Business*, **54** (3), 435–451.

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *Journal of Finance*, **61** (5), 2187–2217.

—, — and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, **22** (9), 3411–3447.

Brunnermeier, M. K., Nagel, S. and Pedersen, L. H. (2008). Carry trades and currency crashes. *NBER Macroeconomics Annual*, **23** (1), 313–348.

Burnside, C., Eichenbaum, M. and Rebelo, S. (2011). Carry trade and momentum in currency markets. *Annual Review of Financial Economics*, **3** (1), 511–535.

Caballero, R. J. and Doyle, J. B. (2012). *Carry trade and systemic risk: Why are FX options so cheap?* Tech. rep., National Bureau of Economic Research.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, **52** (1), 57–82.

Castro, V. (2011). Can central banks' monetary policy be described by a linear (augmented) Taylor rule or by a nonlinear rule? *Journal of Financial Stability*, **7** (4), 228–246.

Christiansen, C., Ranaldo, A. and Söderlind, P. (2011). The time-varying systematic risk of carry trade strategies. *Journal of Financial and Quantitative Analysis*, **46** (4), 1107–1125.

Clarida, R., Davis, J. and Pedersen, N. (2009). Currency carry trade regimes: Beyond the Fama regression. *Journal of International Money and Finance*, **28** (8), 1375–1389.

COCHRANE, J. H. (2017). Macro-finance. *Review of Finance*, **21** (3), 945–985.

COLACITO, R., RIDDIOUGH, S. J. and SARNO, L. (2020). Business cycles and currency returns. *Journal of Financial Economics*, **137** (3), 659–678.

CORNELL, W. B. and DIETRICH, J. K. (1978). The efficiency of the market for foreign exchange under floating exchange rates. *Review of Economics and Statistics*, **60**, 111–120.

COTTER, J., EYIAH-DONKOR, E. and POTÌ, V. (2017). Predictability and diversification benefits of investing in commodity and currency futures. *International Review of Financial Analysis*, **50**, 52–66.

ENGEL, C. and WEST, K. D. (2005). Exchange rates and fundamentals. *Journal of Political Economy*, **113** (3), 485–517.

FAMA, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, **14** (3), 319–338.

— and FRENCH, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, **47** (2), 427–465.

— and — (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33** (1), 3–56.

FARHI, E., FRAIBERGER, S. P., GABAIX, X., RANCIERE, R. and VERDELHAN, A. (2009). *Crash risk in currency markets*. Working paper, National Bureau of Economic Research.

FERSON, W. E., SARKISSIAN, S. and SIMIN, T. (2003). Spurious regressions in financial economics? *Journal of Finance*, **58** (4), 1393–1413.

GALATI, G. and MELVIN, M. (2004). *Why has FX trading surged? Explaining the 2004 triennial survey*. Tech. rep., BIS Quarterly Review, December 2004.

GREENAWAY-MCGREVY, R., MARK, N. C., SUL, D. and WU, J.-L. (2018). Identifying exchange rate common factors. *International Economic Review*, **59** (4), 2193–2218.

HAMILTON, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics*, **100** (5), 831–843.

HAMMERSCHMID, R. and LOHRE, H. (2018). Regime shifts and stock return predictability. *International Review of Economics & Finance*, **56**, 138–160.

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.

JEGADEESH, N. and TITMAN, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, **48** (1), 65–91.

JORDÀ, Ò. and TAYLOR, A. M. (2012). The carry trade and fundamentals: Nothing to fear but FEER itself. *Journal of International Economics*, **88** (1), 74–90.

KROENCKE, T. A., SCHINDLER, F. and SCHRIMPF, A. (2013). International diversification benefits with foreign exchange investment styles. *Review of Finance*, **18** (5), 1847–1883.

LUSTIG, H., ROUSSANOV, N. and VERDELHAN, A. (2014). Countercyclical currency risk premia. *Journal of Financial Economics*, **111** (3), 527–553.

MARK, N. C. (2009). Changing monetary policy rules, learning, and real exchange rate dynamics. *Journal of Money, Credit and Banking*, **41** (6), 1047–1070.

MENKHOFF, L., SARNO, L., SCHMELING, M. and SCHRIMPF, A. (2012a). Carry trades and global foreign exchange volatility. *Journal of Finance*, **67** (2), 681–718.

—, —, — and — (2012b). Currency momentum strategies. *Journal of Financial Economics*, **106** (3), 660–684.

MOLODTSOVA, T., NIKOLSKO-RZHEVSKYY, A. and PAPELL, D. H. (2008). Taylor rules with real-time data: A tale of two countries and one exchange rate. *Journal of Monetary Economics*, **55**, S63–S79.

— and PAPELL, D. H. (2009). Out-of-sample exchange rate predictability with taylor rule fundamentals. *Journal of International Economics*, **77** (2), 167–180.

NEELY, C. J., RAPACH, D. E., TU, J. and ZHOU, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, **60** (7), 1772–1791.

OKUNEV, J. and WHITE, D. (2003). Do momentum-based strategies still work in foreign currency markets? *Journal of Financial and Quantitative Analysis*, **38** (2), 425–447.

ROSSI, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, **51** (4), 1063–1119.

TAYLOR, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, **39**, 195–214.

VERDELHAN, A. (2018). The share of systematic variation in bilateral exchange rates. *Journal of Finance*, **73** (1), 375–418.

WANG, J. and WU, J. J. (2012). The Taylor rule and forecast intervals for exchange rates. *Journal of Money, Credit and banking*, **44** (1), 103–144.

WELCH, I. and GOYAL, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21** (4), 1455–1508.

Chapter 3

Controversial News Detection with Controversy-BERT and Stock Price Reaction to Social Controversies

This project is joint work with Margit Steiner, Carsten Rother and my supervisors, Harald Lohre and Sandra Nolte. We thank the following people for their helpful comments: Viorel Roscovan, and all the members of the 2020 Quarterly Research Meeting of Invesco Quantitative Strategies. This work has partly been supported by funding from the Economic and Social Research Council (UK).

3.1 Introduction

Fuelled by an increasing awareness of the role played by companies in the long-term well-being of the society and growing public scrutiny on firms violating sustainability standards, companies across the world have shown increasing commitment to prioritize sustainability concerns to avoid reputational damage or regulatory action. Given the increased media attention on firms with unethical behaviour, media reports on firms' involvement in controversial activities carry the potential to shape investors' perceptions of the future firm prospects. Moreover, academic literature examining the impact of controversies has identified that negative news published on firms following unethical business practices via traditional news outlets or other social media platforms lead to deterioration in the firms' value and market standing (Krüger, 2015; Aouadi and Marsat, 2018; Capelle-Blancard and Petit, 2019; Cui and Docherty, 2020).

In this paper, we study stock price reactions to involvement of firms in controversial behaviour, especially controversies in the S pillar of environmental, social, and governance (ESG). The social pillar of ESG seeks to map a company's impact on society and its stakeholders, including employees, consumers, suppliers, and other parties directly and indirectly impacted by its operations. Hence, the issues in the social dimension generally pertain to surrounding safe and healthy working conditions, diversity, product standards, employment rights, fair wages to employees, labor practices, and human rights.

The S pillar of ESG encompasses a company's responsibility to its stakeholders beyond just financial performance. Companies that prioritize social responsibility and ethical behavior are more likely to attract and retain talent, build trust with consumers and suppliers, and create long-term sustainable value for its stakeholders. Hence, failure to demonstrate commitment to social responsibility by a company can lead to reputational damage, legal risks, and financial distress. However, while the environmental and governance aspects of ESG have been widely discussed and have more standards definitions for measurement, the social aspect has historically been relatively ignored in comparison. Baid and Jayaraman (2022) argue that the relative lack of focus on the S pillar may be due to lack of consensus around the definition, scope and measurement of the S aspect of ESG which has led to incomparable and fragmented reporting of social factors across sectors and geographies. As more investors and asset managers seek to align their investments with social impacts and choose companies that maximize social value creation, there is a stronger need for robust methodology and metrics for measuring the social dimension of ESG.

Consequently, media attention on firms violating social standards has increased and several companies around the world have come into criticism for violating sustainability standards. For instance, the ongoing scrutiny around Foxconn, one of the world's largest technology manufacturer and service provider has been linked to poor working conditions, unsanitary working conditions and labor abuses at its factories in India and China (Fair Labor Association, 2012). Similarly, child labour allegations in cocoa supply chains have been linked to major chocolate manufacturers such as Mars, Nestlé and Hershey, which stresses the need to address social risks and impacts in a company's operations and supply chains, as well as the need for ESG assessments to reflect these social risks and impacts (Perkiss, Bernardi, Dumay and Haslam, 2021).

Early identification of involvement in controversial practices is important as such behaviour has been shown to affect a firm's market standing and financial performance. Using 33,000 news stories on 100 listed companies between 2002 and 2010, Capelle-Blancard and Petit (2019) find that negative ESG events lead to a drop in a firm's market value, whereas no gain is observed from positive events. Cui and Docherty (2020) document over reaction to negative ESG news and report a decrease in institutional holdings of firms following the release of bad ESG news, consequently leading to firms with higher institutional holding experiencing a negative price reaction to negative ESG news.

To this end, identifying and filtering out companies based on their involvement in social controversial practices is key. However, relying solely on the scores and ratings provided by third-party vendors or rating agencies might be inadequate as there have been concerns around lack of transparency in rating methodology, thereby leading to disagreements in ratings provided by major rating agencies (see Serafeim and Yoon, 2022; Dimson, Marsh and Staunton, 2020; Billio, Costola, Hristova, Latino and Pelizzon, 2021). Such disagreements may relay misleading information on the ESG standing of companies, thereby leading to unintended investments in companies that do not match the ESG goals of investors (Avramov, Cheng, Lioui and Tarelli, 2022; Kotsantonis and Serafeim, 2019).

Leading rating providers have been scrutinised for lack of relationship between the ESG rating given to a company and its environmental and social outcomes (Larcker, Pomorski, Tayan and Watts, 2022). For instance, MSCI, one of the leading rating agencies, has been criticised that their rating upgrades are not reflective of the real improvements in sustainable practices made by companies. Instead, some of MSCI's upgrades are simply a result of their own rating methodological changes (Simpson, Rathi and Kishan, 2021). A recent study by Raghunandan and Rajgopal (2022) reports that companies with high ratings from Sustainalytics, another major ESG rating provider, have worse records for compliance with labor and environmental laws than those companies with lower Sustainalytics ratings in the same period. Conflicts of interests have also been reported as firms affiliated or related to ESG rating providers tend to receive higher ratings than those that are not (Tang, Yan and Yao, 2022). Analyzing such widespread evidence, Larcker, Pomorski, Tayan and Watts (2022) warn that ESG ratings can be a "compass without direction" .

Hence, the poor overlap between controversy scores published by major rating agencies and the lack of consistency and timeliness in controversy scores from rating providers stress the need for looking into alternative or supplementary ways of measuring the involvement of companies in controversies. Therefore, using a novel technique, we extract relevant information from published news articles using deep learning and textual analysis to identify social controversies for analyzing the price impact of unethical business practices.

Given the high volume of news articles published every day, manually going through the news headlines on each company in each portfolio can be overly time-consuming and unfeasible. From a practitioner's perspective, automating this process to extract relevant information from news headlines is more efficient as it minimizes the need for manual processing of news articles. Natural language processing (NLP) and state-of-the-art deep learning-based language modelling techniques have facilitated tapping into unstructured textual data such as news articles, annual reports, corporate regulatory filings, news articles, blogs, forums and tweets and these non-standard forms of financial data have been shown to contain opinions and information unavailable in standard financial or accounting data in numerical format (see Capelle-Blancard and Petit, 2017; Bollen, Mao and Zeng, 2011; Curtis, Richardson and Schmardebeck, 2016).

Extracting information content from textual data via NLP and textual analysis is a burgeoning field. Amongst the standard approaches used for textual analysis, user-defined word lists or dictionaries have had considerable success for sentiment and tone analysis of financial documents. By classifying words into predefined categories (such as positive, negative, neutral, etc.), the dictionary approach involves mapping words in the text to one of those categories for inferring the overall sentiment or tone of the document. However, dictionary approaches have been criticized to contain human/user bias as their accuracy is highly dependent on the list of selected words in each category. For instance, words such as 'outstanding, benefit, beneficial, effective, great, greater, honorable, rewards' which were classified as positive when the Loughran & McDonald (LM) dictionary was published in 2011, have been removed later in subsequent revisions as these words no longer frequently appeared in financial documents anymore and due to the changing context of words in financial texts.

Hence, building a domain adapted language model that is trained explicitly on a specific corpus to recognize the contextual meaning of words in the text can yield higher accuracy than static dictionary approaches. For instance, in the news headline, 'Company XYZ launches policy to tackle gender inequality', whilst the dictionary approach would classify this headline as controversial due to the mere presence of the words gender inequality, a language model would look at the surrounding words for deciphering the context of words and hence correctly classify this headline as uncontroversial. Hence, deep-learning-based NLP models for contextual analysis are being increasingly explored to generate actionable insights from large, noisy textual datasets.

In order to identify companies involved in S related controversies such as gender inequality, poor labour standards, discrimination, etc., we turn to a deep-learning-based language model called BERT (Bidirectional Encoder Representations from Transformers) (Devlin, Chang, Lee and Toutanova, 2018) for building a S controversy domain-specific BERT model that is further trained on company news headlines and is fine-tuned for identifying and categorizing controversial news headlines. Since the BERT model has been pre-trained on general domain texts such as Wikipedia and Book Corpus database, the different word representations in other domain corpora may lead to poor performance of generic BERT models.

In line with previous research (Lee, Yoon, Kim, Kim, Kim, So and Kang, 2020; Huang, Altosaar and Ranganath, 2019a; Beltagy, Lo and Cohan, 2019a; Yang, Uy and Huang, 2020; Liu, Huang, Huang, Li and Zhao, 2021b; Webersinke, Kraus, Bingler and Leippold, 2021), we observe significant benefits in terms of accuracy and loss metrics when adapting a BERT model to S controversy domain vs. a general domain BERT model by further training on a large financial company news corpus. In particular, we build an automated system for screening news headlines to flag companies involved in 8 different dimensions of social related incidents or controversies. We choose these 8 dimensions to capture issues surrounding violation of human rights, gender discrimination, poor labour standards, online data safety, poor safety standards, unsafe working conditions and stakeholder related concerns. Starting with a DistilRoBERTa model, which is an efficient version of the BERT model, we further train this model using 1 million news headlines between 2020 and 2022 provided by RavenPack News Analytics. We fine-tune this domain-adapted model in order to help the model identify and classify controversies into one of 8 pre-defined categories.

To accurately gauge the effectiveness of the BERT approach, we use a hand-tagged dataset for computing classification accuracy. The accuracy of classifications in the different social dimensions of ESG increase from 9% to 72% when fine-tuning a DistilRoBERTa model to S domain, thereby showcasing the benefits of context utilisation and domain adaptation. Using the controversial news headlines identified by Controversy BERT, we build an abnormal news activity metric to capture dramatic controversial events. We use this metric and identify about 1,393 controversial events surrounding 1,142 companies between 2014 and 2022. Using an event study approach, we document negative price impact to controversial news events, consistent with our expectation.

We observe leakage of news in the day preceding the event, leading to a drop in returns by 136 basis points and a further drop by 45 basis points on the day of the event. We also observe significant build up in the controversial news volume leading to the event and note that, the cumulative abnormal returns drops by more than 210 basis points in the week following the event, largely driven by small to medium market capitalization companies. For comparison, we repeat the same procedure using controversy events identified as severe by a leading controversy rating provider, Vigeo Eiris. However, we do not find any significant price reaction neither on the day of the event nor in the week following the event, thereby confirming concerns in the academic literature and amongst practitioners on the subjective nature of controversy events and scores provided by external vendors.

Dissecting the results across the eight different controversy categories, we find that the price impact is higher for issues pertaining to violations of safety standards, consumer safety and online data privacy. Looking across geographies, consistent with the literature, we find that the U.S, Europe, Australia and Emerging market regions react very strongly to social controversial behaviour than Japan or United Kingdom (see de Vincentiis, 2022; De Franco, 2020).

Our paper contributes to the literature in several ways: we demonstrate the relevance, steps involved and advantages of adapting a large language model to identify controversies in the social pillar of ESG. Taking advantage of the contextual understanding of a BERT model, we see large improvements in the accuracy of predicted controversy categories, thereby stressing the importance of adapting a language model from general domain to our domain of interest. Using this novel way of identifying controversies from daily news feed, we develop a metric for identifying dramatic controversial events. Through an event study approach, we confirm that daily news headlines can help guide asset managers and investors to identify dramatic controversial events that impact stock prices. Interestingly, the impact is higher for smaller to medium sized firms in our sample and more pronounced in certain geographies such as Europe, Australia, U.S and Emerging Market areas. Moreover, controversies surrounding violation of safety standards and online data safety have a more pronounced negative effect on firm returns.

The paper proceeds as follows. Section 2 gives an overview of language modelling approaches, highlighting the need for transition to context-specific approaches and outlines the steps involved in building our controversy screening tool using BERT followed by the results and discussion. Section 3 details the event study approach to map the price impact of controversial news events and our findings. Section 4 concludes.

3.2 Controversy BERT for Controversy Screening

Drawing actionable insights from unstructured textual data is the core objective of textual data modelling. One of the most common applications of textual analysis is text classification, in which a text classifier or a text classification algorithm labels texts into predefined categories. Broadly, text classification techniques fall into rule-based classification where the text is classified based on predefined rules and machine learning-based (ML) classification. The latter approach involves training an ML model with representative examples from each of the predefined categories to improve the accuracy of the model for text classification tasks. Common ML-based classifiers include Naive Bayes, Support Vector Machines, K-Nearest Neighbors, Random Forest and, more recently, Deep Learning based approaches which aid in semantic understanding of the text. From a practical perspective, being able to perform large-scale text analytics is one of the key considerations which has made ML and deep learning based approaches more attractive for practitioners in the last decade.

3.2.1 Approaches to textual analysis

The following sub-sections cover the most common approaches to text classification starting from the traditional bag-of-words approach and word-lists to the more recent ML- based techniques that have been prevalent in this field.

Dictionaries/word-lists

Approaches to textual analysis vary from visualizing text as a simple collection of words to extracting the meaning conveyed by the specific combination of words in a sentence. One of the early approaches called the bag-of-words approach relies on an over-simplifying assumption of independence of words in a sentence, thereby ignoring the sequence of words in a text (Tetlock, Saar-Tsechansky and Macskassy, 2008). By counting the occurrence of words across the entire document, such an approach builds a term-document matrix summarizing the total counts of each word in the document. However, the high-dimensionality of text and information that is lost when ignoring the meaning and sequence of words has severely limited the practical applications of the bag-of-words approach.

Rather than a simple count of words in a document, lists of words that share the same sentiment can be created (e.g positive, or negative) to categorize documents based on the frequency of words within the pre-defined sentiment categories. Until the advent of finance-specific dictionaries, the Harvard dictionary and associated General Inquirer software[1] were used for measuring the tone and sentiment of financial text (Tetlock, 2007; Tetlock, Saar-Tsechansky and Macskassy, 2008; Kothari, Li and Short, 2009; Hanley and Hoberg, 2010).

However, Loughran and McDonald (2011) argue that general word lists or word lists from other domains would lead to misclassification of financial text and hence domain-specific word lists are necessary to increase accuracy. They show that 74% of words in the negative word list of the Harvard Dictionary are misclassified. For instance, words such as 'liabilities, tax, excess, capital, board and foreign' which are classified as negative by the Harvard dictionary are indeed not considered negative in financial texts, reinforcing the need for a finance language-specific dictionary. By

[1] General Inquirer is a tool used for content analysis of textual data that was developed by members of the Harvard Laboratory of Social Relations in the 1960s (Stone, Dunphy and Smith, 1966)

examining words with a frequency of at least 5% from 10-K filings between 1994 and 2008, six different word lists (negative, positive, uncertainty, litigious, strong modal, and weak modal word categories) were created based on the meaning of words in a business context. Being more comprehensive than Henry (2008)'s word list, the LM dictionary has been widely used to gauge the tone of financial texts (Gurun and Butler, 2012; Garcia, 2013).

However, dictionaries have come under scrutiny for disagreements in word lists and seemingly misclassification of words. Loughran and McDonald (2015) report a low overlap between the Diction and LM word lists. The disagreement between these two word lists is more pronounced for words in the positive word lists (83%) than for those in the negative word lists (70%). Such divergence of dictionaries is a concern as it could adversely affect the inferences drawn from a given text. Hence, for applications where documents need to be classified according to user-defined categories instead of a simple positive/negative sentiment classification, even established dictionaries such as the LM dictionary might become irrelevant and hence lead to poor classification performance. ML approaches which offer the option of having user-defined categories are better solutions for multi-label text classifications where pre-defined dictionaries are not readily available.

Use of ML in Textual Analysis

The increasing availability and volume of digitized text have made it difficult for practitioners to develop an exhaustive list of words when building niche dictionaries. Evidence from several research articles shows that ML approaches can bypass the said challenges of the dictionary approach (see Antweiler and Frank, 2004; Das and Chen, 2007; Li, 2010). The ease of analyzing large volumes of text and the reduced need for user interference have been put forward as important advantages of ML approaches (Loughran and McDonald, 2016). Some of the most common ML techniques used in the finance literature include the Naive Bayes, Support Vector Machine and Artificial Neural Networks.

The Naive Bayes approach is best suited for information retrieval and opinion extraction in textual analysis. This method follows a statistical approach to classify the textual document into the most likely category based on measuring the prior probabilities between words and categories in the training sample. One of the early adopters of this approach are Antweiler and Frank (2004) who use a small sample of

1,000 stock message postings from Yahoo! Finance and Raging Bull stock message boards to train a Naive Bayes classifier to examine 1.5 million messages from the two boards. By analyzing about 45 companies from the Dow Jones Industrial Average and Dow Jones Internet Index, they use a naive Bayes classifier to automatically generate Buy/Hold/Sell signals. Huang, Zang and Zheng (2014) extract opinions from analyst reports and find that the accuracy of a trained naive Bayes classifier is higher than the one obtained by using the financial dictionaries of Loughran and McDonald (2011) and Henry (2008).

The reproducibility of results remains a challenge for such ML approaches since the accuracy of these techniques is highly dependent on the quality of the labelled dataset used for training the model. Supporting this view, Loughran and McDonald (2016) critique that the rules used by researchers to decipher the context of textual data may remain a mystery. Naturally, the quality of the training dataset plays a huge role in determining the predictive accuracy of the such ML models. Moreover, the need for compiling a manually classified dataset deters researchers and practitioners as it can be very time-consuming to build a considerably large training sample with sufficient size and high accuracy.

However, over the last decade, increased adoption of transfer learning approaches in NLP applications has turned the attention towards large-scale pre-trained language models that facilitate context-based textual analysis. One of the most successful pre-trained language models is BERT aka. the Bidirectional Encoder Representations from Transformers (Devlin, Chang, Lee and Toutanova, 2018). It aims to capture the contextual meaning of words in a text. Given that the grammar and order of words in a sentence play important roles in understanding the meaning of the text, the BERT model has an edge relative to rule-based dictionary approaches by learning the collocations of words in the text.

Proposed by Devlin *et al.* (2018), BERT is trained on 3.3 Billion unique words comprising Wikipedia (2.5 bn words) and Google's BooksCorpus (800 mn words), totalling about 16GB of text data. Similarly, one of its successors, ROBERTA (Liu et al., 2019), which is a robustly optimized version of BERT was trained on 160GB of data from various English-language corpora such as Wikipedia, and Google's BookCorpus, OpenWebText and Stories from Common Crawl. Training on such large unlabelled text allows the model to build knowledge by generating a network of bidirectional word representations that help in capturing the context of words.

The success of machine language models for textual analysis has naturally led to the advent of neural networks-based language modelling. In this regard, pre-trained language models have become quite popular and have been applied for a variety of NLP tasks such as text classification, sentiment analysis, question answering, language translation, etc. By training a model on a pre-chosen dataset, such models show a better understanding of the nuances of the dataset than an untrained language model. In the section that follows, we discuss how we use a BERT model for analyzing news headlines to identify and classify controversies related to the S pillar of ESG by further training and fine-tuning on a large news data set.

3.2.2 Building Controversy BERT

Our goal is twofold. First, we seek to examine the value added of domain-specific pretraining and fine-tuning over an LM trained in the general domain. To this end, we put the further-trained and fine-tuned DistilRoBERTa model to test using binary and multi-label classification tasks for news filtering and controversy screening. Second, we evaluate the ability of the trained BERT model, which considers semantic context, to overcome the flaws of a rule-based approach. In particular, since the dictionary approach scans merely for the presence of certain keywords in news headlines, not taking into account the context of words, this often leads to a lot of poor prediction accuracy.

We use a large news dataset from RavenPack News Analytics, a leading news data vendor, for unsupervised further training and supervised fine-tuning of the DistilRoBERTa model. With an average of 5 million news headlines per month, RavenPack offers worldwide coverage of daily real-time company-specific news headlines from leading news providers across the globe. We use RavenPack's database to build a news corpus by filtering out only company/business-related news headlines between 2020 and 2022, resulting in about 1 million untagged news items. We explore the advantages of further training and fine-tuning for controversy classification tasks by splitting this corpus into an 80% training set and a 20% hold-out set.

To facilitate our understanding of the workings of different LMs, we compare BERT and general domain training vs domain-specific training. Specifically, for the binary text classification exercise, we test the model's ability to recognize whether a given news headline indicates a negative news event, or generally news that can be expected to impact a stock negatively. Preparing multi-label text classification, we rely on broad definitions of the social dimension of ESG to pick 8 well-known categories of corporate controversies and test the model's ability to classify any given controversial news headline into one of these pre-identified categories.

The 8 categories based on social standards are Fundamental Human Rights, Discrimination, Modern Slavery, Labour Standards, Consumer Data Safety and Privacy, Product Safety Standards, Workplace Health/Safety Standards and Stakeholder Opposition. For better understanding of each of the categories, Table 3.2.1 lists out a few example headlines from each of the chosen categories. We also present more examples from different controversy categories towards the end of this section to highlight the advantages of context-aware language models like BERT over context-ignorant dictionary approach.

Table 3.2.1 Examples from hand-tagged dataset- Examples from each category chosen for creating a hand-tagged dataset for the multi-label text classification exercise.

Category	Examples
Human Rights	*'Facebook Is Working Hard To Stifle Your Free Speech'*
	'Lyft has yet to disclose sexual assault incidents as cases grow'
Discrimination	*'McDonald's sued by 52 black ex-franchisees for racial discrimination'*
	'Blizzard president J. Allen Brack steps down amid discrimination lawsuit'
Modern Slavery	*'US senator chides Apple, Nike for using forced labour'*
	'7 Apple suppliers in China linked to forced labour programmes: Report Read More Share'
Labour Standards	*'UPDATE 1-Next, Zalando and Amazon drop Boohoo over worker rights allegations'*
	'Amazon Keeps Getting Sued for Paying Drivers Less Than Minimum Wage'
Consumer Data Safety and Privacy	*'Twitter hack prompts security concerns'*
	'Colonial Pipeline shuts down after cyber attack'
Product Safety Standards	*'Greggs extends food recall as vegetables bakes may contain glass - more batches affected*
	'Australia Drags Mercedes-Benz To The Federal Court For Downplaying Risks Caused By Defective Airbags'
Workplace Health and Safety Standards	*'McDonald's workers in LA, San Jose allege unsafe work conditions amid COVID-19 pandemic'*
	'Troy Resources security worker killed in workplace accident'
Stakeholder Opposition	*'Capital One to Pay \$390m for Money Laundering Failures'*
	'France fines U.S. bank JP Morgan \$29.6 million in tax fraud settlement'

Using news headline corpus from RavenPack and focusing only on news items reported in English, we further-train a DistilRoBERTa model using 1 million untagged news headlines and later for fine-tuning the model for the two text classification tasks. Our hand-tagged dataset for both binary and multi-label text classification were prepared by scraping 10,000 news headlines from the entire sample period and manually classifying them into controversy categories. We ensure that the hand-tagged dataset is fairly balanced across all the chosen categories in order to reduce misclassification errors.

Further Pre-training

Putting language models to practice broadly involves two stages: unsupervised pre-training on a large-scale corpus and supervised fine-tuning of the model on a labelled dataset for task adaption as shown in Figure 3.2.1. However, the corpus used for pre-training is usually from a general domain whereas the fine-tuning task is done by taking text from a specific domain. Hence, to bridge the gap between these two stages, domain-adaptive pre-training (DAPT) and task-adaptive pre-training (TAPT) can help to improve the performance of such models in domain-specific tasks (Gururangan, Marasović, Swayamdipta, Lo, Beltagy, Downey and Smith, 2020). Introducing a further pre-training stage helps the language model learn word representations and patterns in that specific domain or for a specific task, which in turn improves the overall accuracy of the predicted categories and performance of the model. For instance, Table 3.2.2 lists some of the domain adapted BERT models that are further trained on other domains and are more suited for domain-specific tasks. Our dataset is distinct from the one used for pre-training BERT as it comprises several named entities as we focus on company-related news headlines. Most of these names are also concentrated on the beginning of the headlines, such as "Company XYZ reports earnings for the thrid quarter", and usually comprises of specific types of financial news events such as earnings announcements, profit, loss, upgrade, downgrade, etc.

In general, classification accuracy and the success of predictions are evaluated using metrics such as precision, recall and F1 score. Precision is the proportion of true positives to all predicted positives and gives the proportion of positives that were actually correct, implying that a precision of 1.0 can be achieved if there are no false positives. Recall measures the proportion of true positives to actual positives and hence gives the proportion of actual positives that were identified correctly, implying

Fig. 3.2.1 Steps involved in building a custom BERT model using an example headline and the predicted probabilities across output class categories for the example headline.

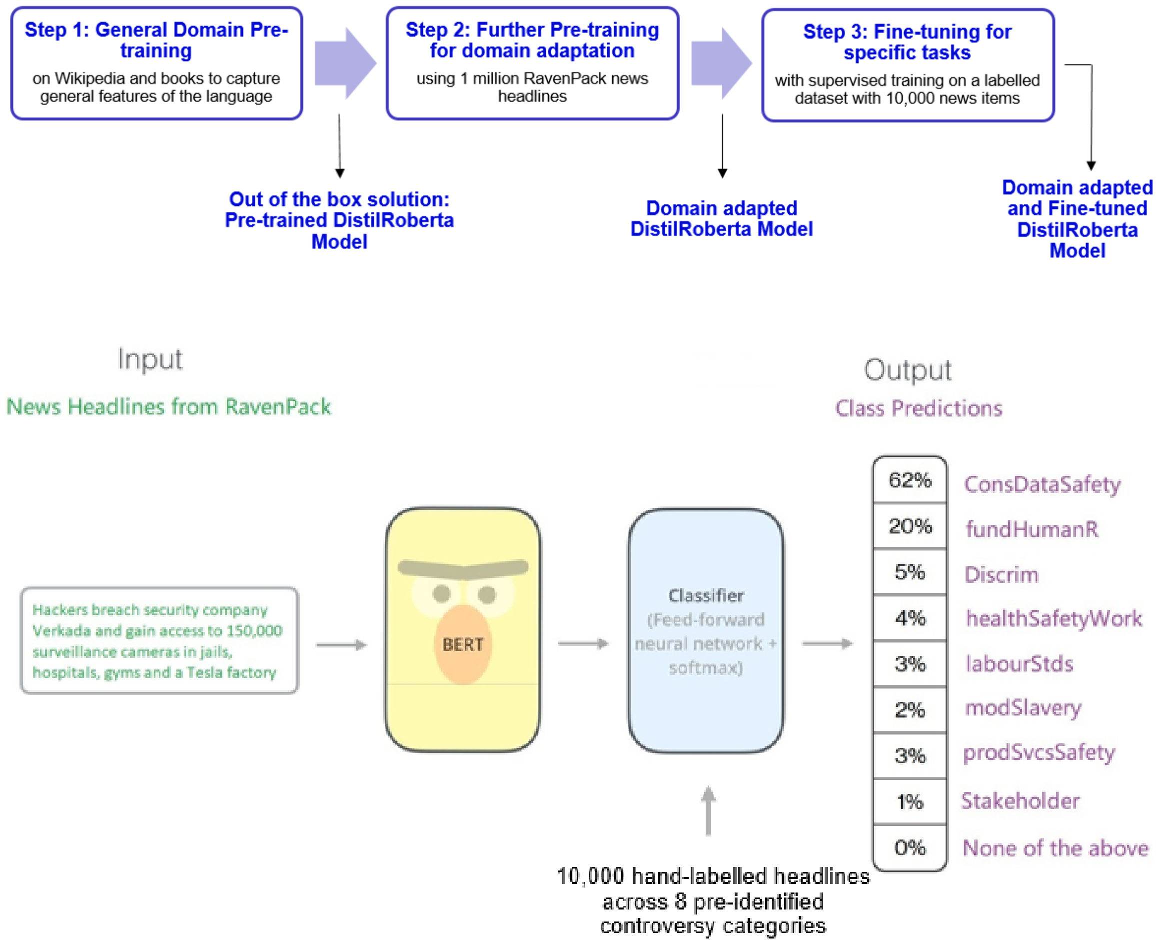


Table 3.2.2 Selected examples of domain-adapted versions of BERT.

Domain-adapted BERT model	Author	Dataset used for domain adaptation
BioBERT	Lee, Yoon, Kim, Kim, Kim, So and Kang (2020)	Biomedical domain corpora - PubMed abstracts and PMC full-text articles
ClinicalBERT	Huang, Altosaar and Ranganath (2019a,b)	Clinical notes and electronic health records
MedBERT	Rasmy, Xiang, Xie, Tao and Zhi (2021)	Patient's visit sequence
MedBERT	Liu, Hu, Xu, Xu and Chen (2021)	Public dataset about Chinese medical information and clinic medical records
FinBERT	Araci (2019)	TRC2-financial, Financial PhraseBank and FiQA Sentiment
FinBERT	Yang, Christopher Siy Uy, and Huang (2020)	10-K, 10-Q reports, Earnings calls transcripts, Analyst reports
TweetBERT	Qudar and Mago (2020)	English tweets
ClimateBERT	Webersinke, Kraus, Bingler and Leippold (2021)	News articles, research abstracts and corporate climate reports

that a recall of 1.0 can be achieved if there are no false negatives. Whereas the F1 score attempts to achieve a balance between precision and recall and is defined as the harmonic mean of precision and recall. Hence, we use F1 scores as a measure of accuracy in order to focus more on the false positives and false negatives rather than true positives and true negatives.

Panel A of Table 3.2.3 reports the results of such a comparison for a binary text classification task that classifies any given news headline into one of the two categories "negative", or, "not negative". Note that the general domain model achieves almost a 44% accuracy, slightly worse than a coin-flip. Hence, the general domain model seems to randomly assign the news headlines into the two categories as it has not been exposed to any representative examples or patterns from any of the two categories. However, when performing domain-adaptive further training, the accuracy of the model increases by about 11 percentage points. Hence, the language model is able to learn the nuances of news headlines, which translates into better accuracy.

Table 3.2.3 Classification Accuracy- Comparison of accuracy from further-training and fine-tuning on binary and multi-label text classification tasks. Panel A reports the results for a binary text classification task that uses two categories negative, or, not negative and panel B for the multi-label classification exercise. The reported weighted-averaged F1 score is the average of the F1 score of each controversy category with weighting depending on the actual occurrences of each controversy category in the test set.

Model	Weighted F1 score
Panel A: Binary text classification	
DistilRoBERTa (Non Further-trained, Non-Fine-tuned)	44%
DistilRoBERTa (Further-trained, Non-Fine-tuned)	55%
DistilRoBERTa (Non Further-trained, Fine-tuned)	82%
Controversy-BERT - DistilRoBERTa (Further-trained, Fine-tuned)	85%
Panel B: Multi-label text classification	
DistilRoBERTa (Non Further-trained, Non-Fine-tuned)	4%
DistilRoBERTa (Further-trained, Non-Fine-tuned)	9%
DistilRoBERTa (Non Further-trained, Fine-tuned)	37%
Controversy-BERT - DistilRoBERTa (Further-trained, Fine-tuned)	72%

Fine-tuning

Whilst out-of-the-box BERT or domain-adapted BERT may be sufficient for some use cases, fine-tuned versions are more suited for task-specific usages. Fine-tuning NLP models for text classification first involves defining the number of categories for the text to be classified into. A multi-label text classification task involves multiple categories and later identifying the most suitable category for the text or sentence. We observe in panel A of Table 3.2.3 that a general domain DistilRoBERTa model fine-tuned on a hand-labelled dataset of 10,000 news headlines has an accuracy of 82%. Introducing a further-training stage before fine-tuning slightly increases the accuracy by 3 percentage points, however still much higher than previous versions. We observe similar findings in panel B for the multi-label classification exercise, strengthening the case for domain adaption and fine-tuning.

Table 3.2.4 gives the controversy category-wise classification from the further-trained, fine-tuned DistilRoBERTa model. The Consumer Data Safety and Privacy category has the highest precision whereas the workplace health/safety standards ranks the lowest. Since all the phishing attacks, scams, data breaches and online privacy based issues are relatively distinct from other controversy categories, the accuracy metrics in this category are relatively high. Whereas in categories such as Modern Slavery, Labour Standards and Fundamental Human Rights, there are several overlapping issues which might lead to easy mis-classification by the Controversy BERT and hence the relatively low accuracy numbers. The category Stakeholders which covers issues such as fraud and money laundering amongst other issues, is most diverse as it covers a wider umbrella of issues compared to the other categories and hence the low accuracy numbers are not surprising.

To foster intuition how the Controversy-BERT does improve upon the dictionary approach, Table 3.2.5 highlights a few examples where taking into account the contextual meaning of words rather than scanning for the mere presence of words in headlines makes a crucial difference: For instance, in the headline, *'FINDING THE BALANCE : 4 ways in which Imerys is tackling gender inequality in the mining industry'*, the mere presence of the words 'gender inequality' is sufficient for the dictionary approach to classify this headline as controversial whereas in reality, the headline only suggests a positive measure made by a company to tackle gender discrimination. Since such false positives are highly undesirable when building an automated screening system, the trained NLP classifier emerges as a better choice over rule-based approaches.

Table 3.2.4 Category-wise Classification Accuracy- Controversy category-wise classification from Controversy-BERT in the multi-label text classification task. We report the category-wise precision, recall and per category F1-Score for comparison across categories.

	Precision	Recall	F1-score
Consumer Data Safety and Privacy	83%	80%	81%
Discrimination	60%	87%	71%
Fundamental Human Rights	58%	52%	55%
Workplace Health/Safety Standards	42%	48%	45%
Labour Standards	63%	63%	63%
Modern Slavery	49%	78%	60%
Product Safety Standards	71%	71%	71%
Stakeholder	46%	81%	58%

Moreover, the limited list of keywords for the dictionary approach renders this approach more error-prone due to several omitted keywords than the Controversy-BERT model. Combining these findings with the results in Table 3.2.3 leads us to conclude that such state-of-the-art LM's can be effectively customized for developing a screening system for identifying the involvement of companies in social controversies. The next section depicts how we use these predicted labels to develop an abnormal controversial activity indicator and map the price impact of controversial news events.

Table 3.2.5 Dictionary vs Controversy-BERT- Examples of false-positives and mis-classifications from the dictionary approach that are correctly classified by Controversy-BERT in the multi-label text classification exercise.

Dictionary Classification	Headline
Consumer Data Safety and Privacy	*'The Transparency Company Launches to Identify and curb Fake Reviews on Google Maps'*
Consumer Data Safety and Privacy	*'CHARLES SCHWAB : Reinforces Its Commitment to Customer Data Protection'*
Consumer Data Safety and Privacy	*'TikTok proposes a global social media coalition to curb harmful content'*
Discrimination	*'FINDING THE BALANCE : 4 ways in which Imerys is tackling gender inequality in the mining industry'*
Discrimination	*'Surveillance cameras at Tesla, many others breached: Report'*
Workplace Health/Safety Standards	*'Walmart announces new health and safety measures for U.S. employees, including temperature checks'*
Product Safety Standards	*'Ford workers express health concerns as COVID-19 pandemic spreads'*
Modern Slavery	*'Coal India slapped Rs 43.25 cr fine for illegal mining in Assam forest'*
Fundamental Human Rights	*'Biometric data privacy battles continue with Shutterfly settlement, university sued'*
Fundamental Human Rights	*'AP Interview: French government to tackle child abuse issue'*

3.3 Stock Price Reaction to Controversial News

We document the price impact of controversial behaviour of firms and rely on daily news headlines for identification of companies involved in socially irresponsible behaviour. We use daily news headlines from RavenPack News Analytics between January 2014 and October 2022 and identify controversial news items as classified by Controversy BERT to measure the volume of controversial news for each day in our sample. Further more, RavenPack's data fields include a metric called relevance, which is an integer score between 0 and 100, which measures how relevant an entity is to the news item, with higher values indicating greater relevance. Since we are interested in mapping the association of companies in controversial behavior, we only pick news items with a relevance scores of more than 90, as suggested by RavenPack, to ensure that the entities and events identified in the news story are prominent and significantly relevant to the underlying news story.

Table 3.3.1 Descriptives- Summary statistics of monthly coverage and news volume of our news dataset spanning sourced from RavenPack from January 2014 to October 2022. Statistics reported on controversial news volume is based on the news items identified as controversial by ControversyBERT.

Number of companies	9,360	
Monthly Averages	**All news**	**News on Controversies**
Number of companies	7,371	2,917
Coverage	79%	31%
Volume	783,589	118,993
News Per company	106	41
Minimum	431,617	42,041
Maximum	1609157	392,902
Standard Deviation	289,313	58,377

As seen from Table 3.3.1, our sample has a monthly average of 9,360 companies, with roughly an average of 7,370 companies having at least one news item published on them every month and about 2,900 companies having at least one controversial news item published on them every month. Whilst this gives us a monthly news coverage of about 78%, on average, there are about 106 news items published per company per month. Whereas out of the 2,917 unique companies with controversial news, these have on average 41 controversial news headlines per company per month.

Breaking this result by size deciles in Table 3.3.2, the average number of monthly headlines grows with size as expected. As the company capitalization increases, the average headline count per company month also increases. This implies that larger companies have more media coverage and hence more news generated on them and smaller companies have much less media attention and hence much lower number of news items published on them. Whereas in terms of controversial news volume, the average number of controversial headlines per company per month remains almost steady, except for the higher mean volume in the top decile. This is in line with expectations as larger companies generate more media attention and hence more news volume than smaller firms when involved in controversial behaviour.The variation in the fraction of controversial news with respect to all news underscores the dynamic nature of corporate behavior and the diverse public perception of different firms. Some companies consistently attract a higher proportion of controversial news coverage, potentially signaling a pattern of controversial actions or heightened scrutiny. In contrast, others maintain a lower level of controversy in their news reports, suggesting a relatively stable or less controversial public image. The distinctive dynamics of news coverage at the individual company level offer valuable insights for tailoring assessments to each company's unique circumstances. Simultaneously, these dynamics deepen our understanding of how market reactions and sentiment fluctuations are influenced by evolving news trends.

In terms of overall news volume in our sample period, Figure 3.3.1 depicts the steady growth in overall news volume and controversial news volume over time. The fraction of controversial news varies This reaffirms the increase in media coverage and attention that firms have received in the last decade. Figure 3.3.1 also depicts significant spikes in controversial news generated post 2020. With the COVID-19 pandemic drawing more attention towards worker safety and labour standards, increased awareness and concerns from public on data privacy, diversity and human rights violations has led to increased media attention and reporting on such violations.

In terms of distribution of controversial news volume across different categories, Table 3.3.3 shows the distribution of news volume across different controversy categories obtained by aggregating the predicted labels from Controversy BERT. The Product Safety Standards category comprises of about 36% of controversial news items followed by Workplace Health/Safety at 19% whereas the Modern slavery category seems to have the lowest number of controversial news items. Given the marked increase in overall news volume since 2020 in Figure 3.3.1, it is unsurprising that news surrounding

Table 3.3.2 News Volume- Summary statistics for news volume-headline count per company month and average company capitalization measured in millions of U.S Dollars by size decile. The top panel gives the size wise distribution for all news items and the bottom panel gives the distribution for only controversial news. First decile comprises of the companies with the lowest market cap in our sample and the tenth decile comprises the largest ones.

Size Decile	1st Qu.	3rd Qu.	Max.	Mean	Median	Min.	Company Cap. (Mn. of U.S$)
All News							
1	6	44	2,212	35	20	1	235
2	4	39	5,198	32	15	1	431
3	4	41	6,021	33	15	1	634
4	4	44	13,857	36	16	1	909
5	5	50	9,431	41	19	1	1,314
6	6	59	3,540	47	23	1	1,944
7	8	68	12,470	55	28	1	2,998
8	9	81	20,265	69	33	1	4,959
9	14	114	8,720	104	48	1	9,761
10	47	336	71,181	461	137	1	53,528
Only Controversial News							
1	1	7	1,785	11	2	1	229
2	1	6	1,522	9	2	1	432
3	1	6	1,971	9	2	1	634
4	1	6	1,915	10	2	1	915
5	1	7	8,662	11	2	1	1,318
6	1	8	1,952	11	3	1	1,958
7	1	8	10,297	13	3	1	3,020
8	1	10	7,313	15	3	1	4,994
9	2	13	7,523	21	4	1	9,946
10	4	48	57,074	120	13	1	60,701

health and workplace safety have dominated the media as there has been an increased focus on general and Covid-related safety protocols across the globe.

Fig. 3.3.1 Growth in overall news volume and controversial news volume over time between January 2014 and October 2022.

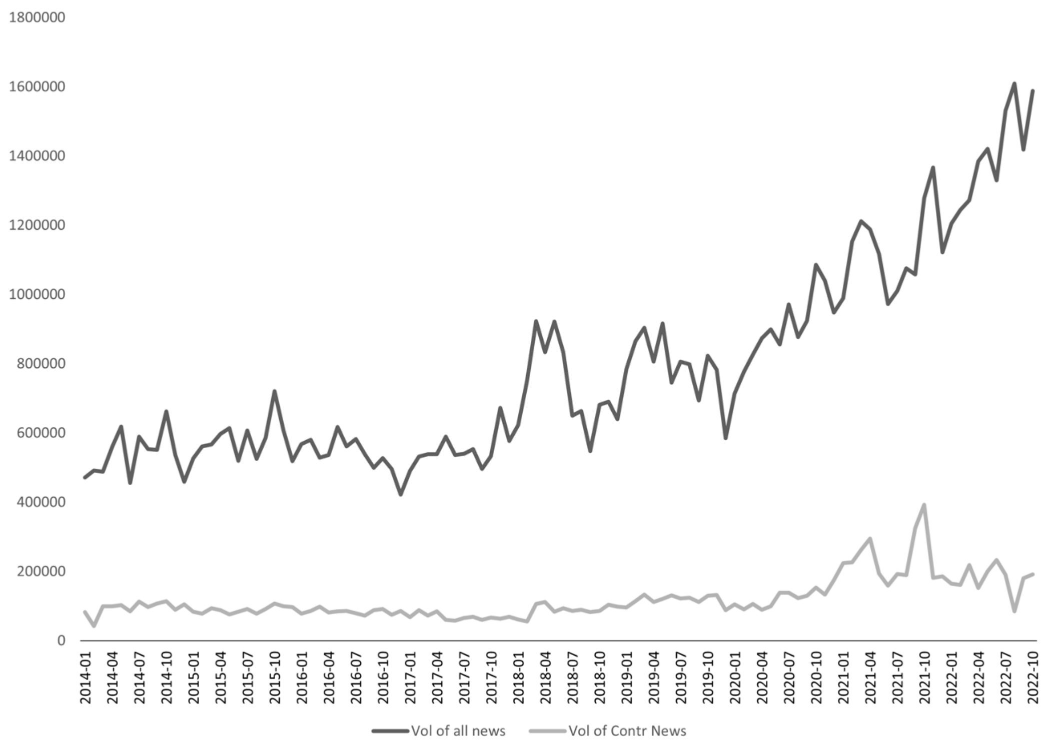


Table 3.3.3 Category-wise controversial news distribution- Average news volume per month between January 2014 and October 2022 across the eight pre-identified controversy categories as classified by Controversy BERT.

Controversy Category	**News Volume**
Product Safety Standards	36%
Workplace Health/Safety Standards	19%
Consumer Data Safety and Privacy	19%
Stakeholder	13%
Fundamental Human Rights	5%
Labour Standards	5%
Discrimination	3%
Modern Slavery	1%

Similarly, Table 3.3.4 gives the category wise distribution of controversies in our sample across size deciles and sectors (based on the Global Industry Classification

Standard (GICS)). The growing regulations on product safety standards and awareness among consumers on the quality, manufacturing standards and composition of products are well reflected in the upper panel of Table 3.3.4, where the Product Safety Standards category dominates in terms of controversial news volume across most of the sectors and sizes. Moreover, given the increasing concerns around online data scams and privacy issues, it is unsurprising that the Telecommunication and Information Technology sectors have higher proportion of controversial news on data privacy/safety related concerns. Sectors such as Energy and Financials have higher proportion of Stakeholder-related controversies as expected as these sectors play a critical role in global economy, impacting the lives of wide range of individuals, businesses and even governments. The Health care and large industry sectors have faced increasing criticism around following safety protocols and meeting safety standards, which is reflected by the higher proportion of controversial news in the Product Safety Standards category in these sectors. Given that the utilities companies are subject to a variety of national and regional regulations that govern product safety, it is not surprising to see that there has been more violations in Product Safety Standards category, which is expected to recieve more media attention than any other category of controversy. The overall results across size deciles in the lower panel of Table 3.3.4 is consistent with the distribution in Table 3.3.3, where the Product Safety Standards category has the most number of controversial news items.

Table 3.3.4 Sector/Size wise controversy distribution- Sector wise and size decile wise distribution of controversial news volume amongst the eight pre-identified categories as classified by Controversy BERT.

Sector	Product Safety Standards	Consumer Data Safety/Privacy	Human Rights	Stakeholder	Discrimination	Workplace Health/ Safety Standards	Labour Standards	Modern Slavery
Consumer Discretionary	48.40%	14.60%	7.00%	10.10%	4.20%	7.40%	7.40%	0.90%
Consumer Staples	50.70%	12.20%	5.50%	9.50%	3.20%	10.10%	6.90%	1.80%
Energy	28.60%	11.40%	2.60%	34.30%	0.30%	15.10%	6.90%	0.80%
Financials	18.70%	23.90%	3.00%	38.80%	1.30%	6.50%	7.30%	0.50%
Health Care	73.50%	5.40%	1.50%	5.80%	0.30%	12.50%	0.60%	0.20%
Industrials	34.00%	13.70%	3.40%	10.50%	1.70%	22.90%	13.00%	0.80%
Information Technology	19.20%	52.50%	8.10%	6.70%	3.90%	5.20%	3.70%	0.70%
Materials	37.20%	5.50%	2.70%	22.80%	0.40%	20.00%	10.10%	1.40%
Telecomm- -unications	14.80%	45.50%	17.70%	6.70%	7.90%	3.40%	3.10%	0.80%
Utilities	42.00%	10.30%	1.40%	16.80%	0.40%	25.30%	3.40%	0.30%
Size Decile								
1	58.90%	13.50%	1.60%	12.40%	0.60%	10.80%	2.00%	0.30%
2	50.40%	15.80%	2.20%	14.20%	0.80%	12.80%	3.20%	0.60%
3	46.20%	18.20%	2.60%	13.50%	0.80%	14.50%	3.30%	0.90%
4	47.70%	15.70%	3.00%	12.50%	0.90%	16.20%	3.20%	0.80%
5	47.20%	18.50%	3.10%	12.10%	1.30%	13.60%	3.60%	0.70%
6	39.70%	18.70%	4.40%	14.00%	1.50%	15.80%	5.10%	0.80%
7	38.80%	20.30%	4.20%	13.10%	1.30%	15.80%	5.40%	1.00%
8	36.20%	22.60%	4.30%	12.80%	1.50%	15.10%	6.60%	0.80%
9	33.70%	23.40%	6.30%	12.30%	2.90%	13.20%	7.40%	0.80%
10	30.30%	31.80%	8.90%	10.80%	4.40%	7.30%	5.60%	0.80%

Event Study Design

To design our event study, defining what qualifies as an event is the key first step, and to this end, we start by looking at the trend in the controversial news volume of each company in our sample. Since the largest companies in our sample have an average of 3 controversial news items per day, we standardise the controversial news volume with a 1 year look back window to identify dramatic events. News outlets serve as the primary means of disseminating information to the public, and they are instrumental in uncovering and reporting on corporate activities, thus fostering transparency and accountability. News reports offer timely and comprehensive insights into the events surrounding a company's actions, making them a valuable tool for assessing public perception of the company's behavior. Consequently, monitoring news headlines and identifying substantial increases in the volume of news reporting on a company's controversial behavior can serve as a reliable indicator of significant events. Therefore, we define our 'events' as those falling within the top percentile of our daily abnormal controversial news activity metric, as they signify noteworthy controversial occurrences. Therefore, we use the controversies classified by Controversy BERT to define a metric that measures abnormal controversial news activity as follows,

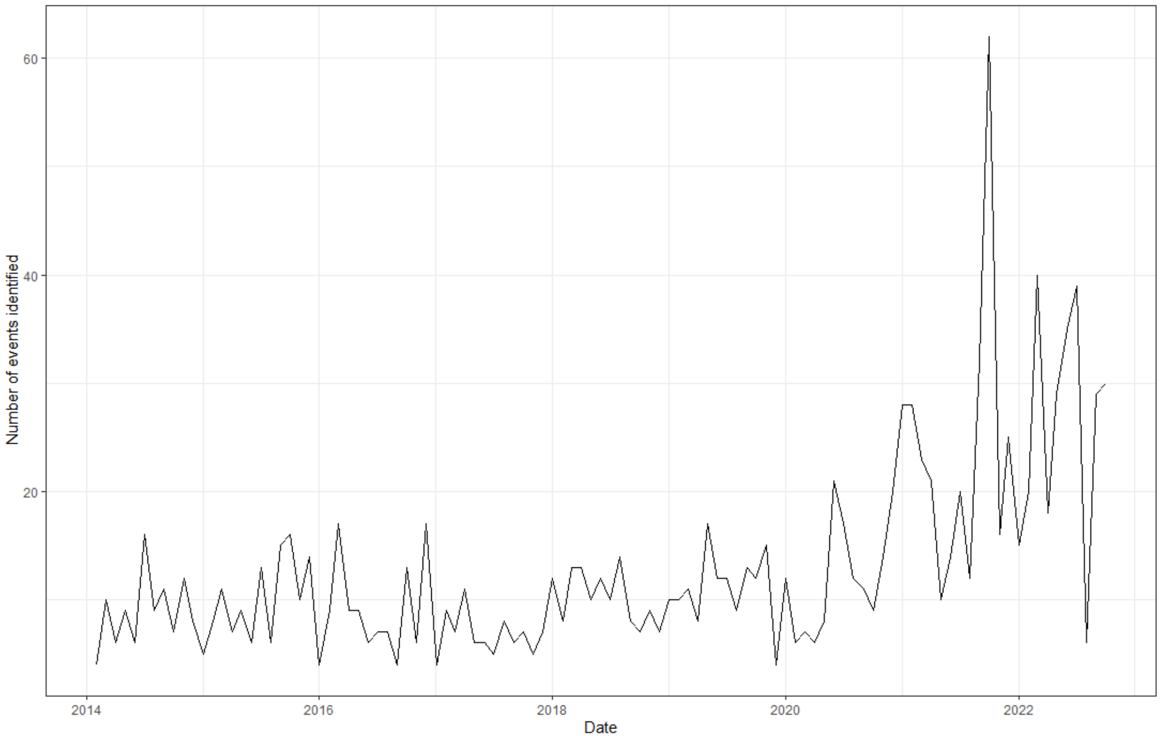
$$A_{it} = \frac{C_{it} - \bar{C}_{i(t-1,t-365)}}{s.d(C_{i(t-1,t-365)})} \tag{3.1}$$

where C_{it} is the total number of controversies identified by Controversy BERT for company i on day t, $\bar{C_{it}}$ is the mean number of controversies over the last one year for company i and $s.d(C_{it})$ is the associated standard deviation over the last one year for company i. Although we use a look-back window of 365 calendar days when computing our abnormal news metric, we only use 252 trading days in our estimation window when estimating expected returns. This means that events are identified based on historical information available up to the day just before the event, ensuring that the analysis is grounded in past data and is not influenced by future developments.

To avoid discrepancy between calendar days and trading days, we account for any controversy breaking on the weekends to the next trading day. Moreover, to capture the impact of significant controversial events, we avoid the inclusion of companies with just one or two controversial news headlines by filtering for events falling in the top quartile of news volume, with at least half of news identified as controversies.

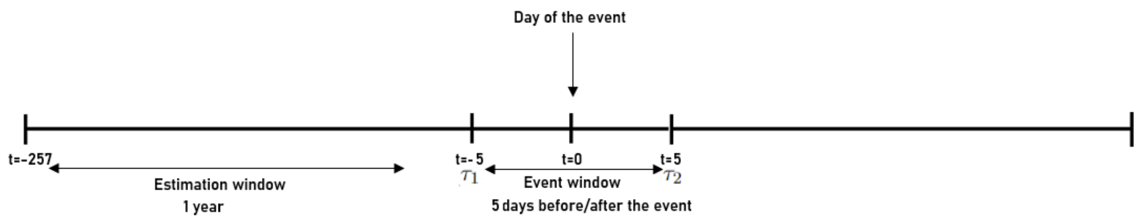
To ensure sufficient data coverage on trading days and to avoid the impact of confounding events, we exclude those companies with coverage of less than 252 trading days and those that have multiple controversial events detected by our abnormal controversy news metric in the last 252 trading days due to the length of our estimation window. Using the threshold above, we identify 1,393 events between January 2014 and October 2022 and Figure 3.3.2 shows the number of events identified across time, which is consistent with the trend observed in Figure 3.3.1.

Fig. 3.3.2 Controversial Events- Time series of events identified using our abnormal controversial news metric between January 2014 and October 2022.



The next key step in an event study analysis is to define the estimation window, event window and abnormal returns. As shown in Figure 3.3.3, we use an estimation window of 252 trading days and observe the price impact in the period building up to the event and the days that follow. Consequently we define the start of our event window as 5 days prior to the event and 5 days after the event to map the short-term and medium-term impact of abnormal controversial behaviour.

Fig. 3.3.3 Event Study Design- Our event study setup using an event window of ±5 days and an estimation window of 252 trading days prior to the start of the event window.



Abnormal returns are measured as the difference between the observed return of company i on day t and the expected return in the absence of the event for company i on day t. Since the expected returns (in the absence of any event) are unobserved, we use the market model to estimate expected returns over our estimation window for each company i in our sample using this model as follows,

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it} \tag{3.2}$$

where R_{it} the daily return of company i on day t of the estimation window, α_i the intercept of the company i, β_i is the systematic risk (or beta) of company i, which measures the sensitivity of company i's return to the market return; R_{mt} is the daily market return on day t of the estimation window; and ϵ_i is the error term for company i on day t of the estimation window.

Using the estimated $\hat{\alpha_i}$ and $\hat{\beta_i}$ from the regression above, we estimate expected returns and abnormal returns for day τ in the event window as follows:

$$\begin{aligned} \widehat{ER_{i\tau}} &= \hat{\alpha_i} + \hat{\beta_i} R_{m\tau} \\ AR_{i\tau} &= R_{i\tau} - \widehat{ER_{i\tau}} \end{aligned} \tag{3.3}$$

where $R_{i\tau}$ the observed daily return of company i. The cumulative abnormal returns are given by,

$$CAR_i[\tau_1, \tau_2] = \sum_{t=\tau_1}^{\tau_2} AR_{i\tau} \tag{3.4}$$

and the cumulative average abnormal returns (CAAR) on each day, τ, of the event window as shown below,

$$CAAR_\tau = \frac{\sum_{t=\tau} CAR_{i\tau}}{N} \tag{3.5}$$

where N is the total number of events identified in the event window, τ_1 is the start of the event window and τ_2 is the end of the event window $[\tau_1, \tau_2]$.

Previous research adopting event study methodology have expressed concerns on event induced variance inflation as the variance of stock returns tends to increase around the day of the event. Hence, using data in estimation window data to estimate the variance of the CAAR in event window would lead to over rejection of null hypothesis (Patell, 1976). Hence, when testing for the significance of CAAR, we compute cross-sectional standard deviation and associated t-statistic within the event window.

Another common issue surrounding event studies is the cross-sectional correlation in abnormal returns when the event day is the same for multiple firms in the sample. This event-date clustering often leads to a downward bias in the standard deviation and thereby overstating the t-statistic. However, we do not find any overlap of event dates among the firms in our sample, thereby avoiding the issues arising from event-date clustering.

Hence, to test the significance of cumulative abnormal returns, we compute the cross-sectional t-statistics as follows,

$$t = \sqrt{N} * \frac{CAAR_\tau}{S.D(CAR_{i\tau})} \tag{3.6}$$

Since our goal is to test price reaction to controversial news outbreak, we test the following hypothesis:

$$H_0 : CAAR_\tau = 0$$
$$H_1 : CAAR_\tau \neq 0$$

Table 3.3.5 and Figure 3.3.4 summarize the results from our event study. Alongside our abnormal news activity metric, we also use controversy events and associated dates identified by an external vendor, Vigeo Eiris, for comparison. Vigeo Eiris offers a database of controversies for companies with a granular breakdown of their severity, frequency and the responsiveness of companies to them. For side-by-side comparison, we use controversies classified as "severe" and related to the social dimension of ESG in the same sample period from January 2014 to October 2022 from the Vigeo Eiris controversy database.

The upper panel of Table 3.3.5 reports the results based on the events identified using our abnormal news activity metric and the lower panel reports the results using controversy event dates identified by Vigeo Eiris. In the upper panel, day 0 is the day

Table 3.3.5 Event Study Results- Mean cumulative abnormal returns and the associated t-statistics on each day of our event window. We use an estimation period of 252 trading days and event window of (-5,5). The upper panel reports the results based on the events identified using our abnormal news activity metric and the lower panel reports the results using event dates identified by Vigeo Eiris. We report the average abnormal returns (AAR), cumulative average abnormal returns (CAAR), the number of controversial events, the proportion of controversial news, the t-statistics of AAR and t-statistics of CAAR. *** denotes significance at 1% level, ** denotes significance at 5% level and * denotes significance at 10% level.

Event Window	**AAR (%)**	**CAAR (%)**	**No. of Events**	**Prop. of Contr.**	**t-stat$_{AAR}$**	**t-stat$_{CAAR}$**
-5	-0.02	-0.02	1,393	0.06	-0.26	-0.26
-4	-0.07	-0.10	1,393	0.06	-1.01	-0.86
-3	-0.06	-0.15	1,393	0.07	-0.88	-1.14
-2	0.15	0.00	1,393	0.09	0.85	-0.05
-1	-1.36***	-1.36***	1,393	0.18	-6.93	-4.55
0	-0.45***	-1.81***	1,393	0.81	-3.12	-5.11
1	-0.12	-1.93***	1,391	0.42	-1.45	-5.11
2	-0.12	-2.05***	1,388	0.26	-1.45	-5.50
3	-0.07	-2.12***	1,385	0.21	-0.88	-5.63
4	-0.02	-2.14**	1,381	0.19	-0.22	-5.64
5	0.03	-2.11***	1,375	0.19	0.39	-5.52
Using Event Dates identified by Vigeo Eiris						
-5	0.10***	0.10***	4,788	0.12	2.98	2.98
-4	0.02	0.12**	4,788	0.12	0.60	2.53
-3	0.01	0.13**	4,788	0.12	0.27	2.30
-2	-0.05	0.08	4,788	0.12	-1.61	1.23
-1	-0.03	0.05	4,788	0.12	-0.74	0.76
0	0.00	0.05	4,788	0.12	0.09	0.74
1	-0.05*	0.00	4,749	0.12	-1.68	-0.08
2	0.01	0.01	4,749	0.11	0.25	0.01
3	0.08**	0.09	4,749	0.11	2.50	0.85
4	0.02	0.11	4,749	0.10	0.61	0.99
5	0.03	0.14	4,749	0.10	0.90	1.21

identified as an event by our abnormal controversial news activity indicator and in the lower panel, day 0 is the day identified as event date by Vigeo Eiris. Since we intend to study the evolution of prices both before the event and in the days following the event, we map the price impact from 5 days before the event to 5 days after the event in both the panels.

Looking at the upper panel of Table 3.3.5 that uses our news based metric, the downward trend in average abnormal returns (AAR) and cumulative average abnormal returns (CAAR) since the beginning of the event window indicates some leakages of news/event before the actual event. Whereas using the event dates from Vigeo Eiris leads to statistically non-significant results as seen from the lower panel of Table 3.3.5, thereby confirming concerns of the subjective nature of controversy events and scores identified by third party vendors.

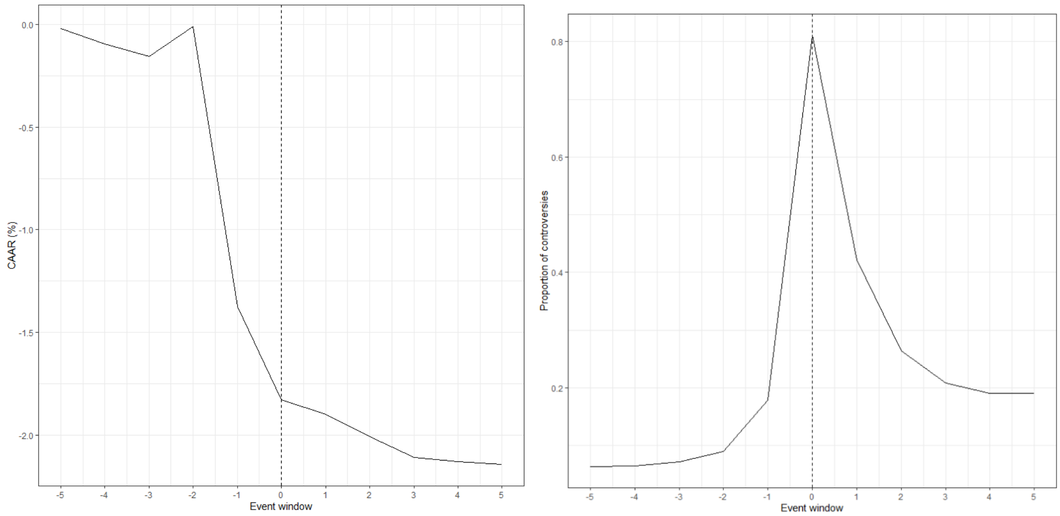
Since we aggregate all the intra-day information into daily data, we notice in the upper panel of Table 3.3.5 that the CAAR drops by 136 basis points on the day preceding the event. As expected, we see a further drop in CAAR by 45 basis points on the event day and about 33 basis points in the week following the event. In line with intuition, this drop in cumulative abnormal returns continues and intensifies on the days following the event in our event window[2]. Figure 3.3.4 summarizes the event study results using our abnormal news metric. The top half of Figure 3.3.4 depicts the downward trend in CAAR as seen in Table 3.3.5 around the days of the event. This is similar to the findings from Cui and Docherty (2020), who observe leakage of information before the event, as the downward tend in CAR begins several days prior to the news release and an increase in abnormal trading volume prior to the news release. More interestingly, the bottom half of Figure 3.3.4 shows how there is a build up in the proportion of company news on controversies leading to event, which dies down slowly afterwards.

Breaking this result down by size deciles based on a company's daily market capitalization, Figure 3.3.5 suggests that the results in Table 3.3.5 and Figure 3.3.4 are largely driven by the small to medium-sized firms in the lower size deciles whereas the firms in the largest size deciles seem to be less affected by an outbreak of controversial news. This is line with findings from Cui and Docherty (2020) who report that the negative reaction to bad ESG news is more pronounced for small cap and mid-cap firms than larger ones as fund managers may prefer to sell smaller cap securities with bad ESG news than larger caps with bad ESG news. Hence, in line with our expectations, Table 3.3.5 and Figure 3.3.4 indicate that controversial behaviour leads to significant drop in daily returns in the days following the outbreak of controversial news and this tends to affect small to medium-sized firms more than larger firms.

We also observe a heterogeneity in results by different geographical regions in Table 3.3.6 similar to the findings from de Vincentiis (2022) who examine stock price reactions

[2]These results are robust across different percentile cut-offs based on our abnormal controversial news activity indicator, and controversial news volume. Results attached in appendix. We report results according to the individual category of controversy used for the event study and also report event study results for longer window of +/- 15 days.

Fig. 3.3.4 The figure plots the evolution of mean cumulative average abnormal returns and mean proportion of controversial news items on each day of our event window. Day 0 is the day of the event and we map the impact from 5 days before the event to 5 days after the event.



to good and bad ESG news in different geographies and also find heterogeneous response of prices to ESG news in different geographies. de Vincentiis (2022) report that Europe, for example, has strong negative price reaction to bad news whereas no such reaction is noted in the US and APAC area. Since divestment strategies are more common in Europe, investors in European markets tend to focus more on mitigating risk, and thereby react more negatively to a firm's controversial behaviour. Supporting this view, De Franco (2020) also report strong negative impact of controversies on stock performance in Europe and USA.

In line with such findings, we also observe in Table 3.3.6 that regions such as Europe, Australia, Emerging Market, and United States of America have highly statistically significant results whereas Japan and United Kingdom do not show such results. These results are not surprising given the higher number of controversial events reported in the U.S, Europe and EMM than the other markets in the sample. Hence, using an event study approach, we successfully document the negative price reaction of controversies across sectors, size deciles and geographies.

Fig. 3.3.5 Mean cumulative average abnormal returns categorized by size deciles based on daily market capitalisation on each day of our event window. Size decile 1 comprises of the lowest market capitalization companies whereas 10 comprises of the largest companies. Day 0 is the day of the event and we map the impact from 5 days before the event to 5 days after the event.

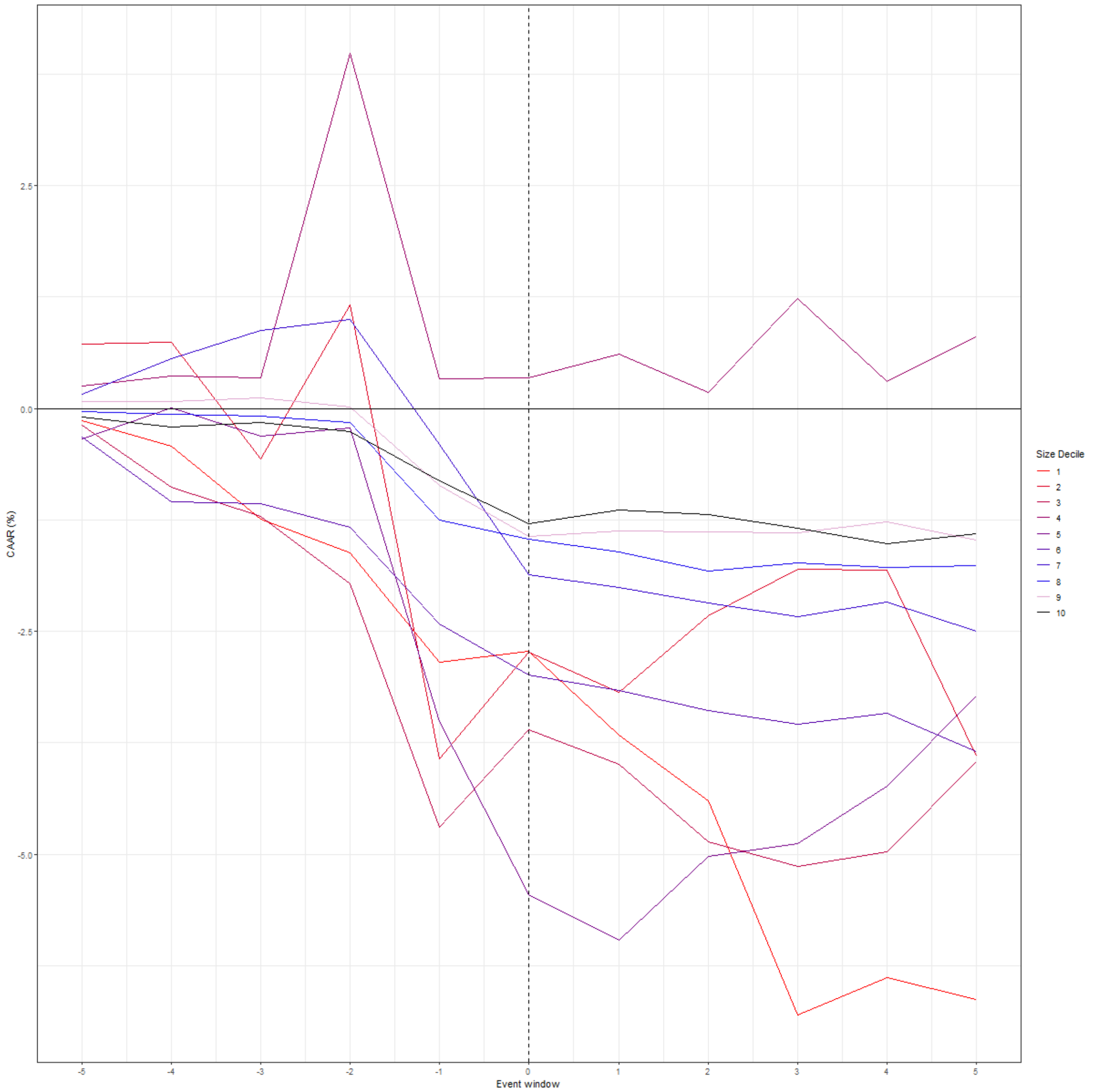


Table 3.3.6 Heterogeneity by Region- Mean cumulative abnormal returns and the associated t-statistic on each day of our event window. We use an estimation period of 252 trading days and event window of (-5,5). We report the average abnormal returns (AAR), cumulative average abnormal returns (CAAR), the number of controversial events, the proportion of controversial news, the t-statistics of AAR and t-statistics of CAAR. ***, ** and ** denotes significance at 1% level, 5% level and * denotes significance at 10% level.

Event Window	AAR (%)	CAAR (%)	No. of Events	PropC	t-stat$_{AAR}$	t-stat$_{CAAR}$	Event Window	AAR (%)	CAAR (%)	No. of Events	Prop.C	t-stat$_{AAR}$	t-stat$_{CAAR}$
Australia							**Japan**						
-5	-0.28	-0.28	39	0.05	-1.02	-1.02	-5	0.76*	0.76*	93	0.04	1.94	1.94
-4	-0.17	-0.45	39	0.09	-0.58	-1.36	-4	-0.17	0.59	93	0.07	-0.61	1.06
-3	-0.31	-0.76	39	0.03	-0.88	-1.49	-3	-0.14	0.45	93	0.17	-0.67	0.79
-2	-0.51*	-1.27**	39	0.04	-1.86	-1.96	-2	-0.02	0.43	93	0.11	-0.11	0.70
-1	-2.11***	-3.38***	39	0.33	-2.79	-3.89	-1	-0.90**	-0.47	93	0.21	-2.38	-0.69
0	-0.27	-3.65***	39	0.82	-0.32	-2.78	0	-0.63	-1.10	93	0.86	-1.50	-1.27
1	0.01	-3.64***	39	0.40	0.03	-2.70	1	-0.26	-1.36	93	0.39	-0.95	-1.41
2	-0.08	-3.72**	39	0.19	-0.31	-2.55	2	-0.17	-1.53	93	0.23	-0.69	-1.45
3	0.14	-3.58**	39	0.31	0.53	-2.41	3	-0.05	-1.58	93	0.22	-0.23	-1.45
4	0.12	-3.46**	39	0.18	0.41	-2.18	4	0.19	-1.39	93	0.17	0.63	-1.29
5	0.37	-3.09**	39	0.24	0.92	-2.00	5	-0.27	-1.66	93	0.15	-0.98	-1.56
United Kingdom							**Europe excluding UK**						
-5	0.66**	0.66**	99	0.06	2.15	2.15	-5	-0.16	-0.16	190	0.04	-1.17	-1.17
-4	-0.26	0.40	99	0.05	-0.73	0.88	-4	-0.06	-0.22	190	0.09	-0.45	-1.14
-3	-0.10	0.30	99	0.04	-0.46	0.56	-3	-0.16	-0.38	190	0.10	-0.95	-1.39
-2	-0.02	0.28	99	0.08	-0.05	0.45	-2	0.12	-0.26	190	0.08	0.86	-0.83
-1	-1.34	-1.06	99	0.18	-1.66	-0.99	-1	-0.86***	-1.12**	190	0.18	-2.93	-2.44
0	0.26	-0.80	99	0.77	0.43	-0.64	0	-0.82***	-1.94***	190	0.80	-3.08	-3.33
1	0.27	-0.53	99	0.48	0.69	-0.37	1	-0.14	-2.08***	190	0.41	-0.76	-3.38
2	-0.40	-0.93	99	0.25	-1.57	-0.68	2	0.08	-2.00***	189	0.23	0.66	-3.30
3	-0.27	-1.20	99	0.15	-0.44	-0.71	3	-0.02	-2.02***	188	0.15	-0.12	-3.38
4	0.45	-0.75	99	0.14	0.98	-0.42	4	-0.04	-2.06***	188	0.18	-0.29	-3.30
5	0.13	-0.62	99	0.16	0.60	-0.35	5	0.06	-2.00***	185	0.16	0.47	-3.03
United States of America							**Emerging Markets**						
-5	-0.06	-0.06	699	0.07	-0.46	-0.46	-5	-0.40**	-0.40**	201	0.08	-2.05	-2.05
-4	-0.01	-0.07	699	0.05	-0.07	-0.38	-4	-0.21	-0.61**	201	0.07	-1.23	-2.31
-3	0.05	-0.02	699	0.06	0.53	-0.05	-3	-0.18	-0.79**	201	0.07	-0.90	-2.29
-2	0.45	0.43	699	0.09	1.37	1.11	-2	-0.58**	-1.37***	201	0.10	-2.20	-3.16
-1	-1.49***	-1.06**	699	0.16	-4.57	-2.04	-1	-1.75***	-3.12***	201	0.21	-3.37	-4.40
0	-0.31	-1.37**	699	0.80	-1.33	-2.27	0	-1.03***	-4.15***	201	0.83	-2.90	-4.60
1	-0.06	-1.43**	699	0.40	-0.49	-2.36	1	-0.32	-4.47***	199	0.49	-1.08	-3.98
2	-0.24**	-1.67**	698	0.27	-2.05	-2.74	2	0.21	-4.26***	198	0.31	0.62	-4.42
3	-0.01	-1.68***	696	0.21	-0.13	-2.78	3	-0.05	-4.31***	198	0.24	-0.23	-4.68
4	-0.15	-1.83***	692	0.20	-1.54	-3.03	4	0.08	-4.23***	198	0.22	0.34	-4.49
5	0.00	-1.83***	690	0.21	-0.02	-3.12	5	0.08	-4.15***	198	0.19	0.49	-4.10

3.4 Conclusion

In this study, we map the price impact of extreme controversial events and explore the potential of NLP techniques for identifying S domain specific controversies from news headlines. Rather than relying on context-ignorant dictionary approaches, we take a novel route of adapting a context-aware language model to identify controversies in the social dimension of ESG. To this end, we further-train a DistilRoBERTa model using a large dataset of 1 million company-specific news headlines. However, since our end goal is to identify controversies along eight different dimension of the social aspect of ESG, we create a hand-labelled dataset with representative examples from each of the 8 categories in order to fine-tune the domain-adapted model.

This Controversy-BERT exhibits marked improvements in the accuracy of predictions for both binary and multi-label headline classification exercises, thereby highlighting the advantages of context-aware language models. Certain categories of controversies such as consumer data safety and privacy and discrimination have higher accuracy of prediction as these categories tend to be more distinctive that the other chosen categories. Overall, we document the usability and adaptability of deep learning based large language models for automating the identification of controversial behaviour from news headlines.

To test the price impact of extreme controversial events, we develop an abnormal news activity metric and detect about 1,393 controversial events between 2014 and 2022. Using an event study approach, we observe leakage of information prior to the outbreak of controversial news as the returns drop by 136 basis points on the day preceding the event. They further drop by 45 basis points following the publication of controversial news, a trend that intensifies in the week following the event. Consistent with the findings in previous literature, this result in largely driven by smaller to medium-sized companies and is more pronounced in Europe, Australia, Emerging markets and the U.S.

Hence, from a practitioner perspective, we propose a novel automated way of identifying social controversies directly from news headlines. This approach is not limited to news headlines and can be easily extended to other textual datasets such as earnings calls transcripts, SEC filings, etc. Given the increased push from investors for investments with social impact, asset managers can use such state-of-the-art language models to build a screening system for identifying social violations by companies in their portfolios.

To summarise, we not only propose a novel automated way of identifying controversies in a specific domain of interest, but also showcase the efficacy of this approach using a hand-labelled dataset. We also demonstrate the potential of our abnormal news activity indicator to detect extreme controversial events that negatively impact firm returns. Using an event study approach, we further quantify the negative impact and document heterogeneity in price reaction across different categories of social controversies, size deciles and geographies.

Appendix 3.A Tables

Table 3.A.1 Overall distribution of daily abnormal controversial news activity indicator, A_{it}.

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-1.57	-0.16	-0.11	0.07	-0.07	2976.05

Table 3.A.2 Size Decile-wise distribution of daily abnormal controversial news activity indicator, A_{it}.

Size Decile	1st Quartile	3rd Quartile	Maximum	Mean	Median	Minimum
1	-0.12	-0.06	2101.49	0.07	-0.09	-0.64
2	-0.12	-0.05	530.82	0.05	-0.09	-0.74
3	-0.12	-0.05	821.06	0.06	-0.09	-0.72
4	-0.12	-0.05	792.76	0.06	-0.09	-0.61
5	-0.13	-0.06	524.60	0.06	-0.09	-0.69
6	-0.14	-0.07	2976.05	0.06	-0.10	-0.83
7	-0.15	-0.07	891.07	0.06	-0.10	-0.97
8	-0.16	-0.07	1012.51	0.07	-0.11	-0.82
9	-0.18	-0.08	839.27	0.07	-0.12	-1.08
10	-0.25	-0.09	508.70	0.07	-0.16	-1.57

Table 3.A.3 Heterogeneity by Controversy Category- Mean cumulative abnormal returns and the associated t-statistic on each day of our event window. We use an estimation period of 252 trading days and event window of (-5,5). We report the average abnormal returns, cumulative average abnormal returns (CAAR), the corresponding standard deviation, the number of controversial events, the proportion of controversial news, the t-statistics of AAR and t-statistics of CAAR. *** denotes significance at 1% level, ** denotes significance at 5% level and * denotes significance at 10% level.

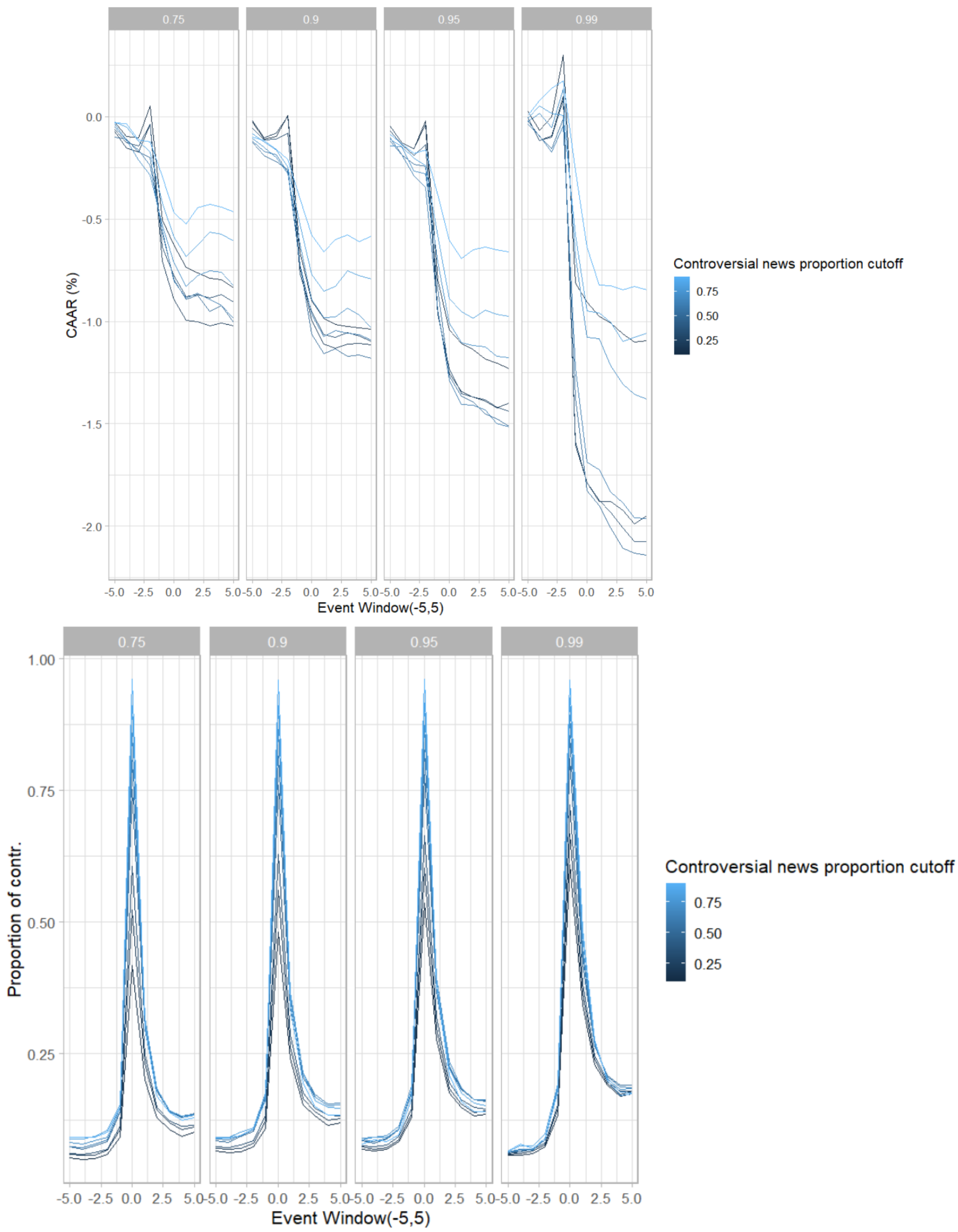
Event Window	AAR (%)	CAAR (%)	No. of Events	PropC	t-stat$_{AAR}$	t-stat$_{CAAR}$	Event Window	AAR (%)	CAAR (%)	No. of Events	Prop.C	t-stat$_{AAR}$	t-stat$_{CAAR}$
Product Safety Standards							Consumer Data Safety and Privacy						
-5	-0.16	-0.16	474	0.07	-1.44	-1.44	-5	0.16	0.16	276	0.06	0.88	0.88
-4	-0.22	-0.38*	474	0.06	-1.45	-1.93	-4	0.15	0.31	276	0.06	1.09	1.30
-3	-0.14	-0.52**	474	0.07	-1.07	-2.04	-3	0.02	0.33	276	0.07	0.13	1.18
-2	0.41	-0.11	474	0.10	0.90	-0.23	-2	0.54**	0.87**	276	0.10	1.97	2.06
-1	-2.53***	-2.64***	474	0.19	-5.59	-3.71	-1	-1.27***	-0.40	276	0.21	-4.49	-0.84
0	-0.92***	-3.56***	474	0.79	-3.35	-4.36	0	-0.81**	-1.21**	276	0.81	-2.51	-2.07
1	0.10	-3.46***	473	0.42	0.59	-4.17	1	-0.52***	-1.73***	276	0.47	-3.40	-2.83
2	-0.30**	-3.76***	472	0.30	-2.38	-4.51	2	-0.06	-1.80***	275	0.3	-0.40	-2.94
3	-0.02	-3.78***	471	0.24	-0.09	-4.43	3	-0.30**	-2.10***	275	0.25	-2.34	-3.27
4	-0.09	-3.97***	469	0.23	-0.79	-4.60	4	-0.12	-2.22***	274	0.20	-0.62	-3.50
5	-0.04	-4.01***	467	0.22	-0.35	-4.59	5	0.01	-2.20***	273	0.20	0.05	-3.24
Stakeholder							Discrimination						
-5	-0.01	-0.01	258	0.06	-0.04	-0.04	-5	0.07	0.07	28	0.04	0.29	0.29
-4	-0.18	-0.19	258	0.07	-1.08	-0.61	-4	-0.44*	-0.37	28	0.03	-1.84	-1.10
-3	-0.06	-0.25	258	0.05	-0.31	-0.65	-3	-0.12	-0.49	28	0.03	-0.58	-1.35
-2	-0.22	-0.47	258	0.07	-1.04	-1.09	-2	-0.90*	-1.39**	28	0.07	-1.90	-2.24
-1	-1.00*	-1.47**	258	0.17	-1.93	-2.14	-1	0.23	-1.16	28	0.21	0.41	-1.48
0	0.39	-1.08	258	0.79	0.92	-1.20	0	0.61	-0.55	28	0.85	0.60	-0.41
1	-0.15	-1.23	257	0.42	-0.60	-1.06	1	-0.04	-0.59	28	0.46	-0.14	-0.44
2	0.28	-0.95	256	0.23	0.93	-0.78	2	-0.38	-0.97	28	0.24	-1.42	-0.68
3	0.09	-0.86	255	0.16	0.46	-0.77	3	0.21	-0.76	28	0.13	0.58	-0.55
4	0.07	-0.79	254	0.15	0.49	-0.65	4	-0.15	-0.91	28	0.25	-0.63	-0.68
5	0.14	-0.65	253	0.16	1.02	-0.73	5	-0.18	-1.09	27	0.15	-0.75	-0.34

Table 3.A.4 Heterogeneity by Controversy Category- Mean cumulative abnormal returns and the associated t-statistic on each day of our event window. We use an estimation period of 252 trading days and event window of (-5,5). We report the average abnormal returns, cumulative average abnormal returns (CAAR), the corresponding standard deviation, the number of controversial events, the proportion of controversial news, the t-statistics of AAR and t-statistics of CAAR. *** denotes significance at 1% level, ** denotes significance at 5% level and * denotes significance at 10% level.

Event Window	AAR (%)	CAAR (%)	No. of Events	PropC	t-stat$_{AAR}$	t-stat$_{CAAR}$	Event Window	AAR (%)	CAAR (%)	No. of Events	Prop.C	t-stat$_{AAR}$	t-stat$_{CAAR}$
Fundamental Human Rights							**Workplace Health/Safety Standards**						
-5	0.48	0.48	55	0.05	1.57	1.57	-5	0.12	0.12	205	0.06	0.67	0.67
-4	-0.11	0.37	55	0.07	-0.37	0.79	-4	0.21	0.33	205	0.06	1.56	1.52
-3	0.09	0.46	55	0.11	0.33	0.86	-3	-0.08	0.25	205	0.06	-0.48	0.91
-2	-0.34	0.12	55	0.12	-1.20	0.20	-2	-0.14	0.11	205	0.06	-0.83	0.31
-1	-0.21	-0.09	55	0.22	-0.49	-0.14	-1	-0.21	-0.10	205	0.09	-0.94	-0.28
0	-0.92***	-1.01	55	0.81	-2.90	-1.41	0	0.08	-0.02	205	0.87	0.38	-0.05
1	0.43	-0.58	55	0.45	1.77	-0.70	1	-0.17	-0.19	205	0.31	-1.09	-0.43
2	-0.07	-0.65	55	0.22	-0.34	-0.77	2	-0.01	-0.20	205	0.17	-0.07	-0.44
3	-0.01	-0.66	55	0.23	-0.04	-0.76	3	-0.23	-0.43	204	0.14	-1.39	-0.88
4	-0.01	-0.67	55	0.19	-0.07	-0.70	4	0.05	-0.38	204	0.12	0.21	-0.73
5	-0.08	-0.75	55	0.19	-0.35	-0.77	5	-0.08	-0.46	203	0.13	-0.55	-0.87
Labour Standards							**Modern Slavery**						
-5	-0.31	-0.31	78	0.07	-0.81	-0.81	-5	-1.08	-1.08	19	0.09	-1.41	-1.41
-4	-0.01	-0.32	78	0.09	-0.01	-0.58	-4	-1.01**	-2.09**	19	0.06	-2.40	-2.19
-3	0.01	-0.31	78	0.16	0.02	-0.57	-3	0.27	-1.82	19	0.18	0.33	-1.46
-2	-0.18	-0.49	78	0.13	-0.62	-0.94	-2	0.34	-1.48	19	0.17	0.49	-0.89
-1	-0.36	-0.85	78	0.19	-0.87	-1.25	-1	-0.73	-2.21	19	0.24	-1.66	-1.25
0	-0.74	-1.59	78	0.78	-2.00	-1.90	0	0.03	-2.18	19	0.88	0.03	-1.65
1	-0.42	-2.01**	78	0.49	-1.24	-2.21	1	0.61	-1.57	19	0.46	1.06	-1.34
2	-0.49	-2.50***	78	0.31	-2.01	-2.85	2	-1.17**	-2.74*	19	0.24	-2.23	-1.93
3	0.13	-2.37**	78	0.20	0.40	-2.57	3	0.06	-2.68*	19	0.24	0.18	-1.91
4	0.24	-2.13	78	0.24	0.86	-2.20	4	0.68	-2.00	19	0.14	1.15	-1.46
5	0.16	-1.97**	78	0.21	0.50	-2.19	5	1.69	-0.31	19	0.27	0.80	-0.10

Appendix 3.B Figure

Fig. 3.B.1 The figure plots the evolution of mean cumulative abnormal returns and mean proportion of controversial news items on each day of our event window for different proportions of controversial news and for different cut-offs of daily news volume. Going from right to left, we vary the threshold percentile based on our abnormal activity indicator A_{it}. Day 0 is the day of the event and we map the impact from 5 days before the event to 5 days after the event.



Appendix 3.C Robustness Check

Fig. 3.C.1 Returns and controversial news volume around the event date
The figure plots the evolution of mean cumulative average abnormal returns and mean proportion of controversial news items on each day of the event window. Day 0 is the event day, and we map the impact from 15 days before the event to 15 days after the event.

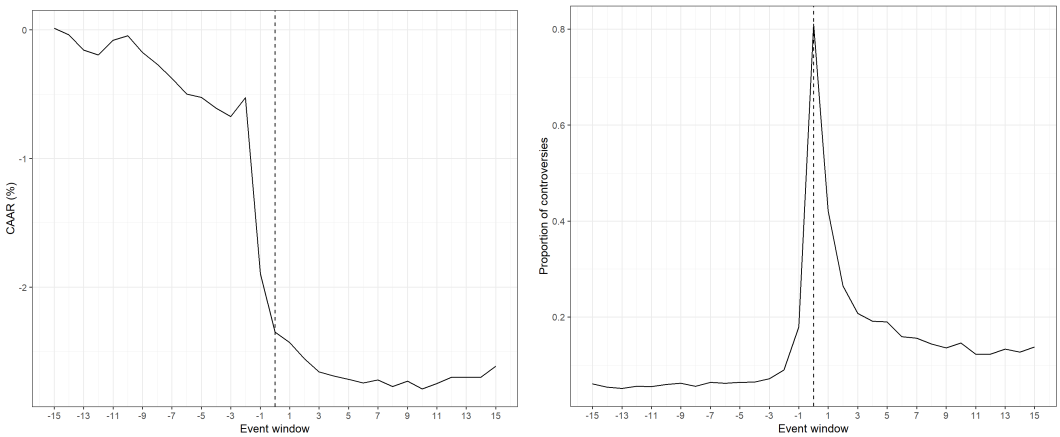


Table 3.C.1 Event study results

Mean cumulative abnormal returns and the associated t-statistics on each event window day are reported based on the events identified using our abnormal news activity metric. We use an estimation period of 252 trading days and an event window of (-15,15) for the events identified between January 2014 and October 2022. We report the average abnormal returns (AAR), cumulative average abnormal returns (CAAR), the number of controversial events, the proportion of controversial news, the t-statistics of AAR and t-statistics of CAAR. *** denotes significance at the 1%level, ** denotes significance at the 5%level and * denotes significance at the 10%level.

Event Window	AAR (%)	CAAR (%)	# of Events	Proportion of controversial news	t-stat$_{AAR}$	t-stat$_{CAAR}$
-15	0.01	0.01	1392	0.06	0.17	0.17
-14	-0.05	-0.04	1392	0.05	-0.65	-0.35
-13	-0.12**	-0.16	1392	0.05	-1.80	-1.25
-12	-0.04	-0.19*	1392	0.06	-0.60	-1.39
-11	0.11	-0.08	1392	0.06	0.73	-0.40
-10	0.03	-0.05	1392	0.06	0.58	-0.21
-9	-0.13**	-0.18	1392	0.06	-2.14	-0.77
-8	-0.09*	-0.27	1392	0.06	-1.37	-1.12
-7	-0.11*	-0.38*	1392	0.06	-1.35	-1.52
-6	-0.12**	-0.50**	1392	0.06	-1.89	-1.94
-5	-0.02	-0.52**	1392	0.06	-0.31	-1.97
-4	-0.08	-0.61**	1392	0.06	-1.12	-2.14
-3	-0.07	-0.67**	1392	0.07	-0.94	-2.28
-2	0.15	-0.53*	1392	0.09	0.85	-1.53
-1	-1.37***	-1.89***	1392	0.18	-6.94	-4.57
0	-0.46***	-2.35***	1392	0.81	-3.13	-5.12
1	-0.12*	-2.43***	1390	0.42	-1.46	-5.13
2	-0.12*	-2.56***	1387	0.26	-1.44	-5.45
3	-0.07	-2.66***	1384	0.21	-0.92	-5.61
4	-0.02	-2.69***	1380	0.19	-0.30	-5.65
5	0.03	-2.71***	1374	0.19	0.38	-5.59
6	-0.06	-2.74***	1372	0.16	-0.87	-5.62
7	0.02	-2.72***	1372	0.16	0.29	-5.38
8	-0.05	-2.77***	1370	0.14	-0.87	-5.43
9	0.00	-2.73***	1366	0.14	0.07	-5.28
10	-0.04	-2.79***	1364	0.15	-0.56	-5.40
11	0.04	-2.75***	1364	0.12	0.66	-5.31
12	0.04	-2.70***	1362	0.12	0.55	-5.17
13	0.00	-2.70***	1360	0.13	0.04	-5.14
14	0.00	-2.70***	1359	0.13	0.01	-5.12
15	0.08	-2.61***	1358	0.14	0.92	-4.89

References

ANTWEILER, W. and FRANK, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, **59** (3), 1259–1294.

AOUADI, A. and MARSAT, S. (2018). Do esg controversies matter for firm value? evidence from international data. *Journal of Business Ethics*, **151**, 1027–1047.

ARACI, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

AVRAMOV, D., CHENG, S., LIOUI, A. and TARELLI, A. (2022). Sustainable investing with esg rating uncertainty. *Journal of Financial Economics*, **145** (2), 642–664.

BAID, V. and JAYARAMAN, V. (2022). Amplifying and promoting the "s" in esg investing: the case for social responsibility in supply chain financing. *Managerial Finance*, (ahead-of-print).

BARTOV, E., FAUREL, L. and MOHANRAM, P. S. (2018). Can twitter help predict firm-level earnings and stock returns? *The Accounting Review*, **93** (3), 25–57.

BELTAGY, I., LO, K. and COHAN, A. (2019a). Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

—, — and — (2019b). Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

BILLIO, M., COSTOLA, M., HRISTOVA, I., LATINO, C. and PELIZZON, L. (2021). Inside the esg ratings:(dis) agreement and performance. *Corporate Social Responsibility and Environmental Management*, **28** (5), 1426–1445.

BOLLEN, J., MAO, H. and ZENG, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, **2** (1), 1–8.

CAPELLE-BLANCARD, G. and PETIT, A. (2017). The weighting of csr dimensions: One size does not fit all. *Business & Society*, **56** (6), 919–943.

— and — (2019). Every little helps? esg news and stock market reaction. *Journal of Business Ethics*, **157**, 543–565.

CHEN, C. Y.-H., GUO, L. and RENAULT, T. (2019). What makes cryptocurrencies special? investor sentiment and return predictability. *Working Paper*.

CUI, B. and DOCHERTY, P. (2020). Stock price overreaction to esg controversies. *Available at SSRN 3559915*.

CURTIS, A., RICHARDSON, V. J. and SCHMARDEBECK, R. (2016). Investor attention and the pricing of earnings news. *Handbook of Sentiment Analysis in Finance, Available at SSRN 2467243*.

DAS, S. R. and CHEN, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, **53** (9), 1375–1388.

Davis, A. K. and Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: Earnings press releases versus md&a. *Contemporary Accounting Research*, **29** (3), 804–837.

De Franco, C. (2020). Esg controversies and their impact on performance. *The Journal of Investing*, **29** (2), 33–45.

de Vincentiis, P. (2022). Do international investors care about esg news? *Qualitative Research in Financial Markets*, (ahead-of-print).

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dimson, E., Marsh, P. and Staunton, M. (2020). Divergent esg ratings. *The Journal of Portfolio Management*, **47** (1), 75–87.

Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements. In *Working paper*.

Fair Labor Association (2012). Independent investigation of apple supplier, foxconn. https://www.fairlabor.org/sites/default/files/documents/reports/foxconn_investigation_report.pdf.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, **68** (3), 1267–1300.

Gentzkow, M., Kelly, B. and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, **57** (3), 535–74.

Gurun, U. G. and Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance*, **67** (2), 561–598.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D. and Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

Hanley, K. W. and Hoberg, G. (2010). The information content of ipo prospectuses. *The Review of Financial Studies*, **23** (7), 2821–2864.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication*, **45** (4), 363–407.

Huang, A. H., Zang, A. Y. and Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, **89** (6), 2151–2180.

Huang, K., Altosaar, J. and Ranganath, R. (2019a). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.

—, — and — (2019b). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.

Jegadeesh, N. and Wu, D. (2013). Word power: A new approach for content analysis. *Journal of financial economics*, **110** (3), 712–729.

KEARNEY, C. and LIU, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, **33**, 171–185.

KOTHARI, S. P., LI, X. and SHORT, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, **84** (5), 1639–1670.

KOTSANTONIS, S. and SERAFEIM, G. (2019). Four things no one will tell you about esg data. *Journal of Applied Corporate Finance*, **31** (2), 50–58.

KRÜGER, P. (2015). Corporate goodness and shareholder wealth. *Journal of financial economics*, **115** (2), 304–329.

LARCKER, D. F., POMORSKI, L., TAYAN, B. and WATTS, E. M. (2022). Esg ratings: A compass without direction. *Rock Center for Corporate Governance at Stanford University Working Paper Forthcoming.*

LEE, J., YOON, W., KIM, S., KIM, D., KIM, S., SO, C. H. and KANG, J. (2020). Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, **36** (4), 1234–1240.

LI, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics*, **45** (2-3), 221–247.

— (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, **29** (1), 143–165.

LIU, N., HU, Q., XU, H., XU, X. and CHEN, M. (2021a). Med-bert: A pretraining framework for medical records named entity recognition. *IEEE Transactions on Industrial Informatics*, **18** (8), 5600–5608.

LIU, Z., HUANG, D., HUANG, K., LI, Z. and ZHAO, J. (2021b). Finbert: A pre-trained financial language representation model for financial text mining. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 4513–4519.

LOUGHRAN, T. and MCDONALD, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, **66** (1), 35–65.

— and — (2015). The use of word lists in textual analysis. *Journal of Behavioral Finance*, **16** (1), 1–11.

— and — (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, **54** (4), 1187–1230.

— and — (2020). Textual analysis in finance. *Annual Review of Financial Economics*, **12**, 357–375.

PATELL, J. M. (1976). Corporate forecasts of earnings per share and stock price behavior: Empirical test. *Journal of accounting research*, pp. 246–276.

PERKISS, S., BERNARDI, C., DUMAY, J. and HASLAM, J. (2021). A sticky chocolate problem: Impression management and counter accounts in the shaping of corporate image. *Critical Perspectives on Accounting*, **81**, 102229.

Qudar, M. M. A. and Mago, V. (2020). Tweetbert: a pretrained language representation model for twitter text analysis. *arXiv preprint arXiv:2010.11091*.

Raghunandan, A. and Rajgopal, S. (2022). Do esg funds make stakeholder-friendly investments? *Review of Accounting Studies*, **27** (3), 822–863.

Rasmy, L., Xiang, Y., Xie, Z., Tao, C. and Zhi, D. (2021). Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, **4** (1), 86.

Serafeim, G. and Yoon, A. (2022). Stock price reactions to esg news: The role of esg ratings and disagreement. *Review of Accounting Studies*, pp. 1–31.

Shiller, R. J. (2017). Narrative economics. *American economic review*, **107** (4), 967–1004.

Simpson, C., Rathi, A. and Kishan, S. (2021). The esg mirage. *Bloomberg Businessweek*.

Stone, P. J., Dunphy, D. C. and Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Tang, D. Y., Yan, J. and Yao, C. Y. (2022). The determinants of esg ratings: Rater ownership matters. In *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, **62** (3), 1139–1168.

—, Saar-Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, **63** (3), 1437–1467.

Webersinke, N., Kraus, M., Bingler, J. A. and Leippold, M. (2021). Climatebert: A pretrained language model for climate-related text. *arXiv preprint arXiv:2110.12010*.

Yang, Y., Uy, M. C. S. and Huang, A. (2020). Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.

Complete References

Aarbu, K. O. and Schroyen, F. (2009). Mapping risk aversion in Norway using hypothetical income gambles. *NHH Dept. of Economics Discussion Paper, 13. Available at SSRN: 1554890*, **13**.

Adams, C. A. (2002). Internal organisational factors influencing corporate social and ethical reporting: Beyond current theorising. *Accounting, Auditing & Accountability Journal*, **15** (2), 223–250.

Admati, A. R. and Pfleiderer, P. (1989). Divide and conquer: A theory of intraday and day-of-the-week mean effects. *Review of Financial Studies*, **2** (2), 189–223.

Agarwal, V. and Taffler, R. (2008). Does financial distress risk drive the momentum anomaly? *Financial Management*, **37**, 461–484.

Aharoni, G., Grundy, B. and Zeng, Q. (2013). Stock returns and the miller modigliani valuation formula: Revisiting the fama french analysis. *Journal of Financial Economics*, **110** (2), 347–357.

Aharony, J. and Swary, I. (1980). Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis. *The Journal of Finance*, **35** (1), 1–12.

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

Akram, Q. F., Rime, D. and Sarno, L. (2008). Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics*, **76** (2), 237–253.

Ali, A., Hwang, L.-S. and Trombley, M. A. (2003). Residual-income-based valuation predicts future stock returns: Evidence on mispricing vs. risk explanations. *Accounting Review*, **78** (2), 377–396.

Amel-Zadeh, A. and Serafeim, G. (2018). Why and how investors use esg information: Evidence from a global survey. *Financial Analysts Journal*, **74**, 87–103.

Amenc, N., Deguest, R., Goltz, F., Lodh, A., Martellini, L. and Shirbini, E. (2014). Risk allocation, factor investing and smart beta: Reconciling innovations in equity portfolio construction. *EDHEC-Risk Institute Working Paper, July 2014. Available at http://docs.edhec-risk.com/mrk/000000/Press/EDHEC-Risk_Publication_Risk_Allocation.pdf.*

Amihud, Y. and Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, **17** (2), 223–249.

Andronoudis, D., Dargenidou, C., Konstantinidi, E. and Pope, P. F. (2019). Conservative accounting and the pricing of risk: the case of research and development. *Available at SSRN 3006537.*

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

— and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial studies*, **15** (4), 1137–1187.

— and — (2004). How regimes affect asset allocation. *Financial Analysts Journal*, **60** (2), 86–99.

— and — (2007). Stock return predictability: Is it there? *Review of Financial Studies*, **20** (3), 651–707.

—, Goetzmann, W. N. and Schaefer, S. (2009a). Evaluation of active management of the Norwegian government pension fund–Global. *report to the Norwegian Ministry of Finance.*

—, Hodrick, R. J., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, **61** (1), 259–299.

—, —, — and — (2009b). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, **91**, 1–23.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, **59** (3), 1259–1294.

Aouadi, A. and Marsat, S. (2018). Do esg controversies matter for firm value? evidence from international data. *Journal of Business Ethics*, **151**, 1027–1047.

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063.*

Aretz, K., Bartram, S. M. and Pope, P. F. (2010). Macroeconomic risks and characteristic-based factor models. *Journal of Banking and Finance*, **34**, 1383–1399.

— and Pope, P. F. (2018). Real options models of the firm, capacity overhang, and the cross section of stock returns. *Journal of Finance*, **73**, 1363–1415.

Arnott, R., Beck, N., Kalesnik, V. and West, J. (2016). How can 'smart beta'go horribly wrong? *Research Affiliates, February.*

Arnott, R. D., Clements, M., Kalesnik, V. and Linnainmaa, J. T. (2021a). *Factor momentum. Available at SSRN 3116974.*

—, Harvey, C. R., Kalesnik, V. and Linnainmaa, J. T. (2021b). Reports of value's death may be greatly exaggerated. *Financial Analysts Journal, forthcoming.*

Asness, C. (2011). Momentum in japan: The exception that proves the rule. *Journal of Portfolio Management*, **37**, 67–75.

—, Chandra, S., Ilmanen, A. and Israel, R. (2017). Contrarian factor timing is deceptively difficult. *Journal of Portfolio Management*, **43**, 72–87.

Asness, C. S. (1997). The interaction of value and momentum strategies. *Financial Analysts Journal*, pp. 29–36.

— (2016). The siren song of factor timing aka "smart beta timing" aka "style timing". *Journal of Portfolio Management*, **42** (5), 1–6.

—, Frazzini, A. and Pedersen, L. H. (2014). Low-risk investing without industry bets. *Financial Analysts Journal*, **70**, 24–41.

—, Friedman, J. A., Krail, R. J. and Liew, J. M. (2000). Style timing: Value versus growth. *Journal of Portfolio Management*, **26** (3), 50–60.

—, Moskowitz, T. J. and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, **68**, 929–985.

Asquith, P. and Mullins Jr, D. W. (1986). Equity issues and offering dilution. *Journal of financial economics*, **15** (1-2), 61–89.

Avramov, D., Cheng, S., Lioui, A. and Tarelli, A. (2022). Sustainable investing with esg rating uncertainty. *Journal of Financial Economics*, **145** (2), 642–664.

—, — and Metzker, L. (2021). *Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability. Available at SSRN 3450322.*

—, —, Schreiber, A. and Shemer, K. (2017). Scaling up market anomalies. *Journal of Investing*, **26**, 89–105.

—, Chordia, T., Jostova, G. and Philipov, A. (2007). Momentum and credit rating. *Journal of Finance*, **62**, 2503–2520.

—, —, — and — (2019). Bonds, stocks, and sources of mispricing. unpublished manuscript.

Bai, J., Bali, T. G. and Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, **131**, 619–642.

Baid, V. and Jayaraman, V. (2022). Amplifying and promoting the "s" in esg investing: the case for social responsibility in supply chain financing. *Managerial Finance*, (ahead-of-print).

Baker, M., Bradley, B. and Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, **67** (1), 40–54.

— and Wurgler, J. (2000). The equity share in new issues and aggregate stock returns. *Journal of Finance*, **55** (5), 2219–2257.

— and — (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, **61** (4), 1645–1680.

— and — (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, **21** (2), 129–151.

—, — and Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, **104** (2), 272–287.

Bakshi, G., Gao, X. and Rossi, A. G. (2017). Understanding the sources of risk underlying the cross section of commodity returns. *Management Science*, **65**, 619–641.

Baku, E., Fortes, R., Hervé, K., Lezmi, E., Malongo, H., Roncalli, T. and Xu, J. (2020). Factor investing in currency markets: Does it make sense. *Journal of Portfolio Management*, **46**, 141–155.

Balakrishnan, K., Bartov, E. and Faurel, L. (2010). Post loss/profit announcement drift. *Journal of Accounting and Economics*, **50** (1), 20–41.

Bali, T. G., Cakici, N. and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, **99** (2), 427–446.

— and Engle, R. F. (2010). The intertemporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics*, **57** (4), 377–390.

—, — and Murray, S. (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

—, Goyal, A., Huang, D., Jiang, F. and Wen, Q. (2020a). The cross-sectional pricing of corporate bonds using big data and machine learning.

—, Subrahmanyam, A. and Wen, Q. (2020b). Long-term reversals in the corporate bond market. *Journal of Financial Economics*.

Ball, R. and Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of accounting research*, pp. 159–178.

Baltas, N. (2019). The impact of crowding in alternative risk premia investing. *Financial Analysts Journal*, **75**, 89–104.

Bansal, R., Dittmar, R. F. and Lundblad, C. T. (2005). Consumption, dividends, and the cross section of equity returns. *Journal of Finance*, **60** (4), 1639–1672.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, **9** (1), 3–18.

Barberis, N., Shleifer, A. and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, **49** (3), 307–343.

Barillas, F. and Shanken, J. (2017). Which alpha? *Review of Financial Studies*, **30**, 1316–1338.

— and — (2018). Comparing asset pricing models. *Journal of Finance*, **73**, 715–754.

Barro, R. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, **121**, 823–866.

Barroso, P. and Santa-Clara, P. (2015). Beyond the carry trade: Optimal currency portfolios. *Journal of Financial and Quantitative Analysis*, **50** (5), 1037–1056.

Barry, C. B., Goldreyer, E., Lockwood, L. and Rodriguez, M. (1985). Robustness of size and value effects in emerging equity markets. *Emerging Markets Review*, **3**, 1–30.

—, —, — and — (2002). Robustness of size and value effects in emerging equity markets, 1985–2000. *Emerging Markets Review*, **3**, 1–30.

Bartov, E., Faurel, L. and Mohanram, P. S. (2018). Can twitter help predict firm-level earnings and stock returns? *The Accounting Review*, **93** (3), 25–57.

Bartram, S. M., Branke, J. and Motahari, M. (2020a). *Artificial intelligence in asset management*. CFA Institute Research Foundation.

—, Brown, G. W. and Conrad, J. (2011). The effects of derivatives on firm risk and value. *Journal of Financial and Quantitative Analysis*, **46**, 967–999.

—, Djuranovik, L. and Garratt, A. (2020b). *Currency anomalies. CEPR Working Paper 15653*.

—, — and — (2020c). *Currency Anomalies*. Working paper, University of Warwick.

— and Dufey, G. (2001). International portfolio investment: Theory, evidence, and institutional framework. financial markets. *Institutions* & *Instruments*, **10**, 85–155.

—, Griffin, J. M., Lim, T.-H. and Ng, D. T. (2015). How important are foreign ownership linkages for international stock returns? *Review of Financial Studies*, **28**, 3036–3072.

— and Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, **128**, 125–147.

— and — (2021). Global market inefficiencies. *Journal of Financial Economics*, **139**, 234–259.

—, — and Nozawa, Y. (2020d). Book-to-market, mispricing, and the cross-section of corporate bond returns.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance*, **32** (3), 663–682.

Bauer, R., Derwall, J. and Molenaar, R. (2004). The real-time predictability of the size and value premium in Japan. *Pacific-Basin Finance Journal*, **12** (5), 503–523.

Bekaert, G. and Panayotov, G. (2020). Good carry, bad carry. *Journal of Financial and Quantitative Analysis*, **55** (4), 1063–1094.

Bektić, D., Wenzler, J.-S., Wegener, M., Schiereck, D. and Spielmann, T. (2019a).

—, —, —, — and Spielmann, T. (2019b). Extending fama–french factors to corporate bond markets. *Journal of Portfolio Management*, **45** (3), 141–158.

Beltagy, I., Lo, K. and Cohan, A. (2019a). Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

—, — and — (2019b). Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

Berk, J. B. (1995). A critique of size-related anomalies. *Review of Financial Studies*, **8**, 275–286.

Berkman, H., Jacobsen, B. and Lee, J. B. (2011). Time-varying rare disaster risk and stock returns. *Journal of Financial Economics*, **101**, 313–332.

—, — and — (2017). Rare disaster risk and the expected equity risk premium. *Accounting & Finance*, **57**, 351–372.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, **43** (2), 507–528.

Bhansali, V. (2007). Volatility and the carry trade. *Journal of Fixed Income*, **17**, 72–84.

Bianchi, D., Büchner, M. and Tamoni, A. (2020). Bond risk premiums with machine learning. *Review of Financial Studies*, **34**, 1046–1089.

Billio, M., Costola, M., Hristova, I., Latino, C. and Pelizzon, L. (2021). Inside the esg ratings:(dis) agreement and performance. *Corporate Social Responsibility and Environmental Management*, **28** (5), 1426–1445.

Bilson, J. F. (1981a). The speculative efficiency hypothesis. *Journal of Business*, **54**, 435–451.

— (1981b). The speculative efficiency hypothesis. *Journal of Business*, **54** (3), 435–451.

Black, F. and Litterman, R. B. (1991). Asset allocation: combining investor views with market equilibrium. *Journal of Fixed Income*, **1** (2), 7–18.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2002). Latent dirichlet allocation. *Advances in Neural Information Processing Systems*, pp. 601–608.

Blitz, D. and De Groot, W. (2014). Strategic allocation to commodity factor premiums. *Journal of Alternative Investments*, **17**, 103–115.

— and Marchesini, T. (2019). The capacity of factor strategies. *Journal of Portfolio Management*, **45**, 30–38.

—, Pang, J. and Van Vliet, P. (2013). The volatility effect in emerging markets. *Emerging Markets Review*, **16**, 31–45.

— and Van Vliet, P. (2007). The volatility effect: Lower risk without lower return. *Journal of Portfolio Management*, **Fall 2007**, 102–113.

Blume, M. E. (1980). Stock returns and dividend yields: Some more evidence. *Review of Economics and Statistics*, **62**, 567–577.

Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, **2** (1), 1–8.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. **8**, 3–62.

Booth, D. G. and Fama, E. F. (1992). Diversification returns and asset contributions. *Financial Analysts Journal*, **48** (3), 26–32.

Borghi, R. and De Rossi, G. (2020). The artificial intelligence approach to picking stocks. *Machine learning for asset management: new developments and financial applications*, pp. 115–166.

Bossaerts, P. and Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies*, **12** (2), 405–428.

Boudoukh, J., Richardson, M. P. and Whitelaw, R. F. (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *Review of Financial Studies*, **7** (3), 539–573.

Bradshaw, M. T., Richardson, S. A. and Sloan, R. G. (2001). Do analysts and auditors use information in accruals? *Journal of Accounting Research*, **39** (1), 45–74.

—, — and — (2006). The relation between corporate financing activities, analysts forecasts and stock returns. *Journal of Accounting and Economics*, **42** (1), 53–85.

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *Journal of Finance*, **61** (5), 2187–2217.

—, — and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, **22** (9), 3411–3447.

Breen, W., Glosten, L. R. and Jagannathan, R. (1989). Economic significance of predictable variations in stock index returns. *Journal of Finance*, **44** (5), 1177–1189.

Brière, M. and Szafarz, A. (2016). Factor-based v. industry-based asset allocation: The contest. *Amundi Working Paper. Available at SSRN: 2615703.*

Brock, W., Lakonishok, J. and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, **47** (5), 1731–1764.

Brooks, J., Gould, T. and Richardson, S. (2020). Active fixed income illusions. *Journal of Fixed Income*, **29**, 3–16.

— and Moskowitz, T. J. (2018). Yield curve premia. available at ssrn 2956411.

—, Palhares, D. and Richardson, S. (2018). Style investing in fixed income. *Journal of Portfolio Management*, **44**, 127–139.

Brown, S. V. and Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year md&a modifications. *Journal of Accounting Research*, **49** (2), 309–346.

Brunnermeier, M. K., Nagel, S. and Pedersen, L. H. (2008). Carry trades and currency crashes. *NBER Macroeconomics Annual*, **23**, 313–348.

Bruno, S., Chincarini, L. B. and Davis, J. (2015). Crowded spaces and copycat risk management.

Bryzgalova, S., Huang, J. and Julliard, C. (2020). Bayesian solutions for the factor zoo: We just ran two quadrillion models.

Burks, J. J., Cuny, C., Gerakos, J. and Granja, J. (2018). Competition and voluntary disclosure: Evidence from deregulation in the banking industry. *Review of Accounting Studies*, **23** (4), 1471–1511.

Burns, P. (2006). Random portfolios for evaluating trading strategies. *Available at SSRN: 881735*.

Burnside, C., Eichenbaum, M., Kleshchelski, I. and Rebelo, S. (2010). Do peso problems explain the returns to the carry trade? *The Review of Financial Studies*, **24** (3), 853–891.

—, —, — and — (2011a). Do peso problems explain the returns to the carry trade? *Review of Financial Studies*, **24**, 853–891.

—, — and Rebelo, S. (2011b). Carry trade and momentum in currency markets. *Annual Review of Financial Economics*, **3** (1), 511–535.

Caballero, R. J. and Doyle, J. B. (2012). *Carry trade and systemic risk: Why are FX options so cheap?* Tech. rep., National Bureau of Economic Research.

Cahan, R. and Luo, Y. (2013). Standing out from the crowd: Measuring crowding in quantitative strategies. *Journal of Portfolio Management*, **39**, 14–23.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, **18** (2), 373–399.

— (1991). A variance decomposition for stock returns. *The Economic Journal*, **101** (405), 157–179.

— (1993). Understanding risk and return. *National Bureau of Economic Research*.

— and Kyle, A. S. (1993). Smart money, noise trading and stock price behaviour. *The Review of Economic Studies*, **60** (1), 1–34.

— and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, **1** (3), 195–228.

— and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, **21** (4), 1509–1531.

— and Vuolteenaho, T. (2004a). Bad beta, good beta. *American Economic Review*, **94**, 1249–1275.

— and — (2004b). Inflation illusion and stock prices. *National Bureau of Economic Research, Working Paper, 10263*.

Capaul, C., Rowley, I. and Sharpe, W. F. (1993). International value and growth stock returns. *Financial Analysts Journal*, **49**, 27–36.

Capelle-Blancard, G. and Petit, A. (2017). The weighting of csr dimensions: One size does not fit all. *Business & Society*, **56** (6), 919–943.

— and — (2019). Every little helps? esg news and stock market reaction. *Journal of Business Ethics*, **157**, 543–565.

Capiński, M. J., Kopp, E. and Ski, M. J. C. (2014). *Portfolio theory and risk management*, vol. 5. Cambridge University Press.

Carhart, M. (1997a). On persistence in mutual fund performance. *Journal of Finance*, **52** (1), 57–82.

Carhart, M. M. (1997b). On persistence in mutual fund performance. *Journal of Finance*, **52**, 57–82.

Castro, V. (2011). Can central banks' monetary policy be described by a linear (augmented) Taylor rule or by a nonlinear rule? *Journal of Financial Stability*, **7** (4), 228–246.

Cazalet, Z. and Roncalli, T. (2014). Facts and fantasies about factor investing. *Available at SSRN: 2524547*.

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N. and Androutsopoulos, I. (2020). Legal-bert: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559*.

Chambers, D., Lohre, H. and Rother, C. (2018). Tail risk management for multiasset multi-factor strategies. *Risk* & *Reward*, pp. 14–20.

Chan, K., Chen, N.-F. and Hsieh, D. A. (1985). An exploratory investigation of the firm size effect. *Journal of Financial Economics*, **14** (3), 451–471.

Chan, K. C. and Chen, N.-F. (1991). Structural and return characteristics of small and large firms. *Journal of Finance*, **46**, 1467–1484.

Chan, L. K., Hamao, Y. and Lakonishok, J. (1991). Fundamentals and stock returns in japan. *Journal of Finance*, **46**, 1739–1764.

Chandrashekar, S. and Rao, R. K. (2009). The productivity of corporate cash holdings and the cross-section of expected stock returns. *McCombs Research Paper Series No. FIN-03-09*.

Chaves, D. B., Hsu, J. C., Li, F. and Shakernia, O. (2011). Risk parity portfolio vs. other asset allocation heuristic portfolios. *Journal of Investing*, **20** (1), 108–118.

Chen, C. Y.-H., Guo, L. and Renault, T. (2019). What makes cryptocurrencies special? investor sentiment and return predictability. *Working Paper*.

Chen, H.-L. and De Bondt, W. (2004). Style momentum within the S&P-500 index. *Journal of Empirical Finance*, **11** (4), 483–507.

Chen, J. V., Nagar, V. and Schoenfeld, J. (2018). Manager-analyst conversations in earnings conference calls. *Review of Accounting Studies*, **23** (4), 1315–1354.

Chen, N.-F. and Zhang, F. (1998). Risk and return of value stocks. *Journal of Business*, **71** (4), 501–535.

Choi, J. and Kim, Y. (2018). Anomalies and market (dis)integration. *Journal of Monetary Economics*, **100**, 16–34.

Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A. and Tong, Q. (2017). Are capital market anomalies common to equity and corporate bond markets? an empirical investigation. *Journal of Financial and Quantitative Analysis*, **52**, 1301–1342.

—, — and Saretto, A. (2020). Anomalies and false rejections. *Review of Financial Studies*, **33**, 2134–2179.

—, Roll, R. and Subrahmanyam, A. (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics*, **101** (2), 243–263.

—, Subrahmanyam, A. and Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, **58**, 41–58.

Christensen, D. M., Serafeim, G. and Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? the case of esg ratings. *The Accounting Review*, **97** (1), 147–175.

Christiansen, C., Ranaldo, A. and Söderlind, P. (2011). The time-varying systematic risk of carry trade strategies. *Journal of Financial and Quantitative Analysis*, **46** (4), 1107–1125.

Chu, Y., Hirshleifer, D. and Ma, L. (2020). The causal effect of limits to arbitrage on asset pricing anomalies. *Journal of Finance*, **75**, 2631–2672.

Chui, A. C., Titman, S. and Wei, K. J. (2010). Individualism and momentum around the world. *Journal of Finance*, **65** (1), 361–392.

Clare, A., Sapuric, S. and Todorovic, N. (2010). Quantitative or momentum-based multi-style rotation? UK experience. *Journal of Asset Management*, **10** (6), 370–381.

Clarida, R., Davis, J. and Pedersen, N. (2009). Currency carry trade regimes: Beyond the Fama regression. *Journal of International Money and Finance*, **28** (8), 1375–1389.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, **21** (4), 1533–1575.

— (2009). *Asset pricing*. Princeton University Press.

— (2011). Presidential address: Discount rates. *Journal of Finance*, **66**, 1047–1108.

— (2017). Macro-finance. *Review of Finance*, **21** (3), 945–985.

Cohen, R. B., Gompers, P. A. and Vuolteenaho, T. (2002). Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. *Journal of Financial Economics*, **66** (2), 409–462.

—, Polk, C. and Vuolteenaho, T. (2003). The value spread. *Journal of Finance*, **58**, 609–641.

Colacito, R., Riddiough, S. J. and Sarno, L. (2020a). Business cycles and currency returns. *Journal of Financial Economics*, **137** (3), 659–678.

—, — and — (2020b). Business cycles and currency returns. *Journal of Financial Economics*, **137** (3), 659–678.

Conrad, J., Gultekin, M. N. and Kaul, G. (1997). Profitability of short-term contrarian strategies: Implications for market efficiency. *Journal of Business & Economic Statistics*, **15** (3), 379–386.

Cooper, M. J., Gulen, H. and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *Journal of Finance*, **63** (4), 1609–1651.

—, — and Vassalou, M. (2001). Investing in size and book-to-market portfolios using information about the macroeconomy: Some new trading rules. *EFMA 2001 Lugano Meetings. Available at SSRN: 265181.*

Copeland, M. M. and Copeland, T. E. (1999). Market timing: Style and size rotation using the VIX. *Financial Analysts Journal*, **55** (2), 73–81.

Cornell, B. (1979). Asymmetric information and portfolio performance measurement. *Journal of Financial Economics*, **7** (4), 381–390.

Cornell, W. B. and Dietrich, J. K. (1978). The efficiency of the market for foreign exchange under floating exchange rates. *Review of Economics and Statistics*, **60**, 111–120.

Corte, P. D., Riddiough, S. J. and Sarno, L. (2016). Currency premia and global imbalances. *Review of Financial Studies*, **29**, 2161–2193.

Cotter, J., Eyiah-Donkor, E. and Potì, V. (2017). Predictability and diversification benefits of investing in commodity and currency futures. *International Review of Financial Analysis*, **50**, 52–66.

Cowles 3rd, A. (1933). Can stock market forecasters forecast? econometrica. pp. 309–324.

Cui, B. and Docherty, P. (2020). Stock price overreaction to esg controversies. *Available at SSRN 3559915.*

Curtis, A., Richardson, V. J. and Schmardebeck, R. (2016). Investor attention and the pricing of earnings news. *Handbook of Sentiment Analysis in Finance, Available at SSRN 2467243.*

Cutler, D. M., Poterba, J. M. and Summers, L. H. (1989). What moves stock prices? *Journal of Portfolio Management*, **15** (3), 4.

Da, Z. (2009). Cash flow, consumption risk, and the cross-section of stock returns. *Journal of Finance*, **64** (2), 923–956.

— and Warachka, M. C. (2009). Cashflow risk, systematic earnings revisions, and the cross-section of stock returns. *Journal of Financial Economics*, **94** (3), 448–468.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, **53** (6), 1839–1885.

—, — and Sun, L. (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies*, **33**, 1673–1736.

— and Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance*, **52** (1), 1–33.

— and — (2006). Market reactions to tangible and intangible information. *Journal of Finance*, **61** (4), 1605–1643.

Daniel, K. D., Hirshleifer, D. and Subrahmanyam, A. (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *Journal of Finance*, **56** (3), 921–965.

Das, S. R. and Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, **53** (9), 1375–1388.

Davis, A. K. and Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: Earnings press releases versus md&a. *Contemporary Accounting Research*, **29** (3), 804–837.

De Bondt, W. F. and Thaler, R. (1985a). Does the stock market overreact? *Journal of Finance*, **40** (3), 793–805.

— and — (1985b). Does the stock market overreact? *The Journal of Finance*, **40** (3), 793–805.

— and Thaler, R. H. (1987). Further evidence on investor overreaction and stock market seasonality. *The Journal of Finance*, **42** (3), 557–581.

De Franco, C. (2020). Esg controversies and their impact on performance. *The Journal of Investing*, **29** (2), 33–45.

De Groot, W., Pang, J. and Swinkels, L. (2012). The cross-section of stock returns in frontier emerging markets. *Journal of Empirical Finance*, **19**, 796–818.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, pp. 703–738.

De Rossi, G. and Brar, G. (2017). Quantamentals – crowding, herding and liquidity. *Working Paper.*

de Vincentiis, P. (2022). Do international investors care about esg news? *Qualitative Research in Financial Markets*, (ahead-of-print).

Dechow, P. M. and Schrand, C. M. (2004). Earnings quality. *Research Foundation of CFA Institute.*

—, Sloan, R. G. and Soliman, M. T. (2004). Implied equity duration: A new measure of equity risk. *Review of Accounting Studies*, **9**, 197–228.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, **22** (5), 1915–1953.

—, Martin-Utrera, A., Nogales, F. J. and Uppal, R. (2017). A portfolio perspective on the multitude of firm characteristics. *Available at SSRN: 2912819.*

—, —, — and — (2020a). A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies*, **33**, 2180–2222.

—, — and Uppal, R. (2020b). *What alleviates crowding in factor investing? Available at SSRN 3392875.*

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

Dichtl, H., Drobetz, W., Lohre, H. and Rother, C. (2021). Active factor completion strategies. *Journal of Portfolio Management*, **47**, 9–37.

—, —, —, — and Vosskamp, P. (2019). Optimal timing and tilting of equity factors. *Financial Analysts Journal*, **75**, 84–102.

Diether, K. B., Lee, K.-H. and Werner, I. M. (2009). Short-sale strategies and return predictability. *Review of Financial Studies*, **22**, 575–607.

Dimson, E. and Marsh, P. (1999). Murphy's law and market anomalies. *Journal of Portfolio Management*, **25**, 53–69.

—, — and Staunton, M. (2020). Divergent esg ratings. *The Journal of Portfolio Management*, **47** (1), 75–87.

Donangelo, A. (2020). Untangling the value premium with labor shares. *Review of Financial Studies*, **34**, 451–508.

Doskov, N. and Swinkels, L. (2015). Empirical evidence on the currency carry trade. *Journal of International Money and Finance*, **51**, 370–389.

El-Haj, M., Rayson, P., Walker, M., Young, S. and Simaki, V. (2019). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. *Journal of Business Finance and Accounting*, **46**, 265–306.

Elton, E. J., Gruber, M. J. and De Souza, A. (2019). Are passive funds really superior investments? an investor perspective. *Financial Analysts Journal*, **75**, 7–19.

Engel, C., Mark, N. C. and West, K. D. (2015). Factor model forecasts of exchange rates. *Econometric Reviews*, **34** (1-2), 32–55.

— and West, K. D. (2005). Exchange rates and fundamentals. *Journal of Political Economy*, **113** (3), 485–517.

Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements. In *Working paper*.

Erb, C. B. and Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, **62**, 69–97.

Fair Labor Association (2012). Independent investigation of apple supplier, foxconn. https://www.fairlabor.org/sites/default/files/documents/reports/foxconn_investigation_report.pdf.

Fairfield, P. M., Whisenant, J. S. and Yohn, T. L. (2003). Accrued earnings and growth: Implications for future profitability and market mispricing. *Accounting Review*, **78** (1), 353–371.

Fama, E. F. (1970). Efficient market hypothesis: A review of theory and empirical work. *Journal of Finance*, **25** (2), 28–30.

— (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, **14**, 319–338.

— (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, **49**, 283–306.

— and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, **22** (1), 3–25.

— and — (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, **25** (1), 23–49.

— and — (1992). The cross-section of expected stock returns. *Journal of Finance*, **47** (2), 427–465.

— and — (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33** (1), 3–56.

— and — (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, **50** (1), 131–155.

— and — (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, **51**, 55–84.

— and — (1998). Value versus growth: The international evidence. *Journal of Finance*, **53**, 1975–1999.

— and — (2006). Profitability, investment and average returns. *Journal of Financial Economics*, **82** (3), 491–518.

— and — (2008). Dissecting anomalies. *Journal of Finance*, **63**, 1653–1678.

— and — (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, **105**, 457–472.

— and — (2015). A five-factor asset pricing model. *Journal of Financial Economics*, **116**, 1–22.

— and — (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies*, **29** (1), 69–103.

— and — (2018). Choosing factors. *Journal of Financial Economics*, **128**, 234–252.

— and Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics*, **5** (2), 115–146.

Farhi, E., Fraiberger, S. P., Gabaix, X., Ranciere, R. and Verdelhan, A. (2009). Crash risk in currency markets. *National Bureau of Economic Research.*

Feng, G., Giglio, S. and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, **75**, 1327–1370.

—, Polson, N. G. and Xu, J. (2019). Deep learning in characteristics-sorted factor models. p. 23527.

Fernandez-Perez, A., Frijns, B., Fuertes, A.-M. and Miffre, J. (2018). The skewness of commodity futures returns. *Journal of Banking and Finance*, **86**, 143–158.

Ferson, W. E., Sarkissian, S. and Simin, T. (2003). Spurious regressions in financial economics? *Journal of Finance*, **58** (4), 1393–1413.

—, — and Simin, T. T. (1999). The alpha factor asset pricing model: A parable. *Journal of Financial Markets*, **2**, 49–68.

Fisher, I. (1930). The theory of interest. *New York*, **43**.

Fombrun, C. and Shanley, M. (1990). What's in a name? reputation building and corporate strategy. *Academy of management Journal*, **33** (2), 233–258.

Frank, M. Z. and Sanati, A. (2018). How does the stock market absorb shocks? *Journal of Financial Economics*, **129** (1), 136–153.

Frazzini, A., Israel, R. and Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies. *Fama-Miller Working Paper*, pp. 14–05.

—, — and — (2014). Trading costs of asset pricing anomalies. working paper.

—, — and — (2018). *Trading costs. Available at SSRN 3229719.*

Freyberger, J., Neuhierl, A. and Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, **33**, 2326–2377.

Frooman, J. (1997). Socially irresponsible and illegal behavior and shareholder wealth: A meta-analysis of event studies. *Business & society*, **36** (3), 221–249.

Froot, K., Kang, N., Ozik, G. and Sadka, R. (2017). What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics*, **125** (1), 143–162.

Froot, K. A. and Thaler, R. H. (1990). Anomalies: Foreign exchange. *Journal of Economic Perspectives*, **4**, 179–192.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics*, **127**, 645–700.

Galati, G. and Melvin, M. (2004). *Why has FX trading surged? Explaining the 2004 triennial survey*. Tech. rep., BIS Quarterly Review, December 2004.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, **68** (3), 1267–1300.

Gebhardt, W. R., Hvidkjaer, S. and Swaminathan, B. (2005). Stock and bond market interaction: Does momentum spill over. *Journal of Financial Economics*, **75**, 651–690.

Gentzkow, M., Kelly, B. and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, **57** (3), 535–74.

Giamouridis, D., Sakkas, A. and Tessaromatis, N. (2017). Dynamic asset allocation with liabilities. *European Financial Management*, **23**, 254–291.

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, **57**, 1121–1152.

Gollier, C. (2001). Wealth inequality and asset pricing. *Review of Economic Studies*, **68**, 181–203.

Gordon, M. J. (1959). Dividends, earnings, and stock prices. *Review of Economics and Statistics*, **41** (2), 99–105.

— (1962). *The investment, financing, and valuation of the corporation*. R.D. Irwin.

Gormsen, N. J. and Lazarus, E. (2020). Duration-driven returns. *Available at SSRN*, **3359027**.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, **49** (5), 639–654.

Graham, B. and Dodd, D. L. (1934). *Security analysis: Principles and techniques*. McGraw-Hill.

Grammig, J. and Jank, S. (2016). Creative destruction and asset prices. *Forthcoming: Journal of Financial and Quantitative Analysis*.

Grant, J., Ahmerkamp, J. D. and Kosowski, R. (2012). *Predictability in carry and momentum strategies across asset classes. Available at SSRN 2024056.*

Green, J., Hand, J. R. and Soliman, M. T. (2011). Going, going, gone? The apparent demise of the accruals anomaly. *Management Science*, **57** (5), 797–816.

—, Hand, J. R. M. and Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies*, **30**, 4389–4436.

Greenaway-McGrevy, R., Mark, N. C., Sul, D. and Wu, J.-L. (2018). Identifying exchange rate common factors. *International Economic Review*, **59** (4), 2193–2218.

Griffin, J. M., Ji, X. and Martin, J. S. (2003). Momentum investing and business cycle risk: Evidence from pole to pole. *Journal of Finance*, **58**, 2515–2547.

—, Kelly, P. J. and Nardari, F. (2010). Do market efficiency measures yield correct inferences? a comparison of developed and emerging markets. *Review of Financial Studies*, **23**, 3225–3277.

Grinold, R. C. and Kahn, R. N. (2000). Active portfolio management.

Griveau-Billion, T., Richard, J.-C. and Roncalli, T. (2013). A fast algorithm for computing high-dimensional risk parity portfolios. *Available at SSRN: 2325255*.

Gu, S., Kelly, B. and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, **33**, 2223–2273.

Guo, H. and Savickas, R. (2006). Idiosyncratic volatility, stock market volatility, and expected stock returns. *Journal of Business & Economic Statistics*, **24** (1), 43–56.

—, —, Wang, Z. and Yang, J. (2009). Is the value premium a proxy for time-varying investment opportunities? some time-series evidence. *Journal of Financial and Quantitative Analysis*, pp. 133–154.

Gupta, T. and Kelly, B. (2019). Factor momentum everywhere. *Journal of Portfolio Management*, **45**, 13–36.

Gurun, U. G. and Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance*, **67** (2), 561–598.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D. and Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

Hamdan, R., Pavlowsky, F., Roncalli, T. and Zheng, B. (2016). A primer on alternative risk premia. *Working Paper. Available at http://www.thierry-roncalli.com/download/Alternative_Risk_Premia.pdf*.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics*, **100** (5), 831–843.

Hammerschmid, R. and Lohre, H. (2018). Regime shifts and stock return predictability. *International Review of Economics & Finance*, **56**, 138–160.

— and — (2020). Integrating time series and crosssectional signals for optimal commodity portfolios. *Available at SSRN*, **3504394**.

Han, Y. and Lesmond, D. (2011). Liquidity biases and the pricing of cross-sectional idiosyncratic volatility. *Review of Financial Studies*, **24** (5), 1590–1629.

Hanley, K. W. and Hoberg, G. (2010). The information content of ipo prospectuses. *The Review of Financial Studies*, **23** (7), 2821–2864.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.

— and Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, **88**, 829–853.

Hansen, S., McMahon, M. and Prat, A. (2017). Transparency and deliberation within the fomc: A computational linguistics approach. *Quarterly Journal of Economics*, **133**, 801–870.

Hanson, S. and Sunderam, A. (2013). The growth and limits of arbitrage: Evidence from short interest. *Review of Financial Studies*, **27**, 1238–1286.

Härdle, W. K., Hoffmann, L. and Moro, R. (2011). Learning machines supporting bankruptcy prediction. In *Statistical Tools for Finance and Insurance*, Springer, pp. 225–250.

Hart, R. P. (1984). Systematic analysis of political discourse: The development of diction. *Political communication yearbook*, **1**, 97–134.

Harvey, C. R. and Liu, Y. (2019). A census of the factor zoo. *Available at SSRN 3341728*.

—, — and Zhu, H. (2016a). ... and the cross-section of expected returns. *Review of Financial Studies*, **29**, 5–68.

—, — and — (2016b). . . . and the cross-section of expected returns. *Review of Financial Studies*, **29**, 5–68.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *Journal of Finance*, **64**, 1445–1477.

Hatchondo, J. C. (2008). A quantitative study of the role of wealth inequality on asset prices. *FRB Richmond Economic Quarterly*, **94**, 73–96.

Haugen, R. A. (1995). The new finance: The case against efficient markets.

— and Baker, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *Journal of Portfolio Management*, **17** (3), 35–40.

— and — (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, **41**, 401–439.

— and — (2012). Low risk stocks outperform within all observable markets of the world.

— and Heins, A. J. (1972). On the evidence supporting the existence of risk premiums in the capital market.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication*, **45** (4), 363–407.

Hillier, D., Grinblatt, M. and Titman, S. (2011). *Financial markets and corporate strategy*. McGraw–Hill.

Hirshleifer, D., Hou, K. and Teoh, S. H. (2009). Accruals, cash flows, and aggregate stock returns. *Journal of Financial Economics*, **91** (3), 389–406.

—, — and — (2012). The accrual anomaly: Risk or mispricing? *Management Science*, **58** (2), 320–335.

Hodges, P., Hogan, K., Peterson, J. R. and Ang, A. (2017). Factor timing with cross-sectional and time-series predictors. *Journal of Portfolio Management*, **44** (1), 30–43.

Hoffmann, M. and Studer-Suter, R. (2017). Systematic consumption risk in currency returns. *Journal of International Money and Finance*, **74**, 187–208.

Hong, H., Li, F., Ni, S., Scheinkman, J. and Yan, P. (2016). Days to cover and stock returns. *Working Paper*.

— and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, **54** (6), 2143–2184.

Hottinga, J., van Leeuwen, E. and van Ijserloo, J. (2001). Successful factors to select outperforming corporate bonds. *Journal of Portfolio Management*, **28**, 88–101.

Hou, K., Karolyi, G. A. and Kho, B.-C. (2011). What factors drive global stock returns? *Review of Financial Studies*, **24** (8), 2527–2574.

—, Mo, H., Xue, C. and Zhang, L. (2021). An augmented q-factor model with expected growth. *Review of Finance*, **25** (1), 1–41.

— and Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies*, **18** (3), 981–1020.

—, Xue, C. and Zhang, L. (2012). Digesting anomalies: An investment approach. *National Bureau of Economic Research, Working Paper, 18435*.

—, — and — (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, **28**, 650–705.

—, — and — (2020). Replicating anomalies. *Review of Financial Studies*, **33**, 2019–2133.

hsani, S. and Linnainmaa, J. T. (2019). Factor momentum and the momentum factor. *National Bureau of Economic Research*.

Hsu, J., Kalesnik, V. and Kose, E. (2019). What is quality? *Financial Analysts Journal*, **75**, 44–61.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177.

Huang, A. H., Zang, A. Y. and Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, **89** (6), 2151–2180.

Huang, C.-f. and Litzenberger, R. H. (1988). *Foundations for Financial Economics*. Elsevier.

Huang, E., Liu, V., Ma, L. and Osiol, J. (2010). Methods in dynamic weighting. *Capital IQ Working Paper*.

Huang, K., Altosaar, J. and Ranganath, R. (2019a). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.

—, — and — (2019b). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.

Ilmanen, A. (2011). *Expected returns: An investor's guide to harvesting market rewards*. John Wiley & Sons.

— and Kizer, J. (2012). The death of diversification has been greatly exaggerated. *Journal of Portfolio Management*, **38** (3), 15–27.

Israel, R. (2014). Tell it like it is: Disclosed risks and factor portfolios.

—, Palhares, D. and Richardson, S. A. (2018). Common factors in corporate 65 bond returns. *Journal of Investment Management*, **16**, 17–46.

Jacobs, H. and Müller, S. (2016). Anomalies across the globe: Once public, no longer existent? *Working paper. Available at SSRN: 2816490*.

— and Müller, S. (2020). Anomalies across the globe: Once public, no longer existent. *Journal of Financial Economics*, **135**, 213–230.

Jaffe, J., Keim, D. B. and Westerfield, R. (1989). Earnings yields, market values, and stock returns. *Journal of Finance*, **44** (1), 135–148.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, **58** (4), 1651–1684.

— and Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, **51** (1), 3–53.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, **45** (3), 881–898.

— and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, **48** (1), 65–91.

— and — (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance*, **56** (2), 699–720.

— and Wu, D. (2013). Word power: A new approach for content analysis. *Journal of financial economics*, **110** (3), 712–729.

Jensen, M. C., Black, F. and Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests. *Studies in the Theory of Capital Markets. Praeger Publishers.*

Jensen, T. I., Kelly, B. T. and Pedersen, L. H. (2021). Is there a replication crisis in finance?

Jordà, Ò. and Taylor, A. M. (2012). The carry trade and fundamentals: Nothing to fear but FEER itself. *Journal of International Economics*, **88** (1), 74–90.

Jorion, P. *et al.* (2007). *Financial risk manager handbook*, vol. 406. John Wiley & Sons.

Jostova, G., Nikolova, S., Philipov, A. and Stahel, C. W. (2013). Momentum in corporate bond returns. *Review of Financial Studies*, **26**, 1649–1693.

Jotikasthira, C., Lundblad, C. and Ramadorai, T. (2012). Asset fire sales and purchases and the international transmission of funding shocks. *Journal of Finance*, **67**, 2015–2050.

Jurek, J. W. (2014). Crash-neutral currency carry trades. *Journal of Financial Economics*, **113**, 325–347.

Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns: An asset-allocation perspective. *Journal of Finance*, **51** (2), 385–424.

Kao, D.-L. and Shumaker, R. D. (1999). Equity style timing. *Financial Analysts Journal*, **55** (1), 37–48.

Kapadia, N. (2011). Tracking down distress risk. *Journal of Financial Economics*, **102**, 167–182.

Karolyi, A. (2016). Home bias, an academic puzzle. *Review of Finance*, **20**, 2049–2078.

Kearney, C. and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, **33**, 171–185.

Keim, D. B. and Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics*, **17** (2), 357–390.

Kelly, B. T., Palhares, D. and Pruitt, S. (2020). Modeling corporate bond returns.

—, Pruitt, S. and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, **134**, 501–524.

Khan, M. (2008). Are accruals mispriced? Evidence from tests of an intertemporal capital asset pricing model. *Journal of Accounting and Economics*, **45** (1), 55–77.

Khandani, A. and Lo, A. (2007). What happened to the quants in August 2007? *Journal of Investment Management*, **5** (4), 29–78.

Khandani, A. E. and Lo, A. W. (2011). What happened to the quants in august 2007? evidence from factors and transactions data. *Journal of Financial Markets*, **14** (1), 1–46.

Kim, D. and Lee, B.-J. (2019). Shorting costs and profitability of long-short strategies.

Kimball, M. S., Sahm, C. R. and Shapiro, M. D. (2008). Imputing risk tolerance from survey responses. *Journal of the American Statistical Association*, **103** (483), 1028–1038.

Kogan, L., Tian, M. *et al.* (2013). Firm characteristics and empirical factor models: A data-mining experiment. *International Finance Discussion Papers. Available at http://www.federalreserve.gov/pubs/ifdp/2012/1070/ifdp1070.pdf.*

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H. and Vrugt, E. B. (2013). Carry.

—, —, — and — (2018). **127**, 197–225.

Kong, A., Rapach, D. E., Strauss, J. K., Tu, J. and Zhou, G. (2009). How predictable are components of the aggregate market portfolio. *Research Collection Lee Kong Chian School of Business. Available at https://core.ac.uk/download/pdf/13243655.pdf.*

Korajczyk, R. A. and Sadka, R. (2004). Are momentum profits robust to trading costs. *Journal of Finance*, **59**, 1039–1082.

Kothari, S. P., Li, X. and Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, **84** (5), 1639–1670.

Kothe, J., Lohre, H. and Rother, C. (2021). Rates factors and global asset allocation. *Journal of Fixed Income*, **30**, 6–25.

Kotsantonis, S. and Serafeim, G. (2019). Four things no one will tell you about esg data. *Journal of Applied Corporate Finance*, **31** (2), 50–58.

Kozak, S., Nagel, S. and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, **135**, 271–292.

Kremer, D. (2016). Das fiese Fonds-Spiel. *Newspaper article, Frankfurter Allgemeine Zeitung, 27 July 2016. Available at http://www.faz.net/aktuell/finanzen/fonds-mehr/mogelpackung-das-fiese-fonds-spiel-14354724-p3.html.*

Kroencke, T. A., Schindler, F. and Schrimpf, A. (2013). International diversification benefits with foreign exchange investment styles. *Review of Finance*, **18** (5), 1847–1883.

—, — and — (2014). International diversification benefits with foreign exchange investment styles. *Review of Finance*, **18**, 1847–1883.

Krueger, P., Sautner, Z. and Starks, L. T. (2020). The importance of climate risks for institutional investors. *Review of Financial Studies*, **33**, 1067–1111.

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of financial economics*, **115** (2), 304–329.

La Porta, R. (1996). Expectations and the cross-section of stock returns. *Journal of Finance*, **51** (5), 1715–1742.

—, Lakonishok, J., Shleifer, A. and Vishny, R. (1997). Good news for value stocks: Further evidence on market efficiency. *Journal of Finance*, **52** (2), 859–874.

Lakonishok, J., Shleifer, A. and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, **49**, 1541–1578.

Lam, S. S. and Zhang, W. (2014). Does policy uncertainty matter for international equity markets?

Lamont, O. A. (2000). Investment plans and stock returns. *Journal of Finance*, **55** (6), 2719–2745.

Lane, G. C. (1984). Lane's stochastics. *Technical Analysis of Stocks and Commodities*, **2** (3), 80.

Larcker, D. F., Pomorski, L., Tayan, B. and Watts, E. M. (2022). Esg ratings: A compass without direction. *Rock Center for Corporate Governance at Stanford University Working Paper Forthcoming*.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H. and Kang, J. (2020). Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, **36** (4), 1234–1240.

Lee, W. (2011). Risk-based asset allocation: A new answer to an old question? *Journal of Portfolio Management*, **37** (4), 11.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics*, **105** (1), 1–28.

Leippold, M. and Lohre, H. (2012a). Data snooping and the global accrual anomaly. *Applied Financial Economics*, **22**, 509–535.

— and — (2012b). International price and earnings momentum. *European Journal of Finance*, **18**, 535–573.

Leitch, G. and Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review*, pp. 580–590.

Lesmond, D. A., Schill, M. J. and Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics*, **71**, 349–380.

Lettau, M., Ludvigson, S. C. and Wachter, J. A. (2008). The declining equity premium: What role does macroeconomic risk play? *Review of Financial Studies*, **21**, 1653–1687.

— and Wachter, J. A. (2007). Why is long-horizon equity less risky? a duration-based explanation of the value premium. *Journal of Finance*, **62**, 55–92.

— and — (2011). The term structures of equity and interest rates. *Journal of Financial Economics*, **101**, 90–113.

Leung, E., Lohre, H., Mischlich, D., Shea, Y. and Stroh, M. (2021). The promises and pitfalls of machine learning for predicting stock returns. *Journal of Financial Data Science*.

Levis, M. and Liodakis, M. (1999). The profitability of style rotation strategies in the United Kingdom. *Journal of Portfolio Management*, **26** (1), 73–86.

Lewellen, J. (2002). Momentum and autocorrelation in stock returns. *The Review of Financial Studies*, **15** (2), 533–564.

— and Nagel, S. (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of financial economics*, **82** (2), 289–314.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, **45**, 221–247.

— (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, **29** (1), 143–165.

Liang, H. and Renneboog, L. (2020). Corporate social responsibility and sustainable finance: A review of the literature. *Oxford Research Encyclopedia of Economics and Finance*.

Liew, J. and Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics*, **57** (2), 221–245.

Linnainmaa, J. T. and Roberts, M. R. (2018). The history of the cross-section of stock returns. *Review of Financial Studies*, **31**, 2606–2649.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, **47** (1), 13–37.

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), 163–195.

Liu, L. X. and Zhang, L. (2008). Momentum profits, factor pricing, and macroeconomic risk. *Review of Financial Studies*, **21** (6), 2417–2448.

Liu, N., Hu, Q., Xu, H., Xu, X. and Chen, M. (2021a). Med-bert: A pretraining framework for medical records named entity recognition. *IEEE Transactions on Industrial Informatics*, **18** (8), 5600–5608.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of Financial Economics*, **82** (3), 631–671.

Liu, Z., Huang, D., Huang, K., Li, Z. and Zhao, J. (2021b). Finbert: A pre-trained financial language representation model for financial text mining. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 4513–4519.

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, **1**, 41–66.

— and — (1990a). Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, **3**, 431–467.

— and — (1990b). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, **3** (2), 175–205.

Lohre, H., Papenbrock, J. and Poonia, M. (2014). The use of correlation networks in parametric portfolio policies. *Available at SSRN: 2505732*.

Lopez-Lira, A. (2019). Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns.

Lou, D. and Polk, C. (2013). *Comomentum: Inferring arbitrage activity from return correlations*. Paul Woolley Centre for the Study of Capital Market Dysfunctionality.

— and — (2014a). Comomentum: Inferring arbitrage activity from return correlations.

— and — (2014b). Comomentum: Inferring arbitrage activity from return correlations. *Working Paper*.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, **66** (1), 35–65.

— and — (2015). The use of word lists in textual analysis. *Journal of Behavioral Finance*, **16** (1), 1–11.

— and — (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, **54** (4), 1187–1230.

— and — (2020). Textual analysis in finance. *Annual Review of Financial Economics*, **12**, 357–375.

— and Ritter, J. R. (1995). The new issues puzzle. *Journal of Finance*, **50** (1), 23–51.

— and — (1997). The operating performance of firms conducting seasoned equity offerings. *Journal of Finance*, **52** (5), 1823–1850.

Lucas, A., van Dijk, R. and Kloek, T. (2002). Stock selection, style rotation, and risk. *Journal of Empirical Finance*, **9** (1), 1–34.

Ludvigson, S. and Ng, S. (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics*, **83** (1), 171–222.

— and — (2009). Macro factors in bond risk premia. *Review of Financial Studies*, **22** (12), 5027–5067.

Lustig, H., Roussanov, N. and Verdelhan, A. (2011). Common risk factors in currency markets. *Review of Financial Studies*, **24**, 3731–3777.

—, — and — (2014). Countercyclical currency risk premia. *Journal of Financial Economics*, **111** (3), 527–553.

— and Verdelhan, A. (2007). The cross section of foreign currency risk premia and consumption growth risk. *American Economic Review*, **97**, 89–117.

Maillard, S., Roncalli, T. and Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management*, **36** (4), 60–70.

Mandelker, G. (1974). Risk and return: The case of merging firms. *Journal of financial economics*, **1** (4), 303–335.

Manela, A. and Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, **123** (1), 137–162.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *European Physical Journal B—Condensed Matter and Complex Systems*, **11** (1), 193–197.

Mark, N. C. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *The American Economic Review*, pp. 201–218.

— (2009). Changing monetary policy rules, learning, and real exchange rate dynamics. *Journal of Money, Credit and Banking*, **41** (6), 1047–1070.

— and Sul, D. (2001). Nominal exchange rates and monetary fundamentals: evidence from a small post-bretton woods panel. *Journal of International Economics*, **53** (1), 29–52.

Martellini, L. and Milhau, V. (2015). Factor investing: A welfare-improving new investment paradigm or yet another marketing fad. *EDHEC-Risk Institute Publication. Available at http://professoral.edhec.com/_medias/fichier/edhec-publication-factor-investing-a-welfare_1441888737064-pdf.*

Martens, M., Beekhuizen, P., Duyvesteyn, J. and Zomerdijk, C. (2019). Carry investing on the yield curve. *Financial Analysts Journal*, **75**, 51–63.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability. *Journal of Finance*, **71**, 5–32.

Mech, T. S. (1993). Portfolio return autocorrelation. *Journal of Financial Economics*, **34** (3), 307–344.

Medvedev, A. and Vaucher, B. (2017). Efficient integration of risk premia exposures into equity portfolios. *Journal of Asset Management*, **18** (7), 538–546.

Meese, R. A. and Rogoff, K. (1983). Do they fit out of sample? *Journal of international economics*, **14**, 3–24.

Menkhoff, L., Sarno, L., Schmeling, M. and Schrimpf, A. (2012a). Carry trades and global foreign exchange volatility. *Journal of Finance*, **67**, 681–718.

—, —, — and — (2012b). Currency momentum strategies. *Journal of Financial Economics*, **106**, 660–684.

—, —, — and — (2017). Currency value. *Review of Financial Studies*, **30**, 416–441.

Merton, R. C. (1973a). An intertemporal capital asset pricing model. *Econometrica*, **42**, 867–887.

— (1973b). An intertemporal capital asset pricing model. *Econometrica*, **41** (5), 867–887.

— (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, **8** (4), 323–361.

— (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance*, **42** (3), 483–510.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is "optimized" optimal? *Financial Analysts Journal*, **45** (1), 31–42.

Miffre, J. (2016). Long-short commodity investing: A review of the literature. *Journal of Commodity Markets*, **1**, 3–13.

— and Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking and Finance*, **3**, 1863–1886.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *Journal of Finance*, **32**, 1151–1168.

Mitton, T. and Vorkink, K. (2007). Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies*, **20** (4), 1255–1288.

Molodtsova, T., Nikolsko-Rzhevskyy, A. and Papell, D. H. (2008). Taylor rules with real-time data: A tale of two countries and one exchange rate. *Journal of Monetary Economics*, **55**, S63–S79.

— and Papell, D. H. (2009). Out-of-sample exchange rate predictability with taylor rule fundamentals. *Journal of International Economics*, **77** (2), 167–180.

Montagu, C., Krause, H., Jalan, R., Murray, J., Chew, D. and Yusuf, Z. (2016). Searching for alpha: Dynamic style weighting. *Citigroup Working Paper*.

Moskowitz, T. J. and Grinblatt, M. (1999). Do industries explain momentum? *Journal of Finance*, **54** (4), 1249–1290.

Mossin, J. (1966). Wages, profits, and the dynamics of growth. *Quarterly Journal of Economics*, **80** (3), 376–399.

Muijsson, C., Fishwick, E. and Satchell, S. (2014). Taking the art out of smart beta.

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

Nakamura, E., Steinsson, J., Barro, R. and Ursúa, J. (2009). Crises and 58 recoveries in an empirical model of consumption disasters. working paper.

NEELY, C. J., RAPACH, D. E., TU, J. and ZHOU, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, **60** (7), 1772–1791.

NELSON, C. R. (1976). Inflation and rates of return on common stocks. *Journal of Finance*, **31** (2), 471–483.

NEWEY, W. K. and WEST, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, **28** (3), 777–787.

NOVY-MARX, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, **108** (1), 1–28.

— (2014). The quality dimension of value investing.

— and VELIKOV, M. (2016a). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, **29**, 104–147.

— and — (2016b). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, **29** (1), 104–147.

OKUNEV, J. and WHITE, D. (2003). Do momentum-based strategies still work in foreign currency markets? *Journal of Financial and Quantitative Analysis*, **38** (2), 425–447.

ORLITZKY, M. (2013). Corporate social responsibility, noise, and stock market volatility. *Academy of Management Perspectives*, **27** (3), 238–254.

OU, J. A. and PENMAN, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, **11** (4), 295–329.

PARK, H. (2019). Intangible assets and the book-to-market effect. *European Financial Management*, **25**, 207–236.

PARKER, A. S., HAYES, B. T., ORTEGA, A. and NAHA, C. (2016). The 2016 factor timing quantroversy. *Morgan Stanley US Quant Research: April 27, 2016.*

PÁSTOR, L. and STAMBAUGH, R. F. (2009). Predictive systems: Living with imperfect predictors. *Journal of Finance*, **64** (4), 1583–1628.

— and VERONESI, P. (2020). Political cycles and stock returns. *Journal of Political Economy*, **128** (11), 4011–4045.

PATELL, J. M. (1976). Corporate forecasts of earnings per share and stock price behavior: Empirical test. *Journal of accounting research*, pp. 246–276.

PATTON, A. J. and TIMMERMANN, A. (2010). *Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts*, **98**, 605–625.

— and WELLER, B. M. (2020). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Eonomics*, **137**, 515–549.

Perkiss, S., Bernardi, C., Dumay, J. and Haslam, J. (2021). A sticky chocolate problem: Impression management and counter accounts in the shaping of corporate image. *Critical Perspectives on Accounting*, **81**, 102229.

Pesaran, M. H. and Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, **9** (5), 495–510.

Petkova, R. (2006). Do the Fama–French factors proxy for innovations in predictive variables? *Journal of Finance*, **61** (2), 581–612.

— and Zhang, L. (2005). Is value riskier than growth? *Journal of Financial Economics*, **78** (1), 187–202.

Pfaff, B. (2013). *Financial Risk Modelling and Portfolio Optimization with R*. John Wiley and Sons.

Plyakha, Y., Uppal, R. and Vilkov, G. (2014). Equal or value weighting? Implications for asset-pricing tests. *Working paper. Available at SSRN: 1787045*.

Podkaminer, E. L. (2013). Risk factors as building blocks for portfolio diversification: The chemistry of asset allocation. *Investment Risk and Performance Newsletter*, **2013** (1).

Pojarliev, M. and Levich, R. M. (2008). Do professional currency managers beat the benchmark? *Financial Analysts Journal*, **64**, 18–32.

— and — (2010). Detecting crowded trades in currency funds. *Working Paper*.

Pollet, J. and Wilson, M. (2010). Average correlation and stock market returns. *Journal of Financial Economics*, **96** (3), 364–380.

Pontiff, J. and Woodgate, A. (2008). Share issuance and cross-sectional returns. *Journal of Finance*, **63** (2), 921–945.

Pozzi, F., Di Matteo, T. and Aste, T. (2013). Spread of risk across financial markets: Better to invest in the peripheries. *Scientific Reports*, **3**, 1–7.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal*, **36** (6), 1389–1401.

Pukthuanthong, K., Roll, R. and Subrahmanyam, A. (2019). A protocol for factor identification. *Review of Financial Studies*, **32**, 1573–1607.

Qian, E. (2011). Risk parity and diversification. *Journal of Investing*, **20**, 119–127.

Qudar, M. M. A. and Mago, V. (2020). Tweetbert: a pretrained language representation model for twitter text analysis. *arXiv preprint arXiv:2010.11091*.

Raghunandan, A. and Rajgopal, S. (2022). Do esg funds make stakeholder-friendly investments? *Review of Accounting Studies*, **27** (3), 822–863.

Ranganathan, A., Lohre, H., Nolte, S. and Braham, H. (2023). An integrated approach to currency factor management. *Forthcoming in the Journal of Systematic Investing*.

RASMY, L., XIANG, Y., XIE, Z., TAO, C. and ZHI, D. (2021). Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, **4** (1), 86.

RATCLIFFE, R., MIRANDA, P. and ANG, A. (2017). Capacity of smart beta strategies from a transaction cost perspective. *Journal of Index Investing*, **8**, 39–50.

RICHARDSON, S. A., SLOAN, R. G., SOLIMAN, M. T. and TUNA, I. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, **39** (3), 437–485.

RIETZ, T. A. (1988). The equity risk premium a solution. *Journal of Monetary Economics*, **22**, 117–131.

RITTER, J. R. (1991). The long-run performance of initial public offerings. *Journal of Finance*, **46** (1), 3–27.

ROLL, R. (1977). A critique of the asset pricing theory's tests part I: On past and potential testability of the theory. *Journal of Financial Economics*, **4** (2), 129–176.

— (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, **39** (4), 1127–1139.

ROMANO, J. P. and WOLF, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, **73**, 1237–1282.

— and — (2013). Testing for monotonicity in expected asset returns. *Journal of Empirical Finance*, **23**, 93–116.

ROSENBERG, B., MARATHE, V. *et al.* (1976). *Common factors in security returns: Microeconomic determinants and macroeconomic correlates*. Tech. rep., University of California at Berkeley.

—, REID, K. and LANSTEIN, R. (1985). Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, **11** (3), 9–16.

ROSS, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13** (3), 341–360.

ROSSI, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, **51** (4), 1063–1119.

ROUWENHORST, K. G. (1998). International momentum strategies. *Journal of Finance*, **53**, 267–284.

— (1999). Local return factors and turnover in emerging stock markets. *Journal of Finance*, **54**, 1439–1464.

ROZEFF, M. S. (1984). Dividend yields are equity risk premiums. *Journal of Portfolio Management*, pp. 68–75.

SADKA, R. (2006). Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics*, **80** (2), 309–349.

Sakkas, A. and Tessaromatis, N. (2020). Factor based commodity investing. *Journal of Banking & Finance*, **115**, 105807.

Sanh, V., Debut, L., Chaumond, J. and Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Scherer, B. (2007). *Portfolio Construction and Risk Budgeting*. Riskbooks, 3rd edn.

Schröder, D. and Esterer, F. (2016). A new measure of equity and cash flow duration: The duration-based explanation of the value premium revisited. *Journal of Money, Credit and Banking*, **48** (5), 857–900.

Schumpeter, J. A. (1961). *Capitalism, socialism and democracy*. Allen & Unwin.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, **1**, 939–974.

Serafeim, G. and Yoon, A. (2022). Stock price reactions to esg news: The role of esg ratings and disagreement. *Review of Accounting Studies*, pp. 1–31.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19** (3), 425–442.

Shiller, R. J. (2017). Narrative economics. *American economic review*, **107** (4), 967–1004.

— (2019). *Narrative economics*. Princeton University Press Princeton.

— *et al.* (1981). Alternative tests of rational expectations models: The case of the term structure. *Journal of Econometrics*, **16** (1), 71–87.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *Journal of Finance*, **52**, 35–55.

Sias, T., Richard, Harry and Zykaj, B. (2016). Hedge fund crowds and mispricing. *Management Science*, **62**, 764–784.

Simpson, C., Rathi, A. and Kishan, S. (2021). The esg mirage. *Bloomberg Businessweek*.

Siriwardane, E. (2013). The pricing of disaster risk.

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, **71** (3), 289–315.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The Accounting Review*, **83** (3), 823–853.

Spinu, F. (2013). An algorithm for computing risk parity weights. *OMERS Capital working paper. Available at SSRN: 2297383*.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, **54** (3), 375–421.

—, Yu, J. and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, **104**, 288–302.

— and Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, **30**, 1270–1315.

Stein, J. C. (2009a). Presidential address: Sophisticated investors and market efficiency. *Journal of Finance*, **64**, 1517–1548.

— (2009b). Sophisticated investors and market efficiency. *Journal of Finance*, **64**, 1517–1548.

Stone, P. J., Dunphy, D. C. and Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Sullivan, R., Timmermann, A. and White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, **54**, 1647–1691.

—, — and — (2001). Dangers of data-mining: The case of calendar effects in stock returns. *Journal of Econometrics*, **105**, 249–286.

Tang, D. Y., Yan, J. and Yao, C. Y. (2022). The determinants of esg ratings: Rater ownership matters. In *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, **39**, 195–214.

Teoh, S. H. and Wong, T. J. (2002). Why new issues and high-accrual firms underperform: The role of analysts' credulity. *Review of Financial Studies*, **15** (3), 869–900.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, **62** (3), 1139–1168.

—, Saar-Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, **63** (3), 1437–1467.

Titman, S., Wei, K. J. and Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, **39** (04), 677–700.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies*, **25** (2), 65–86.

Treynor, J. L. (1961). Market value, time, and risk. *Unpublished manuscript*.

Van der Hart, J., Slagter, E. and Van Dijk, D. (2003). Stock selection strategies in emerging markets. *Journal of Empirical Finance*, **10**, 105–132.

Van Gelderen, E. and Huij, J. (2014). Academic knowledge dissemination in the mutual fund industry: can mutual funds successfully adopt factor investing strategies. *Journal of Portfolio Management*, **40**, 157–167.

—, — and Kyosev, G. (2019). Factor investing from concept to implementation. *Journal of Portfolio Management*, **45**, 125–140.

Vanguard Investment Strategy Group, . (2019). Cycle-based factor timing: Clear in hindsight but hazy ahead. valley forge.

Vassalou, M. and Xing, Y. (2004). Default risk in equity returns. *Journal of Finance*, **59** (2), 831–868.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, **30**.

Verdelhan, A. (2018). The share of systematic variation in bilateral exchange rates. *Journal of Finance*, **73** (1), 375–418.

Wahal, S. (2019). The profitability and investment premium: Pre-1963 evidence. *Journal of Financial Economics*, **131**, 362–377.

Wang, H. and Yu, J. (2013). Dissecting the profitability premium. *AFA 2013 San Diego Meetings Paper. Available at SSRN: 1711856*.

Wang, J. and Wu, J. J. (2012). The Taylor rule and forecast intervals for exchange rates. *Journal of Money, Credit and banking*, **44** (1), 103–144.

Warstadt, A., Cao, Y., Grosu, I., Peng, W., Blix, H., Nie, Y., Alsop, A., Bordia, S., Liu, H., Parrish, A. *et al.* (2019). Investigating bert's knowledge of language: five analysis methods with npis. *arXiv preprint arXiv:1909.02597*.

Webersinke, N., Kraus, M., Bingler, J. A. and Leippold, M. (2021). Climatebert: A pretrained language model for climate-related text. *arXiv preprint arXiv:2110.12010*.

Wei, Z., Chia, C.-P. and Katiyar, S. (2015). Harvesting equity yield. Understanding factor investing. *MSCI Research Insight Working Paper. Available at https://www.msci.com/documents/10199/e7a57428-106e-4a33-90e4-6c4c596378d3*.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21** (4), 1455–1508.

Williams, J. B. (1938). *The theory of investment value*, vol. 36. JSTOR.

Xu, Y. (2004). Small levels of predictability and large economic gains. *Journal of Empirical Finance*, **11** (2), 247–275.

Yang, C. and Zhou, L. (2016). Individual stock crowded trades, individual stock investor sentiment and excess returns. *The North American Journal of Economics and Finance*, **38**, 39–53.

Yang, Y., Uy, M. C. S. and Huang, A. (2020). Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.

Zaremba, A. (2016). Risk-based explanation for the country-level size and value effects. *Finance Research Letters*, **18**, 226–233.

Zhang, L. (2005). The value premium. *Journal of Finance*, **60** (1), 67–103.

Zhou, G. (2010). How much stock return predictability can we expect from an asset pricing model? *Economics Letters*, **108** (2), 184–186.