Macroeconomic and geopolitical impact on asset returns and risk premia

Three essays on asset pricing

by

Matthias Apel
Matriculation number 1456123

Inaugural Dissertation
in partial fulfillment of the requirements for the degree of
Doctor rerum oeconomicarum

Oberursel, Germany
March 12, 2023

This inaugural dissertation was submitted to the chairman of the Doctoral Candidate Admissions Board Professor Dr. Stefan Bock at the Schumpeter School of Business and Economics at the University of Wuppertal in partial fulfillment of the requirements for the degree of Doctor rerum oeconomicarum (Dr. rer. oec.) following the examination regulations dated August 28, 2019.

First examiner
Dr. Bernhard Scherer
Chief Investment Officer at LBBW Asset Management
Research Associate EDHEC

Second examiner
Professor Dr. Nils Crasselt
Dean
Chair of Managerial Accounting
Schumpeter School of Business and Economics
University of Wuppertal

This doctoral thesis is based on the following three published papers:

Scherer, B., & Apel, M. (2020). Business Cycle–Related Timing of Alternative Risk Premia Strategies. The Journal of Alternative Investments, 22(4), 8-24.

Apel, M., Betzer, A., & Scherer, B. (2022). Real-time transition risk. Finance Research Letters, 103600.

Apel, M., Betzer, A., & Scherer, B. (2023). Point-in-Time Language Model for Geopolitical Risk Events. The Journal of Financial Data Science, 5(1), 65-75.

Acknowledgement

This work represents the culmination of my efforts and the dedication of those who have supported and inspired me throughout my academic journey.

I would like to start by expressing my deepest gratitude to Professor André Betzer for the supervision of my dissertation. His encouragement and personal commitment were crucial to complete this dissertation. Professor Betzer was always poised to provide his full assistance and support at any time of day, especially in periods of setbacks. Special thanks go to my former boss, mentor, and co-author Dr. Bernhard Scherer. Dr. Scherer's enthusiasm for research is unmatched, working in the finance industry for so many years. His dedication and invaluable expertise have been an inspiration to me from day one we met. We share a special relationship that helped me to reach another level in my professional life as well as in academical research. His insightful feedback, and unrelenting belief in me have been the driving force behind my success.

Furthermore, I wish to thank my fellow researchers at the University of Wuppertal, my former and current colleagues at Lampe Asset Management GmbH and Quoniam Asset Management GmbH that have shared countless discussions with me on different research topics. In particular, I want to thank my former colleague and friend Pawel Piechaczek for pushing me to do a doctorate.

My heartfelt gratitude goes to my family. There were times in my life when this academical and professional path seemed very unlikely. Without their endless support in any given situation none of my achievements would have been possible. No words can express how truly grateful I am to have you in my life.

Finally, I want to express my sincerest thanks to my caring wife. Working full-time during the entire time of writing my dissertation was challenging for our partnership. Her sacrifices and unwavering belief in me have allowed me to pursue my dreams and achieve my goals. Her encouragement throughout my academic journey has meant the world to me and has played an integral role in my success.

Matthias Apel

Oberursel, February 2023

Executive Summary

This thesis is comprised by three studies that contribute to the field of asset pricing by expanding our knowledge how macroeconomic and (geo-)political factors drive time variation in asset returns and risk premia.

It is widely agreed upon in finance literature that macroeconomic and (geo-)political conditions impact asset returns. Yet, there is still relatively little-known how cross-asset risk premia vary with business cycle regimes. Time variation in risk premia is not a violation of market efficiency but rather a reflection of time-varying economic rewards. By analyzing macroeconomic sensitivities, the first study reveals that time-varying returns of certain alternative risk premia strategies are significantly related to economic conditions.

Geopolitical and regulatory risks are challenging to measure. However, the broad emergence and accessibility of so-called alternative data sources like machine-readable texts, give cause to take new approaches to approximate latent variables.

The second study introduces an agnostic language model designed to generate domain- and period-specific vocabulary from raw news data to identify topic-related articles. Here, the framework is utilized to develop a point-in-time index that approximates changes in transition risk from climate-related news events. The index is applied to evaluate return sensitivity of publicly available green minus brown portfolio proxies. Based on investors' climate objectives, different approaches to measure a firms' environmental performance are considered for portfolio construction. The study shows that short-term transition risk tends to affect stock prices based on firms' business activity but not emissions.

In the third study another application of the language model is presented by constructing a real-time news index to measure country-specific, geopolitical risk. The approximation of a latent risk variable with media attention is in line with previous research on different unobservable risk measures utilizing news flow. However, avoiding even subtle look-ahead biases is essential for evaluating systematic investment propositions. The study documents that the proposed model can resemble the results of other more heavily curated methods.

Contents

List of Tables

List of Figures

1. Introduction

This thesis is comprised by three studies that contribute to the field of asset pricing by expanding our knowledge how macroeconomic and different political factors drive time variation in asset returns and risk premia. Besides the contextual overlap, the studies share an important commonality in methodology as all considered risk factors are latent variables. A relevant aspect of each research study is the development of an innovative approach that confronts any look-ahead bias to approximate non-observable factors. Given its general importance for the validity of research findings, this aspect earns particular attention.

The introduction gives an in-depth overview of the economical foundation of each considered factor. Previous studies that relate to the presented research on risk factors are discussed thoroughly. Based on the literature review, the motivation for the topics covered in this thesis is derived.

1.1. Macroeconomic risk factors

Various observable, economic data like GPD growth, retail sales or inflation is released on a regular basis. While this kind of data provides market participants with some information about the economic conditions, the actual state of an economy is a latent variable composed of various dimensions. In an attempt to reduce complexity, the state of economic condition is often expressed as one of the idealized regimes (expansion, contraction, recession, and recovery) related to business cycle theory. These periods are characterized by consistent market conditions influenced by external macroeconomic trends. The identification of regimes is important for making informed decisions about asset allocation and portfolio construction, as the specific conditions of each regime may have a significant influence on investment returns (e.g., Chen et al., 1986; Gertler and Gilchrist, 1994; Kiyotaki and Moore, 1997; Berk, Green and Naik, 1999; Perez-Quiros and Timmerman, 2000; Ang and Bekaert, 2004). More recently, several studies emerged that propose fundamental approaches which combine economic data to identify the different regimes of a business cycle. This section presents an overview of relevant research that focus on different methodologies to approximate macroeconomic risk and its impact on asset returns or risk premia.

Generally, there are various challenges to address with the development of a composite macroeconomic risk framework. The selection of potential input data is based on several criteria. Specifically, a sufficient data history that ranges across multiple business cycles is a prerequisite. Ideally, data was subject to previous research and found to be significant in the approximation of macroeconomic risk. The data should be available at monthly or higher frequency. Further, data that is subject to revisions after its initial release needs adjustment to prevent any look-ahead bias in derived findings.

The approach presented in Chapter 2 is inspired by the research of Vliet and Blitz (2011). The authors make use of economic data with more than 50 years of history to construct an indicator for explicitly identifying the prevailing economic regime. In particular, the aggregated regime indicator should grasp the future development of the business cycle and infer the characteristic regime according to the economic cycle classification by the National Bureau of Economic Research (NBER). Vliet and Blitz (2011) consider a regime model which applies four economic variables (the credit spread, earnings yield, ISM and the unemployment rate) to identify the different cycle phases. The authors investigate the attractiveness of various asset classes and investment styles. In addition to equities, bonds and cash, small caps, value, growth, credit and commodities are included into the analysis. Vliet and Blitz (2011) find that the risk and return properties of asset classes are highly dependent on the prevailing economic regime. Based on these findings, the authors derive a dynamic strategic asset allocation approach to stabilize absolute portfolio risk and simultaneously enhance portfolio returns by exploiting economic regimes.

More recently, several studies have examined the impact of macroeconomic risks on a broader range of equity factors. Cooper et al. (2016) find that value and momentum returns are to some extent explained by their loadings on global macroeconomic risk factors. The loadings are supposed to describe the observed negative correlation between value and momentum. The findings hold across both countries and asset classes.

Hodges et al. (2017) provide an analysis of historical factor performance across different business and economic regimes. Observed patterns are in line with economic intuition and previous findings. Value tends to underperform during economic troughs, as these companies have relatively inflexible capital structures while minimum-volatility strategies have generally outperformed because of their risk mitigation properties (Ang et al., 2006). Conditional Sharpe ratios indicate that momentum strategies perform well when the economic growth is approaching its cycles peak. During the following phase quality stocks are in high demand as the probability of a recession increases.

Amenc et al. (2019) propose a method for selecting macroeconomic state variables that reflect changes in expectations about the overall economy. The authors show that returns of common equity factors are significantly impacted by these state variables. Additionally, factor returns also depend on aggregate macroeconomic regimes reflecting good and bad times. Amenc et al. (2019) point out that popular multifactor allocations do not effectively address macroeconomic dependency and that the simple combination of factors may not reduce macroeconomic risks. The authors highlight the importance to understand macroeconomic risks to improve diversification with equity factor investments.

Looking at these findings, it is widely agreed upon that factors affecting equities are cyclical and influenced by macroeconomic conditions. An increasing number of investors consider their portfolio returns as a reward for exposure to different risk factors. However, the relationship between time-variation in cross-asset factor premia and the regimes of the business cycle is relatively unexplored. Most research studies focus on a limited set of common equity factors. The presented work builds on Ilmanen et al. (2014). The authors explicitly analyze the difference in performance for various asset classes and risk premia across different macroeconomic regimes. Ilmanen et al. (2014) take a rather simple approach to approximate the current state of economic conditions by defining four regimes derived from the interaction of growth and inflation. Based on historical context, data values for growth and inflation are classified to be either related to an "up" or "down" environment. Findings of Ilmanen et al. (2014) are in line with previous research of Vliet and Blitz (2011) as both document significant sensitivities between macroeconomic regimes and asst class performance. On the other hand, Ilmanen et al. (2014) report only small macroeconomic exposure for the considered risk and style premia (momentum, value, defensive, carry and trend). A possible explanation for these findings may be the use of aggregated style premia strategies combining different asset classes. The investigation of asset class-specific risk premia strategies increases the probability of identifying meaningful return patterns. Besides sharing similar rational, strategy returns are assumed to be specifically related to the underlying assets.

Previous research studies mainly focus on either equity or aggregated asset class factor premia. Chapter 2 introduces a consistent fundamental framework for the evaluation of asset class-specific risk premia during different business cycle regimes. Based on the time-varying returns of certain cross-asset risk premia strategies, a dynamic factor allocation approach is introduced and validated across different investment universes.

1.2. Climate-related risk factors

The literature distinguishes the economic effects of two major categories of climate change-related risk factors: risks related to the transition to a lower-carbon economy and risks related to the physical impacts of climate change (e.g., TCFD, 2017).

Physical climate risks directly affect economic activity from changes in the climate. The risks from the physical impact of global warming can be either event driven or longer-term shifts in climate patterns. Event driven risks may result from natural disasters linked to climate change like wildfires, storms, or floods that can cause impairment to productive assets and disrupt the supply chain. Alternatively, long-term shifts in climate patterns like the threat of damage from rising sea levels to exposed firms' production facilities, and the associated destruction of real estate values, would be considered a chronic risk (Dietz et al., 2016).

Transition risks encompass a wide range of climate-related effects on business' activities and models that result from an anticipated transition to a low carbon economy. These risks may arise from efforts to address global warming, including but not limited to abrupt or disorderly shifts in climate policy and environmental regulation, the development of disruptive technology, or changes in consumers' or investors' climate awareness (Cambridge Centre for Sustainable Finance, 2016). Depending on the intensity, speed, and driver of these changes, transition risks have varying levels of financial impact on organizations (Bolton & Kacperczyk, 2021a). For institutions and investors with exposure to fossil fuels, transition risk is expressed in evolving beliefs about the shift away from fossil fuels to renewable energy resulting in the devaluation of carbon-intensive assets or even the emergence of so-called “stranded assets” (NFGS, 2020a).

The realizations of physical and transition risks are likely to move independent from each other with respect to time and direction (Giglio et al., 2020). The growing number of natural disasters is likely to increase climate awareness and pressure for mitigation policies, hence leading to an aligned movement of physical and transition risk realizations. By contrast, effective measures to tackle the effects of global warming shall reduce the physical threats of climate change, which results in physical and transition risks changing in the opposite direction. For the upcoming analysis, the focus is solely on transition risks and its return impact on potential hedging instruments.

The integration of climate-related risks into the investment process faces persistent challenges. Research on transmission channels between climate change and financial sectors is still originating, resulting in an insufficient understanding in how climate change and the transition towards a low-carbon economy affect sectors, regions, markets and ultimately the financial system (BCBS, 2020). Consequently, with investment tools and best practices not yet well established, many market participants struggle to price and hedge climate-related risk as recent studies and surveys among institutional investors indicate (Dyck et al., 2019; Krueger et al., 2020; Bolton & Kacperczyk, 2021b). The lack of consensus on how to measure climate risk and a shortage of suitable hedging instruments are just a few of several obstacles to overcome (Andersson et al., 2016). Climate risks are complex and with many dimensions to consider, including the extent how the risks vary depending on the time horizon, the risk distribution and the potential impacts from events with limited historical experience (NGFS, 2020b).

The adverse, physical effects of climate change are expected to inflict direct damages to society and assets by natural hazards. The physical risks of climate change are commonly measured by changes in temperatures, frequency in natural disasters over time, intensity of hurricanes, and changes in sea levels (e.g., Hauer et al., 2016; Gibson et al., 2017; Giglio et al., 2021).

In contrast, transition risk is a purely latent variable. Transition risk is related to the shift to a lower-carbon, more sustainable economy and involve different risk drivers, organizations should consider (e.g., NGFS, 2020b; BCBS, 2021): Policy and legislation (e.g., changes in environmental and emission standards), technology (e.g., the emergence of disruptive alternative energy sources), and climate awareness or company reputation (e.g., shifts in consumer preferences may change buying behavior or effect the overall perception of the companies' business operations). The impact of the different transmission channels on transition risk are not directly observable. The path and frequency of realizations of transition risk is unknown.

Existing approaches to measure climate risk can be mainly divided into three categories: Weather-based measures, fundamental measures, and measures based on the textual analysis of news. Weather-based measures focus on the physical aspect of climate risk like heat waves, floods, drought, and storms. For example, Sheng et al. (2022) analyze the impact of physical climate risks on the persistence of economic and policy-related uncertainty. The authors find that shocks to uncertainty in the United States are generally more pronounced in regimes of high temperature growth. The results imply that the physical effects of climate change induce additional uncertainty.

Fundamental measures are commonly based on return differentials between portfolios sorted on assumed factor expressions (e.g., environmental ratings of third-party agencies, emission levels) of climate risk. Returns of so-called "Green Minus Brown" (GMB) portfolios are considered to explain the varying importance of climate risk to investment decisions. Thereby, investors are assumed to take into account the adverse effects of the transition to a lower-carbon economy on the firm level. Bolton & Kacperczyk (2021a) evaluate the economic importance investors attach to transition risk, by analyzing the stock prices of a large set of companies with different degrees of emission levels. The authors consider market participants to incorporate climate risk expectations into investment behavior with evolving beliefs about the transition to an economy based on less carbon-intensive energy. From an individual firm's perspective, transition risk is assumed to reflect the uncertain path towards carbon neutrality. Hence, firms with high exposure to fossil fuel production or consumption are expected to require an increased risk premium. Bolton & Kacperczyk (2021a) estimate the market-based premium associated with this transition risk at the firm level in the cross-section across various countries as well as sectors. The authors document a carbon risk premium as companies with higher levels of carbon emissions (and higher annual changes) are associated with higher stock returns.

Görgen et al. (2020) develop a proxy of carbon risk from return differences between self-defined "brown" and "green" stocks. The authors take a fundamental approach to quantify carbon risk and calculate an aggregated score from various emission-based metrics for each stock in a global universe. Emission data are derived from four major ESG databases. Görgen et al. (2020) define a long-short portfolio of "brown" versus "green" stocks that is supposed to mimic a factor related to carbon risk. Based on their environmental performance firms in the top tercile have high emissions scores and are classified as "brown" while firms in the bottom tercile are classified as "green". The authors apply the factor-mimicking portfolio to understand carbon risk through the lens of a factor-based asset pricing model. While carbon risk seems to explain systematic return variation well, results provide no evidence of a carbon risk premium.

The described empirical measures select firm characteristics to approximate risk factor exposure a priori. Instead, text-based measures attempt to grasp transition risk explicitly to investigate which firm characteristics relate to risk exposure in the next step. For example, Engle et al. (2020) construct two complementary indices that measure the extent to which climate change is discussed in the news media. The first is computed as the share of texts in the Wall Street Journal (WSJ) that are devoted to issues on climate change. Topic classification is based on a stationary climate change vocabulary, which is constructed from a list of manually selected, topic-specific climate articles and reports. The WSJ index associates increased climate change reporting with news about elevated climate risk, based on the idea that climate change primarily rises to the media's attention when there is a cause for concern. The authors construct a second news-based climate index based on a variety of media sources that is designed to focus specifically on bad news about climate change. This index applies sentiment analysis to climate-related articles to measure the intensity of negative climate news in a given month. The index is calculated as the percent of all news articles that are related to climate change and have been labeled with negative sentiment. No distinction is made between the physical and transition risks of climate change for both indices. However, the authors admit that this aspect is a potential shortcoming in the index construction as these two risk measures might move independent from each other.

Ardia et al. (2020) construct a Media Climate Change Concern (MCCC) index to capture unexpected increases in climate change concerns from 2003 to mid-2018. The authors consider news published by major US-based newspapers. Article data are retrieved from the Dow Jones Factiva, the ProQuest, and the LexisNexis databases. Climate-related news articles are selected based on topic categorization provided by these data bases. The authors take a dictionary-based approach for sentiment classification and define a score that measures concerns from the informational content of news articles based on two lexicons: A risk lexicon to determine the level of discussion about risk events and a second lexicon to assess the increase in the perception of risk. By definition, only negative expressions of risk are considered for index construction. Positive news events that infer progress in the effort to confront the adverse effects of climate change and potentially relief concerns are not explicitly considered. Ardia et al. (2020) apply their MCCC index to document a negative relationship between the firms' exposure to approximated climate concerns and the firms' relative greenhouse gas emissions. Their results imply that an unexpected increase in concerns about climate change induces a rise in stock prices for firms with low emissions, while stock prices of firms with high emissions decrease.

Previous research on a potential, time-varying effect of climate-related risks do either not distinguish between risk types or take simplifying assumptions in measuring the complexity of transition risk. Hence, innovation in the field of risk approximation is a prerequisite to investigate the potential effects of transition risks on asset prices.

Chapter 3 introduces an innovative approach to approximate changes in transition risk with climate-specific news sentiment. Here, the potential impacts of changes in regulatory climate risk on the returns of portfolios that incorporate different climate objectives are thoroughly analyzed.

1.3. Geopolitical risk factors

Like climate transition risk, geopolitical risk is another latent variable that is not directly observable. Depending on the scope of considered macroeconomic and political events, different definitions of geopolitical risk exist. Some of those will be covered in this section. For this thesis, geopolitical risk associated with terrorism, social unrest, or any conflict between states and political institutions is of particular interest. The more specific a risk's definition, the more isolated the effects of the potentially adverse risk events can be analyzed. Irrespective of the exact definition, geopolitical risk is increasingly recognized by investors and policymakers as a relevant aspect of international economics as global trade relations intensify and supply chains become more and more connected.

From an economic perspective, adverse geopolitical events may influence macroeconomic variables via different transmission channels, such as civilian casualties, the isolation of states, or erosion of capital stock. The potential influence of geopolitical risk on macroeconomic outcomes and asset prices have been previously analyzed in economics literature and will be discussed next. However, the main challenge derives from the need for a risk proxy that is consistent over time, and grasps real-time geopolitical tensions as perceived by market participants and central bank officials. Existing ways in academical research to approximate geopolitical risk differ in their methodologies and data sources.

Karagozoglu et al. (2022) provide a comprehensive overview of methodologies and applications of geopolitical risk measures. Construction techniques can be mainly divided into three categories: asset price-based, analyst rating-based measures, and measures based on the textual analysis of news.

The approach of Engle & Campos-Martins (2021) falls into the first category. The authors introduce a statistical model to measure geopolitical risk based on the magnitude of common volatility shocks to a wide range of financial assets, defined as GEOVOL. Thereby, Engle & Campos-Martins (2021) follow the assumption of fully efficient markets with asset prices to incorporate all available information. Hence, empirical measures are supposed to incorporate changes in geopolitical risk in a timely manner as those estimations are based on the actual change of market prices. Investors are expected to react to new information irrespective of the actual data source by adjusting their portfolio, incorporating changes in geopolitical risk into asset prices accordingly. While modelling volatility factors using numerical methods is rather straightforward to implement and replicate, results of GEOVOL show that approximated financial distress is not limited to our definition of geopolitical risks. Instead, GEOVOL interpretates geopolitical risk in a wider sense as the exposure of countries to political actions, regime shifts, regulatory events, trade affairs and even climate change. The isolated measurement of geopolitical risk in relation to terrorism, social unrest, war, or political conflict is more specific and not easily achieved with price-based measures. Here, the other considered categories of risk approximation techniques come into play. These measures, either based on analyzing texts or expert opinions, are more likely to reflect the current beliefs of market participants and potentially their expectations of future developments.

Industry-wide risk measures based on expert opinions commonly share the rationale to forecast the probability of the realization of geopolitical events while measures based on ratings may rely on ex-post information (e.g., Brecher & Wilkenfeld, 2000).

Given the advances in natural language processing and the thorough coverage of geopolitical events in news media, approaches based on textual analysis have gained popularity in recent years. Baker et al. (2016) are among the first to approximate latent macroeconomic and policy risks with text-based methods. The authors construct a broad index of economic policy uncertainty (EPU) for the United States and examine its development since 1985. The index is defined to reflect the frequency of articles in 10 US newspapers that contain a pre-defined combination of terms related to macroeconomic and political distress. Baker et al. (2016) document a relationship between their EPU index and other measures of economic and policy risk, like implied stock market volatility and the frequency with which the Federal Reserve System's Beige Books mention policy uncertainty.

Ahir et al. (2022) follow the approach of Baker et al. (2016) and construct an index to measure global uncertainty for almost 150 individual countries on a quarterly basis from 1952. The so-called World Uncertainty Index (WUI) reflects the frequency of the word "uncertainty" in the quarterly Economist Intelligence Unit country reports. The authors scale the raw counts by the total number of words in each report to make the WUI comparable across countries. The authors show that the index is associated with economic policy uncertainty (as measured by the EPU index), stock market volatility, and tends to indicate lower GDP growth. Like the EPU index from Baker et al. (2016), the WUI index is a broad measure of geopolitical risks. By definition, risk approximation is not restricted to uncertainty related to geopolitical tension among states and political actors. As a result, the global index spikes around a diverse set of major macroeconomic and political events like the Euro debt crisis, the Brexit vote and the COVID pandemic. Ahir et al. (2022) apply a vector autoregressive model to international panel data and document that innovations in the WUI predict declines in output. The authors find that the observed effect of uncertainty is more pronounced in countries with relatively low institutional quality.

Caldara & Iacoviello (2022) are among the first who attempt to measure geopolitical risk explicitly as the threat, realization, and escalation of adverse events associated with wars, terrorism, and any tensions that impact international relations between political institutions. Therefore, Caldara & Iacoviello (2022) construct a news-based index that approximates geopolitical risk by the share of articles that discuss geopolitical events and related threats. Their most recent GPR index version starts in 1985 and is based on the automated text-searches on ten, mainly U.S.-based newspapers. To identify topic-related articles the authors use a dictionary-based approach. Caldara & Iacoviello (2022) manually pick a set of terms and expressions (uni- and bigrams) from different sources whose occurrence in text articles is assumed to be associated with the coverage of geopolitical events and threats in news media.

The authors apply vector autoregressive models to the data of their most recent GPR index and document that major index upswings of approximated geopolitical risk relate to prolonged slowdowns in investment and employment in the United States. Additionally, Caldara & Iacoviello (2022) also construct an index with 120 years of data from searches of the historical archives of the Washington Post, the Chicago Tribune, and the New York Times. For the longer index version, the authors document lower expected GDP growth and an increased probability of economic decline for high values of country-specific GPR indices.

Given its potential to specifically analyze the effect of geopolitical threats and tensions, the GPR indices are subject of various applications in academical research. Cheng and Chiu (2018) examined the impact of shocks to geopolitical risk on the business cycle for emerging markets. The authors apply vector autoregressive models for 38 different countries and document that shocks to the GPR index induce economic contractions. While the average share of output variation explained by global geopolitical risk shocks is between 13% and 22%, there is significant cross-country dispersion.

Baur and Smales (2020) investigate the potential of different assets to hedge against geopolitical risks. The authors focus specifically on the precious metals gold, silver, platinum, and palladium along with other assets (e.g., S&P 500 Index, US Treasuries, USD Index) representative for other classes (equities, bonds, FX). Baur and Smales (2020) apply a regression model to asset returns, the lagged log of the GPR index and a set of control variables to evaluate the interaction between asset returns and geopolitical risks. Results indicate that the relationship between precious metals and geopolitical risk differs substantially from that of other assets. While each considered precious metal shows at least some ability to hedge against geopolitical risk, only gold and silver preserve its safe haven character for elevated levels of geopolitical risk. In contrast, findings for platinum and palladium are less conclusive. Baur and Smales (2020) explain observed differences across precious metals by the diverse demand characteristics. Gold and silver are historically sought-after during times of high uncertainty, whereas the demand of palladium and platinum is primarily industrial.

In Chapter 4, a novel news-based approach is introduced that follows Caldara and Iacoviello (2022) in their definition of geopolitical risk. However, instead of manually selecting presumably domain-specific words to identify news events associated with geopolitical risk, the methodology is designed to overcome any look-ahead bias by generating topic-related terms point-in-time.

2. Business Cycle-related Timing of Alternative Risk Premia Strategies

2.1. Introduction

Theoretical asset pricing model conjecture that returns on risky assets depend on economic states of the world that resemble business cycle related risks. Inflation, real rates or term spreads are known as state variables that help predict time-varying risk premia. Time variation in risk premia is not a violation of market efficiency but rather a reflection of the fact that economic rewards for taking on risk must be large when economic times are bad. There is ample empirical evidence that the financial payoffs of asset classes such as stocks and bonds vary with the business cycle. Much work has been put in determining cross-sectional variation in expected returns of traditional assets using macroeconomic variables (e.g., Gertler and Gilchrist, 1994; Kiyotaki and Moore, 1997; Berk, Green and Naik, 1999; Perez-Quiros and Timmerman, 2000).

More recently the attention of research has also turned to explaining time-varying returns of equity style factors by approximating state and development of the business cycle (Ferson and Harvey, 1991). Chordia and Shivakumar (2002) analyze dynamic return patterns of momentum strategies. Using a set of lagged standard macroeconomic variables to forecast one-month-ahead stock returns, the authors show that the predicted part of returns is the primary cause of the observed momentum premium. Hodges, Hogan, Peterson and Ang (2017) search for sources of cross-sectional and time-series information to predict the premiums of equity factor strategies using smart beta indices and conclude that timing equity factors based on macroeconomic conditions can generate excess returns to a passive allocation approach.

Macroeconomic indicators as explanatory variables for observable cross-sectional return differences are well established. In this study we examine the transferability on alternative risk premia. Few studies emerged relating the behavior of documented alternative risk premia (ARP) to the macroeconomic environment. Christiansen, Ranaldo and Söderlind (2011) investigate time-varying systematic risk of FX carry trade strategies across different market regimes. To distinguish between high and low risk environments, proxies commonly used to measure market and liquidity risk are adopted. The authors show that in turbulent times, carry trade strategies record a coincident increase in volatility and exposure to other risky assets. Ang, Israelov, Sullivan and Tummala (2018) analyze the volatility risk premium in the stock market which refers to the phenomenon that option-implied volatility tends to exceed realized volatility of the same underlying asset over time. Given the nature of a volatility selling strategy, major drawdowns are recorded when the underlying asset experiences large sudden losses as investors revise their expectations. Studying the bond market, Asvanunt and Richardson (2016) confirm the existence of a positive premium for bearing exposure to default risk. The authors construct a time series of corporate bond returns in excess of Treasury bond returns (adjusted for any duration differences). The measured excess return of corporate bonds, referred to as credit risk premium, is more pronounced in regimes of economic growth and negative in periods of increasing inflation and economic slowdown.

It is common practice to use different risk parity models individually or in combination to strategically optimize the allocation of ARP strategies. In this study we explicitly analyze the macroeconomic sensitivities and present an approach for actively allocating ARP strategies conditional on the prevailing economic regime. The presented findings give an implication for the performance potential in a tactical framework.

Our work builds on Ilmanen, Maloney and Ross (2014) who examine macroeconomic sensitivities of various asset classes. Considering different macroeconomic variables and inflation/growth scenarios the authors report return patterns for traditional asset classes fairly in line with economic intuition. Additionally, the authors include five simulated long/short style premia composites in their analysis. However, given the presented macroeconomic sensitivities for momentum, value, carry, defensive and trend-following strategies the authors find little evidence to suggest style premia performance relate to the economic environment as results for conditional returns are insignificant.

We enhance the approach from Ilmanen, Maloney and Ross (2014) in numerous ways. While the authors use aggregated style premia strategies combining different asset classes, we investigate asset class-specific risk premia strategies to increase the probability of identifying meaningful return patterns. Besides sharing similar rational, strategy returns are assumed to be specifically related to the underlying assets. Furthermore, instead of regarding a set of common economic variables we construct a slightly more complex but intuitive business cycle model to approximate and predict the state of the economic environment. Whereas many papers use statistical properties of the assets themselves for regime definitions, we adopt a fundamental approach which uses economic data for explicitly identifying the prevailing regime in contrast to estimating probabilities. The model enables us to analyze conditional returns and to deduce transparent regime-based allocation scenarios.

We extend the existing literature for alternative risk premia by analyzing macroeconomic sensitivities for a diversified basket of tradeable ARP strategies. The key issue addressed in this chapter is whether such dependencies can be observed and exploited in a portfolio construction context. The approach closest to ours is from Blin et al. (2018). The authors use a (parsimoniously described) proprietary nowcasting procedure to identify business cycles that makes extensive use of large datasets. They also find positive timing abilities in the same order of magnitude and significance as our less demanding procedure. In contrast to our approach no cross validation across related universes takes place.

The rest of the chapter is organized as follows. In the next section the methodology of the business cycle model is described in detail. Section 2.3 discusses sample and data. Section 2.4 presents the findings concerning the macroeconomic sensitivities of ARP strategies before analyzing the return potential of an active portfolio allocation approach. To validate conclusions made based on our economic model, we investigate conditional returns of different investment classes and compare these with previous findings and economic intuition in Section 2.5. In the final section results are put into perspective and suggestions are given for future studies.

2.2. Methodology

The purpose of a business cycle model is to capture the global economic environment by combining various macroeconomic data. Given the amount of literature and research on economic models existing we opt for adopting and adjusting a model suited for our requirements. Our approach is an enhancement of the economic model developed by Van Vliet and Blitz (2011) as the methodology is intuitive, robust and easily applicable. The model makes use of economic data to construct an indicator for explicitly identifying the prevailing economic regime. In particular, the aggregated regime indicator should grasp the future development of the business cycle and infer the characteristic regime according to the economic cycle classification by the National Bureau of Economic Research (NBER). It enables us to analyze conditional asset returns and to deduce transparent regime-based allocation scenarios.

Besides the necessary predictive power of the aggregated indicator, the potential of the business cycle model relies on its sensitivity to allow for early signal generation without generating too much false positives. Highly reactive trading indicators typically result in inefficient asset allocation and excessive turnover. To achieve the desired trade-off, we combine lagging indicators with more volatile but leading indicators. Details on the selected variables are provided in the next section.

We use an aggregated indicator to derive the characteristic regime. Each phase of the ideal-type economic cycle is defined by two criteria – the current level of the indicator and its change from the last observation. To ensure comparability and combinability of the regarded macroeconomic data the variables have been standardized. The Z-scores are calculated at the end of each month under the expanding window approach to avoid look-ahead bias.[1] We make use of the entire historical data given at the point in time to benefit from the inherent "learning effect".[2] To limit the influence of individual variables we cap the Z-score to three standard deviations on either side. An equal-weighted Z-Score of the selected macroeconomic variables equates to the level of the overall regime indicator. While the state of the economic environment is represented by the sign of the aggregated Z-Score its trend shall be captured by the change of the level over the defined reference period.

To account for swift changes in Z-Scores calculated on few observations, we use at least a three-year training period before we start generating trading signals based on the regime indicator. The four business cycle regimes are defined as follows:

[1] Median is used instead of mean for standardization to reduce the impact of outliers.

[2] For data samples with a much longer lookback it might be necessary to operate with a rolling window approach to account for structural changes in the long run.

- Expansion: Z-Score positive and increasing
- Peak: Z-Score positive but decreasing
- Recession: Z-Score negative and decreasing
- Recovery: Z-Score negative but increasing

In order to limit the amount of regime switches (and thereby the turnover), two consecutive periods of uniform changes or a monthly change of more than one standard deviation above its average are required to signalize a sustainable change in the business cycle regime.

We are fully aware that there might be more dynamic or sophisticated approaches for predicting the economic business cycle but with the main purpose of this study being analyzing and exploiting possible macroeconomic sensitivities of alternative risk premia, the model considered absolutely meets the requirements.

2.3. Data

2.3.1. Business Cycle

We use monthly data from March 1998 as this is the longest available common data history of the used macroeconomic variables.

Besides complying with the characteristic requirements related to the methodology mentioned above, the variables should capture information from different *macroeconomic dimensions* influencing the development of the business cycle. Therefore, we consider five indicators represented by the following macroeconomic data:

- *Unemployment* is defined by the seasonally adjusted U.S. unemployment rate (USURTOT Index).

- The OECD Leading Indicator (OEOTKLAC Index) is an approximation of economic activity itself to reflect adjustments to expectations concerning the *economic growth*. As the officially announced GDP growth is a lagging indicator, substitute variables with a higher frequency are selected to provide a monthly estimate of an otherwise quarterly released data set.[3]

- *Producer sentiment* is expressed by combining Markit Global Manufacturing PMI (MPMIGLMA Index) and ISM Non-Manufacturing NMI (NAPMNMI Index). Both survey-based indices try to forecast business activity and climate by polling purchasing and supply managers in the manufacturing as well as in the service sector.

- The Conference Board Consumer Confidence Index (CONCCONF Index)[4] and the University of Michigan Consumer Sentiment Index (CONSSENT Index)[5] are selected to measure *consumer sentiment* as the degree of optimism that consumers are expressing through their economic expectations and financial activities.

- *Financial market stress* can be thought of as an interruption to its normal functioning often related to either increased uncertainty about fundamentals value of assets or about the behavior of other investors. We adapt the approach of the Kansas City Financial Stress Index (KCFSINDX Index) to construct a financial stress indicator independent of the release date. The combined indicator includes different

---

[3] Used data are end-of sample figures and could differ from the preliminary figures published earlier. However, direction and magnitude of the monthly change should rarely be affected. Due to standardization small adjustments to the absolute value are of less importance. Hence, the impact on the business cycle indicator should be rather negligible. For detailed analysis on the effect of revisions see Nilsson and Guidetti (2008).
Additional information on the methodology of the Leading Indicator CLI is released on the OECD Homepage (http://www.oecd.org/sdd/leading-indicators/41629509.pdf).

[4] https://www.conference-board.org

[5] https://data.sca.isr.umich.edu/

spread, volatility and correlation data to reflect possible market frictions. With the actual KCFS Index being released only once at the beginning of each month and our business cycle indicator being calculated at its end, we significantly reduce the information lag provided by an otherwise delayed data processing.[6]

Each of the *five dimensions* contributes equally to the overall regime indicator. All macroeconomic data are retrieved from Bloomberg and the FRED database of the Federal Reserve Bank of St. Louis.[7] Unemployment, financial stress and all sentiment data are based on their initial releases to reflect the information that would have been available at the time of the forecast.[8] To account for aligned interpretation the inverse of the unemployment rate and the financial stress indicator with opposite sign are used for further analysis.

We are aware of the fact that besides regarding global asset returns, we have a minor focus on U.S. macroeconomic data due to data limitations. Practically, the effect may be negligible because of the impact the U.S. economy has on the development of global capital markets.

2.3.2. Alternative Risk Premia

For our analysis we use a well-diversified basket of 25 third-party risk premia strategies developed by a leading investment bank. The data set covers tradeable ARP strategies for multiple asset classes including equities, bonds, currencies, commodities and rates. Price data for the strategies are provided by Bloomberg.

---

[6] It is beyond the scope of this article to explain the complete data generating process. For further details on the selected financial variables and on how they are combined see exemplarily Hakkio and Keeton (2009).
[7] https://fred.stlouisfed.org/
[8] Data for the ISM Non-Manufacturing NMI are final figures until 2008. From then on initial release data are available and used.

As the strategies are at least partially back-tested we apply a haircut on the Sharpe ratio of the simulated sample data based on the approach of Harvey and Liu (2015). The rationale for this is the following. The more optimistic we are on unconditional Sharpe ratios (which are the results of back-tests and hence almost surely overestimate future returns), the more pessimistic we will be on timing. Deviating from a diversified portfolio of ARP creates a diversification loss that successful timing needs to overcome. This loss is larger the higher Sharpe ratios on individual ARP are. Using unadjusted back-test performance will make investors expect unrealistically high payoffs from a static strategy and at the same time impose an unrealistically high burden for the viability of timing strategies. To ensure comparability and account for differences in the risk profile of the underlying strategies, each ARP strategy is scaled to a volatility of 8% per annum (p.a.).

Further details on the considered ARP strategies concerning their functionality are given in Table A2.1. We note that the ability to directly trade these vehicles makes them an option for individual investors to use in portfolio construction.

2.3.3. Global Multi Asset Classes

The global portfolio is a representative mix of broad and diversified asset classes, including global equity, rates, credit, and commodity. For practitioners, the model provides a straightforward approach of implementing an approximation of the global portfolio. All indexes are easily investable using exchange-traded funds.

The indexes representing global equity are the MSCI North America, MSCI Europe, MSCI Japan, MSCI Pacific ex Japan, and MSCI Emerging Markets total return indexes. The remaining asset classes are replicated using the BB Barclays Global Government, BB Barclays Global Credit, BB Barclays Global Securitized, BB Barclays Global High Yield, and BB Barclays World Govt Inflation Linked Bonds total return indexes and the gold spot price.

We calculate the daily excess return for EUR-denominated asset classes using DB EONIA TR index as a cash equivalent. To reduce the influence of currency movements on otherwise noisy bond returns we use currency hedged bond benchmarks. All daily data are retrieved by Bloomberg from January 2001 to August 2018.

2.3.4. Equity Style Factors

Finally, we look at equity style factors to cross-validate our business cycle model. While it is well known that variation in stock market risk exposures (dumb beta) can be associated with successful timing, we extend that line of thought and test whether time-varying returns of equity risk factors (smart beta) relate to changes in the prevailing economic regime. For this purpose, we address style factors identified by Fama and French (1993, 2015) and the corresponding global MSCI smart beta indices. Our universe of choice is given by the Fama–French five-factor model as an alternative investment environment that augments the original three-factor model of market (MKT), size (SMB), and value (HML) with profitability (RMW) and investment factors (CMA). We obtain daily return data for the modeled risk factors from January 2001 to August 2018 published on the website of Kenneth French.[9]

Further we investigate the economic dependencies of equity style factors momentum, size, value, quality, and minimum volatility provided by the MSCI style indices.[10] We calculate the daily active returns of MSCI World smart beta indices using Bloomberg data

[9] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[10] According to the mentioned style factors we use the MSCI World Momentum, MSCI World Size, MSCI World Enhanced Value, MSCI World Sector Neutral Quality, and MSCI World Minimum Volatility as related smart beta indices.

for the same look-back period. The given data sets are transparent and investable for individuals via exchange traded funds.

The used style factors enhance our other data set for analyzing conditional returns of risk premia by common equity style factors from different sources.

2.4. Empirical Results

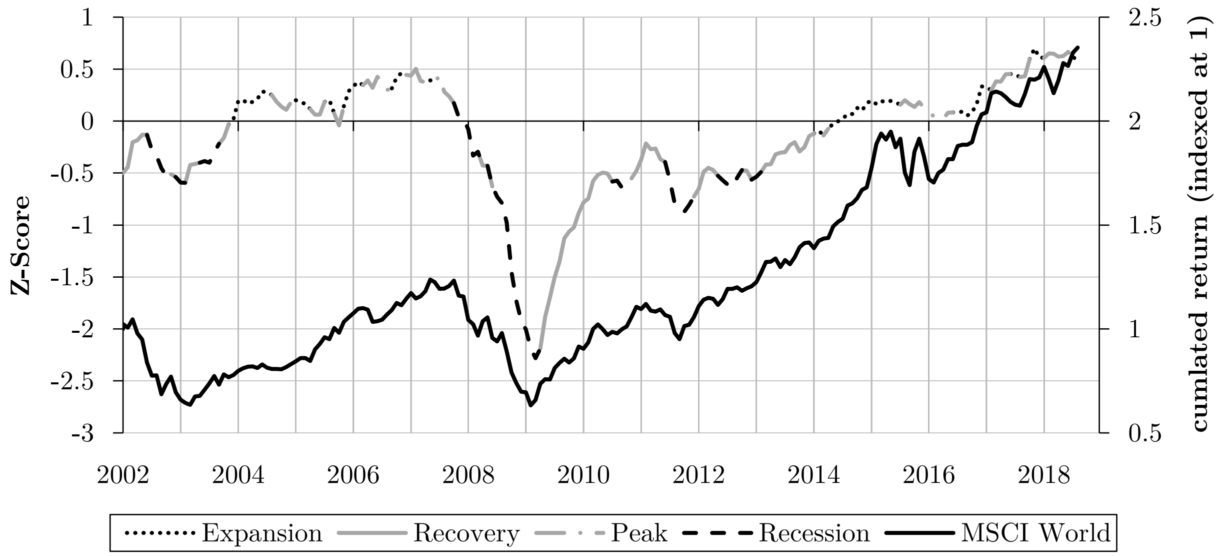
In this section, we report our main findings analyzing conditional return patterns of alternative risk factors and exploiting these for a timing strategy based on the economic regime. To examine the return dependency on the macroeconomic environment, reported conditional performance data are the result of an in-sample analysis. A possible relationship should not be overshadowed by the degree of predictive power the business cycle indicator possesses. Therefore, the monthly returns of the ARP strategies are assigned to the business cycle regime indicated at the end of the same month. In addition, we assess the statistical significance of the observed differences in regime-related performance. The robustness of our findings is critical for the possible applicability of the results to portfolio construction.

Finally, we analyze the performance of a long/short portfolio, which is based on the identified conditional return patterns to give an indication for the potential of regime-related risk premia timing.

2.4.1. Macroeconomic Sensitivities

Figure 2.1 displays the business cycle indicator we estimate from January 2002 to August 2018 given the end-of-sample Z-score. The line-type of the chart corresponds to one of the four economic regimes provided by the business cycle model. The regimes accord with economic intuition, and the development of the indicator is correlated, as expected, with the global equity market.

Figure 2.1: Development of the business cycle indicator from 2002 to 2018



On the primary x-axis the end of sample Z-Score of the aggregated business cycle indicator is reported. The line type corresponds to the regime indicated at the point in time. The cumulated monthly performance of the MSCI World AC as proxy of the global equity market is displayed on the secondary x-axis. Performance is indexed at 1 starting in January 2002. Monthly data are reported since the end of August 2018.

In Table 2.1, we report the annualized Sharpe ratio of the ARP strategies conditional on each regime of the business cycle. While the conditional Sharpe ratios of some strategies barely diverge from their overall, unconditional Sharpe ratios; others look to have characteristic return patterns in line with economic expectations concerning cycle-related asset behavior.

For example, the observed return patterns for the Interest Rate Curve strategy might be explained by the dependency of the strategies' performance on the shape of the expected yield curve, which is closely related to the regime of the economic cycle. With the strategy investing in the short end of the curve and selling the long end while maintaining duration neutrality, performance is negative in expansion and peak phases, in which inflation tends to increase, and investors expect the yield curve to steepen. However, conditional strategy performance is especially positive in the recession phase as market expectations might turn to future interest rate cuts as monetary policy focuses on stimulating economic growth (Evans and Marshall, 2007).

Table 2.1: Conditional Sharpe ratio of alternative risk premia strategies

	Annualized Sharpe ratio					**ANOVA**		
Risk premia strategy	**Expansion**	**Peak**	**Recession**	**Recovery**	**Overall**	**Obs.**	***F*-statistics**	***p* value**
Equity Multi Factor	1.04	0.01	0.23	1.68	0.76	166	1.762	0.157
Equity Vol. Carry	1.87	0.54	1.13	3.02	1.68	198	4.433	0.005***
Equity Trend	0.71	0.32	0.84	0.01	0.44	202	0.648	0.585
Equity Imbalance (1)	0.80	0.05	1.28	1.06	0.86	174	3.747	0.012**
Equity Imbalance (2)	1.53	-0.08	1.35	2.60	0.96	151	2.403	0.070*
Equity Defensive	1.63	0.18	1.03	3.31	1.37	166	2.449	0.066*
Equity Carry	0.93	0.72	-0.12	2.06	0.82	121	1.748	0.161
Credit Carry	0.48	0.25	-0.15	1.78	0.71	123	2.426	0.069*
Credit Curve	0.22	0.30	0.01	1.77	0.59	132	2.327	0.078*
Credit Vol. Carry	0.69	0.82	1.74	1.99	1.52	126	3.217	0.025**
FX Trend	0.89	0.22	0.09	0.80	0.56	202	0.472	0.699
FX Value	0.30	0.06	0.12	0.84	0.28	202	0.344	0.811
FX Vol. Carry	0.13	-0.05	0.29	2.47	0.65	202	4.127	0.007***
FX Carry	0.88	1.29	0.80	0.89	1.00	202	0.402	0.751
Interest Rate Trend	0.47	0.57	1.45	1.16	0.92	202	1.386	0.248
Interest Rate Carry	0.62	1.77	0.89	0.88	0.96	202	0.670	0.570
Interest Rate Vol. Carry	1.22	1.16	-0.47	1.71	0.60	202	3.612	0.014**
Interest Rate Curve	-0.65	-1.01	1.29	0.42	0.26	202	5.228	0.002***
Interest Rate Value	0.53	0.25	1.47	0.78	0.73	202	1.533	0.208
Commodity Vol. Carry	0.44	0.74	1.07	2.33	1.21	186	3.677	0.013**
Commodity Curve	1.67	2.97	0.96	0.72	1.40	202	2.276	0.079*
Commodity Imbalance (1)	0.00	0.65	1.25	1.19	0.81	202	1.802	0.146
Commodity Imbalance (2)	0.84	1.27	0.44	0.62	0.75	202	0.190	0.904
Commodity Trend	0.95	0.51	0.45	-0.26	0.41	202	1.212	0.303
Cross Asset Trend	1.08	0.69	0.61	1.71	1.01	202	0.388	0.759

Vol. = Volatility; Obs. = Number of monthly observations given for the data sample. *,** and *** indicate statistical significance at the 10%, 5% and 1%, respectively.

Equally eye-catching is the regime-related performance of various volatility carry strategies. Remarkable, high statistical significance is achieved by the volatility carry strategies of equity and FX. In line with economic intuition, both strategies outperform in the recovery phase as implied volatility tends to trade at a high premium to subsequent realized volatility (Ang, Israelov, Sullivan and Tummala, 2018). With the uncertainty of market participants and volatility of risky assets typically increasing in the peak phase, this premium likely disappears when implied volatility at which options are sold underestimates future realized volatility. Both carry strategies underperform in this regime.

However, return patterns for trend strategies show no sign of relationship to the economic cycle. Trend strategies typically underperform in non-directional markets. As the transition phase between regime shifts is often accompanied by changes in cross-asset market trends, the period of below-average strategy performance will last until new trends are established. Hence, inconclusive return patterns might be the result of regime shifts affecting strategies' performance. We will explicitly address that issue in the example of the equity momentum factor in later sections.

We use analysis of variance (ANOVA) to test whether the observed differences in the conditional monthly returns are robust.[11] Overall, almost 50% of the considered ARP strategies have a statistical significance at the 10% level or lower (see Table 2.1). These results are rather impressive since several strategies are designed to deliver a continuous and unconditional performance contribution. Furthermore, the business cycle indicator is not a perfect match for an underlying economic trend, and the economic condition is not the only expected return dependency.

[11] We correct for unequal variances across economic regimes by calculating a nonparametric (bootstrapped) test statistic.

To confront concerns about the beta neutrality of the risk premia strategies we calculate the conditional betas against global equity and bonds returns. Results of the multivariate regression are reported in Table 2.2. While some strategies record high betas in certain regimes, only two of them experience significant, uniform beta exposure in each regime, with one of them being Equity Volatility Carry. It is of little surprise that this strategy has significant equity exposure over all regimes considering the nature of its composition. However decisive, for the successful implementation of a regime-related timing strategy, is not the coincident development of subsequent risk factor returns and the economic environment, but the capability of predicting future factor returns by forecasting the business cycle.

2.4.2. Portfolio Construction

To assess the possibility of exploiting the observed regime-related return dependencies in portfolio allocation, we set up an active timing strategy. Ahead of portfolio simulation, we need to define active allocation scenarios for each of the business cycle regimes based on our previous findings for macroeconomic sensitivities. To ensure suitability only ARP strategies with a statistical significance at the 10% level or better are considered for portfolio construction. In addition to the restrictions for robustness, we exclude risk premia strategies with excessive rebalancing costs.[12] The remaining strategies are ordered by the conditional Sharpe ratio in each regime. We select the four strategies with highest respectively the lowest performance to form the regime-specific long/short allocation scenarios.

[12] The chosen limit results from a trade-off between performance contribution and transaction costs. We consider swap-based rebalancing costs above 50 bps as excessive.

The regime indicator is applied to the timing strategy as follows. Referring to the methodology described in the Section 2.2, we calculate the indicator at the final trading day of each month using the macroeconomic data available at that point in time. According to the phase signaled by the regime indicator, the corresponding allocation scenario is selected. The considered ARP strategies receive either a long or a short signal. We implement the timing strategy as a tactical overlay portfolio. Hence, long/short positions are defined as over- and underweight positions relative to a strategic asset allocation. By combining the timing strategy with a risk parity approach, we measure explicitly the performance enhancement of intentionally deviating from an already risk optimized portfolio. The initial portfolio strategically allocates the twenty-five ARP strategies to contribute equally to the ex-ante portfolio volatility of 5% p.a.[13]

Finally, the tactical weights result from an optimization process with the tracking error target set at 1% p.a. and the extend of the underweights being constrained by the strategic weights of considered ARP strategies. The calculated weights w^{TAA} define the dollar-neutral overlay portfolio for the upcoming period. We use an implementation lag of one trading day and rebalance the portfolio on a monthly basis. The output of this processing is a long/short portfolio that incorporates the conditional information assigned to the current market regime R. The daily return of the tactical overlay portfolio r^{TAA} is defined as follows:

$$r^{TAA} = \sum_{k=1}^{n} w_k^{TAA} * r_k, \text{ with } n \text{ being the number of ARP strategies considered.}$$

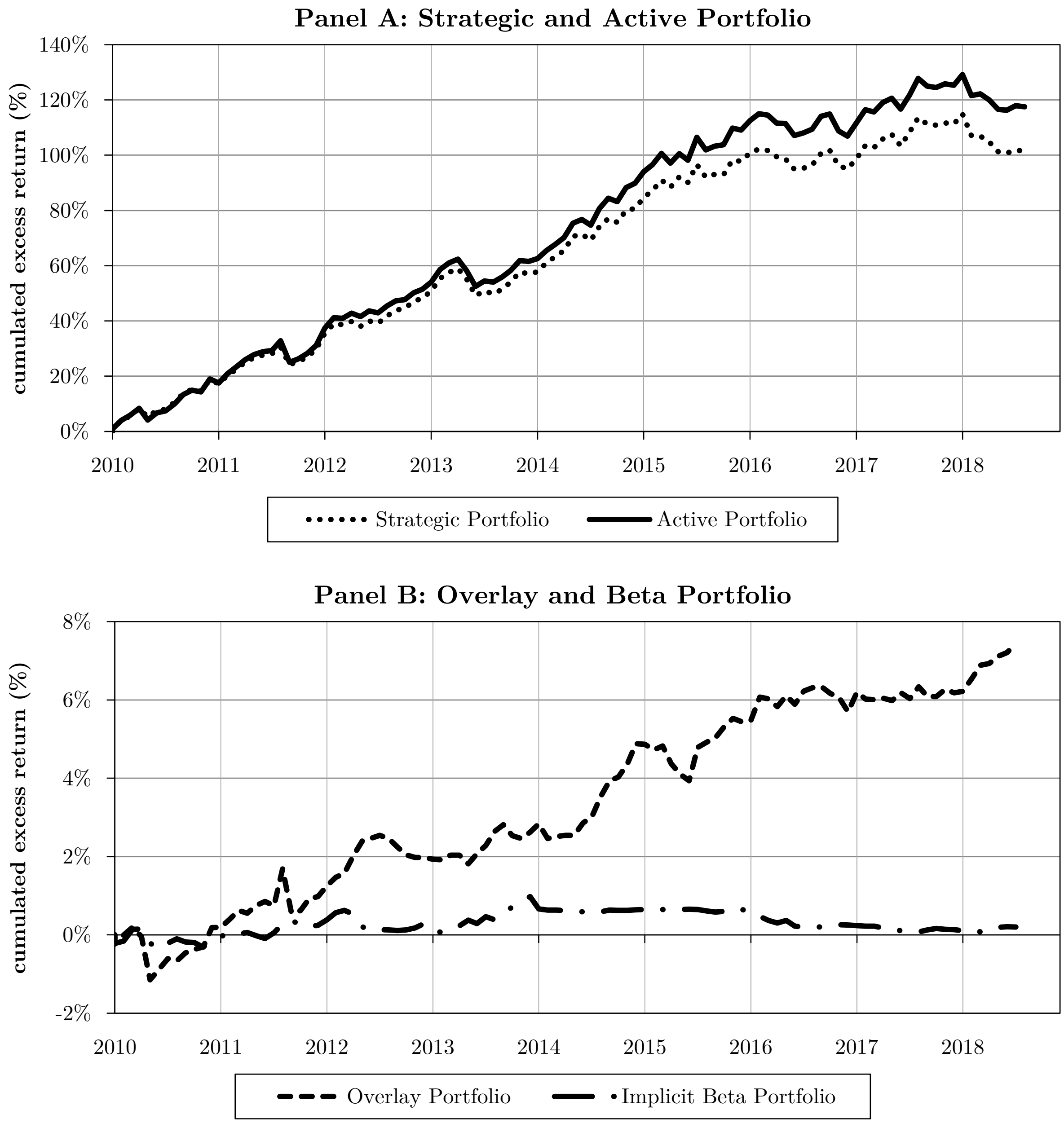
[13] With each ARP strategy having a scaled volatility of 8% p.a. the long-only risk parity portfolio can be subject to leverage.

Table 2.2: Conditional beta exposure of alternative risk premia strategies

	Global Equity					Global Bonds				
Risk premia strategy	**Expansion**	**Peak**	**Recession**	**Recovery**	**Overall**	**Expansion**	**Peak**	**Recession**	**Recovery**	**Overall**
Equity Multi Factor	-0.142	-0.016	0.155*	-0.040	0.083	0.187	-0.173	0.097	-0.218	-0.053
Equity Vol. Carry	0.412***	0.484***	0.214**	0.340***	0.290***	-0.502**	-0.302	0.163	0.026	-0.049
Equity Trend	0.728***	0.397**	-0.205***	0.170*	0.015	-0.315*	-0.305	-0.081	-0.100	-0.085
Equity Imbalance (1)	-0.162	0.052	-0.179	-0.188**	-0.163**	0.001	0.125	0.426	0.473**	0.324*
Equity Imbalance (2)	0.241**	0.238	-0.124	-0.017	0.005	-0.581***	-0.102	0.420	0.144	0.024
Equity Defensive	0.244**	0.206	-0.610**	0.062	-0.255**	-0.495**	-0.143	0.399	0.105	0.022
Equity Carry	0.209	0.177**	0.384*	0.354***	0.337***	-0.235	0.003	-0.286	0.109	-0.092
Credit Carry	0.244*	-0.097	0.292*	0.188	0.196**	-0.437*	-0.106	-0.870	0.047	-0.352*
Credit Curve	-0.005	-0.047	0.365***	0.361***	0.281***	-0.168	-0.087	-0.436	0.045	-0.164
Credit Vol. Carry	0.205*	-0.056	0.155	0.088	0.078	-0.423**	-0.349	-0.166	0.469	0.073
FX Trend	0.007	0.158	-0.085	0.024	0.001	0.225	0.029	0.632**	0.242	0.324**
FX Value	-0.004	-0.061	0.166*	0.031	0.081*	-0.159	0.252	-0.670**	-0.124	-0.278**
FX Vol. Carry	0.317*	0.212*	0.052	0.167*	0.130**	-0.176	-0.040	0.278	0.016	0.091
FX Carry	-0.087	0.294*	0.002	0.150	0.067	0.056	-0.922**	-0.240	-0.057	-0.229*
Interest Rate Trend	-0.036	-0.077	-0.242***	-0.153*	-0.172***	0.335	-0.418	1.056***	0.087***	0.658***
Interest Rate Carry	-0.328*	-0.172	-0.143**	-0.249***	-0.168***	0.742**	0.621*	0.901***	0.943***	0.803***
Interest Rate Vol. Carry	0.100	0.085	0.064	-0.026	0.074*	0.132	0.130	0.089	0.525**	0.192*
Interest Rate Curve	0.004	-0.073	-0.180*	-0.135	-0.167***	0.169	0.240	0.282	0.629**	0.402***
Interest Rate Value	-0.125	-0.056	-0.205***	-0.238***	-0.187***	0.334	0.825**	0.882***	0.694***	0.706***
Commodity Vol. Carry	0.100	0.206	0.211*	0.398***	0.229***	-0.051	-0.019	-0.236	-0.018	-0.070
Commodity Curve	-0.035	0.006	-0.111	-0.105	-0.078	-0.063	-0.084	0.181	0.106	0.056
Commodity Imbalance (1)	0.181	-0.014	-0.052	-0.106	-0.046	-0.152	0.138	0.512*	0.015	0.222
Commodity Imbalance (2)	-0.142	0.068	-0.058	0.211*	0.015	0.181	-0.165	-0.145	-0.343	-0.142
Commodity Trend	-0.191	0.037	-0.042	0.099	-0.009	0.245	-0.208	0.272	-0.163	0.051
Cross Asset Trend	0.192	0.276*	-0.107*	0.014	0.012	0.298	-0.104	0.732***	0.499**	0.452***

This table presents beta coefficients and t-statistics obtained when risk premia strategy returns are regressed against global equity and bonds returns. MSCI World AC is used as benchmark for global equity and Barclays Multiverse as global bonds proxy. Beta estimation is based on monthly returns. *,** and *** indicate statistical significance at the 5%, 1% and 0.1% level, respectively.

Figure 2.2: Performance of simulated portfolios based on ARP strategies



Monthly cumulated return data used from Jan 2010 to Aug 2018.

In Panel A of Figure 2.2, we graph the cumulated excess returns of the long-only risk parity portfolio with and without a tactical overlay. Panel B of Figure 2.2 displays the isolated performance of the tactical long/short portfolio. The reported performance data are net of transaction costs and calculated from January 2010 to August 2018 as we are limited as a result of the data availability of certain risk premia strategies. Further, we need one year of common data history to calculate the strategic weights of the risk parity portfolio and the ex-ante volatility.

Table 2.3 gives an overview of performance figures for the constructed portfolios. Applying a tactical overlay based on economic information increases the Sharpe ratio of the strategic portfolio further from 1.46 to 1.57. Investigating the tactical portfolio itself, the strategy achieves a consistent performance over the whole sample, generating a Sharpe ratio of 0.88. Observed returns are significantly positive according to the one-sided Welch test. To account for a non-Gaussian distribution of portfolio returns we further conduct the non-parametric Wilcoxon test, which leads to the same indication. The tactical portfolio had a period of weak performance from September 2012 to June 2014, recording its largest drawdown of 1.6%.

Table 2.3: Performance measurements of ARP portfolios

	Strategic	**Active (total)**	**Tactical (long/short)**
Excess return p.a.	8.20%	9.14%	0.82%
Volatility p.a.	5.61%	5.82%	0.94%
Sharpe Ratio	1.46	1.57	0.88
Max. Drawdown	8.6%	8.5%	1.6%
Value at Risk	-2.44%	-2.59%	-0.43%
Skewness	-0.69	-0.79	-0.88
t-statistics	4.064***	4.246***	2.615***

Returns are net of transaction costs. Value at Risk is calculated for monthly return data on a 95% confidence level. t-statistics pertain to the one-sided null-hypotheses that mean portfolio returns not being above 0% using Welch test. Corresponding p values are reported in parentheses. *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

We further investigate the possibility of explaining observed conditional return patterns by underlying beta exposure. The importance of making the distinction between performance as a result of true alpha instead of implicit beta in the context of tactical asset allocation is also stressed by Lee (2000). If conditional beta exposure was the driving force behind identified patterns, then the approach would ultimately transform a factor timing strategy to a rather expensive beta timing strategy. In order to account for that attempted explanation we calculate the implicit strategy returns based on the regime dependent betas ($\beta_R^{EQ}, \beta_R^{FI}$) reported in Table 2.2. We use the return series and construct an implicit beta portfolio combining the implicit returns with the tactical weights w^{TAA} of the business cycle model. The daily return for the implicit beta portfolio r^{BETA} is defined as follows:

$$r^{BETA} = \sum_{k=1}^{n} w_k^{TAA} \, * (\beta_{k,R}^{EQ} * r^{EQ} + \beta_{k,R}^{FI} * r^{FI})$$

Assuming the conditional betas would perfectly explain observed strategy returns, explicit and implicit returns would be identical. However, as can easily be seen from Panel B of Figure 2.2 returns of the implicit beta portfolio hardly distinguish from zero and return differences between both portfolios are significantly different from zero. On the basis of this analysis, we can exclude underlying beta exposure of actively allocated ARP strategies as an explanation for documented performance patterns.[14]

We are aware that we present only a short backtest horizon with few observations for each regime as we are limited by the available data history. In addition, several of the risk premia strategies have a major stack of back-tested data themselves. Therefore, having a somewhat idealized performance is not overly surprising. However, the major purpose of this chapter is not to implement a fully sophisticated portfolio timing strategy.

[14] We further extend the multivariate regression by adding commodity returns as explanatory variables. However, as conditional betas have only a marginally effect on implicit strategy returns, we withdraw from reporting the results.

We aim to give a realistic impression of macroeconomic sensitivities and the performance potential in exploring these dependencies in a tactical framework, considering a real investable set of risk premia strategies and rebalancing costs.

2.5. Cross Validation

Calculated macroeconomic sensitivities and the derived allocation scenarios are based on the full sample because of data limitations. An approach based on in-sample strategy development followed by an out-of-sample test is practically infeasible as the number of observations per regime is relatively small. In applying a data-splitting approach, the in-sample set would already require most of our available sample, leaving hardly any remaining data for an out-of-sample test. In an attempt to validate our previous findings without having sufficient data history, we repeat the described procedure using global multi asset classes and style factors to compare results with idealized asset behavior on the one hand and to examine the predictive power of the business cycle model given a different set of investments on the other hand.

2.5.1. Global Multi Asset Classes

First, just as we did with risk premia strategies, we investigate the conditional performance of regarded multi asset classes. Results are reported in Table 2.4. In line with the assumptions we have made, equities have strong differences in conditional excess returns. Performance in expansion and particularly in the recovery phase is highly positive while the opposite is true in recession. There is a considerable drop in returns for equities in the peak phase compared with the expansion phase. While most of the conditional equity performance in the peak phase remains positive, an expected increase in dispersion can be observed.[15] Documented results are highly significant at the 1% level.

Table 2.4: Conditional excess return of multi asset strategies

	Annualized excess return in %					**ANOVA**		
Multi Asset strategy	**Expansion**	**Peak**	**Recession**	**Recovery**	**Overall**	**Obs.**	***F*-statistics**	***p* value**
Equity North America	14.55	3.87	-19.76	16.37	3.81	199	11.666	<0.001***
Trading signal	+1	+1	-1	+1				
Equity Europe	12.95	7.28	-24.50	16.19	2.73	199	12.761	<0.001***
Trading signal	+1	+1	-1	+1				
Equity Japan	16.59	-2.07	-19.37	13.42	2.30	199	10.722	0.002***
Trading signal	+1	-1	-1	0				
Equity Pacific ex Japan	16.68	9.31	-17.49	18.17	6.55	199	7.657	0.007***
Trading signal	+1	+1	-1	+1				
Equity Emerging Mkts.	21.40	8.24	-19.53	17.67	6.84	199	8.068	0.006***
Trading signal	+1	+1	-1	+1				
Global Treasuries	1.12	2.46	4.98	1.62	2.32	199	3.164	0.078*
Trading signal	-1	-1	+1	-1				
Global Credit	1.99	1.29	1.62	6.56	2.86	199	4.606	0.034**
Trading signal	-1	-1	+1	-1				
Securitized	1.30	1.00	3.97	3.64	2.35	199	2.541	0.114
Trading signal	-1	-1	+1	-1				
Global High Yield	5.84	3.41	-5.15	21.30	6.42	199	10.776	0.001***
Trading signal	-1	0	0	+1				
Gold	7.40	6.53	16.80	-3.03	5.90	199	1.980	0.163
Trading signal	0	+1	+1	-1				
Inflation-linked Bonds	3.39	1.81	5.05	3.51	3.31	199	0.572	0.451
Trading signal	-1	-1	+1	-1				

Mkts. = Markets; Obs. = Number of monthly observations given for the data sample. *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Findings for the bond market are equally strong with global Treasuries outperforming in recession while underperforming in expansion, an economic environment of steady growth and rate hike anticipation (Ludvigson and Ng, 2009). In regard to the regime-related return differences between Global High Yield and Global Credit, results match

[15] Results for regime-related volatility are not reported here.

the conditional performance of the Credit Carry premium strategy. While the return difference and the strategy performance are especially positive in recovery and expansion, Global High Yield relatively underperforms in recession, when investors likely look to diminish default risks (see Table 2.2 and 2.4). In addition, safe-haven assets Gold and inflation-linked Bonds perform best in recession, when the demand from market participants for risk-averse assets is peaking.

However, while return patterns for Gold fit the expectation, results are not statistically significant. This finding might be partially explained by the temporary indication of a recession in the context of the dotcom crisis in the early 2000s. While equity markets were hit, this was not an idealized type of recession in which inflation and unemployment typically pick up while growth expectations decrease. Hence, the markets demand for Gold was rather low and so was the conditional return during this period.

Finally, we construct a long/short portfolio using a straight quantitative approach to set up allocation scenarios for the traditional multi asset classes. We rank the assets according to the conditional excess returns and form the phase-related portfolios accordingly (see Table 2.4). It is noteworthy that while most of the constructed allocation scenarios fit the economic intuition of conditional asset behavior, we acknowledge that the resulting allocation scenario of the peak phase is biased on the data sample used. Assuming that the return of equities will eventually decline with progressing duration of the economic slowdown, the origin of any global recession remains dynamic. Therefore, relying solely on the observed return patterns, particularly in this regime, is not recommended for practical application.

The reported performance is calculated from January 2002 to August 2018. We graph cumulated excess returns of the timing strategy as the bold line in Figure 2.3. The active strategy has a consistent active performance over the whole sample, generating a Sharpe ratio of 0.79 without transaction costs.

We withdraw from accounting for costs as most of the assets can be traded at single digit basis points nowadays. In any case, the impact on the overall performance should be rather low at a two-way turnover of only 68% p.a. The portfolio had a long period of flat performance from October 2002 to May 2005 as inconclusive signaling from the economic regime indicator in the aftermath of the mentioned dotcom crisis affected the active performance. Maximum drawdown was at 1.6% for an ex-ante portfolio volatility of 1% p.a.

2.5.2. Equity Style Factors

In Panel A of Table 2.5, we report the active performance of MSCI smart beta style indices conditional on each economic regime. During the expansion phase well-established investment trends in equity markets tend to support momentum stocks. Typically, only a small share of the overall stocks accounts most of the markets performance as the economy reaches its turning point (Bessembinder, 2018). In peak, this mechanism seems to intensify as the outperformance of momentum stocks is increasing even further.

However, a growing sense of risk awareness among market participants might be responsible for minimum volatility and quality strategies having above-average returns in the peak phase as well, with the economic development starting to slow down (Ang et al., 2006). This trend continues in recessions when the economy is exposed to different types of shocks. Investors' demand is high for crisis-proof assets, and companies with relatively low leverage, stable earnings, and high profitability (RMW) might be favored (Hodges et al., 2017). Finally, when the economy recovers from its trough, size and value strategies witness the highest relative performance, with Sharpe ratios at about one.

Table 2.5: Conditional Sharpe ratio of equity style factors

	Annualized Sharpe ratio						ANOVA	
Equity style factors	**Expansion**	**Peak**	**Recession**	**Recovery**	**Overall**	**Obs.**	***F*-statistics**	***p* value**
Panel A: MSCI World smart beta style indexes								
Momentum	0.75	1.26	-0.34	0.54	0.32	199	0.774	0.510
Trading signal	+1	+1	-1	0				
Size	0.47	0.17	0.09	1.06	0.34	199	2.680	0.048**
Trading signal	0	-1	0	+1				
Value	1.57	-0.31	-0.30	0.94	0.47	199	1.957	0.122
Trading signal	+1	-1	-1	+1				
Quality	-0.37	1.30	0.58	0.13	0.34	199	3.303	0.021**
Trading signal	-1	+1	+1	-1				
Minimum Volatility	-0.33	0.38	0.40	-0.22	0.11	199	1.438	0.233
Trading signal	-1	0	+1	-1				
Panel B: Fama/French five-factors								
MKT	1.50	0.71	-0.71	1.81	0.48	199	4.746	0.003***
Trading signal	+1	+1	-1	+1				
SMB	0.28	-0.29	0.15	1.18	0.34	199	1.430	0.235
Trading signal	-1	0	0	0				
HML	0.48	-0.51	-0.64	0.79	0.04	199	2.197	0.090*
Trading signal	+1	-1	-1	-1				
RMW	0.46	0.13	0.99	0.13	0.43	199	0.983	0.402
Trading signal	0	+1	+1	-1				
CMA	-0.62	-0.66	0.39	1.61	0.26	199	5.364	0.001***
Trading signal	-1	-1	+1	+1				

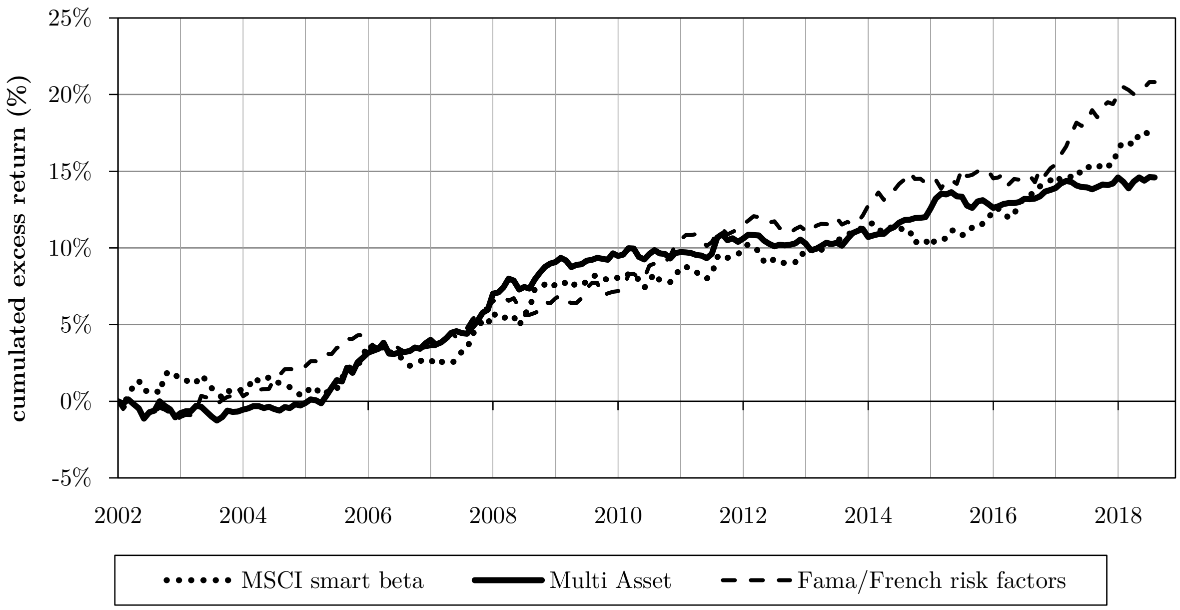
Obs. = Number of monthly observations given for the data sample. *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Although, besides fitting economic intuition, reported results for momentum are statistically non-significant (Chordia and Shivakumar, 2002).[16] Noisy performance might explain that effect, particularly in recovery and recession phases. Momentum strategies suffer most from swift changes in investment trends typically occurring in these regimes (Daniel and Moskowitz, 2016). That leads to a temporary underperformance, but with an

[16] Conditional performance of the momentum style factor shows little conformity with return patterns of the Equity Trend strategy (see Table 2.2 and 2.5). While being insignificant, this can at least partially be attributed to the diverging investment style. The momentum factor selects past winners versus past losers on the single-stock level, in contrast the Equity Trend strategy operates on the index level.

extending duration of the prevailing regime, the strategy will eventually adjust its allocation, and so will the conditional performance.

Figure 2.3: Performance of L/S portfolios based on multi asset classes and equity style factors



Monthly cumulated return data used from Jan 2002 to Aug 2018.

Observed return patterns for size and value can be validated with the equivalent SMB and HML factors defined by Fama-French. Compliant with the conditional active performance of the MSCI smart beta indices, both factors record below-average Sharpe ratios in the peak phase as well as in recession while being highest in recovery (Panel B of Table 2.5). Stocks with a conservative investment policy seem to underperform when risk appetite is high, as in expansion. With investors rather cautious in recession and maybe still cautious in the recovery phase, low investment companies tend to be relatively attractive in these regimes. The regime-related extent of the equity premium (MKT) delivers supporting evidence for the conditional attractiveness of global equities reported in the prior section. Our findings are very nearly in line with the economic regime–related factor returns analyzed in previous studies (e.g., Amenc et al., 2018). However, observed return differences are significant for only half of the considered equity style factors. With the sample being rather short because of restrictions for the macroeconomic data, evidence for the timing capability of equity style factors remains scarce. In particular, the time-varying relationship between indicators and factors and the existence of temporary investment trends make a dynamic approach a necessity (Bender et al., 2018). Therefore, the general application for active allocation and interpretation of equity style factors is less straightforward as for much of the considered ARP strategies.

Table 2.6: Performance measurements of L/S portfolios

	Multi Asset	**MSCI smart beta**	**Fama/French risk factors**
Excess return p.a.	0.80%	0.94%	1.07%
Volatility p.a.	1.01%	1.03%	1.24%
Sharpe Ratio	0.79	0.91	0.86
Max. Drawdown	1.6%	2.1%	2.2%
Value at Risk	-0.35%	-0.41%	-0.42%
Skewness	0.20	0.59	-0.03
t-statistics	3.598***	3.362***	4.292***

Value at Risk is calculated for monthly return data on a 95% confidence level. t-statistics pertain to the one-sided null-hypotheses that mean portfolio returns not being above 0% using Welch test. *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

We use the reported regime-related performance to set up allocation scenarios for both sets of equity style premia. When active performance is back-tested according to the methodology already used for the multi asset classes, both timing portfolios record highly positive Sharpe ratios at 0.91 and 0.86, respectively, using the MSCI smart beta indices and Fama-French risk factors (see Table 2.6). The reported performance is gross of all transaction costs and calculated from January 2002 to August 2018 for an ex-ante volatility of 1% p.a. Cumulated returns of both long/short portfolios are illustrated as dotted and dashed lines in Figure 2.3. According to Table 2.6 maximum drawdowns were at 2.1% and 2.2%.

2.6. Conclusion

By analyzing macroeconomic sensitivities, we show that time-varying returns of ARP strategies are significantly related to economic conditions. Our results distinguish from the findings made by Ilmanen, Maloney and Ross (2014). The discrepancy can be at least partially explained by the composition of analyzed strategies. We construct long/short style premia strategies by combining different asset classes. However, as the strategies' performance is at least to some degree assumed to be related to the underlying asset behavior, and with asset classes not being perfectly correlated, it is in line with expectations that differences in regime-related aggregate returns are insignificant. To increase the probability of identifying meaningful return patterns we analyze macroeconomic sensitivities of asset class–specific ARP strategies. On the basis of our findings, we select strategies and construct a timing portfolio that illustrates the return potential of a regime-related allocation approach.

As transaction costs for ARP strategies remain relatively high compared with traditional asset classes, using a timing strategy with a low frequency of trading signals is preferable. In contrast to many other proposed active investment strategies with reported high Sharpe ratios but also high turnover, our business cycle model achieves a tradeoff between signal frequency and reactiveness to changing economic circumstances.

Given the real data set of tradable ARP strategies and the accounting of transaction costs, the calculated performance of the active portfolio gives at least a realistic impression of the possible profitability of using factor timing as a tactical overlay. To test the hit ratio of our business cycle model we investigate the conditional returns of different asset classes and style factors. Results are in line with theoretical assumptions and the findings of previous studies (e.g., Hodges et al., 2017; Blin et al., 2018).

With our given look-back period being relatively short because of data limitations, further research analyzing different data sets of ARP with an extended history to validate our findings might be necessary.

However, since the underlying causal links might be time-varying, a dynamic approach to select macroeconomic data for predicting the business cycle and to control for adjustments in sensitivities of ARP strategies is a must. Therefore, future studies might focus on developing models capable of grasping the ever-changing structural macroeconomic environment in selecting explanatory data in a continuous approach.

3. Real-Time Transition Risk

3.1. Introduction

Despite the growing empirical evidence that investors should consider the economic effects of climate change, the integration of climate-related risks into the investment process faces persistent challenges (Dyck et al., 2019; Krueger et al., 2020; Bolton & Kacperczyk, 2021b). The lack of consensus on how to measure climate risks and their expected impact on asset returns are two obstacles to overcome for practitioners and academics alike. Generally, the literature on climate finance distinguishes the economic effects of two broad categories of climate-related risk factors: risks related to the possible transition to a lower-carbon economy and risks associated with the physical impacts of climate change. It is essential to acknowledge that the realizations of both risk types are likely to move independently regarding time and direction (Giglio et al., 2020). In this context, it is indispensable to measure the effect of climate-related events, such as natural disasters or emission limits, on physical and transition risks separately when inferring climate-related risk.

Our research contributes to the emerging literature on climate-related risk in two ways. First, we create a point-in-time index to isolate changes in transition risk from climate-related news events. Previous approaches of news-based risk index construction (Engle et al., 2020; Ardia et al., 2020) focus either on media attention towards climate change or solely on negative climate-related news flow as a proxy for climate risk. Thereby, the authors follow the premise “no news is good news on climate”. While the saying may often hold true in the past, news of climate initiatives (like the Paris Climate Agreement or Green New Deal), the emergence of disruptive clean technology, or an increase in consumers’ or investors’ climate awareness do not support approaches based on this simplifying assumption. In this study, we address the possible shortcomings of this assumption in index construction and explicitly consider news to signal an increase or a decrease in the external pressure towards a shift to a lower-carbon economy. By doing so, we enhance the approach for climate news-based risk approximation innovated by Engle et al. (2020) and develop, to the best of our knowledge, the most comprehensive framework to approximate changes in real-time transition risk from news events so far.[1]

We manually label thousands of climate-related news items and train a BERT-based classification model to predict the implied impact of a news event on transition risk. Specifically, we consider news to signal an increase or a decrease in the external pressure towards a shift to a lower-carbon economy. To ensure the economic foundation of the classification process, we link our approach to the drivers of transition risk as established by various climate risk frameworks (e.g., TCFD, 2017; NGFS, 2020a; BCBS, 2021) and to existing academic research on the different transmission channels. To identify climate-related news articles from general news flow, we apply domain-specific dictionaries with point-in-time vocabulary. Look ahead bias inherent in most existing approaches of news indices is thus avoided. We believe the correct specification of information sets is of utmost importance for the evaluation of investors' reactions to climate news. From the aggregated sentiment of all news identified to be climate-related, we ultimately construct a global Transition Risk Index (TRI).

Second, we analyze the exposure of presumably climate-friendly portfolio returns towards innovations in transition risk - without making a-priori assumptions on the “green credentials” of firm characteristics. Our work seeks clarification which type of current climate investment approaches provide exposure to the risk and opportunities related to

[1] The Transition Risk Index data are available at https://www.finance.uni-wuppertal.de/de/research-data/transition-risk-index.html

climate transition. Investors that want to hedge transition risk will desire portfolios that perform well (poor) if the public demand to confront the adverse effects of climate change is rising (falling). Therefore, we select various, publicly available proxies of "green" investment portfolios and distinguish two different construction approaches to measure a firms' environmental performance: Decarbonized portfolios that appeal to benchmark-oriented investors who want to mitigate carbon price risk. Pure-play approaches refer to investors who wish to directly allocate to companies that generate a significant share of revenues from products and services related to environmental challenges (Andersson et al., 2016). We run regression models considering a broad set of factors and find a significant contemporaneous relationship between the active weekly returns of pure-play portfolios and innovations in the Transition Risk Index. From an economic perspective, these findings make sense. The potential inclusion in pure-play portfolios is based on a firm's involvement in business activities like producing alternative energy, pollution prevention, waste management, or the provision of clean technology and equipment. The business development of these activities is supposed to be causally related to the environmental policy and consumer preferences. Therefore, changes in transition risk (e.g., induced by subsidies for energy-efficient products or the mitigation of environmental standards) shall alter the business outlook for pure-play companies with a corresponding effect on investors' return expectations.

In contrast, we find no contemporaneous relationship between the active weekly returns of decarbonized portfolios and changes in transition risk. Carbon-tilted portfolio optimization does not address companies' exposure to environmentally beneficial business activities and even may result in counterintuitive portfolio allocation (UNEP FI, 2015). For example, in its most simple form, excluding energy stocks may also cause the exclusion of renewable energy stocks due to sector affiliation. Further, most companies from sectors with relatively low emission exposure (e.g., healthcare, information technology and financials) will not profit directly from more restrictive environmental regulations. Neither will a weakening of environmental policies result in an expected decline in demand of their products and services. Thus, future revenues and consequently return expectations should be hardly related to changes in transition risk, consistent with regression results.

Carbon metrics are the main contributor to environmental scores (Berg et al., 2019). Consequently, ESG and decarbonized portfolios should be correlated to some extent, given the partial similarity in portfolio construction. ESG-related capital flows are found to be driven by investor preferences (e.g., Pastor et al., 2020; Hartzman & Sussman, 2019). Hence, we also expect trends in sustainable investment to impact active returns of decarbonized indices. We incorporate both climate news indices from Engle et al. (2020) to account for a media attention-based explanation of active index returns. In this context, relative news coverage of climate-related topics is considered as a proxy for the salience of climate change to investors' attention. We find no significant coefficients for both index approaches with previous indications generally remain unchanged. Hence, we are unable to provide evidence for an attention-based explanation of carbon-tilted portfolio returns. However, for this analysis we only considered contemporaneous returns with a limited number of monthly observations.

The rest of the chapter is structured as follows. In the next section the framework for inferring implied transition risk from climate-related news events is introduced. In section 3.3 presents an unsupervised language model designed to generate domain- and period-specific vocabulary from point-in-time data. This model is used to identify climate-related news out of sample. Section 3.4 describes the applied sentiment model for the construction of the Transition Risk Index. In Section 3.5, the index is compared to existing news-based risk approaches. Different regression model specifications are used to analyze the return sensitivity of commonly used portfolio approaches towards transition risk. The chapter concludes with a discussion of the findings.

3.2. Transition risk framework

In the following section, we present our approach to approximate changes in transition risk with climate-specific news. To provide profound guidance on how to manually label climate-related news events for their inferred change in transition risk, we align our approach to widely accepted climate risk frameworks and existing academic research.

In 2017, the Task Force on Climate-related Financial Disclosures (TCFD) established a common framework for climate-related financial disclosures to support stakeholders in assessing the potential financial impacts of climate change on business activities. Since its release, the TCFD recommendations have been strongly supported as industry-standard on climate-related risk disclosure and incorporated into various sustainability-related disclosure frameworks.[2] An essential element of these frameworks is the consistent categorization of climate-related risks and opportunities. We combine the insights from various climate frameworks and identify three different drivers of transition risk that organizations should consider: *Policy and legislation* (e.g., environmental and emission standards), *technology* (e.g., decrease of production costs for renewable energy), and *climate awareness* (e.g., shifts in consumer preferences).[3] These drivers represent climate-related adjustments that could generate, increase or reduce transition risks via different

[2] https://www.fsb-tcfd.org/supporters/

[3] Generally, there are assumed to be strong interaction effects between the different risk drivers. In the initial report of the TCFD (2017) market risk is identified as another dimension of transition risk that describes the risk of shifts in supply and demand for certain products, and services as climate-related risks and opportunities are increasingly considered. Here, we follow the assessment of the UNEP FI (2019) to treat market risk within the other dimensions, as policy and legislation, technology, and climate awareness are seen to be the underlying risk drivers of shifts in existing supply-demand patterns.

transmission channels. Transmission channels include the causal chains by which risk drivers affect companies. Next, we describe the three drivers of transition risk and its transmission channels in detail. We provide examples of news events that we identify to imply changes in transition risk based on economic intuition and academic findings for the different transmission channels.

Policy seeking to regulate the carbon externality is a crucial driver of transition risk. Generally, government policies to protect the climate either attempt to constrain actions that contribute to the adverse effects of climate change or to encourage the shift to a low-carbon economy (BCBS, 2021). The central element of most perspective climate change mitigation mechanisms is incentive-based regulation that prices emissions via taxation or trading systems. Hence, effective mitigation policies could significantly increase industry and consumer prices for high-carbon products in the future (Hepburn, 2006). Examples of news events we expect to signal an increased policy risk are planned sales limits of combustion engine cars (Burch & Gilchrist, 2018), mandatory carbon emission caps, or the phase-out of fossil fuels from energy production (Monasterolo & Raberto, 2019). Vice versa, a weakening of planned or existing environmental standards, the enactment of insufficient regulation (e.g., Gössling et al., 2014), the lifespan extension of coal plants, or political resistance against climate measurements is assumed to reduce future transition risk. Restrictive regulations are just one transmission channel of policy risk. Public subsidies and investments can reduce low-carbon products' production costs or consumer-end prices and create transition opportunities (Semieniuk et al., 2021). Pro-climate policy initiatives such as the European Green Deal (European Commission, 2019) support sustainable growth by mandating and incentivizing low-carbon products or by adopting energy-efficient solutions. With policies creating markets and innovation clusters or providing investment into much-needed infrastructure (e.g., charging points for electric vehicles), low-carbon products become more competitive over time. Exemplary news events that imply transition opportunities focus on the application of renewable energy quotas, tax benefits for carbon-neutral revenues, or consumer incentives for solar panels.

Litigation or *legal risk* strongly relates to policy risk, as changes to transition risk may result from legislation or jurisdiction. Climate-related litigation claims are being brought before courts by property owners, municipalities, states, insurers, shareholders, and public interest organizations and are closely tracked by news media.[4] Reasons for such litigation may include the reluctance of organizations and authorities to mitigate the effects of climate change, failure to adapt to climate change, and insufficient disclosure around material financial risks (TCFD, 2017). However, lawsuits against enacted environmental guidelines or the appeal of the competencies in the elicitation of regulations may - in case of success - ease pressure on affected corporations and polluters to align with environmental standards, and therefore reduce transition risk. Another transmission channel of legal risk is the sanctioning or penalizing of misbehavior by firms, organizations, or member states for breaching climate rules or emission caps. A lack of consequences in case of misconduct will likely lead to a deteriorating willingness of violators to comply with the political framework. In contrast, serve punishment is more likely to alter climate actions of former and future violators (Aldy & Stavins, 2012).

The second transition risk driver consistently identified by various climate risk frameworks refers to *technological* improvements or innovations as developments that promote the transition to an energy-efficient economy can drastically impact industries. For example, the application of new technologies such as alternative energy sources and

[4] Courts worldwide are adjudicating a growing number of disputes over actions (or inaction) related to climate change mitigation and adaptation efforts. In recent years, both the number of cases and the number of countries where global warming lawsuits have been filed have increased significantly (UNEP, 2017). Overall, more than 1,500 cases of climate litigation were registered between 1986 and May 2020 with commercial disputes not included (Setzer & Byrnes, 2020).

battery storage will affect the competitiveness of organizations with exposure to fossil fuels, their production and distribution costs, and ultimately the demand for their products and services. To the extent that new technology displaces old systems and disrupts some parts of the existing economic system, winners and losers will emerge from this development cycle (TCFD, 2017). Most drivers affect green business operations positively as technological adjustments, possibly subsidized by former climate policies, alter relative prices in favor of low-carbon products (Kavlak et al., 2018). By implication, technological innovations related to climate change negatively threaten expectations about carbon-intensive industries as corporates' existing business models may be based on technologies that are likely to become effectively superseded (BCBS, 2021). With business models of firms that operate in traditional industries under threat, transition risk is supposed to rise in the future. The realization of technological risk results in pressure to adapt and invest in order to remain competitive. Increased (public or private) investments into research and development of environmental technology or the progressive adoption of green innovations within an industry are common, exemplary news events that imply increasing technology risk.

Besides the assumption that the transition to a low-carbon economy is mainly driven by the supply side (due to regulations or technological innovations), academic research intensifies on how shifts in consumer preferences for certain goods and services could accelerate the transition and contribute to achieve climate goals (NGFS, 2020b). As defined by the TCFD (2017), *Reputational risks* can be a leading force for transition risk regardless of political or technical initiatives. These risks arise from changing customer perceptions of an organization's contribution to or detraction from the transition to a more energy-efficient economic system and may trigger shifts of consumer sentiment to less carbon-intensive products. Consequently, firms that continue to operate in carbon-intensive sectors risk the impairment of their products even in the absence of regulatory pressure. Other typologies (e.g., Cambridge Centre for Sustainable Finance, 2016; BCBS, 2021) mainly focus on investor sentiment as a complementary driver of transition risk to consumer sentiment. Here, we adopt the more comprehensive approach of the NFGS (2020a) by extending the scope of the risk dimension to cover the effect of shifts in overall *climate awareness*.

Climate awareness is considered the amalgamation of climate-related shifts in public and private sentiment, demand patterns, and preferences and expectations. We expect climate awareness to transmit transition risk in different ways. First, we suppose that more public attention to climate change (e.g., due to the materialization of physical risk in the form of natural disasters) indicates a shift in consumer preferences and, therefore an expected growth in demand for green products (Wells et al., 2011). Additionally, high public awareness may translate to political actions. As public awareness of the consequences of inaction on climate change improves, so should the acceptance of tough measurements to fight global warming. Awareness may turn to public demand (as expressed by widespread climate protests) for regulations at a specific point. With climate policies becoming essential for voting decisions, policymakers are likely to act on public pressure, which significantly increases the probability of environmental regulations being adopted.[5]

Changes in consumer preferences and perceptions are difficult to observe directly. However, news about surveys that analyze consumer acceptance of higher prices for green products, or the importance of environmental aspects for buying or investment decisions give insight into how consumer preferences may alter. News media reporting about

[5] Based on their general equilibrium model, Pastor & Veronesi (2012) find that a new policy is more likely to be adopted if the authority derives an unexpectedly large political benefit from changing its policy, even if the regulations that will be replaced, worked well in the past.

shareholders filing climate resolutions or pension funds pressuring firms to disclose emission risk are considered a positive shift in investor sentiment.[6] Concerning public attention on climate change, media coverage of protests for swift climate action strongly indicates that climate awareness is rising (Ramelli et al., 2021). Resistance against the construction of coal plants or the call for a boycott of products from firms with bad environmental records is seen as a reputational risk. Accordingly, we assume news that imply a surge in climate awareness to signal an increase in transition risk. In contrast, news about the widespread denial of human-induced climate change and its adverse effects or consumer reluctance to adjust buying behavior in favor of more environment-friendly products are expected to be negative for a transition to a low-carbon economy.

We consider external pressure to be instrumental for a successful shift to a low-carbon economy and for the attempt to reach climate goals. Consequently, news events that infer an increasing transition risk are defined to be positive on climate. News events that imply a decrease in external pressure and respectively transition risk are deemed to be negative for achieving emission targets. We label news samples accordingly.

3.3. Climate-related news and point-in-time language model

Before we provide details on the construction of our sentiment classification model to approximate changes in transition risk, we present a language model designed to select climate-related news from raw text data, utilizing only point-in-time, domain-specific vocabulary. For this purpose, we use the services of the news data analytics provider *RavenPack*. We apply the API document search to screen more than one hundred million

[6] Flammer, et al. (2019) find that environmental shareholder activism (regardless of mandated disclosure requirements) increases the voluntary disclosure of climate change risks, especially if initiated by institutional investors.

of news articles, starting in January 2000. RavenPack covers news articles and social media posts from a variety of sources. We choose a subset of the most relevant media outlets, including the Dow Jones Newswires, Reuters, BBC, WSJ, The New York Times, The Washington Post, MSN, and CNN. Further, news sources are selected to cover the international news flow as we analyze the potential impact of transition risk on global stocks. Figure A3.1 shows the share of news articles by media source and country for the news data sample from January 2000 to December 2020. The country-related share of news topics roughly aligns with the country weights of a market capitalization-weighted, global stock universe (e.g., MSCI ACWI). Finally, we filter the data sample using metadata to ensure only non-corporate news flow is included. We take a macroeconomic perspective to approximate changes in transition risk from relevant news articles.

We divide the process of climate news index development into three parts: Domain-specific vocabulary construction, topic identification, and sentiment classification. For each task, we utilize either information from the news article texts or headlines. The decision of applying either full-text news articles or headlines for text analysis is based on the use case and a trade-off between noise reduction (Nassirtoussi et al., 2015) and the risk of incongruency (Ecker et al., 2014). While incongruent headlines do not accurately represent the information contained in the article e.g., due to click-bait or sensationalism (Molek-Kozakowska, 2013), this effect is less prevalent for fact-based, non-emotional news reports (Thomson et al., 2008) and reputable, balanced news media outlets (Dor, 2003; Lindemann, 1990). In terms of sentiment classification, one has to differentiate between a lexicon-based classification approach that scores on domain-specific or emotionally charged expressions (Loughran & McDonald, (2011, 2016)) and a contextual-based classification approach. The latter is modeled on the word representation of domain-specific texts and a customized methodology to analyze the deeper semantic meaning of a given text by considering the positional relationships and dependencies of expressions (Kraus & Feuerriegel, 2017). Lexicon-based approaches are commonly used to approximate the conveyed sentiment and tonality on the article level. In contrast, transfer learning or contextual-based approaches by construction classify sentiment on short text elements like sentences. However, by aggregating sentiment on the full-text news article, one risks deflating the impact of the most relevant event as an article is often enriched with additional and potentially noisy information, e.g., by referencing quotes, discussions, or historical context. We specifically opt for headlines here to increase the interpretability and simplicity of our contextual, domain-specific sentiment classification approach as we want to identify regulatory risk based on relevant, headline-grabbing climate events. Therefore, we assume the article's most important information or event to be summarized in the headline. To confront concerns regarding incongruency of headlines, we select only accountable and reliable news sources given the highest ranking of trustworthiness by news analytics provider RavenPack.

With this significant degree of freedom, we will either take the most straightforward approach or provide a reasonable explanation to our design choices in index construction.

3.3.1. Domain-specific vocabulary

In their paper, Engle et al. (2020) use a topic-specific dictionary to measure the similarity between a “Climate Change Vocabulary” and news articles in the Wall Street Journal (WSJ). To generate their topic-specific dictionary, the authors manually select white papers and glossaries concerning climate change from 1990 to 2017 and select the most frequent terms from the merged text corpus. While this straightforward approach likely leads to the identification of the most relevant vocabulary related to climate change, it has two disadvantages. First, defining a dictionary containing information from the same sample period that is also used for the subsequent topic identification, results in a look-ahead bias. Time-dependent events and developments shape specific terms and vocabulary (e.g., "Kyoto protocol", "Fukushima", "Paris Agreement", "Green New Deal"). Given their approach, contextual terms will be considered for topic identification before they are mentioned in news media and thus deemed related to climate change. Secondly, time dependence can be a disadvantage itself within a stationary dictionary. By classifying news based on a dictionary that aggregates information from more than 20 years of data, time-dependent terms may be underrepresented during periods in which these terms were actually of increased relevance to the related topic.[7]

Instead of manually selecting topic-related documents for dictionary construction, we develop an unsupervised algorithm that utilizes the information from millions of unclassified news items. Figure A3.2 illustrates our approach for automated dictionary construction. The objective of the algorithm is the generation of a dictionary that strongly relates to a domain-specific buzzword. We provide the term "Climate Change" as input parameter.[8] Then, the algorithm selects news (from a given period) that contain the buzzword in the full-text article. The most frequent terms are calculated from the headlines of the selected news articles (in the following, referred to as the headline corpus). However, rigorous text normalization and cleaning raw headline data is required to generate a dictionary with a high degree of pureness in domain-specificity. At first, language-based stop words[9] and text elements with less than three characters are removed

[7] We consider this to be an important aspect that could even improve the out-of-sample accuracy of a time dependent approach compared to a stationary approach that does not account for any look-ahead bias.
[8] Generally, the number of provided buzzwords is unrestricted. A single expression refers to the minimum input required.
[9] Stop words are predominantly the most used words of a given language. Generally, stop words are removed because they are not relevant for dictionary construction and distort the word frequency analysis.

from each headline text. Next, Named Entity Recognition (NER)[10] is performed to identify entity-related terms that should not be considered for dictionary construction. Specifically, names of persons, organizations, and locations are dropped as their inclusion may lead to an undesired bias towards entities and, hence, inaccuracy in the subsequent topic identification task. After lemmatization[11], terms are formed from a contiguous sequence of up to three text elements (unigram to trigram) for each cleaned headline. Next, we calculate the term frequency by counting the appearance of each term, divided by the total number of terms in the cleaned headline corpus. The frequency of terms allows us to compare the relevance of specific terms over dictionaries with different sample sizes. Finally, term frequency calculation is repeated for all unselected news items in the sample (period) to derive a list of frequently used terms in general news. The most commonly used terms (e.g., "stock", "market", "rate") are automatically withdrawn from the topic-related dictionary as specific vocabulary is desired. [12]

This process of buzzword-based vocabulary generation may be repeated multiple times if the initial buzzword has synonyms or strongly related terms. The total number of iterations I defines the amount of most frequent terms that are applied for dictionary construction itself. Re-running the process of vocabulary generation by extending the list of considered, topic-specific buzzwords – by the most frequent and unused terms of the initial search process ($i = 1$) – results in more observations and possibly noise reduction.

[10] Named Entity Recognition (NER) is the process of locating named entities in unstructured text and then classifying them into pre-defined categories. We use the application from the Stanford NLP Group (Stanza) for all NER tasks in this paper.

[11] Lemmatization is the process of reducing inflected forms of a word while ensuring that the reduced form belongs to the language. The initial word is stored to retrieve the original expression from the reduced form. These untrimmed words are needed for the content search process based on the full-text article.

[12] We set the number of most commonly used terms that are removed from the buzzword-related dictionary to 500. Given our approach this step is expandable and only applied for illustrative purposes. We use *tfidf*-scores for topic identification, consequently unspecific and commonly used terms will be of low relevance anyway.

However, there is a trade-off to be considered as an extending number of iterations also leads to an increased risk of incorporating non-related or less specific buzzwords while diluting the impact of the initial news search process. Hence, the number of iterations may be chosen with respect to the news sample size of the initial search process. Generally, the smaller the number of observations and the more synonyms is available for the initial buzzword, the more iterations may be beneficial. Given multiple iterations, the algorithm stops after the last iteration ($i = I$) and generates the terminal, domain-specific dictionary from the equally weighted term frequencies of the I buzzword-specific dictionaries. The aggregation of multiple dictionaries and the selection of overlapping terms lead to noise reduction.[13]

We run the process of vocabulary generation with five iterations at the end of each year to create a climate-specific dictionary that is used for topic identification in the following year. Each dictionary consists of the period- and domain-specific terms and their respective frequency. The uni- to trigrams are sorted by term frequency. We limit the number of terms in a dictionary to the 500 most frequent expressions. Afterward, the sum of term frequencies is scaled to 1. Finally, the resulting normalized term frequency (tf) represents the relative relevance of each term to the dictionary.[14]

The presented approach tackles the look-ahead bias and the problem of underrepresentation inherent in fixed dictionary construction. With topic-related vocabulary possibly changing over time, we generate period-specific dictionaries to account for the time-dependent relevance of terms. Each period-specific dictionary defines

[13] A threshold of minimum appearances is applied on each term at the dictionary level. By doing so, certain terms that are frequently but almost exclusively used in headlines for only some of the considered buzzwords, get removed from the aggregated dictionary. We set the threshold at 0.5, requiring a term to appear in at least half of the buzzword-specific dictionaries.

[14] The presented results are not sensible to the specific number of iterations in the vocabulary generation process.

the topic-related vocabulary that is used for next year's news classification.[15] Figure A3.3 exemplary shows word cloud summaries of period-specific climate change vocabulary. The upper word cloud shows the vocabulary used for topic identification in 2006. The vocabulary of the second word cloud is generated from news data of 2019. The size of each term refers to its respective frequency. The figure illustrates how the relevance of vocabulary is varying over time for “Climate Change”. While some terms seem to be of continued relevance over time (e.g., “Emission”, “Environment”), other frequent terms of one period are not to be found in the dictionary of the other period. For example, “Kyoto” is a widespread expression in the early 2000's referring to the ratification and adoption of the Kyoto protocol at the beginning of the century. However, since its ratification, more than 15 years have passed. With other climate treaties having replaced the Kyoto protocol, it is of little surprise that the term is not found in the vocabulary based on climate-specific news data from 2019. On the other hand, expressions like “Green New Deal”, “Extinction Rebellion” or “Sustainability” that strongly relate to the current climate debate were hardly used in the context of climate change-related discussions in the early 2000's. These findings indicate how topic identification and news index construction may benefit from a time dependent dictionary generation.

The main advantages of the presented approach are the opportunity to derive a domain-specific dictionary from news by providing as little input as a single term. Hardly any human intervention or supervision is necessary (with only a few parameters adjustable). This opens the possibility to extend this technique to different languages.

[15] Available data history starts in the beginning of the year 2000. For the first period only, we use the same (in-sample) data for dictionary generation and for topic identification.

3.3.2. Topic identification

Based on the domain-specific vocabulary, we want to score unseen news articles by their relation to the topic of climate change. Therefore, we apply the domain-specific dictionaries to approximate the similarity between the climate-related vocabulary of year t and any news article text of year $t+1$, for all years $T-1$ in the data sample. We follow the approach of Engle et al. (2020) and use a score based on the "term frequency-inverse document frequency" (*tf-idf*), which is often applied in information retrieval and text mining. The *tf-idf* is composed by two functions: (i) the normalized term frequency (*tf*), which we derive directly from the domain-specific dictionary (ii) the inverse document frequency (*idf*), computed as the logarithm of the number of articles in the corpus (of year t) divided by the number of articles in which the considered term appears.

Certain terms, such as "energy" or "climate", are commonly used in news articles but convey no specific information. Independent of their relevance to the topic-specific dictionary (as measured by *tf*), these highly frequent expressions are penalized by multiplying with a low *idf*. Hence, the *tf-idf* identifies the most representative terms that appear infrequently overall but frequently in domain-related documents. We calculate the *tf-idf* for each term in the dictionary of year t. The *tf-idf* scores are stored in a $n \times 1$ vector $\mathbf{v}$, where n refers to the number of terms in the dictionary:

$$\mathbf{v}_{t+1} = \begin{pmatrix} tfidf_1 \\ tfidf_2 \\ \vdots \\ tfidf_n \end{pmatrix}_t$$

We normalize vector $\mathbf{v}$ so that $\hat{\mathbf{v}} = {}^{\mathbf{v}}/_{\|\mathbf{v}\|}$ and $\|\hat{\mathbf{v}}\| = 1$.

Next, we want to score news articles of the following period for their usage of climate-specific vocabulary. Therefore, we construct a $m \times n$ matrix $\mathbf{A}$, where n refers to the number of terms in the dictionary (of year t) and m is the number of news articles in the corpus of year $t+1$:[16]

$$\mathbf{A}_{t+1} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \iddots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

The matrix stores binary information on whether an article text contains a given term, with any $a \in \{0,1\}$. Afterward, we perform multiplication of matrix $\mathbf{A}$ and vector $\hat{\mathbf{v}}$ to calculate the sum of *tf-idf* scores over all terms for each news article in year $t+1$, resulting in a $m \times 1$ vector $\mathbf{w}$:

$$\mathbf{w}_{t+1} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} = \mathbf{A}_{t+1} \cdot \hat{\mathbf{v}}_{t+1}$$

We define vector $\mathbf{w}$ to contain the article-based relevance scores with $0 \leq w \leq 1$. The relevance score w measures the weighted intersection in vocabulary between the article text and the domain-specific dictionary. A value close to 1 refers to the inclusion of the most representative terms in the article text. We implement two conditions for noise reduction in the process of topic identification. First, for a given news article to be considered (at least to some extent) climate-related we set a threshold of $w \geq 0.05$. The threshold for topic identification is selected by the inspection of the training sample and is based on a trade-off between sensitivity and specificity. Second, we require the article headlines to include one of the 20% most representative terms. By doing so, we want to exclude less specific article types like news summaries, highlights, and market roundups.

[16] The dictionary used for the first period is based on news data from the same period and part of the training sample.

Additionally, we reduce the risk of considering incongruent headlines in the following sentiment classification task.

Ultimately, we derive a $k \times 1$ vector $\widehat{\mathbf{w}}$ that contains the relevance scores for all news articles in the data sample, defined to be climate-related by satisfying the mentioned conditions. The identified news articles will be subject of the upcoming sentiment analysis.

3.4. Sentiment classification

Sentiment analysis refers to the identification of the tonality of a given text document. Generally, sentiment analysis is performed by the application of either rule-based or machine learning methods. Rule-based approaches commonly use a pre-defined lexicon of emotionally charged (or domain-specific) terms to approximate sentiment by measuring the polarity towards "positive" or "negative" expressions. In contrast, machine learning approaches allow for the adaption and creation of language models for specific purposes and contexts by utilizing pre-labeled data samples that relate text elements to (human) perceived sentiment. Contextual, domain-specific sentiment classification significantly differs in terms of the objective. For example, sentiment analytics of financial news are usually constructed to measure a news event's expected financial (or subsequent return) effect on a mentioned stock.[17] Given our task at hand, we particularly want the sentiment to approximate the impact of a given news event on transition risk. Therefore, the sentiment model needs to be trained on a specific dataset that allows the model to infer the intended classification methodology concerning the objective (as described in Section

[17] Consequently, news analytics providers, such as Bloomberg, Thomson Reuters, RavenPack or Alexandria, provide clients with a variety of stock- and market-specific news sentiment measures related to different target variables like expected subsequent return impact, change in short-term volatility or investor perception. See Coqueret (2020) for an overview of existing research on stock-specific sentiment.

3.2) from the provided text samples.[18] For this purpose, we choose a machine learning model based on Bidirectional Encoder Representations from Transformers (BERT), which is a popular choice for a wide range of domain-specific natural language processing (NLP) applications in economics.[19]

BERT is a well-established, deep neural language model capable of learning word representations from large volumes of unannotated text (Devlin et al., 2018). Compared to earlier approaches that forward text input sequentially (left-to-right or right-to-left), BERT embeddings are highly contextual due to its bidirectional (left-to-right and right-to-left) training. BERT pre-training uses a masked language model that learns from an unprecedentedly large corpus to predict words randomly masked from a sequence, and whether one of two sentences is subsequent to the other. We use the off-the-shelf, base version of BERT, which was trained using English Wikipedia and the BookCorpus (Zhu et al., 2015), accounting for approximately 3,300M words.[20]

[18] We will particularly focus the differences of tonality-based and domain-specific sentiment classifiers to approximate transition risk when we evaluate our transition risk index in comparison with other existing approaches of news-based climate risk measurements in the following section.

[19] An early domain-specific adoption of BERT in the scientific field is "SciBERT". Based on large-scale labeled scientific data, Beltagy et al. (2019) create a pre-trained model based on BERT that leverages unsupervised pretraining on a large multi-domain corpus of scientific publications to improve performance in sequence tagging, sentence classification and dependency parsing. To support computational analysis of financial language, Araci (2019) presents "FinBERT", a variant of BERT trained on the purely financial corpus of Reuters TRC2 to achieve domain language adaptation by exposing the model to financial jargon. Afterwards, the model was fine-tuned on the dataset of Financial Phrasebank for a sentence-based sentiment classification task, achieving higher test set accuracy than previous state-of-the-art models. To our knowledge, the application that comes closest to ours in terms of domain language adaption is "ClimateBERT". Bingler et al. (2021) design a contextual-based pre-trained model variant of BERT on thousands of sentences related to climate-risk disclosures aligned with the TCFD recommendations. By analyzing the disclosures of TCFD-supporting firms, the authors conclude that the firms' TCFD support is mostly cheap talk and that firms tend to cherry pick.

[20] The base version of BERT consists of 12 encoder layers, 768 hidden units, 12 attention heads, and a total of 110M parameters.

The model needs to be trained, i.e., fine-tuned on a labeled dataset to perform a sentiment classification task. For data sample creation, we select all news identified to be climate-related from 2000 to the end of 2008, accounting for more than 25,000 news headlines.[21] Extracted news items are labeled with respect to their implied impact on transition risk and in accordance with the framework defined in Section 3.2. News data of the year 2008 is reserved as an out-of-sample test set, totaling almost 10,000 observations. The remaining data from 2000 to 2007 is used for model training and validation.

Before starting the training process, two methods for model performance enhancement are applied to the initial data sample – *data augmentation* and *entity masking*.

3.4.1. Data augmentation

With annotated data for supervised learning tasks that remain generally scarce, *data augmentation* originated in computer vision to artificially increase the variety of data for model training without additional observations. Data augmentation for textual data is of particular interest when language from a different subject domain as the pre-trained model is used. A common approach to applying data augmentation on textual information is back-translation (Edunov et al., 2018). Given an input text in some source language A, the text is translated temporarily to a second language B before it is translated back into source language A. This process enables diverse samples to be generated that preserve the semantic meaning of the input text.

[21] We utilize a dictionary of climate-specific vocabulary. As a result, news about the physical effects of climate change is likely to be identified as climate-related. However, any news about physical risks in the training sample is labeled neutral with regard to the implied change in transition risk. Given a sufficient accuracy in out-of-sample sentiment classification, we expect no significant impact from news about physical risk on transition risk approximation.

We use the "fairseq" algorithm (released by Facebook AI Research)[22] for the English-German and German-English models from WMT'19[23] to perform back-translation on each headline in our training set (Ng et al., 2019). Synthetic texts are created by applying a random sampling strategy. We control the likelihood of low probability words being included in the generated sample with the so-called temperature of the softmax (Holtzman et al., 2019). A parameter value close to zero will likely result in samples identical to the original text, while a value of 1 results in highly diverse samples that risk altering the semantic meaning. We generate one augmented headline for each considered softmax setting (0.5, 0.6, 0.7) to yield enhanced variety without sacrificing fluency and coherence. Examples of back-translated headlines for different parameter values are provided in Table 3.1. Duplicates resulting from exact back-translating are dropped. Adding the remaining artificial data to the initial training set more than doubles the training sample size to almost 50,000 different news headlines.

Table 3.1: Examples of back-translated headlines for different softmax settings

Value	Text sample
-	Kyoto protocol creates the climate for new ideas to cool down warming. Quest to lift efficiency, cut emissions sparks a shift in technological thinking.
0.5	Kyoto Protocol creates the climate for new ideas for cooling warming, and the pursuit of efficiency and emission reductions is transforming technological thinking.
0.6	Kyoto Protocol creates the climate for new ideas to slow warming. The drive to increase efficiency and reduce emissions leads to a change in technological thinking.
0.7	Kyoto Protocol creates the climate for new ideas on cooling warming. The pursuit of efficiency and emission reductions leads to a change in thinking about technology.

The table provides examples of back-translated headlines generated using a range of softmax temperature settings. The first example is the original text input.

[22] https://github.com/pytorch/fairseq

[23] http://www.statmt.org/wmt19/translation-task.html

3.4.2. Entity masking

Next, we confront the risk of overfitting in model training with *entity masking.* As for any classification task, the risk of overfitting arises from overestimating the polarity of specific terms in a text document for sentiment calculation. With a limited dataset of diverse samples, the generalization of particular terms that are expected to have no (standalone) impact on sentiment classification will reduce complexity. Terms or so-called entities we want to be neutral concerning their sentiment contribution are categorized as person names [PER], organizations [ORG], or locations [LOC].[24] To extract the sequences of words in the text that relate to one of the considered entity categories, we use Named Entity Recognition. The identified named entities are substituted with the entities' category label. By doing so, the complexity of non-essential information will decrease while retaining the semantics of the original text. Most importantly, an undesired sentiment bias towards specific entities is prevented. While the entity category (e.g., "organization") is of relevance for the sequential learning process, the entity name (e.g., "European Union" or "G20") should not. A given news event should be consistently classified irrespective of the specific entity value. For example, in the training sample news headlines containing the entity name "US" are negative, whereas headlines with the term "California" are significantly positive. These findings most likely result from the respective political agenda during the considered period. While the US administration under President George W. Bush was somewhat reluctant to adopt climate-friendly policies and refused to join the Kyoto Protocol, California under Senator Arnold Schwarzenegger followed a progressive environmental policy in the early 2000s. By labeling "US" and "California" with the same entity type (i.e., "LOC"), we hinder the

[24] Quantities, dates, monetary values, or percentages are also subject of entity masking. We do however differentiate between the relationship of values if there are multiple entities of the same type.

emergence of biases that may reduce out-of-sample performance - especially when previous biases switch (e.g., due to a change of political direction). Entity masking is performed for the whole data sample. With humans being as (unwittingly) prone to potential biases as trained models, only masked sentences are subject to the manual labeling process.

3.4.3. Model performance

We train different model specifications with respect to the adjustments applied to the input data. To evaluate the potential performance enhancement by the described data optimization techniques, we use the different model specifications to predict the sentiment label on the test set. Generally, the output of the classification model is the log odds for the different class labels: negative, neutral, and positive. We apply a softmax activation function to normalize the log odds into a probability distribution. Finally, any news item is assigned to the sentiment label with the highest predicted probability. We report results for the different model specifications in Table 3.2.

Table 3.2: Performance of different model specifications on the test dataset

Model	Precision	Recall	F1 score
$BERT_{BASE}$	0.75	0.74	0.74
$BERT_{AUG}$	0.79	0.78	0.79
$BERT_{ENT}$	0.79	0.78	0.78
$BERT_{AUG-ENT}$	**0.82**	**0.82**	**0.82**

The table reports performance metrics of the different model specifications that are evaluated on the test set data. The presented model performance results from equally weighting the scores, calculated for each class individually.

All model specifications are run using identical parameter settings, based on the recommendations given for model configuration in the initial paper of Devlin et al.

(2018).[25] We deliberately optimize classification performance solely on input data adjustments not on model configuration. By doing so, we want to choose the most straightforward approach, confront concerns with respect to design choices and limit complexity. Three performance metrics are considered for the comparison of model specifications: (i) Recall, or sensitivity, is the proportion of actual positives that are correctly predicted positive, (ii) precision, also referred to as positive predictive value, denotes the proportion of predicted positives that are true positives, and (iii) the F1 score, which is defined as the harmonic mean of precision and recall. For a classification task with more than two labels, model performance is calculated by averaging the scores of the individual classes to account for imbalanced classes (Koyejo et al., 2015). All three metrics are commonly used to evaluate the accuracy of machine learning-based classification models.[26]

We define the pre-trained base version of BERT ($BERT_{BASE}$) as a benchmark for performance evaluation. $BERT_{BASE}$ achieves a F1 score of 0.74, which is a considerable performance for a baseline approach, probably due to the already high number of actual training observations. By increasing the size of the training data set with augmented data, model accuracy ($BERT_{AUG}$) improves to 0.79. The performance is comparable to the one of the fine-tuned models ($BERT_{ENT}$) that is based on the data set with entity masking. Both approaches are outclassed by the fine-tuned model ($BERT_{AUG-ENT}$) that applies both methods for performance enhancement. $BERT_{AUG-ENT}$ achieves an impressive F1 score of 0.82. Overall, our results align with previous findings for textual data

[25] For the sentiment classification task, we use a dropout probability of p = 0.3, a maximum sequence length of 64 tokens, a learning rate of 2e-5 and a batch size of 16. We train the model for 6 epochs and choose the iteration with the highest accuracy for the validation set.

[26] See James et al. (2013), pp. 145ff.

augmentation (e.g., Nugent et al., 2020).[27] With an increased diversity of examples using domain-specific language, generalization improves in line with model accuracy.

Table 3.3: Confusion matrix

Actual sentiment \ Predicted sentiment	Positive	Neutral	Negative
Positive	7.4	1.8	0.2
Neutral	2.0	76.6	2.3
Negative	0.4	2.1	7.1

The table shows the confusion matrix for the fine-tuned $BERT_{AUG-ENT}$ model and the test dataset.

For further evaluation, we report the confusion matrix for $BERT_{AUG-ENT}$ in Table 3.3. By inspecting the off-diagonal values, we see that only a small fraction of misclassified observations results from mistaking positive for negative sentiment and vice versa. Approximately 95% of the failures happen between labels positive and negative. These findings make intuitive sense as it is easier to differentiate between positive and negative than between positive and neutral or neutral and negative. Just as for human decision-making, the difference in boundaries can be marginal for specific observations.

Ultimately, we use $BERT_{AUG-ENT}$ as the fine-tuned model for the out-of-sample sentiment classification.

[27] Nugent et al. (2020) apply back-translation to improve the accuracy of a fine-tuned BERT model for ESG controversy classification.

3.4.4. Sentiment score

Sentiment classification is performed on all news that are identified to be climate-related. This results in a k-vector **SENT** that contains the sentiment score for each considered news item, where k refers to the number of climate-related news articles in the data sample:

$$\mathbf{SENT} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix},$$

with any $s \in \{-1, 0, 1\}$ for either negative, neutral or positive sentiment. The vector **SENT** and the vector $\widehat{\mathbf{w}}$ (containing the relevance scores) have the same dimension. We calculate the element-wise product of vector $\widehat{\mathbf{w}}$ and vector **SENT** to calculate the weighted sentiment score for all climate-related news articles:

$$\mathbf{wSENT} = \begin{pmatrix} w_1 s_1 \\ w_2 s_2 \\ \vdots \\ w_k s_k \end{pmatrix} = \widehat{\mathbf{w}} \odot \mathbf{SENT},$$

with $-1 \leq ws \leq 1$. We use the weighted scores to emphasize the article's relevance to the topic of climate change. The weighted sentiment scores lay the foundation for the calculation of our Transition Risk Index (TRI).

3.5. News Index Validation

3.5.1. Index calculation

We run index calculations on a daily, weekly, and monthly frequency.[28] The start and end times of each period are based on the NYSE trading dates and hours. News that is released after closing are considered in the next period. This includes news on weekends and holidays that are ascribed to the next trading day. The index score results from the simple aggregation of sentiment for climate-related news articles, divided by the total number of news articles n. Given the weighted sentiment score $wSENT$ for climate-related news articles $k = 1, \dots, K$, the index $score$ of a given period p is defined as:

$$score_p = \frac{\sum_{k=1}^{K_p} wSENT_{k,p}}{n_p}$$

Figure 3.1 shows the monthly time series of the Transition Risk Index since 2000. The out-of-sample period starts in 2008 as prior news data are utilized to train the sentiment model. We annotate influential climate news events to make sense of the index scores. To provide further intuition on sentiment classification, Table A3.1 reports the three months with the highest and lowest TRI scores. For each period, the most contributing news are reported as measured by the relevance score.

We begin the evaluation of the Transition Risk Index with a qualitative inspection of the monthly index scores. Therefore, we will discuss the behavior of the Transition Risk

[28] We do not report results for daily frequency due to a considerable proportion of days with an index score of zero. Index scores of zero occur when either no non-neutral climate news is identified, or the values of positive and negative climate news are equal. However, to further validate our index, we ran the upcoming regressions also for a daily index frequency and came basically to the same statistical findings and conclusions as for weekly and monthly periodicity.

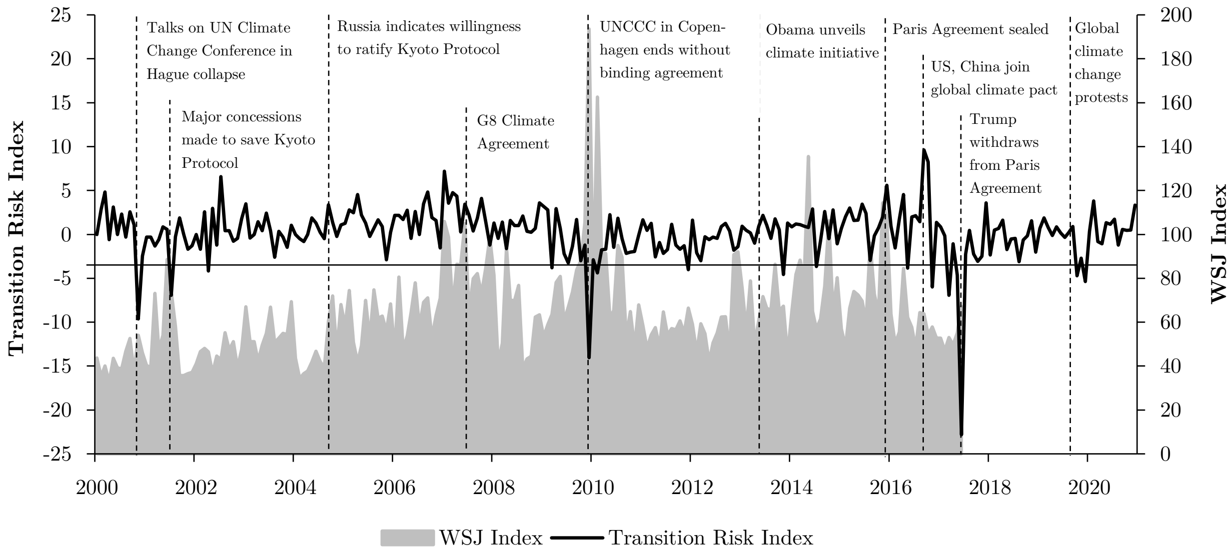
Index in the context of some particular events and in comparison to other existing approaches on news-based climate risk indices.

3.5.2. Comparison of existing news-based climate risk indices

In their influential paper, Engle et al. (2020) construct two complementary indices to measure the extent at which climate change is discussed in the news media. The first index is designed to measure the fraction of news coverage in the Wall Street Journal related to climate change as the correlation between the newspaper content and a fixed vocabulary. The authors assume that the more attention is paid to climate change by news media, the more salient the topic is for investors. With news media being considered as the most essential sources of investor sentiment, an alteration of investment behavior is expected with changes in news coverage on specific topics. The WSJ Climate Change News Index (going forward referred to as WSJ Index) associates increased climate change reporting with news about elevated climate risk. This is based on the idea that climate change primarily rises to the media's attention when there is a cause for concern. Accordingly, the index spikes in times of attention-grabbing events like climate conferences and summits. We plot the time series of the WSJ Index in addition to the Transition Risk Index on the secondary y-axis of Figure 3.1 for comparison.

Concerning its construction, the WSJ Index implicitly assumes that news media covers solely events that infer an increase in climate risk. To overcome the possible shortcomings of the WSJ Index in terms of news differentiation and source dependence, Engle et al. (2020) develop an alternative climate news index that is designed to focus specifically on negative news about climate change. The authors use the services of the data analytics vendor Crimson Hexagon (CH) to identify climate-related news articles concerning negative sentiment from more than a thousand media outlets. Here, sentiment is classified based on the tonality of a given text document.[29] The resulting CH Negative Climate Change News Index (referred to as CH Index) measures the percentage of negative climate news in all news articles.[30]

Figure 3.1: Global Transition Risk Index and WSJ Climate Change News Index



On the primary y-axis the Transition Risk Index is reported. The WSJ Climate Change News Index from Engle et al. (2020) is shown on the secondary y-axis. Annotations are provided for climate-relevant news events. Monthly data are reported from January 2000 to December 2020 for the Transition Risk Index. The out-of-sample period starts in January 2008. The available time series for the WSJ Climate Change News Index ends in June 2017. The WSJ Climate Change News Index data are accessible on both Stefano Giglio's website at https://sites.google.com/view/stefanogiglio and Johannes Stroebel's website at http://pages.stern.nyu.edu/~jstroebe. We scale both indices by a factor of 10,000 to allow the interpretation of the magnitudes of innovations.

In a related paper, Ardia et al. (2020) take a similar approach to Engle et al. (2020) to capture unexpected increases in climate change concerns. The authors design a score to measure concerns of climate-related articles from eight different US newspapers. Therefore, Ardia et al. (2020) make use of two lexicons to identify risk-related terms and

[29] See Section 3.4 for a more detailed discussion of different sentiment classification techniques.

[30] The times series data of the WSJ Climate Change News Index and CH Negative Climate Change News Index are accessible on both Stefano Giglio's website at https://sites.google.com/view/stefanogiglio and Johannes Stroebel's website at http://pages.stern.nyu.edu/~jstroebe.

emotionally charged words.[31] The "concern score" is defined as the level of discussion about risk-events (measured by the share of risk-related expressions in the text) multiplied by the negativity (measured by the difference of negative and positive expressions in relation to the sum of all emotionally charged terms). The authors combine both terms to differentiate between negative texts about risk from those that are positive. Ultimately, their Media Climate Change Concerns Index (going forward referred to as MCCC Index) results from the daily aggregation of article-level concern scores by taking the sum of scores.[32] The cross-correlation between the Transition Risk Index and the WSJ, CH, MCCC Index ranges from -0.2 to +0.2 for each index.

Next, we exemplary highlight two substantial climate events where the indications of the considered news-related climate risk indices are remarkably similar or divergent and discuss possible explanations based on the thematical background and the rationale of the different construction methods:

(I) The Copenhagen Climate Change Conference in December 2009 raised climate change policy to the highest political agenda. More than 100 world leaders attended the summit, making it one of the largest gatherings of world leaders ever outside UN headquarters in New York. Given the international conference's importance, news media coverage was intense, with the attention-based WSJ Index on its peak. An accord was ultimately reached on the long-term goal of limiting the maximum global average temperature increase. The conference ended, however, without an agreement on how to achieve this target in practical terms.[33] Tough talks between

[31] The referred lexicons are retrieved from the LIWC2015 software. The academic version is available at https://liwc.wpengine.com/.

[32] The time series data of the MCCC index is available at https://sentometrics-research.com/.

[33] See the official website of the United Nations on the conference communique: https://unfccc.int/process-and-meetings/conferences/past-conferences/copenhagen-climate-change-conference-december-2009/copenhagen-climate-change-conference-december-2009.

industrial and developing nations and member states' reluctance towards compulsory actions increased concerns about a common political agenda on the international level. Due to the lack of concrete measurements, the accord was predominantly considered a disappointment, resulting in more articles with negative tonality.[34] Considering the negativity in news coverage, it makes intuitive sense that a concern-based sentiment index like the CH index spike during this period. The month around this event records the historical low point of the Transition Risk Index up until then. In line with our defined transition risk framework, news about stiff negotiations resulting in tenuous compromises and non-binding agreements is expected to weaken the international accord, imply resistance towards adopting environmental-friendly policies and thus soften regulatory pressure. Consequently, transition risk is supposed to decline, as the index indicates.

(II) In December 2015 probably the most comprehensive international climate agreement was concluded at the UN Climate Change Conference in Paris, France. The participating 196 countries agreed, by consensus, to reduce emissions for climate mitigation. While there have been discussions about the realization of global ambitions, the accord was widely celebrated as a landmark deal from politicians and journalists alike. If not for the case of immediate actions, then for the irrefutable signal that the age of fossil fuels has started drawing to a close.[35] Based on the overwhelmingly positive media reaction to the conference's successful conclusion, the Transition Risk Index records its third-highest monthly score (in

[34] See the following news examples for the immediate assessment of the conference outcome from Reuters, BBC, The New York Times and The Guardian.

[35] See the following link for an extensive overview of media reaction on the Paris Climate Agreement: https://www.carbonbrief.org/the-paris-agreement-on-climate-change-the-world-reacts

the out-of-sample period). News about the acknowledgment of global warming in combination with the commitment to fight climate change on the international level transmits transition risk via an increase in public awareness and related pressure for the adoption of climate policies on the national level to reach agreed targets.

In contrast, the high levels of climate concern and risk as measured by the CH Index, and the MCCC Index come at a surprise. While there is a number of negative news in the build-up to the Paris Climate Agreement due to uncertainty at preceding negotiations, these news articles are outweighed by the positive news flow related to the final climate accord. Hence, the focus of both concern-based index approaches on negative news flow probably causes these high index scores. By construction, the CH Index is defined to measure the share of all news articles about "climate change" and classified to convey negative sentiment. While the MCCC Index accounts for the polarity on the article-level (with scores from zero, the most positive text, to one, the most negative text), it aggregates article sentiment by the sum, not mean to calculate index values. Considering the intensity of topic-related news coverage - as reported by the attention-based WSJ Index - the high concern score likely results from the sheer number of domain-specific articles (with at least partial negative tonality)[36] without accounting for a fraction of mainly positive climate-related articles.[37]

Given the differences in implication on climate policy and media reaction to the outcome of the Copenhagen and the Paris Climate Conferences, it seems counterintuitive

[36] An article-level concern score above zero indicates the inclusion of some negative expression in the article content.

[37] See Figure 2 in Ardia et al. (2020) for a visual comparison of the 30-day moving average concern scores at both events.

that the periodical indication on climate risk around the events (I) and (II) is supposed to be relatively similar.[38] While the saying that *no news is good news on climate* often did hold true in the past, news of the Paris Climate Agreement just as other subsequent climate initiatives like the Green New Deal do not support approaches based on this assumption. The given examples illustrate that risk indices utilizing media attention or concern-based news sentiment are rather inapt to capture the complexity of transition risk, while making a case for a more sophisticated approach to infer changes in transition risk from news events. The complexity of classifying changes to transition risk relates not only to the consideration of positive and negative sentiment but also to the semantics. Context not tonality alone is decisive, as we interpret bad news for carbon-intensive activities (e.g., "Sales limits of combustion engines will hurt carmakers.", "Coal industry to suffer from dwindling governmental support.") as positive for the climate in terms of increasing transition risk. Accounting for this specificity, we developed the most comprehensive framework to approximate changes in transition risk from news events to our knowledge yet.

3.5.3. Data

Recent studies and surveys among institutional investors show the rising importance of integrating climate risk into the investment process (Dyck et al., 2019; Krueger et al., 2020; Bolton & Kacperczyk, 2021b). We want to provide some insight into whether changes in transition risk are incorporated in asset returns. Therefore, we pick a variety of presumed "green" market indices to analyze the sensitivity towards transition risk.

[38] It is fair to admit that the value of the CH Index for the period of event (II) is not as high as for the period of event (I). Oddly, the CH Index is still higher at the time of the Paris Climate Agreement than in June 2017 when the U.S. administration announced to withdraw from the Paris Climate Pact. An event which is supposed to induce a decent amount of negative climate-related news flow.

Generally, a distinction is made between two different portfolio construction types based on investors' climate objectives. Benchmark-oriented investors who simply want to mitigate climate risk (in terms of unexpected changes in carbon price) commonly adopt *decarbonized* approaches. These strategies typically reduce carbon exposure while preserving diversification by reweighting constituents on emissions relative to a financial metric (e.g., revenues or sales) or by excluding companies with fossil fuel-related activities. Alternatively, *pure-play* approaches refer to investors who want to directly allocate to companies that accelerate the transition to a lower-carbon economy with their business activities. Selection criteria are usually related to the share of company revenues associated with products and services that deliver solutions to environmental challenges. Due to the specific requirements for portfolio inclusion, only a limited number of companies come into consideration. Consequently, the tracking error is high. Pure-play approaches are most likely used as a satellite in portfolio construction for investors seeking exposure to environmental-friendly investments.

Selection criteria of third-party indices are based on quantitative and qualitative measures. The first criterium is a fit in construction methodology that ensures that portfolio deviation from the parent benchmark is exclusively resulting from the optimization with regard to different environmental metrics. Next, we sort considered indices by (1) history of live index calculation and (2) assets under management benchmarked by the index. Both measures provide some validation for the applied index construction approach by market participants. Indices are scored by their average rank of both measures. Finally, we limit the number of considered indices per provider to secure diversity among data sources. Detailed information on selected indices is provided in Table A3.2. Daily data on indices' prices is collected directly from Datastream, Bloomberg, or the index providers. We calculate the active returns from its benchmark for each climate stock index to isolate the return contribution resulting from portfolio deviation.

We report the cross-correlation among the indices in Table A3.3. The table shows that the active returns of pure-play and decarbonized indices are uncorrelated, providing initial evidence that both approaches result in independent portfolio specifications. The correlation across pure-play indices ranges between 0.2 and 0.8. While the consistently positive correlation indicates a certain similarity in the construction methods and an overlap in companies covered, there is sufficient variability to suggest the usefulness of a comprehensive analysis of various indices.[39] The cross-correlation of decarbonized indices is (while being positive on average) significantly lower and less consistent compared to the correlation between pure-play indices. This is an important observation as the low correlation seems at odds with the uniformity in the goal of carbon exposure reduction, the almost perfectly correlated underlying benchmarks, and the overall low tracking error of the decarbonized indices.[40] However, the divergence across ESG ratings is a well-documented fact (e.g., Chatterji et al., 2016; Berg et al., 2019; Dimson et al., 2020). With each index provider applying its own data analytics (e.g., MSCI ESG Research, S&P Trucost, ISS ESG), differences in carbon metrics (e.g., scope of emissions) likely lead to a variation in active constituents' weights, providing a possible explanation for the relatively low cross-correlation among decarbonized index returns. Due to a lack of standardization,

[39] Furthermore, we find that active renewable index returns (i.e., of the MSCI Global Alternative Energy Index and S&P Global Clean Energy Index), and sector returns of oil, gas and coal are notably positive correlated - depending on the considered period (results not shown here). While several explanations come into question for this observation, we want to specifically stress the subsumption of clean energy on the sector level as a possible contributing factor. Investors, who explicitly allocate energy stocks regularly trade a basket of companies with possibly contrasting sensitivities towards transition risk. This investor behavior might result in the time-dependent correlation of active "green" and "brown" energy sector returns.

[40] The cross-correlation between the global parent indices is above 99% on average. The ex-post tracking error of active weekly returns ranges between 0.5% and 1.2% p.a. for the data sample.

a companies' 'green' credentials will vary over different metrics, so will constructed portfolios and ultimately academic findings.

Here, we opt against creating our own green minus brown (GMB) portfolios for multiple reasons. First, the construction of long-short portfolios offers significant degrees of freedom. Instead, we seek to limit design choices predominantly to the configuration of our language model. Second, we facilitate the reproducibility of results with a detailed description of selected market indices and the accessible time-series data of the TRI. Most importantly, by choosing a multitude of index providers that use different climate metrics for portfolio construction, our findings become less reliant on specific measures to identify high/low carbon emitters and pure-play companies.

3.5.4. Methodology

Our first analysis focuses on the contemporaneous relationship between innovations (changes, unexpected by investors) in the Transition Risk Index and weekly active returns for different decarbonized and pure-play indices. We calculate values of TRI as residuals from an AR(1) model to capture innovations and confront autocorrelation in the stationary index. We consider a multivariate time series regression framework to control for other factors potentially driving active index returns. Therefore, we regress the active returns $r_{i,t}$ of each index i on the innovations in the Transition Risk Index TRI_t and the widely used five Fama-French factors (Fama and French, 1993; Fama and French, 2015), i.e., (i) MKT_t, the excess market return, (ii) SMB_t, the small minus big factor, (iii) HML_t, the high minus low factor, (iv) RMW_t, the robust minus weak factor, and (v) CMA_t, the conservative minus aggressive factor. We also incorporate the momentum factor of Carhart (1997), (vi) MOM_t which yield in the following regression model:[41]

$$\begin{gathered} r_{i,t} = \alpha_i + \beta_{1,i} MKT_t + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \beta_{4,i} RMW_t + \\ \beta_{5,i} CMA_t + \beta_{6,i} MOM_t + \beta_{7,i} TRI_t + \varepsilon_{i,t}, \end{gathered} \tag{1}$$

where α_i is a constant and $\varepsilon_{i,t}$ is an i.i.d. error term. We scale the active index returns to a volatility of 5% p.a. to allow for the comparison of coefficients. The innovations in the Transition Risk Index TRI_t are winsorized at the 1% level to limit the influence of extreme climate news events on regression results.[42]

Additional to the six-factor specification (MKT_t, HML_t, SMB_t, RMW_t, CMA_t, MOM_t), we define two further model specifications:

$$\begin{gathered} r_{i,t} = \alpha_i + \beta_{1,i} MKT_t + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \beta_{4,i} RMW_t + \\ \beta_{5,i} CMA_t + \beta_{6,i} MOM_t + \beta_{7,i} CC_t^{WSJ} + \beta_{8,i} TRI_t + \varepsilon_{i,t}, \end{gathered} \tag{2}$$

where CC_t^{WSJ} denotes the innovation in WSJ Climate Change News Index, and

$$\begin{gathered} r_{i,t} = \alpha_i + \beta_{1,i} MKT_t + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \beta_{4,i} RMW_t + \\ \beta_{5,i} CMA_t + \beta_{6,i} MOM_t + \beta_{7,i} CC_t^{NegNews} + \beta_{8,i} TRI_t + \varepsilon_{i,t}, \end{gathered} \tag{3}$$

where $CC_t^{NegNews}$ denotes the innovation in CH Negative Climate Change News Index as defined by Engle et al. (2020).[43] We exemplify both climate news indices to account for media attention- and concern-based approaches in explaining contemporaneous returns. Here, monthly return and factor data are used due to the availability of time series data

---

[41] All daily factor data are taken from Kenneth French's data library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html) and converted to weekly and monthly frequency.

[42] Instead of winsorization, we also perform a robust regression with Huber weights in an alternative approach to handle outliers. The significance of coefficients and implications on findings are consistent with presented results.

[43] Engle et al. (2020) average the daily values for the WSJ Climate Change News Index and CH Negative Climate Change News Index to the monthly level and measure innovations in climate news as the residuals from an AR(1) model.

for both climate news indices. Possible multicollinearity is not an issue. The innovations in the Transition Risk Index are uncorrelated to both innovations in the WSJ Climate Change News Index and innovations in the CH Negative Climate Change News Index.

3.5.5. Results

We report estimation results for the first regression model (1) in Table A3.4. The number of observations N (degrees of freedom) varies with respect to the maximum available common data history of active index returns and factors. Further, we only consider data points within the range of the out-of-sample period starting in January 2008 and ending in December 2020.

First, we analyze the findings for each pure-play index, reported in Panel (A) columns. Coefficients for the *TRI* are favorable for every pure-play index and in fact, significant at the 1% level. The estimation results deliver a strong indication for the exposure of active returns to transition risk as measured by our news-based index. In line with the positive cross-correlation between pure-play indices, we find a clear tendency in the indices' factor exposures. Coefficients for *SMB*, the size factor, are consistently positive for all pure-play indices and significant at the 1% level for the MSCI Global Environment Index (MSCI GLENV), the Solactive Climate Change Index (Solactive CC) and the FTSE Environmental Technology Index (FTSE ENVT), which hint at exposure to firms with relatively low market capitalization. This seems intuitive as companies that derive the lion's share of their revenues from environmentally beneficial products and services are generally specialized and did make up only for a small market share for most of the past. In contrast, coefficients for *HML*, the value factor, are more divergent. The coefficients for *RMW* are predominantly negative and significant at the 1% level for the MSCI Global Alternative Energy Index (MSCI GLALT) and S&P Global Clean Energy Index (S&P

GLCE), indicating relatively low profitability for companies within the renewable energy sector. Results for the coefficients of *CMA*, the investment factor, are to some extent comparable. The investment factor measures the difference between the returns of firms that invest conservatively and firms that invest more aggressively. Given these findings, companies with substantial involvement in environmental business activities tend to be less profitable, probably due to some expansionary growth policy. Finally, coefficients for the momentum factor, *MOM*, are inconclusive and insignificant. Overall, the explained variance R^2 ranges from relatively low 9% to almost 30% for the FTSE ENVT and S&P GLCE Index.

Next, we analogously analyze the findings for each decarbonized index. Results for those indices are reported in the columns of Panel (B) from Table A3.4. In contrast to the findings for the pure-play indices, coefficients for the *TRI* are insignificant for each decarbonized index. The estimation results show no exposure of active decarbonized index returns to contemporaneous innovations in transition risk. Coefficients for *SMB*, the size factor, are also inconclusive and barely significant for any decarbonized index. Generally, this makes intuitive sense as companies' carbon emissions are measured as relative value with respect to sales or per dollar of market capitalization. Coefficients for *HML*, the value factor, are negative for five of the six decarbonized indices, with the MSCI World Climate Change Index (MSCI WCC), the S&P Global LargeMidCap Carbon Efficient Index (S&P LMCCE) and the STOXX Global 1800 Low Carbon (STOXX 1800 LC) reporting significant coefficients at the 1% level. These findings provide a first indication that companies with substantial value exposure may tend to have relatively high carbon emission levels and thus be underweighted. The same five indices that report negative coefficients for *HML* have positive coefficients for *RMW*, the profitability factor, these are significant at the 1% level for the MSCI WCC, S&P LMCCE and S&P Global 1200

Fossil Fuel Free Index (S&P 1200 FFF). Coefficients of *CMA*, the investment factor, are negative on average with varying statistical significance, indicating that companies with rather conservative investment policies tend to be overweighted in decarbonized indices. Interestingly, coefficients for *MOM*, also show exposure to the momentum factor, which is significantly positive for the MSCI WCCC, S&P 1200 FFF, S&P LMCCE and MSCI World Low Carbon Leaders Index (MSCI WLCL) at the 1% respectively 5% level. The explained variance R^2 ranges from only 4% to over 35%.

We repeat the regression (1) for the pure-play indices with a rolling window of five years to investigate the development of coefficients over time. Figure A3.4 shows the t-value for the *TRI* coefficients and each pure-play index. The dashed red lines illustrate the t-values for the respective p-values at the 10%, 5% and 1% significance level. For the pure-play indices, we can infer from the t-values that the exposure to transition risk is constantly significant over time with hardly any exception. Few events seem to induce major shifts in the significance of transition risk. Overall, most indices experience an increasing significance of the *TRI* coefficient over time until the emergence of the corona crisis. A dominating event with far-reaching effects drawing most of the investors' attention. Unsurprisingly, the volume of climate-related news flow was relatively low during this period.

We run regression on two model specifications to further investigate previous findings and account for different news-based climate risk factors in explaining contemporaneous active index returns. The maximum available common data history of return and factor data is used. The number of observations is quite limited by utilizing monthly data and by the end of the time series for the WSJ Climate Change News Index and the CH Negative Climate Change News Index in June 2017 and May 2018 respectively. We analyze the findings for the pure-play indices in Panel (A) of Table A3.5. Results for the model specifications (2) and (3) are shown in the columns for each index. While there are slight differences in the coefficients of five Fama-French factors, we expect them to be mainly driven by the reduced data sample. The coefficients for TRI, the innovations in the Transition Risk Index remain positive and highly significant, even after incorporating two alternative news-based climate risk measures.[44] In contrast, the coefficients for CC^{WSJ}, the innovations in the WSJ Climate Change News Index are consistently positive for all pure-play indices but barely significant. Hence, we cannot provide evidence that relative media attention to climate change as measured by the WSJ Climate Change News Index is related to pure-play indices' contemporaneous active monthly returns. The coefficients for $CC^{NegNews}$, the innovations in the CH Negative Climate Change News Index are inconsistent and insignificant.

Results for the decarbonized indices are reported in Panel (B) of Table A3.5. For both model specifications, the coefficients of TRI are somewhat inconsistent and not significant. Looking at the findings for model specification (2) we find that the coefficients for CC^{WSJ} are on average surprisingly negative but again insignificant. The coefficients for $CC^{NegNews}$, in the model specification (3) are mainly positive but still hardly significant.

In line with existing research, we expect active returns of decarbonized indices or carbon-tilted portfolios to be driven by investment trends that may be approximated by the relative media attention to climate change. Engle et al. (2020) use a mimicking portfolio approach to build climate change hedge portfolios by using third-party ESG scores. They find that these portfolios perform well in hedging innovations in the WSJ Climate Change News Index and the CH Negative Climate Change News Index. As previously discussed, both indices are constructed to measure the share in (negative)

[44] Coefficients for CC^{WSJ} and $CC^{NegNews}$ do not significantly change if TRI is not included in the regression model.

climate-related news. With the potential effects of climate change being intensely discussed in news media, it can be assumed that climate risk should rise in investors' awareness. Hence, as investors' demand for climate i.e., carbon, risk mitigation increases, the demand for low carbon stocks. To account for changes in investors' risk preferences, we incorporate the innovations in the attentions-based WSJ Climate Change News Index and the CH Negative Climate Change News Index into our regression model. However, we do not find significant coefficients for CC^{WSJ} and $CC^{NegNews}$ for either pure-play or decarbonized indices during the considered period. We do not want to overemphasize these findings, given the relatively low number of observations. Additionally, we use indices with a global benchmark, while Engle et al. (2020) focus most notably on U.S. news flow. Further, we only analyze a contemporaneous relationship between active returns and factors, which we consider reasonable for incorporating transition risk. However, changes in investor risk preferences may also materialize over more extended periods (Pastor et al., 2020; Hartzmark & Sussman, 2019).

Overall, the estimation results show no exposure of active decarbonized index returns to contemporaneous innovations in transition risk as measured by the TRI. Instead, we document a systematic negative exposure to value and positive exposure to momentum and quality (as implied by the profitability factor). These findings may result from the actual composition of labeled 'green' portfolios. Carbon-tilted optimization that does not substantially address companies' exposure to environmentally beneficial business activities may result in counterintuitive portfolio allocation (UNEP FI, 2015). For example, in its most simple form, excluding energy stocks, may also cause the exclusion of renewable energy stocks due to sector affiliation. Cohen et al. (2020) present evidence that recent green patenting is not driven by highly rated ESG firms, but instead by energy-producing firms. Paradoxically, these (value) firms are precisely those to which capital is often restricted by carbon optimized portfolio approaches whose directive is to solve investment challenges linked to climate risk. Further, most companies from sectors with relatively low emission exposure (e.g., healthcare, IT) will neither profit directly from more restrictive environmental regulations nor will a weakening result in an expected decline in demand of their products and services. Thus, future revenues and consequently return expectations should be hardly related to changes in transition risk, consistent with our regression results.

The documented findings let us conclude that a significant dependency between stock returns and innovations in short-term transition risk is most likely observed for pure-play companies that economically benefit directly from an increase in transition risk. Their business development is closely tied to the environmental policy and consumer preferences as covered in news media. Subsidies for energy-efficient products, or emission limits will lead to an expected, sector-wide increase in demand for services that accelerate the transition to a low carbon economy. Therefore, these companies can be considered as winners of intensifying transition risk and pressure induced by the adverse effects of climate change. In contrast, the mitigation of environmental standards and continuous support for carbon-intensive activities reduces external pressure on firms and consumers to alter behavior in favor of environmental criteria.

3.6. Conclusion

Capturing transition risks is challenging. In this study, we contribute to the nascent but growing literature on climate risk in multiple ways. First, we provide a rigorous approach for the approximation of changes in transition risk from climate-related news. We start with developing a language model designed to generate domain-specific vocabulary from millions of news items. With topic-related vocabulary possibly changing over time, we utilize period-specific dictionaries to account for the time-dependent relevance of terms. Our presented approach tackles the look-ahead bias and the problem of underrepresentation inherent in fixed dictionary construction. The main advantage of the presented methodology is the ability to derive a domain-specific dictionary from the news by providing as little input as a single term. Hardly any human intervention or supervision is necessary. Given the capability to extend this technique to construct domain- and period-specific vocabulary for a variety of topics and even different languages, we specifically want to stress the opportunity for future research. Here, we use the period- and domain-specific dictionaries to identify climate-related news out of sample. Afterward, we manually label thousands of climate-related news items for their inferred change in transition risk. To ensure the economic foundation of the classification process, we align our approach to the different drivers of transition risk established by the TCFD and other climate risk frameworks. In an extension to previous media attention- or concern-based climate risk proxies, we use the extensive data set to train a sophisticated sentiment model that specifically predicts the implied impact of a news event on transition risk. Ultimately, we construct a global transition risk index from the aggregated news sentiment.

From our analysis of the contemporaneous relationship between presumably “green” stock portfolio returns and innovations in transition risk, we derive two major implications for investors’ climate objectives: Benchmark-oriented investors who simply want to mitigate carbon price risk with decarbonized portfolio approaches, will experience no significant return exposure to short-term transition risk. Hence, portfolios solely optimized on (backward-looking) emission data are unlikely to provide a hedge when transition risks materialize as public and political pressure to confront the adverse effects of climate change intensifies.

Instead, investors who look to hedge transition risk in extent to carbon price risk may seek portfolio approaches that consider a more diverse set of metrics to measure a firms' environmental performance. Given the current set of common investment vehicles, pure-play approaches provide impact-oriented investors with an opportunity to allocate capital to companies that accelerate the transition to a lower-carbon economy with their business activities. Our findings show that these portfolio construction methods will most likely offer exposure to the risk and opportunities related to climate transition.

4. Point-in-Time Language Model for Geopolitical Risk Events

4.1. Introduction

The invasion of the Ukraine by Russia at the beginning of 2022 renewed the interest of global investors in objective measures for the build-up of geopolitical risks. Geopolitical risk measures have been subject to past research and fall into three categories. Market-based measures where asset prices discount risks, textual measures where newspaper articles anticipate risks and ratings-based measures where analyst/expert opinions foresee risks.[1]

This study focuses on a novel text-based methodology for risk approximation. We develop a language model to select news related to geopolitical risk from raw text data with point-in-time, domain-specific vocabulary. Avoiding even subtle look-ahead biases is essential for evaluating (backtesting) systematic investment propositions. By approximating a latent risk variable with media attention, we follow the approach of previous research on different unobservable risk measures utilizing news flow. Our model builds on existing research and improves current methods on different dimensions. Baker et al. (2016) constructed an index to measure economic uncertainty. The authors' research employed a static and curated list of keywords created with perfect knowledge from the past. Rather than using modern text mining techniques, they apply Boolean text mining. Specifically, the authors compute the percentage of articles that satisfy a Boolean logic. Articles must contain "uncertain" or "uncertainty" AND "economic" or "economy" AND "congress" or "deficit" or "federal reserve" or "white house" or "legislation".

[1] See Karagozoglu et al. (2022) for a comprehensive review of existing approaches to measure geopolitical risks.

Ahir et al. (2022) develop country uncertainty indexes using the Economist Intelligence Unit country reports, where the authors count the percentage of occasions (also using Boolean logic) that the word uncertainty (or its variants) are used in a given country report. While the data source looks narrow, the country reports themselves are created by experienced researchers using a standardized research template.

Caldara and Iacoviello (2022) use a dictionary-based method, prespecifying a much wider collection of words whose mention in newspaper articles is related to the coverage of geopolitical events and threats. The authors manually picked around 100 search terms associated with a geopolitical risk to identify relevant news articles based on the joint occurrence of terms. Their most recent Geopolitical Risk (GPR) index started in 1985 and is based on automated text searches on the electronic archives of ten, mainly US-based newspapers. While this approach likely leads to selecting the most specific vocabulary related to geopolitical risk, it has two disadvantages. First, manually defining a dictionary always compromises the risk of a look-ahead bias. Secondly, the evolvement of domain-specific expressions over time can be a disadvantage within a stationary dictionary. Time-dependent events and developments shape specific terms and vocabulary (e.g., “embargo”, “drone strike”, “chemical weapon”). By classifying news based on a stationary dictionary, time-dependent terms may be underrepresented during periods in which these terms were actually of increased relevance to the related topic. Table 4.1 provides an overview of research methodologies to approximate geopolitical risks on text-based measures. Similarities and differences between the various approaches are summarized.

Table 4.1: Text-based methodologies for measuring geopolitical risks

Research	**Data Source**	**Keywords**	**Updates**	**Methodology**
Baker et al. (2016)	Leading (10) US newspapers	small, curated list, boolean	in sample, manual	% of articles with dictionary terms
Ahir et al. (2022)	Economists intelligence unit country reports	small, curated list, boolean	in sample, manual	% of articles with dictionary terms
Caldara and Iacoviello (2022)	Leading (20) US newspapers	dictionary based	in sample, manual	% of articles with dictionary terms
Our approach	Global news provider, accessed via RavenPack	dictionary based	out of sample, algorithmic	% of articles with intensity-weighted dictionary terms

The table provides a summary of research methodologies to approximate geopolitical risks on text-based measures. Boolean text mining combines a small list of keywords with logical (and/or) operators to form search patterns.

This chapter is organized as follows. The next section describes our index construction process, where the development of the Geopolitical Risk Event (GRE) index is divided into two parts: Construction of the point-in-time language model and index calculation. Afterwards, the framework is applied in a case study. We build a global risk index from country-by-country data to identify different dimensions of geopolitical risk (latent components). Instead of employing a principal component model, we account for the autoregressive nature of our news-based series (as newspapers copy or follow up competitors' stories) by estimating a latent dynamic factor model in Section 4.3.

4.2. Index construction

4.2.1. Point-in-time language model

This study defines geopolitical risk as the emergence, realization, and escalation of events associated with terrorism, social unrest, or any conflict between states and political institutions that affect global trade, security, and international relations. Our GRE index approximates real-time geopolitical tensions as media attention, where we measure media attention as the share of news articles attributed to geopolitical risk events. For this purpose, we use the services of the news data analytics provider *RavenPack*. Document search allows us to screen more than one hundred million news articles, starting in January 2000. RavenPack covers news articles and social media posts from a variety of sources. We choose a subset of the most relevant media outlets, including Dow Jones Newswires, Reuters, BBC, WSJ, The New York Times, The Washington Post, MSN, and CNN. News sources are selected to cover the international news flow as we analyze global geopolitical risk. Figure A4.1 shows the news articles' share by media outlet and country. Our data sample stretches from January 2000 to May 2022. Instead of manually selecting topic-related terms, we apply the unsupervised search algorithm, introduced in Chapter 3, that utilizes the information from millions of unclassified news items. The algorithm's objective is to generate a time-sensitive dictionary related to domain-specific buzzwords. We provide five general terms associated with geopolitical risk events as input parameter: “war”, “conflict”, “tension”, “attack” and “terror”. The algorithm searches news (from a given period) that contain any of the buzzwords in the full-text article. The most frequent terms are calculated from the headlines of the selected news articles (in the following, referred to as the headline corpus). Here, rigorous text normalization and cleaning of raw text data are required to generate a dictionary with a high degree of pureness in domain-specificity. At first, language-based stop words[2] and text elements with less than three characters are removed from each headline text. Next, Named Entity Recognition (NER)[3]

[2] Stop words are predominantly the most used words of a given language. Generally, stop words are removed because they are not relevant for dictionary construction and distort the word frequency analysis.
[3] Named Entity Recognition (NER) is the process of locating named entities in unstructured text and then classifying them into pre-defined categories. We use the application from the Stanford NLP Group (Stanza) for all NER tasks in this paper.

is performed to identify entity-related terms that should not be considered for dictionary construction. Specifically, names of persons, organizations, and locations are dropped as their inclusion may lead to an undesired bias towards entities and, hence, inaccuracy in the subsequent topic identification task. After lemmatization[4], terms are formed from a contiguous sequence of up to two text elements (unigram and bigram) for each cleaned headline. Next, we calculate the term frequency by counting the appearance of each term, divided by the total number of terms in the cleaned headline corpus. Finally, term frequency calculation is repeated for all unselected news items in the sample (period) to derive a list of frequently used terms in the overall news flow. The most commonly used terms are general expressions and automatically withdrawn from the topic-related dictionary as specific vocabulary is desired.[5]

We run the process of vocabulary generation at the end of each year to create a dictionary that is related to geopolitical risk and used for topic identification in the following year. Each dictionary consists of the period- and domain-specific terms and their respective frequency. The uni- and bigrams are sorted by term frequency. We limit the number of terms in a dictionary to the 250 most frequent expressions. The sum of term frequencies is scaled to 1. The resulting normalized term frequency (*tf*) represents the relative relevance of each term to the dictionary.[6]

[4] Lemmatization is the process of reducing inflected forms of a word while ensuring that the reduced form belongs to the language. The initial word is stored to retrieve the original expression from the reduced form. These untrimmed words are needed for the content search process based on the full-text article.

[5] We set the number of most commonly used terms that are removed from the buzzword-specific dictionary to 500. Given our approach this step is expandable and only applied for illustrative purposes. We use *tfidf*-scores for topic identification, consequently unspecific and commonly used terms will be of low relevance anyway.

[6] The presented results are not sensible to the specific number of iterations in the vocabulary generation process.

The presented approach tackles the look-ahead bias and the described problem of underrepresentation inherent in fixed dictionary construction. With topic-related vocabulary possibly changing over time, we generate period-specific dictionaries to account for the time-dependent relevance of terms. Each period-specific dictionary defines the topic-related vocabulary that is used for next year's news classification.[7] Figure A4.2 exemplary shows word cloud summaries of period-specific vocabulary related to geopolitical risk events. The upper word cloud is generated from news data of the year 2001. The second word cloud shows the vocabulary used for topic identification almost twenty years later. The size of each term refers to its respective frequency. The figure illustrates how the relevance of vocabulary varies over time. While some terms seem to be of continued relevance over time (e.g., “soldier”, “bomb”), other frequent terms of one period are hardly to be found in the dictionary of the other period. For example, expressions related to terror events (e.g., “terrorism”, “terrorist”, “terror suspect”) are widespread in the early 2000’s initialized by the terrorist attacks on September 9th in 2001. While terrorism remains a global threat over time, the most representative vocabulary of period-specific dictionaries changes with the geopolitical scenario. From 2018, the nascent trade war between the United States and China resulted in a shift in vocabulary related to geopolitical risk. With the economic conflict gathering pace and increased media attention, more expressions to tariffs and other trade barriers have gained importance to the dictionary. Additionally, terms that reflect developments in modern warfare will be represented in dictionaries at the time of their emergence. While expressions referring to digital warfare (e.g., “cyber attack”, “cyber security”) are of

[7] Available data history starts in the beginning of the year 2000. Accordingly, the out-of-sample period starts in 2001.

increased relevance to dictionaries of recent years, these terms are missing in the vocabulary of 2001.

These findings indicate how topic identification and news index construction may benefit from a time-dependent dictionary generation. The main advantages of the presented approach are the opportunity to derive a domain- and period-specific dictionary from news by providing as little input as a few buzzwords. Hardly any human intervention or supervision is necessary.

4.2.2. Computation

Based on the generated vocabulary, we want to score unseen news articles by their relation to geopolitical risk events. Therefore, we apply the domain-specific dictionaries to approximate the similarity between the vocabulary of year t and any news article text of year $t+1$, for all years $T-1$ in the data sample. We use a score based on the "term frequency-inverse document frequency" (*tf-idf*), which is often applied in information retrieval and text mining. The *tf-idf* is composed by two functions: (i) the normalized term frequency (*tf*), which we derive directly from the domain-specific dictionary (ii) the inverse document frequency (*idf*), computed as the logarithm of the number of articles in the corpus (of year t) divided by the number of articles in which the considered term appears.[8]

We calculate the *tf-idf* for each term in the dictionary of year t. Next, we score news articles of the following period $t+1$ for their usage of vocabulary related to geopolitical risk events. Therefore, we calculate the sum of *tf-idf* scores over all terms for each news article in year $t+1$. The sum of *tf-idf* scores measures the weighted intersection in vocabulary between the article text and the domain-specific dictionary and is defined as

[8] For a detailed description on how the score is derived, see Section 3.3.

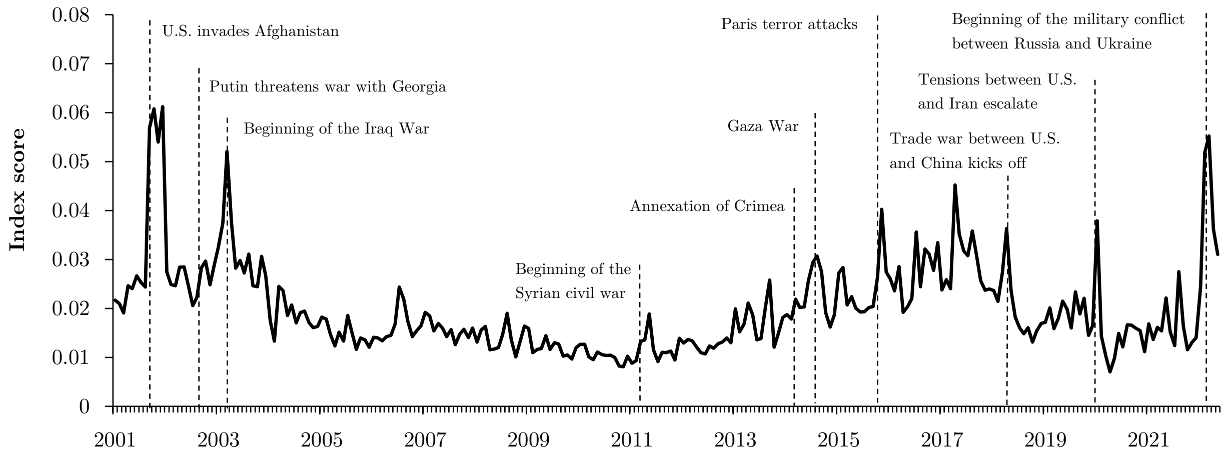
the relevance score w, where $0 \leq w \leq 1$. A value close to 1 refers to the inclusion of the most representative terms in the article text. We implement two conditions for noise reduction in the process of topic identification. First, for a given news article to be considered related to a geopolitical risk event, we set a threshold of $w \geq 0.05$ and a minimum term appearance in the article headline of one. Second, we require either the article headlines to include one of the 20% most representative bigrams or the article texts to include at least ten different dictionary terms. By doing so, we want to reduce false-positives and exclude less specific article types like news summaries, highlights, and market roundups. The criteria for topic identification are selected by the inspection of the in-sample period of the year 2000.

We run index calculations on a daily and monthly frequency for different regions and countries. The start and end times of each period are based on the NYSE trading dates and hours. News that is released after closing is considered in the next period. This includes news on weekends and holidays that is ascribed to the next trading day. The index score on media attention results from the sum of news articles identified to be related to a geopolitical risk event, divided by the total number of news articles in a given period.

4.2.3. Index calculation

Figure 4.1 shows the monthly time series of the Global Geopolitical Risk Event (GRE) index since 2001. The index value translates to the percental share of published news articles related to geopolitical risk events. We annotate influential risk events to make sense of the index scores. We begin the evaluation of the Geopolitical Risk Event index with a qualitative inspection of the monthly index scores and a discussion of the index behavior in context of some particular events.

Figure 4.1: Global Geopolitical Risk Event Index



On the primary y-axis the Geopolitical Risk Event Index is reported. Annotations are provided for news events related to geopolitical risk. Monthly data are reported from January 2001 to May 2022.

The index surged and peaked shortly after the 9/11 events when the U.S. decided to invade Afghanistan in an almost instant reaction to the terrorist attacks in 2001. The index spiked again at the beginning of the Iraq War in 2003 before recording relatively low index values for most of the following ten years. With the Russian annexation of Crimea in 2014 the global index entered a regime of elevated scores. Several terror attacks in Europe shaped this period, the Gaza War, escalating tensions between the US and Iran, threating gestures from North Korea and ultimately, the emerging trade war between the U.S. and China. With the beginning of the corona crisis in early 2020 the index slumped dramatically as a shift in news coverage drew global media attention towards the effects of the unfolding pandemic. Unfortunately, the index significantly increased most recently with the beginning of the military conflict between Russia and Ukraine. We observe that in several cases, the perceived risk, as measured by relative media attention, fades quickly in comparison to the duration of the events (e.g., the Iraq war and the Russian invasion of Ukraine).

4.3. Decomposition of Geopolitical Risk

We illustrate the usefulness of our geopolitical risk event index with a case study. We aim to build a global risk index from country-by-country data to identify different dimensions of geopolitical risk (latent components). Instead of employing a principal component model, we account for the autoregressive nature of our news-based series (as newspapers copy or follow up competitors' stories) by estimating a latent dynamic factor model. To account for the persistence of local geopolitical risks, we postulate in line with the econometric literature that

$$X_t = Cf_t + e_t, e_t \sim N(0, S) \tag{1}$$

$$f_t = Af_{t-1} + n_t, n_t \sim N(0, W) \tag{2}$$

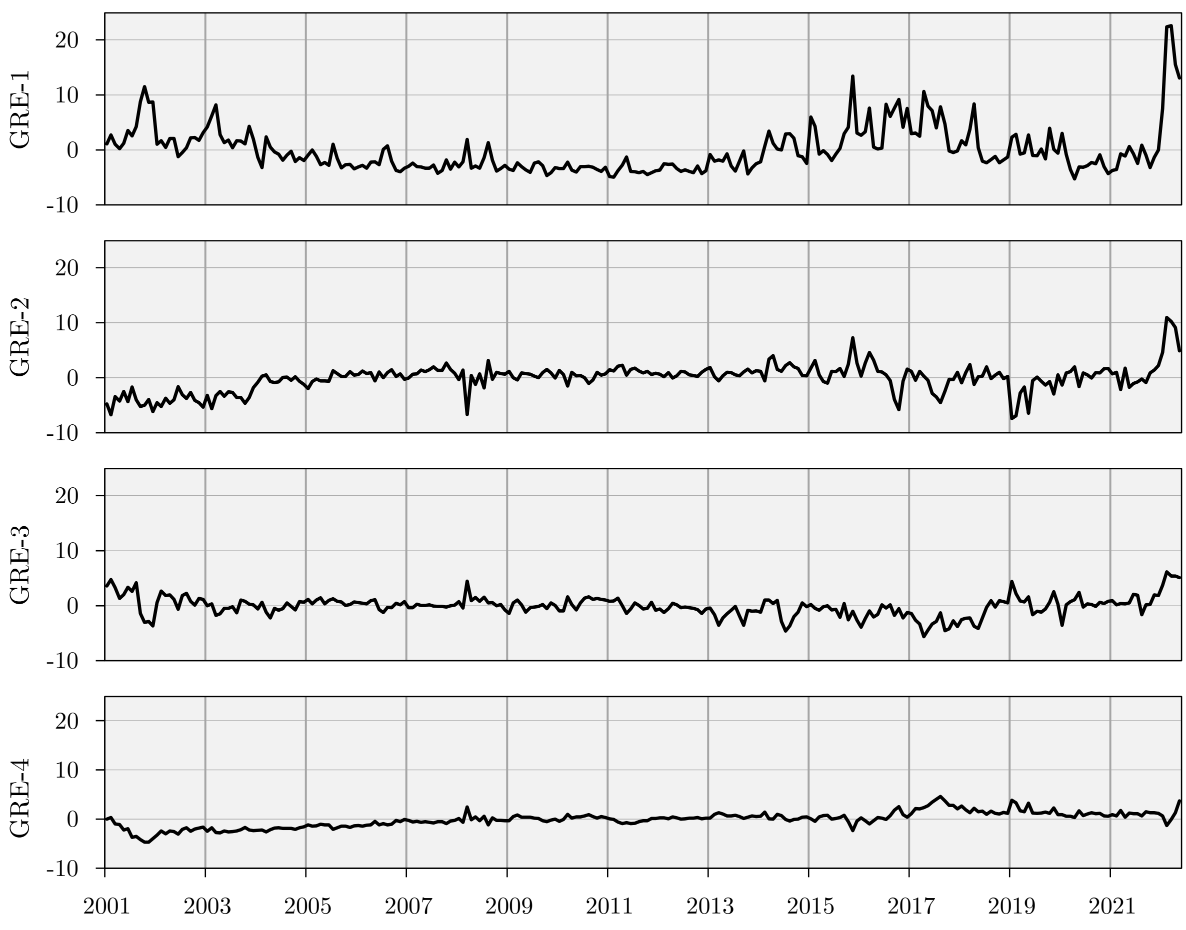
Here X_t is a $p'1$ vector of country-by-country geopolitical risk indices at time t, C is a $p'q$ factor loading matrix (q denotes the number of latent factors), f_t is a $q'1$ vector of latent factors at time t, e_t is a $p'1$ vector of measurement errors with a diagonal covariance matrix S, A denotes a $q'q$ matrix describing the auto- and cross regressive nature of latent factors and n_t is a random shock vector with the covariance matrix W. Estimation of equations (1) and (2) for $q = 4$ yields a time series for each latent factor as shown in Figure 4.2.[9]

[9] We estimate (1) and (2) using EM maximization by Doz et al., 2012.

We find substantial autocorrelation in our latent factors but little cross-regressive correlation, as can be seen by our estimates for matrix A.

$$A = \begin{bmatrix} 0.82 & 0.05 & 0.14 & 0.02 \\ 0.02 & 0.72 & 0.1 & 0.27 \\ -0.02 & 0.07 & 0.68 & -0.12 \\ 0.01 & 0.12 & -0.05 & 0.87 \end{bmatrix}$$

Figure 4.2: Latent Risk Factors



The figure shows the time series of latent risk factors with estimates of f_t in equation (1) for $q = 4$. The variance of each factor varies due to its importance.

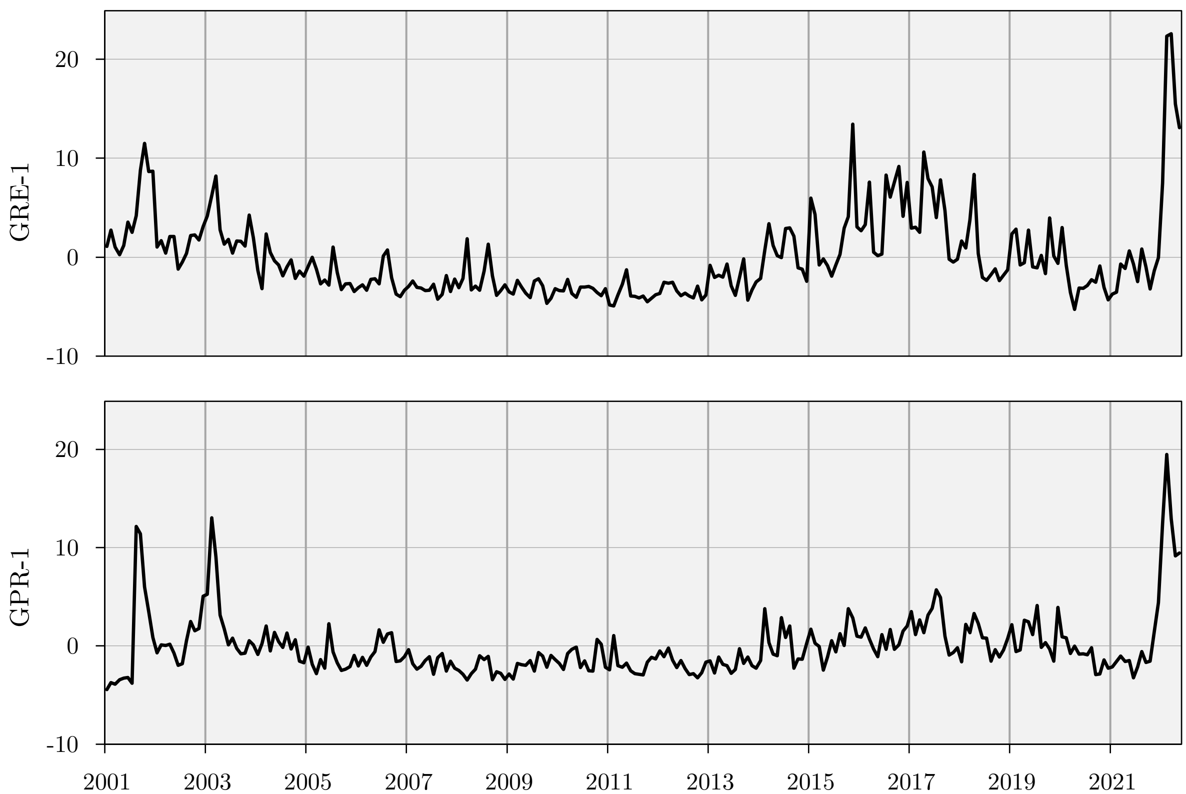
Table 4.2: Exposure to Latent Risk Factors

Continent	Country	Factor 1 (46%)	Factor 2 (61%)	Factor 3 (69%)	Factor 4 (74%)
South America	Argentina	0.13	-0.3	0.17	0.07
	Brazil	0.13	-0.29	0.18	0.05
	Chile	0.13	-0.3	0.19	0.07
	Colombia	0.09	-0.24	0.09	-0.15
	Mexico	0.04	-0.09	0.14	0.1
	Peru	0.12	-0.31	0.18	0.04
	Venezuela	0.07	-0.08	0.21	0.14
North America	Canada	0.1	-0.02	0.03	-0.09
	USA	0.14	-0.16	-0.12	-0.22
Europe	Belgium	0.22	0.13	0	-0.06
	Denmark	0.22	0.15	0.05	-0.05
	Finland	0.22	0.15	0.06	-0.03
	France	0.18	0.1	-0.07	-0.06
	Germany	0.22	0.1	0.03	-0.07
	Italy	0.21	0.12	0.04	-0.13
	Netherlands	0.22	0.14	0.05	-0.05
	Norway	0.22	0.13	0.05	-0.07
	Poland	0.21	0.15	0.08	-0.03
	Portugal	0.22	0.14	0.04	-0.04
	Russia	0.18	0.03	0.08	0
	Spain	0.2	0.07	0.03	-0.14
	Sweden	0.21	0.15	0.06	-0.01
	Switzerland	0.22	0.14	0.04	-0.03
	Turkey	0.11	0.02	-0.12	0.08
	Ukraine	0.13	0.09	0.06	0.09
	United Kingdom	0.2	0.1	0.03	-0.07
Asia	China	0.11	-0.03	0.1	0.16
	Hong Kong	0.05	-0.1	0	-0.1
	India	0.08	-0.21	-0.11	-0.22
	Indonesia	0.13	-0.18	-0.27	-0.13
	Israel	0.11	-0.06	-0.02	-0.23
	Japan	0.12	0.02	-0.07	0.21
	Malaysia	0.14	-0.12	-0.38	0.01
	Philippines	0.13	-0.14	-0.32	0.04
	Saudi Arabia	0.13	-0.16	-0.24	-0.14
	South Korea	0.13	-0.04	-0.24	0.3
	Taiwan	0.12	-0.06	-0.2	0.19
	Thailand	0.11	-0.09	-0.34	0.03
Africa	South Africa	0.12	-0.28	0.17	-0.02

The table shows the exposure of latent risk factor by continent and country. Estimates of factor loadings result from matrix C for $q = 4$ in equation (1). The cumulative explained variance for each factor is shown in brackets.

We document the factor loadings in Table 4.2. Our estimated loadings show that the first latent factor resembles a global factor, as all countries load positively. The second factor is long Europe versus the rest of the world and picks up stress in Europe, i.e., the various inner European issues. Factors three and four become increasingly difficult to interpret and likely resemble noise, given the low additional variance explained by these factors.

Figure 4.3: Comparison between GPR and GRE Index



We show the first latent factor from our dynamic factor model applied to data derived from our methodology and by Caldara and Iacoviello (2022). Data period from January 2011 to May 2022. The Geopolitical Risk (GPR) Index data are accessible on Matteo Iacoviello's website at https://www.matteoiacoviello.com/gpr.htm.

For further evaluation of our news-based methodology of geopolitical risk approximation, we compare its behavior with the widely used GPR index data as calculated by Caldara and Iacoviello (2022). For the U.S., the authors find that a shock to geopolitical risk induces a persistent decrease in investment and employment using vector autoregressive (VAR) models. Caldara and Iacoviello (2022) conclude that the decline in economic activity is due to both the threat and the realization of adverse geopolitical events.

Figure 4.3 shows the time series of first latent factor from our dynamic factor model applied to our GRE index data and the GPR index data. Our methodology creates a highly correlated time series (correlation of 0.75) for the first latent factor. Given the significance of previous findings with geopolitical risk proxies, we think this is a solid result for using a point-in-time, unsupervised technique (essentially pure machine learning) without relevant domain knowledge. Of course, we still are by no means experts in geopolitical risk.

Our index shows increased geopolitical risks from 2013 to 2018, while the GPR index does not. We believe this is due to our language model picking up alternative geopolitical tensions like the European refugee crisis, a fallout from the emerging Syrian war, and the build-up of the trade war between the US and China.

4.4. Conclusion

We use unsupervised learning techniques to build a real-time dictionary for war related media attention to measure geopolitical tensions. Our method applies to all languages and topics and does not require domain knowledge. Despite the lack of using expert knowledge, we are able to create a media attention series that not only closely resemble more heavily curated methods but also pick up changes in the dimensions of warfare from armed conflict to trade and cyber war.

5. Concluding remarks

This thesis comprises three studies that contribute to the field of asset pricing by expanding our knowledge how macroeconomic and different political factors drive time variation in asset returns and risk premia.

An important part of each study is the development of an innovative approach to approximate the non-observable variables. More frequent or alternative data are used to provide improved measures of latent risk factors. Real-time data utilized in predictive models support the estimation of current economic conditions. Alternative data sources can provide unique insights into the drivers of asset prices that traditional data sources may not capture. Thereby, alternative data has the potential to improve the validation of research hypothesis. Additionally, rapid advances in machine learning allow for the analysis of large volumes of this unstructured information.

This thesis provides novel techniques to approximate different risk factors based on sophisticated language models. At the time of writing, the release of ChatGPT, a state-of-the-art model to generate human-like text, receives widespread media attention as the next milestone in natural language processing. The steady technological progress will also inspire future research in the field of asset pricing. The opportunity to systematically access alternative data sources may help to explain and predict the behavior of asset returns and to enhance the accuracy of factor models by the reduction of noise. Investors are able to make decisions with increased confidence and to reduce the risk to make decisions based on outdated or incomplete information.

However, given the naturally myriad degrees of freedom resulting from possible model specifications, the rigorous confrontation of any bias in data handling is a prerequisite in the application of these tools.

Bibliography

Ahir, H., Bloom, N., & Furceri, D. (2022). The World Uncertainty Index. National bureau of economic research. Working paper.

Aldy, J. E., & Stavins, R. N. (2012). The promise and problems of pricing carbon: Theory and experience. The Journal of Environment & Development, 21(2), 152-180.

Amenc, N., Esakia, M., Goltz, F., & Luyten, B. (2019). Macroeconomic risks in equity factor investing. The Journal of Portfolio Management, 45(6), 39-60.

Amenc, N., Goltz, F., & Lodh, A. (2018). Mind the gap: On the importance of understanding and controlling market risk in smart beta strategies. The Journal of Portfolio Management, 44(4), 60-70.

Andersson, M., Bolton, P., & Samama, F. (2016). Hedging climate risk. Financial Analysts Journal, 72(3), 13-32.

Ang, A., & Bekaert, G. (2004). How regimes affect asset allocation. Financial Analysts Journal, 60(2), 86-99.

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

Ang, I., Israelov, R., Sullivan, R. N., & Tummala, H. (2018). Understanding the volatility risk premium. Working Paper.

Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. Working Paper.

Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2020). Climate change concerns and the performance of green versus brown stocks. Working paper.

Asvanunt, A., & Richardson, S. (2016). The credit risk premium. The Journal of Fixed Income, 26(3), 6-24.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. Quarterly Journal of Economics, 131(4), 1593-1636.

Baur, D. G., & Smales, L. A. (2020). Hedging geopolitical risk with precious metals. Journal of Banking & Finance, 117, 105823.

BCBS (2020). Climate-related financial risks: a survey on current initiatives. Basel Committee on Banking Supervision.

BCBS (2021). Climate-related risk drivers and their transmission channels. Basel Committee on Banking Supervision.

Beltagy, I., Lo, K., & Cohan, A. (2019). Scibert: A pretrained language model for scientific text. Working paper.

Bender, J., Sun, X., Thomas, R., & Zdorovtsov, V. (2018). The promises and pitfalls of factor timing. The Journal of Portfolio Management, 44(4), 79-92.

Berg, F., Koelbel, J. F., & Rigobon, R. (2019). Aggregate confusion: The divergence of ESG ratings. MIT Sloan School of Management. Working paper.

Berk, J. B., Green, R. C., & Naik, V. (1999). Optimal investment, growth options, and security returns. The Journal of Finance, 54(5), 1553-1607.

Bessembinder, H. (2018). Do stocks outperform treasury bills? Journal of Financial Economics, 129(3), 440-457.

Bingler, J., Kraus, M. and Leippold, M. (2021). Cheap Talk and Cherry-Picking: What ClimateBert has to say on Corporate Climate Risk Disclosure. Working paper.

Blin O., Ielpo, F., Lee, J., & Teiletche, J. (2018). Factor Timing Revisited: Alternative Risk Premia Allocation Based on Nowcasting and Valuation Signals. Working Paper.

Bolton, P., & Kacperczyk, M. (2021a). Global pricing of carbon-transition risk. Working Paper.

Bolton, P., & Kacperczyk, M. (2021b). Do investors care about carbon risk? Journal of Financial Economics.

Brecher, M., & Wilkenfeld, J. (1997). A study of crisis. University of Michigan Press.

Burch, I., & Gilchrist, J. (2018). Survey of global activity to phase out internal combustion engine vehicles. Center of Climate Protection. Working paper.

Caldara, D., & Iacoviello, M. (2022). Measuring Geopolitical Risk. American Economic Review, 112(4), 1194-1225.

Cambridge Centre for Sustainable Finance (2016). Environmental risk analysis by financial institutions: a review of global practice. Cambridge Institute for Sustainability Leadership.

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

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. Strategic Management Journal, 37(8), 1597-1614.

Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. Journal of business, 383-403.

Cheng, C. H. J., & Chiu, C. W. J. (2018). How important are global geopolitical risks to emerging countries. International economics, 156, 305-325.

Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. The Journal of Finance, 57(2), 985-1019.

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

Cooper, I., Mitrache, A., & Priestley, R. (2016). A Global Macroeconomic Risk Explanation for Momentum and Value. Working paper.

Cohen, L., Gurun, U. G., & Nguyen, Q. H. (2020). The ESG-innovation disconnect: Evidence from green patenting. National Bureau of Economic Research. Working Paper.

Coqueret, G. (2020). Stock-specific sentiment and return predictability. Quantitative Finance, 20(9), 1531-1551.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. Journal of Financial Economics, 122(2), 221-247.

Demirer, R., Lien, D., & Zhang, H. (2015). Industry herding and momentum strategies. Pacific-Basin Finance Journal, 32, 95-110.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. Working paper.

Dietz, S., Bowen, A., Dixon, C., & Gradwell, P. (2016). 'Climate value at risk' of global financial assets. Nature Climate Change, 6(7), 676-679.

Dimson, E., Marsh, P., & Staunton, M. (2020). Divergent ESG ratings. The Journal of Portfolio Management, 47(1), 75-87.

Dor, D. (2003). On newspaper headlines as relevance optimizers. Journal of Pragmatics, 35(5), 695-721.

Doz, C., Giannone, D., & Reichlin, L. (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. Review of economics and statistics, 94(4), 1014-1024.

Dyck, A., Lins, K. V., Roth, L., & Wagner, H. F. (2019). Do institutional investors drive corporate social responsibility? International evidence. Journal of Financial Economics, 131(3), 693-714.

Ecker, U. K., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. Journal of Experimental Psychology: Applied, 20(4), 323.

Edunov, S., Ott, M., Auli, M., & Grangier, D. (2018). Understanding back-translation at scale. Working paper.

Engle, R. F., & Campos-Martins, S. (2020). Measuring and hedging geopolitical risk. Working Paper.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging climate change news. The Review of Financial Studies, 33(3), 1184-1216.

European Commission. (2019). The European Green Deal. COM(2019) 640 Final.

Evans, C. L., & Marshall, D. A. (2007). Economic determinants of the nominal treasury yield curve. Journal of Monetary Economics, 54(7), 1986-2003.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1-22.

Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. Journal of Political Economy, 99(2), 385-415.

Flammer, C., Toffel, M. W., & Viswanathan, K. (2019). Shareholder activism and firms' voluntary disclosure of climate change risks. Strategic Management Journal.

Gertler, M., & Gilchrist, S. (1994). Monetary policy, business cycles, and the behavior of small manufacturing firms. The Quarterly Journal of Economics, 109(2), 309-340.

Gibson, M., Mullins, J. T., & Hill, A. (2017). Climate change, flood risk, and property values: Evidence from New York City. Working paper.

Giglio, S., Kelly, B., & Stroebel, J. (2020). Climate Finance. Annual Review of Financial Economics, forthcoming.

Giglio, S., Maggiori, M., Rao, K., Stroebel, J., & Weber, A. (2021). Climate change and long-run discount rates: Evidence from real estate. The Review of Financial Studies, 34(8), 3527-3571.

Görgen, M., Jacob, A., Nerlinger, M., Riordan, R., Rohleder, M., & Wilkens, M. (2020). Carbon risk. Working paper.

Gössling, S., & Cohen, S. (2014). Why sustainable transport policies will fail: EU climate policy in the light of transport taboos. Journal of Transport Geography, 39, 197-207.

Hakkio, C. S., & Keeton, W. R. (2009). Financial stress: what is it, how can it be measured, and why does it matter. Economic Review, 94(2), 5-50.

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. The Journal of Finance, 74(6), 2789-2837.

Harvey, C. R., & Liu, Y. (2015). Backtesting. Journal of Portfolio Management, 42(1), 13-28.

Hauer, M. E., Evans, J. M., & Mishra, D. R. (2016). Millions projected to be at risk from sea-level rise in the continental United States. Nature Climate Change, 6(7), 691-695.

Hepburn, C. (2006). Regulation by prices, quantities, or both: a review of instrument choice. Oxford review of economic policy, 22(2), 226-247.

Hodges, P., Hogan, K., Peterson, J. R. & Ang, A. (2017). Factor timing with cross-sectional and time-series predictors. The Journal of Portfolio Management, 44(1), 30-43.

Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2019). The curious case of neural text degeneration. Working paper.

Ilmanen, A., Maloney, T., & Ross, A. (2014). Exploring macroeconomic sensitivities: How investments respond to different economic environments. The Journal of Portfolio Management, 40(3), 87-99.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R. New York: Springer.

Karagozoglu, A. K., Wang, N., & Zhou, T. (2022). Comparing Geopolitical Risk Measures. The Journal of Portfolio Management, 48(10), 226-257.

Kavlak, G., McNerney, J., & Trancik, J. E. (2018). Evaluating the causes of cost reduction in photovoltaic modules. Energy Policy, 123, 700-710.

Kiyotaki, N., & Moore, J. (1997). Credit cycles. Journal of political economy, 105(2), 211-248.

Koyejo, O., Natarajan, N., Ravikumar, P., & Dhillon, I. S. (2015). Consistent Multilabel Classification. Advances in Neural Information Processing Systems, 29, 3321-3329.

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. Decision Support Systems, 104, 38-48.

Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. The Review of Financial Studies, 33(3), 1067-1111.

Lee, W. (2000). Theory and methodology of tactical asset allocation. John Wiley & Sons (Vol. 65).

Lindemann, B. (1990). Cheap thrills we live by-some notes on the poetics of tabloid headlinese. Journal of Literary Semantics, 19(1), 46-59.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 54(4), 1187-1230.

Ludvigson, S.C., & Ng, S. (2009). Macro factors in bond risk premia. The Review of Financial Studies, 22(12), 5027-5067.

Molek-Kozakowska, K. (2013). Towards a pragma-linguistic framework for the study of sensationalism in news headlines. Discourse & Communication, 7(2), 173-197.

Monasterolo, I., & Raberto, M. (2019). The impact of phasing out fossil fuel subsidies on the low-carbon transition. Energy Policy, 124, 355-370.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2015). Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. Expert Systems with Applications, 42(1), 306-324.

Ng, N., Yee, K., Baevski, A., Ott, M., Auli, M., & Edunov, S. (2019). Facebook FAIR's WMT19 News Translation Task Submission. Working paper.

NGFS (2020a). Guide for supervisors: integrating climate-related and environmental risks into prudential supervision. Network for Greening the Financial System.

NGFS (2020b). Guide to Climate Scenario Analysis for Central Banks and Supervisors. Network for Greening the Financial System.

Nilsson, R., & Guidetti, E. (2008). Current Period Performance of OECD Composite Leading Indicators. Journal of Business Cycle Measurement and Analysis, 2007(2), 235-266.

Nugent, T., Stelea, N., & Leidner, J. L. (2020). Detecting ESG topics using domain-specific language models and data augmentation approaches. Working paper.

Pastor, L., & Veronesi, P. (2012). Uncertainty about government policy and stock prices. The Journal of Finance, 67(4), 1219-1264.

Pastor, L., Stambaugh, R. F., & Taylor, L. A. (2020). Sustainable investing in equilibrium. Journal of Financial Economics.

Perez-Quiros, G., & Timmermann, A. (2000). Firm size and cyclical variations in stock returns. The Journal of Finance, 55(3), 1229-1262.

Ramelli, S., Ossola, E., & Rancan, M. (2021). Stock Price Effects of Climate Activism: Evidence from the First Global Climate Strike. Working paper.

Semieniuk, G., Campiglio, E., Mercure, J. F., Volz, U., & Edwards, N. R. (2021). Low-carbon transition risks for finance. Wiley Interdisciplinary Reviews: Climate Change, 12(1).

Setzer, J & Byrnes, R. (2020). Global trends in climate change litigation: 2020 snapshot, Grantham Research Institute on Climate Change and the Environment. Working Paper.

Sheng, X., Gupta, R., & Cepni, O. (2022). Persistence of state-level uncertainty of the United States: The role of climate risks. Economics Letters, 215, 110500.

Task Force on Climate-related Financial Disclosures (2017). Recommendations of the Task Force on Climate-related Financial Disclosures (TCFD).

Thomson, E. A., White, P. R., & Kitley, P. (2008). "Objectivity" and "hard news" reporting across cultures: Comparing the news report in English, French, Japanese and Indonesian journalism. Journalism studies, 9(2), 212-228.

United Nations Environment Programme - Finance Initiative (2015): Climate Strategies and Metrics: Exploring Options for Institutional Investors (UNEP FI).

United Nations Environment Programme - Finance Initiative (2017): The status of climate change litigation - a global review (UNEP FI).

United Nations Environment Programme - Finance Initiative (2019): Changing course: UNEP FI and twenty institutional investors launch new guidance to implement TCFD Recommendations (UNEP FI).

Van Vliet, P., & Blitz, D. (2011). Dynamic strategic asset allocation: Risk and return across the business cycle. Journal of Asset Management, 12(5), 360-375.

Wells, V. K., Ponting, C. A., & Peattie, K. (2011). Behaviour and climate change: Consumer perceptions of responsibility. Journal of Marketing Management, 27(7-8), 808-833.

Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. Proceedings of the IEEE international conference on computer vision 2015, 19-27.

Appendices

Appendix to Chapter 2

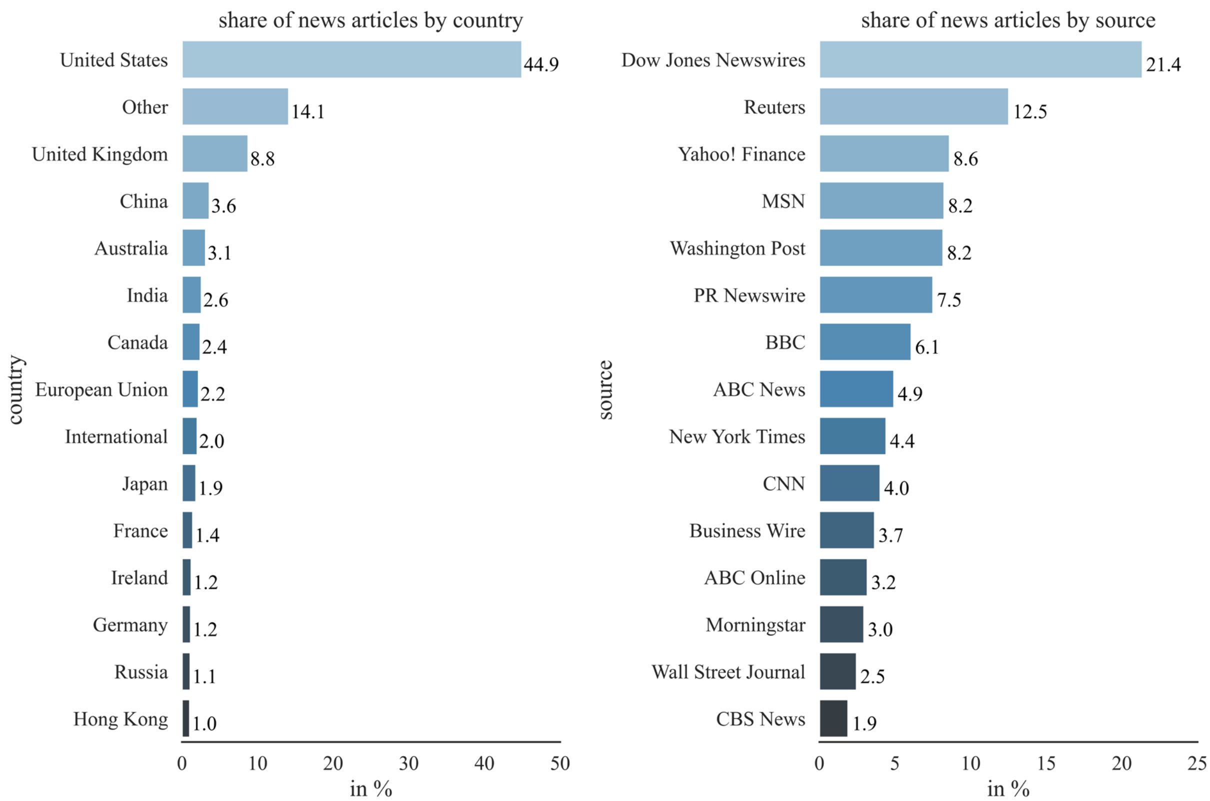
Table A2.1: Detailed information on considered risk premia strategies

Risk premia strategy	Rational
Equity Multifactor	Multifactor strategy aims to provide exposure to number of well-known equity risk premia such as Momentum and Value.
Equity Defensive	Defensive strategies are designed to provide defensive characteristics during stress scenarios.
Equity Trend FX Trend Interest Rate Trend Commodity Trend Cross Asset Trend	Trend strategies aim to exploit time series trends in asset prices.
Equity Carry Credit Carry FX Carry Interest Rate Carry	Carry strategies overweight high-yielding assets and underweight or short low-yielding assets.
FX Value Interest Rate Value	Value strategies seek premia from overweighting undervalued assets and underweighting or shorting overvalued assets.
Credit Curve Interest Rate Curve Commodity Curve	Curve strategies seek to earn a premium by going long and short different maturities on a yield/futures curve.
Equity Vol. Carry Credit Vol. Carry FX Vol. Carry Interest Rate Vol. Carry Commodity Vol. Carry	Volatility carry strategies aim to harvest volatility risk premium which arises from supply and demand imbalances for options.
Equity Imbalance (1) (2) Commodity Imbalance (1) (2)	Imbalance strategies seek to profit from structural imbalances in markets.

The table provides details on the considered Alternative Risk Premia strategies regarding their rationale.

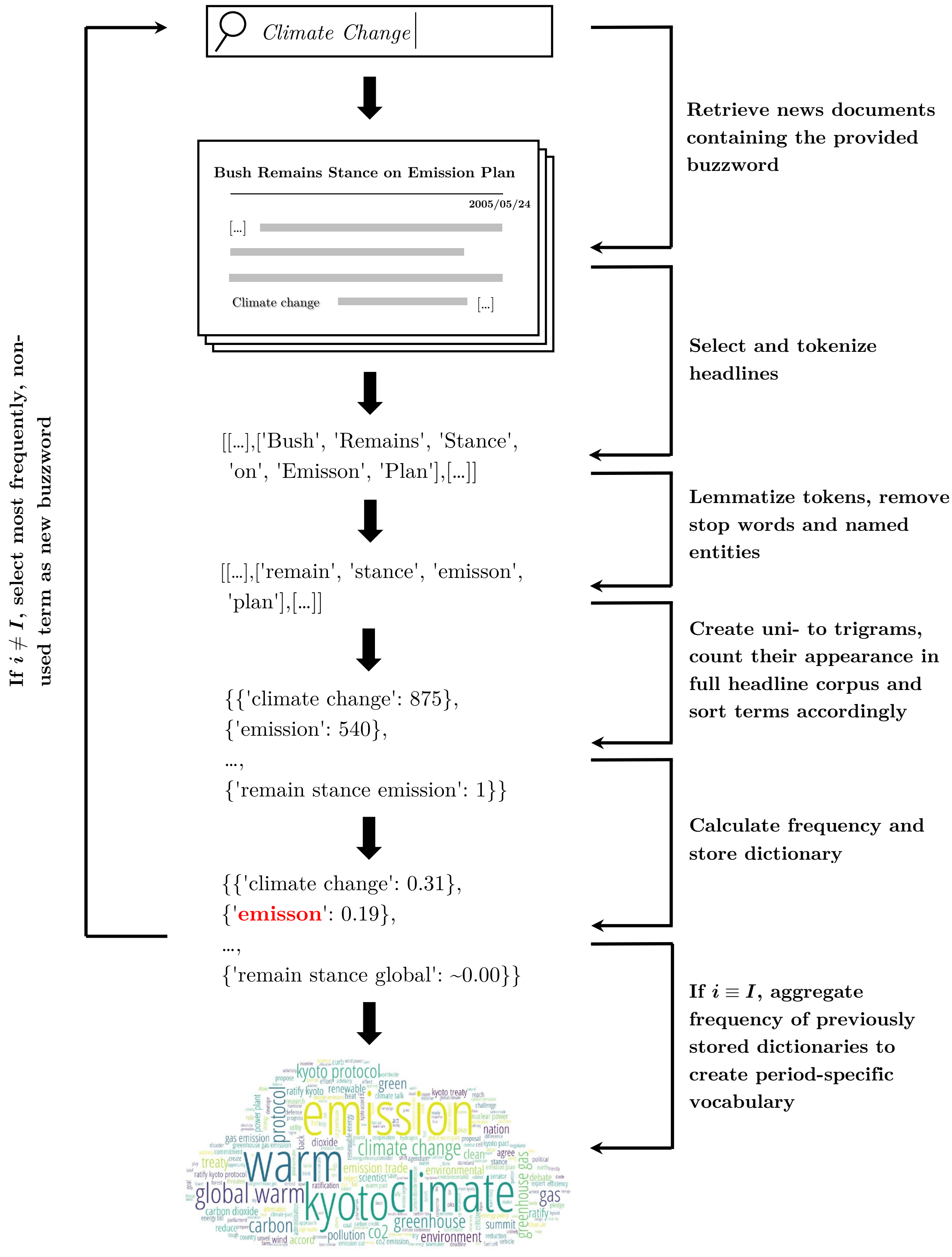
Appendix to Chapter 3

Figure A3.1: Share of news articles by country and source



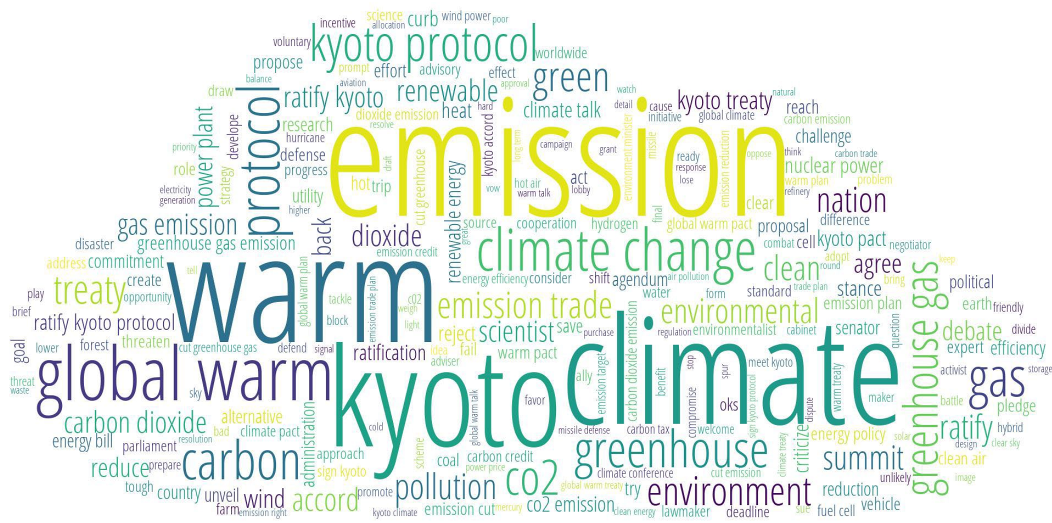
This figure shows the share of news articles by media source and country. Figures are based on the news data sample from January 2000 to December 2020.

Figure A3.2: Iterative approach of vocabulary generation



Used for illustrative purposes only. Numerical values are exemplary and not real observations.

Figure A3.3: Word cloud summary of period-specific Climate Change Vocabulary



2005

2019

The figure shows word cloud summaries of period-specific climate change vocabulary. The upper word cloud utilizes news of 2005. The vocabulary of the bottom word cloud is generated from news in 2019. Term sizes are proportional to their frequency in the text corpus.

Table A3.1: Evaluation of periods with highest and lowest transition risk in the out of sample data

Period	TRI Score	Most contributing headlines	Sentiment model prediction			Date	Source
			Negative	Neutral	Positive		
2016-09	9.61	EU states agree UN's 2015 climate deal ratification	0.0004	0.0006	0.9990	2016-09-30	Reuters
		Paris climate deal expected to receive major boost at UN	0.0000	0.0001	0.9998	2016-09-20	CBS News
		Finally, USA and China ratify Paris climate deal	0.0001	0.0001	0.9999	2016-09-04	MSN
		USA, China join climate deal in turning point for planet	0.0001	0.0002	0.9997	2016-09-03	Yahoo! Finance
2016-10	8.24	With EU backing, Paris climate deal clears last hurdle to taking effect	0.0000	0.0021	0.9978	2016-10-04	Reuters
		The world is about to get tough on aviation emissions.	0.0000	0.0002	0.9998	2016-10-14	Washington Post
		200 nations reach landmark climate deal	0.0000	0.0003	0.9996	2016-10-15	CNN
		Paris Climate Accord to Take Effect; Obama Hails Historic Day	0.0001	0.0006	0.9993	2016-10-05	New York Times
2015-12	5.60	We have a Paris climate agreement. Now what?	0.0000	0.0001	0.9999	2015-12-14	CNN
		Climate accord is a big win for Obama, even as dangers loom	0.0029	0.0076	0.9895	2015-12-13	Washington Post
		Even Republicans will uphold climate deal	0.0000	0.0001	0.9999	2015-12-01	Yahoo! Finance
		Leaders urge breakthrough at climate talks	0.0000	0.0000	1.0000	2015-12-01	Reuters
2017-03	-6.93	Trump's big environmental regulation rollback is all kinds of unpopular	1.0000	0.0000	0.0000	2017-03-29	Washington Post
		Trump puts anti-global warming projects on chopping block	1.0000	0.0000	0.0000	2017-03-28	Yahoo! Finance
		Trump tosses Obama's clean energy plan, embraces coal	0.5211	0.4789	0.0000	2017-03-28	Yahoo! Finance
		Trump to scrap Obama climate policies	1.0000	0.0000	0.0000	2017-03-28	BBC
2009-12	-14.05	Rich nations slam climate draft, thousands protest	0.6854	0.3114	0.0032	2009-12-12	Washington Post
		Big developing states reject Copenhagen climate plan	1.0000	0.0000	0.0000	2009-12-02	Washington Post
		Negotiators at Climate Talks Face Deep Set of Fault Lines	0.9194	0.0805	0.0001	2009-12-06	New York Times
		Emission Curbs Are Weak; USA Calls Compromise With China	1.0000	0.0000	0.0000	2009-12-19	Dow Jones Newswires
2017-06	-22.82	USA quitting the Paris climate agreement would be a moral disgrace	0.9999	0.0000	0.0001	2017-06-01	MSN
		Trump under fire over expected global climate deal withdrawal	0.9999	0.0000	0.0001	2017-05-31	Yahoo! Finance
		Trade rivals have limited armoury as USA quits climate deal	0.9998	0.0002	0.0000	2017-06-02	Yahoo! Finance
		Leaving climate deal likely wouldn't add USA jobs	1.0000	0.0000	0.0000	2017-05-31	ABC News

The table shows exemplary the three periods with the highest respectively the lowest TRI score from January 2008 to December 2020, sorted by the TRI score. The initial TRI score was scaled by factor of 10,000 for visualization. For each period, the most contributing headlines are reported as measured by the relevance score. The sentiment model prediction is the softmax function output of the classification model. Sentiment is labeled according to the highest probability of the classification labels. Additionally, date and source of the selected news headlines are reported. The out of sample period for the sentiment prediction starts in January 2008.

Table A3.2: Index description

(A) Pure-play indices

Name	Benchmark	History	Launch	Constituents	Rationale
MSCI Global Environment Index	MSCI ACWI IMI	Nov 08	Jan 09	varying >200	Index is comprised of companies that derive at least 50% of their revenues from environmentally beneficial products and services. The index is based on various key themes like green building, pollution prevention or clean technology. Constituent selection is based on data from MSCI ESG Research.
MSCI Global Alternative Energy Index	MSCI ACWI IMI	Jan 09	Jan 09	varying >50	Thematic sub-index of the MSCI Global Environment Index and includes companies that derive 50% or more of their revenues from products and services in Alternative Energy.
S&P Global Clean Energy Index	S&P Global BMI	Nov 03	Feb 07	fixed 30	Index inclusion is based on factors like company's business description and most recent reported revenue by segment. Companies are identified as being in the clean energy business for their involvement in the production of clean energy or provision of clean energy technology & equipment. Companies which exceed a certain carbon emissions threshold are excluded.
Solactive Climate Change Index	Solactive GBS Dev Mkts Large Mid Cap Index	Nov 05	Oct 07	fixed 30	Index includes the 30 largest companies active in the sectors agribusiness and biofuel, CO2 reduction, water, waste management, solar and wind energy. Companies are classified according to the percentage of total revenues associated with activities that generate CO2 avoidance.
FTSE EO Renewable and Alternative Energy Index	FTSE Global All Cap	Nov 08	Jun 09	fixed 50	The Environmental Opportunities (EO) Index requires companies to have at least 20% of their revenues derived from significant involvement in business activities related to Renewable & Alternative Energy as defined by the FTSE Environmental Markets Classification System (EMCS).
FTSE Environmental Technology Index	FTSE Global All Cap	Oct 03	Jan 08	fixed 50	Index consists of companies that are required to have at least 50% of their business derived from environmental markets and technologies as defined by the FTSE EMCS. The Index was first launched by Impax AM in 1999, making it the longest-running environmental technology index available.

(B) Decarbonized indices

Name	Benchmark	History	Launch	Constituents	Rationale
MSCI World Climate Change Index	MSCI World	Nov 13	Jun 19	varying >1000	The index uses the MSCI Low Carbon Transition score to re-weight benchmark constituents to increase exposure to companies participating in opportunities and decrease exposure to companies exposed to risks associated with transition, while seeking to minimize exclusions from the parent index.
MSCI World Low Carbon Leaders Index	MSCI World	Nov 10	Nov 14	varying >1000	Index is designed to address two dimensions of carbon exposure – carbon emissions and fossil fuel reserves. By selecting companies low carbon emissions (relative to sales) and those with low potential carbon emissions (per dollar of market capitalization), the index aims to achieve at least 50% reduction in its carbon footprint while minimizing the tracking error.
MSCI World Low Carbon Target Index	MSCI World	Nov 10	Sep 13	varying >1500	By overweighting companies with low carbon emissions and those with low potential carbon emissions the index aims to minimize the carbon exposure subject to a tracking error constraint of 30 basis points relative to the parent index. It uses MSCI ESG CarbonMetrics data from MSCI ESG Research Inc.
S&P Global 1200 Fossil Fuel Free Index	S&P Global 1200	Dec 11	Aug 15	varying >1000	Index is based on its respective underlying index and consists of companies that do not own fossil fuel reserves as measured by S&P Trucost Limited.
S&P Global LargeMidCap Carbon Efficient Index	S&P Global LargeMidCap	Mar 09	Jul 18	varying >2000	Index measures the performance of companies in the underlying index, excluding companies classified as high non-disclosing carbon emitters, while overweighting or underweighting those companies that have lower or higher levels of GHG emissions per unit of revenue as defined in eligibility criteria.
STOXX Global 1800 Low Carbon	STOXX Global 1800	Dec 11	Feb 16	varying >1500	Index closely track the underlying benchmark while offering a reduction in carbon emissions by overweighting lower carbon emitters and underweighting higher carbon emitters. STOXX uses CDP and ISS ESG as data sources. Data considered comprise Scope 1 and Scope 2 emissions.

The table provides descriptive information for the pure-play indices in panel (A) and decarbonized indices in panel (B) considered in this paper. The column "History" provides the earliest date of available index price data, while "Launch" is the date of actual index inception. The number of constituents is either fixed per definition or varying over time while the given value relates to the historical average. The benchmark was selected based on the universe used for constituent selection. All index information is retrieved and accessible from the providers' websites and index factsheets for MSCI, S&P, Solactive, FTSE Russell and STOXX.

Table A3.3: Correlation matrix of active index returns

	Pure-play indices						Decarbonized indices					
	MSCI GLENV	MSCI GLALT	S&P GLCE	Solactive CC	FTSE EORAN	FTSE ENVT	MSCI WCC	MSCI WLCL	MSCI WLCT	S&P 1200 FFF	S&P LMCCE	STOXX 1800 LC
MSCI GLENV	1.000 (631)											
MSCI GLALT	0.422 (623)	1.000 (623)										
S&P GLCE	0.388 (631)	0.733 (623)	1.000 (893)									
Solactive CC	0.247 (631)	0.586 (623)	0.600 (764)	1.000 (764)								
FTSE EORAN	0.249 (631)	0.647 (623)	0.485 (633)	0.611 (633)	1.000 (633)							
FTSE ENVT	0.641 (631)	0.580 (623)	0.695 (893)	0.560 (764)	0.338 (633)	1.000 (939)						
MSCI WCC	0.208 (370)	-0.050 (370)	-0.069 (370)	-0.239 (370)	-0.196 (370)	0.124 (370)	1.000 (370)					
MSCI WLCL	0.089 (526)	0.068 (526)	0.056 (526)	0.025 (526)	0.086 (526)	0.130 (526)	0.305 (370)	1.000 (526)				
MSCI WLCT	0.028 (526)	0.012 (526)	0.004 (526)	0.003 (526)	-0.048 (526)	0.088 (526)	0.526 (370)	0.384 (526)	1.000 (526)			
S&P 1200 FFF	-0.041 (470)	-0.093 (470)	-0.089 (470)	-0.154 (470)	-0.156 (470)	0.025 (470)	0.714 (370)	0.211 (470)	0.503 (470)	1.000 (470)		
S&P LMCCE	-0.018 (615)	-0.086 (615)	-0.157 (615)	-0.103 (615)	0.006 (615)	-0.050 (615)	0.036 (370)	0.011 (526)	-0.007 (526)	-0.054 (470)	1.000 (615)	
STOXX 1800 LC	0.019 (471)	0.029 (471)	0.024 (471)	-0.085 (471)	0.008 (471)	0.035 (471)	0.018 (370)	-0.098 (471)	-0.092 (471)	-0.060 (471)	0.213 (471)	1.000 (471)

The table shows the cross-correlations of active weekly index returns for all pure-play and decarbonized indices considered in this paper (see Table 2). Pearson correlation and pairwise complete observations are used for calculation. The number of pairwise observations used for calculation is reported in the parentheses.

Table A3.4: Regression of weekly active index returns

(A) Pure-play indices

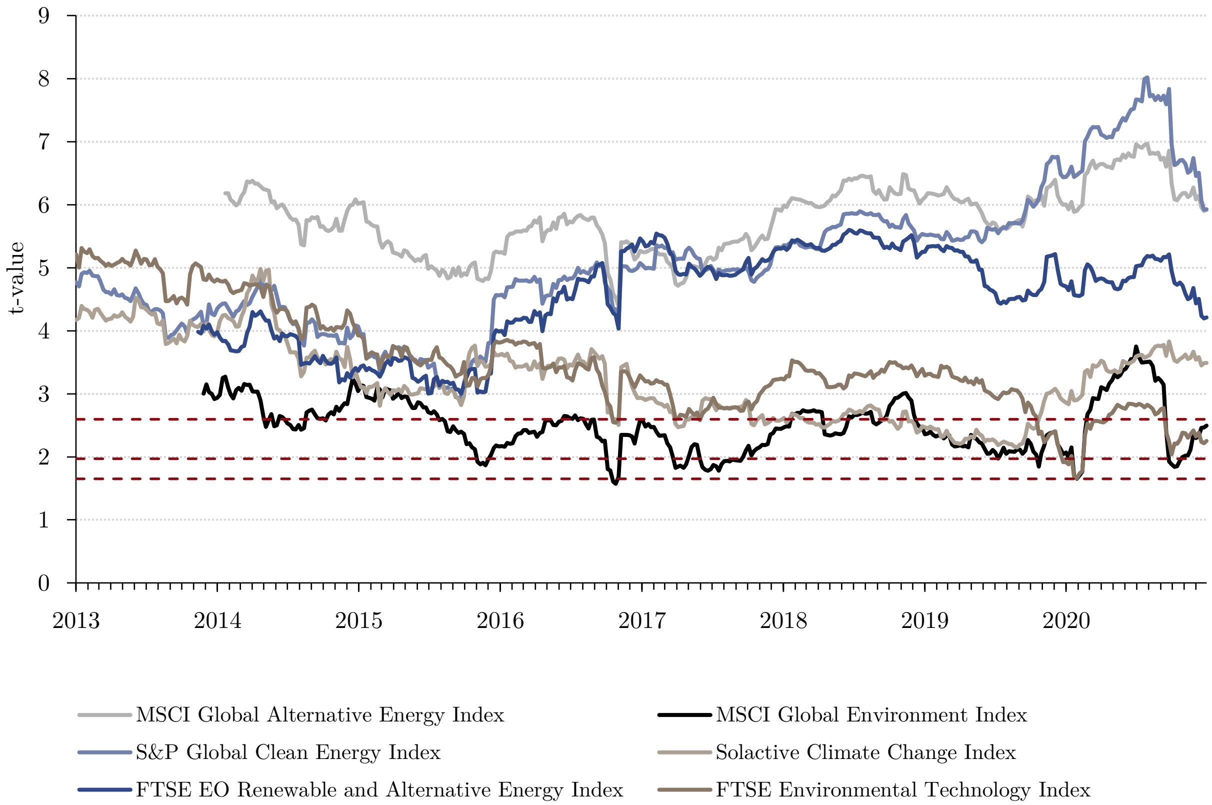
	MSCI GLENV	MSCI GLALT	S&P GLCE	Solactive CC	FTSE EORAE	FTSE ENVT
MKT	0.094***	0.032***	0.077***	0.018*	-0.026**	0.076***
	(0.012)	(0.012)	(0.011)	(0.011)	(0.013)	(0.01)
SMB	0.084***	0.037*	0.025	0.075***	0.025	0.111***
	(0.022)	(0.022)	(0.02)	(0.021)	(0.023)	(0.018)
HML	-0.068***	-0.008	0.019	0.011	0.016	-0.035*
	(0.022)	(0.023)	(0.02)	(0.021)	(0.023)	(0.019)
RMW	-0.053	-0.134***	-0.158***	0.012	-0.034	-0.056**
	(0.033)	(0.034)	(0.029)	(0.03)	(0.033)	(0.027)
CMA	0.012	-0.021	-0.143***	-0.080**	0.038	-0.154***
	(0.042)	(0.043)	(0.039)	(0.04)	(0.043)	(0.036)
MOM	-0.008	-0.019	-0.02	0.002	0.016	0.0001
	(0.014)	(0.015)	(0.013)	(0.013)	(0.015)	(0.012)
TRI	2.650***	6.363***	5.327***	4.199***	5.061***	3.814***
	(0.676)	(0.678)	(0.647)	(0.669)	(0.706)	(0.602)
Constant	0.0002	-0.0003	-0.0003	-0.0003	-0.0002	-0.0001
	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0002)
N	631	623	679	679	633	679
R-squared	0.173	0.176	0.283	0.09	0.097	0.26

(B) Decarbonized indices

	MSCI WCC	MSCI WLCL	MSCI WLCT	S&P 1200 FFF	S&P LMCCE	STOXX 1800 LC
MKT	-0.011	-0.007	-0.039***	-0.029**	0.097***	0.037**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)	(0.015)
SMB	0.013	0.055**	-0.008	0.029	-0.046**	0.021
	(0.026)	(0.028)	(0.028)	(0.027)	(0.023)	(0.029)
HML	-0.078***	0.104***	-0.003	-0.03	-0.066***	-0.108***
	(0.026)	(0.029)	(0.029)	(0.027)	(0.024)	(0.03)
RMW	0.162***	-0.069*	0.024	0.175***	0.205***	0.083*
	(0.04)	(0.041)	(0.041)	(0.04)	(0.035)	(0.044)
CMA	-0.300***	-0.102*	-0.222***	-0.219***	0.013	-0.087
	(0.048)	(0.053)	(0.052)	(0.05)	(0.044)	(0.054)
MOM	0.092***	0.050**	-0.01	0.110***	0.050***	-0.043**
	(0.019)	(0.021)	(0.02)	(0.02)	(0.015)	(0.021)
TRI	-0.914	0.35	0.333	-1.399*	0.448	-1.295
	(0.741)	(0.833)	(0.829)	(0.776)	(0.692)	(0.847)
Constant	-0.001***	-0.003***	-0.004***	0.001***	-0.0005*	0.003***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
N	370	526	526	470	615	471
R-squared	0.36	0.045	0.055	0.227	0.152	0.075

This table shows results from the multivariate time-series regression (1) for the pure-play indices in panel (A) and decarbonized indices in panel (B). The independent variables are shown in rows. Control variables are defined as Fama-French factors (Fama and French, 1993; Fama and French, 2015), i.e., (i) *MKT*, the excess market return, (ii) *SMB*, the small minus big factor, (iii) *HML*, the high minus low factor, (iv) *RMW*, the robust minus weak factor, (v) *CMA*, the conservative minus aggressive factor, and the momentum factor of Carhart (1997), (vi) *MOM*. Innovations in the Transition Risk Index are denoted as *TRI*.The unit of observation is a week, and each sample runs for the maximum common data history of the independent and dependent variables. Standard errors are presented in parentheses. *p $<$.1; **p $<$.05; ***p $<$.01.

Figure A3.4: Rolling 5-year regression of active weekly index returns



The figure shows the t-value for the *TRI* coefficients and pure-play indices for the regression model (1) using a rolling window of 5 years. Weekly data are used. The out of sample period starts in January 2008 and with only out of sample data used for calculation, t-values are reported from January 2013 to December 2020 for a rolling window of 5 years. The dashed red lines show the t-values for the respective p-values at the 10%, 5% and 1% significance level.

Table A3.5: Regression of monthly active index returns

(A) Pure-play indices

	MSCI GLENV		MSCI GLALT		S&P GLCE		Solactive CC		FTSE EORAE		FTSE ENVT	
	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)
MKT	0.059**	0.061**	0.104**	0.089**	0.128***	0.121***	0.072**	0.064*	0.017	0.002	0.113***	0.116***
	(0.025)	(0.024)	(0.045)	(0.042)	(0.031)	(0.03)	(0.036)	(0.034)	(0.04)	(0.038)	(0.029)	(0.027)
SMB	0.147***	0.144***	-0.026	-0.016	0.078	0.075	-0.019	-0.009	-0.081	-0.095	0.246***	0.238***
	(0.039)	(0.036)	(0.072)	(0.065)	(0.054)	(0.051)	(0.061)	(0.057)	(0.064)	(0.058)	(0.05)	(0.046)
HML	-0.056	-0.029	0.1	0.101	-0.012	-0.013	-0.113*	-0.110*	-0.031	-0.012	-0.132**	-0.117**
	(0.045)	(0.042)	(0.096)	(0.086)	(0.057)	(0.054)	(0.064)	(0.061)	(0.074)	(0.067)	(0.053)	(0.05)
RMW	-0.03	-0.024	-0.033	-0.08	0.012	-0.015	0.026	-0.009	-0.08	-0.113	0.074	0.085
	(0.059)	(0.056)	(0.11)	(0.102)	(0.079)	(0.076)	(0.09)	(0.086)	(0.096)	(0.09)	(0.074)	(0.07)
CMA	0.054	-0.004	-0.203	-0.189	-0.302***	-0.282***	-0.131	-0.119	-0.061	-0.068	-0.151	-0.171**
	(0.077)	(0.068)	(0.147)	(0.127)	(0.1)	(0.091)	(0.113)	(0.103)	(0.125)	(0.109)	(0.093)	(0.084)
MOM	-0.019	-0.008	0.004	0.009	0.002	0.0001	-0.011	-0.013	-0.005	-0.006	0.0002	0.004
	(0.021)	(0.02)	(0.052)	(0.048)	(0.028)	(0.027)	(0.032)	(0.03)	(0.035)	(0.032)	(0.026)	(0.024)
CC^{WSJ}	0.477		0.805		0.394		1.010*		0.355		0.113	
	(0.386)		(0.661)		(0.535)		(0.606)		(0.625)		(0.497)	
$CC^{NegNews}$		-0.14		0.942		0.423		1.698		1.971		-0.632
		(1.133)		(1.946)		(1.397)		(1.581)		(1.809)		(1.28)
TRI	9.553**	8.947**	21.345***	20.819***	19.537***	20.736***	16.397**	16.358***	19.916***	20.770***	16.823***	15.949***
	(3.927)	(3.667)	(7.19)	(6.566)	(5.671)	(5.266)	(6.426)	(5.959)	(6.367)	(5.853)	(5.27)	(4.825)
Constant	-0.001	-0.001	-0.004***	-0.004***	-0.005***	-0.004***	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**	-0.003***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	103	114	94	105	114	125	114	125	103	114	114	125
R-squared	0.341	0.306	0.205	0.192	0.368	0.357	0.183	0.165	0.136	0.163	0.436	0.426

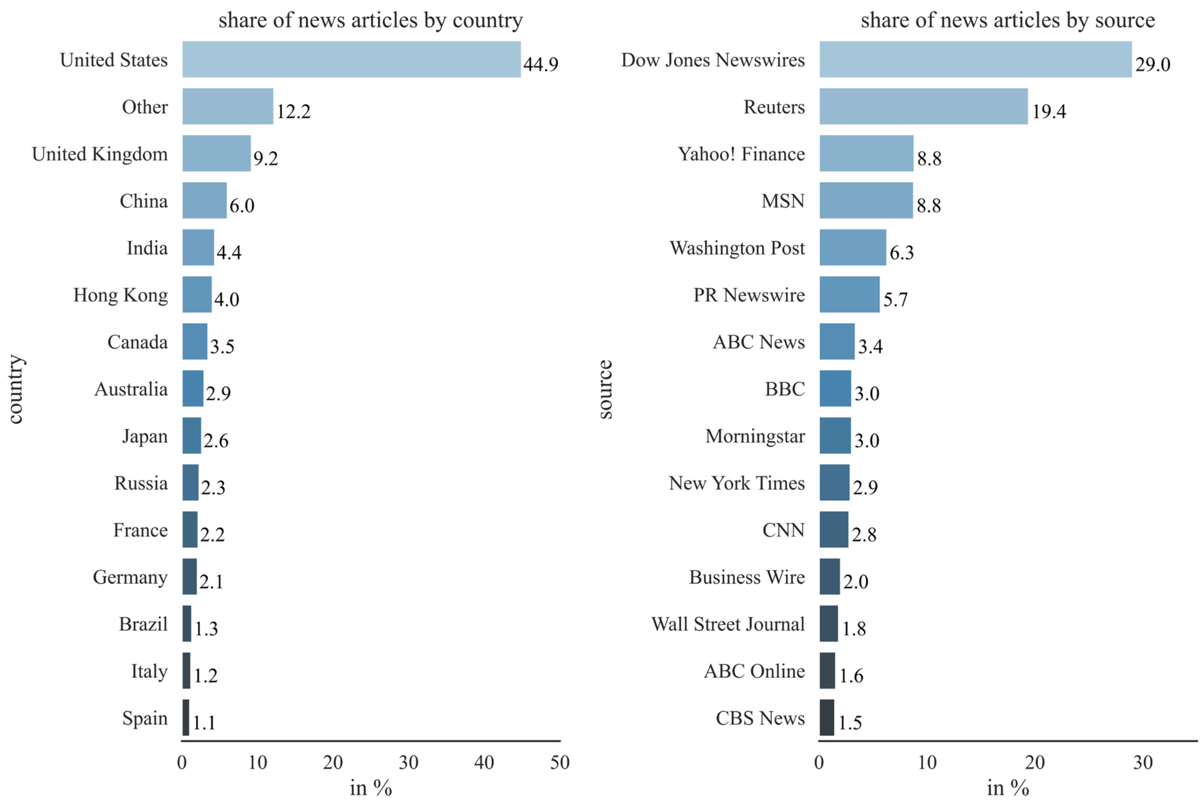
(B) Decarbonized indices

	MSCI WCC		MSCI WLCL		MSCI WLCT		S&P 1200 FFF		S&P LMCCE		STOXX 1800 LC	
	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)
MKT	-0.021	-0.002	0.037	0.053	0.054	0.048	0.105*	0.088	0.114***	0.093**	0.158**	0.127*
	(0.077)	(0.064)	(0.058)	(0.055)	(0.059)	(0.055)	(0.063)	(0.057)	(0.043)	(0.04)	(0.072)	(0.064)
SMB	0.023	-0.051	0.007	-0.074	-0.015	-0.079	0.015	0.006	-0.141**	-0.102*	0.046	0.044
	(0.094)	(0.079)	(0.086)	(0.079)	(0.087)	(0.079)	(0.088)	(0.077)	(0.065)	(0.059)	(0.101)	(0.087)
HML	-0.197*	-0.14	0.09	0.174*	-0.18	-0.163	-0.209*	-0.199**	0.02	-0.024	0.007	-0.065
	(0.116)	(0.101)	(0.11)	(0.104)	(0.111)	(0.104)	(0.109)	(0.098)	(0.089)	(0.08)	(0.124)	(0.111)
RMW	0.087	0.138	-0.362***	-0.299**	-0.066	-0.035	0.09	0.146	0.095	0.113	0.419***	0.260*
	(0.16)	(0.13)	(0.124)	(0.117)	(0.125)	(0.117)	(0.135)	(0.12)	(0.097)	(0.092)	(0.154)	(0.135)
CMA	-0.219	-0.303*	-0.019	-0.094	-0.12	-0.071	-0.017	-0.026	-0.124	-0.053	-0.538***	-0.397**
	(0.212)	(0.168)	(0.16)	(0.145)	(0.161)	(0.144)	(0.174)	(0.147)	(0.135)	(0.117)	(0.199)	(0.166)
MOM	0.101	0.073	0.03	0.029	-0.025	-0.064	0.234***	0.171***	0.107***	0.091***	-0.057	-0.065
	(0.073)	(0.067)	(0.061)	(0.06)	(0.062)	(0.06)	(0.061)	(0.058)	(0.037)	(0.035)	(0.07)	(0.065)
CC^{WSJ}	-0.357		-0.239		-0.691		0.364		-0.782		0.259	
	(1.099)		(1.014)		(1.023)		(0.982)		(0.62)		(1.123)	
$CC^{NegNews}$		3.431		2.119		2.282		5.667**		-3.140*		-0.302
		(2.762)		(2.782)		(2.777)		(2.573)		(1.844)		(2.895)
TRI	12.278	10.216	11.714	4.839	7.645	-0.542	7.228	5.88	8.354	6.955	-3.931	-2.401
	(8.867)	(7.601)	(8.175)	(7.679)	(8.247)	(7.665)	(8.123)	(7.299)	(6.775)	(6.106)	(9.291)	(8.212)
Constant	-0.004*	-0.006*	-0.013***	-0.013***	-0.017***	-0.018***	0.004**	0.003*	-0.001	-0.002	0.009***	0.010***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
N	43	54	79	90	79	90	66	77	99	110	66	77
R-squared	0.397	0.394	0.252	0.191	0.141	0.102	0.392	0.358	0.218	0.185	0.306	0.235

This table shows results from regressions (2) and (3) for the pure-play indices in panel (A) and decarbonized indices in panel (B). Monthly data are used, and each sample runs for the maximum common data history of the independent and dependent variables. SE are presented in parentheses. *p <.1; **p <.05; ***p <.01.

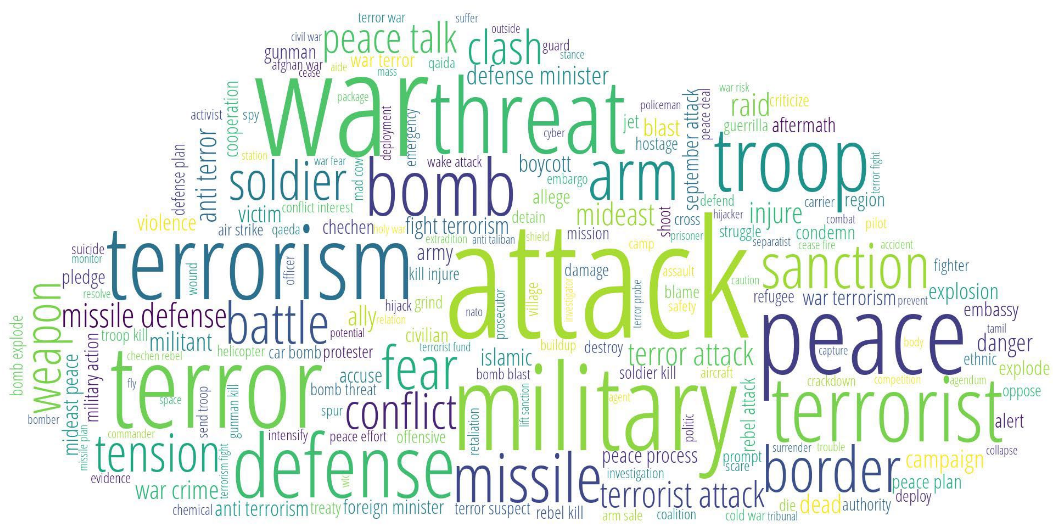
Appendix to Chapter 4

Figure A4.1: Share of news articles by country and source



This figure shows the share of news articles by media source and country. Figures are based on the news data sample from January 2000 to May 2022.

Figure A4.2: Word cloud summary of period-specific vocabulary related to geopolitical risk events



2001

2019

The figure shows word cloud summaries of period-specific vocabulary related to geopolitical risk events. The upper word cloud utilizes news of 2001. The vocabulary of the bottom word cloud is generated from news in 2019. Term sizes are proportional to their frequency in the text corpus.