



HAL
open science

Five essays on regulation, capital market financing and asset pricing in the era of sustainable finance (2005 - 2019)

Romain Berrou

► To cite this version:

Romain Berrou. Five essays on regulation, capital market financing and asset pricing in the era of sustainable finance (2005 - 2019). Business administration. Université Panthéon-Sorbonne - Paris I, 2021. English. NNT: 2021PA01E013 . tel-03461290

HAL Id: tel-03461290

<https://theses.hal.science/tel-03461290v1>

Submitted on 1 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS I- PANTHEON SORBONNE
Institut d'Administration des Entreprises de Paris

École doctorale de Management Panthéon-Sorbonne - ED 559
Equipe de Recherche GREGOR - EA 2474

**Five essays on regulation, capital market
financing and asset pricing in the era of
sustainable finance (2005 – 2019)**

THESE
présentée et soutenue publiquement le **11 janvier 2021**
en vue de l'obtention du

DOCTORAT EN SCIENCES DE GESTION

par

Romain BERROU

JURY

Directeur de recherche : **Monsieur Philippe Dessertine**
Professeur à l'IAE de Paris

Rapporteurs : **Madame Céline Barredy**
Professeur à l'Université Paris Nanterre

Monsieur Pascal Barneto
Professeur à l'IAE de Bordeaux

Suffragants : **Monsieur Franck Bancel**
Professeur à l'Ecole des Ponts ParisTech

Monsieur Eric Larmarque
Professeur à l'IAE de Paris

L'université de PARIS I – Panthéon Sorbonne n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

Five essays on regulation, capital market financing and asset pricing in the era of sustainable finance (2005 - 2019)

Our work consists in an exploration of the integration of sustainability in the financial sphere, with a particular focus on capital markets and its actors, centered on the following research question : “are financial markets putting a price on the positive sustainable and environmental performance of firms?”. Our research is carried out within the theoretical framework of asset pricing, which is currently at the crossroads of two major revolutions. On the one hand is the sustainability revolution, which brings a challenge to these actors in the form of a massive global systemic risk which can only be compared in its possible global impact to the threat once posed by weapons of mass destruction. As financial actors are starting to understand the economic threats – amongst others – that represent climate action failure, biodiversity loss, extreme weather events and environmental disasters, financial markets will inevitably change. The five essays in this PhD dissertation represent an attempt, using the most reliable data at hand, to start understanding how these financial markets could be evolving. On the other is the data revolution bringing both massive and constant flows of new forms of data to financial market actors, which are dealt with through constant improvements in computational power and the corresponding highly demanding statistical and machine learning methods needed to analyze these continuously growing datasets.

The first step of our work consists in exploring how practitioners, academics and regulators are trying to define what constitutes sustainable finance and allows us to understand how consensus on this central issue has still not been reached to date. In a second step, we then look at climate-aligned firms – companies that originate more than 75% of their turnover from climate-aligned activities – and analyze how capital markets have been financing both traditional and climate-aligned US and EU firms in the last decades. This analysis allows us to understand that corporate bond markets have a much larger role to play in financing these firms than stock markets. In a third step, we then choose to focus our attention on corporate bond markets and on the very specific challenge of building reliable financial datasets on traditional corporate bonds, green corporate bonds and bonds that are issued by climate-aligned firms. In a fourth step, using this detailed data treatment procedure, we apply traditional asset pricing methodologies to a dataset of both traditional bonds and climate bonds going from 2005 to 2019, finding some differences between these two products. Finally, our last step constitutes an attempt to understand, with these same datasets, how newly developed factors and machine learning methods could also be used to price corporate bonds and to differentiate traditional bonds from climate bonds.

Key words: Sustainable Finance, Green Finance, SRI, ESG, EU Taxonomy, Climate Bonds, Corporate Bonds, Asset Pricing, Factor Models, Machine Learning

Cinq essais sur la régulation, le financement par les marchés et l'évaluation des actifs financiers dans l'ère de la finance responsable (2005 - 2019)

Notre travail consiste à étudier l'intégration des problématiques de développement durable dans la sphère financière en se focalisant principalement sur les marchés de capitaux et leurs acteurs et en se posant la question de recherche suivante: « est-ce que les marchés financiers mettent un prix sur les performances responsables et environnementales des entreprises ? ». Notre recherche est effectuée dans le cadre théorique de l'évaluation des actifs financiers, cadre théorique qui fait actuellement face à deux révolutions majeures. D'une part, la révolution « durable », qui pose une problématique nouvelle aux acteurs financiers sous la forme d'un risque systémique global qui ne peut être comparé dans son impact potentiel qu'à la menace que représentait les armes de destruction massives. Alors que les acteurs financiers commencent à comprendre le risque économique - entre autres - que représenterait un échec à atténuer le changement climatique, la perte de biodiversité, les événements climatiques extrêmes et les catastrophes naturelles correspondantes, les marchés financiers vont inévitablement changer. Les cinq essais de cette thèse doctorale représentent une tentative, en utilisant les données les plus fiables disponibles aujourd'hui, de comprendre comment ces marchés financiers peuvent être en train d'évoluer. De l'autre, la révolution des données apportant des flux massifs et constants de nouvelles formes de données aux acteurs des marchés financiers, traitées à l'aide d'ordinateurs dont la puissance augmente constamment ainsi que de nouvelles méthodes très exigeantes de traitement statistiques et de *machine learning* avancées nécessaires à l'analyse de ces bases de données grandissantes.

La première étape de notre travail consiste à explorer comment les praticiens, les académiques et les régulateurs essaient de définir ce que constitue la finance responsable et nous permet de comprendre qu'un consensus sur ce sujet n'a toujours pas été atteint à ce jour. Dans une seconde étape, nous observons les entreprises climatiques – ces entreprises dont 75% du chiffre d'affaire provient d'activités alignées avec les objectifs climatiques de la conférence de Paris – et analysons comment les marchés de capitaux financent les entreprises traditionnelles et climatiques aux Etats-Unis et en Europe sur les dernières décennies. Cette analyse nous permet de comprendre que les marchés obligataires ont un rôle plus important de financement pour ces entreprises que les marchés des actions. Dans une troisième étape, nous choisissons de nous concentrer sur le marché des obligations d'entreprise et sur la problématique spécifique de construire une base de données financière fiable sur les données des obligations d'entreprise américaines, sur les obligations vertes et les obligations climatiques. Une quatrième étape consiste, en utilisant cette base de données, à appliquer les méthodes d'évaluation d'actifs classiques sur une base de données d'obligations d'entreprise traditionnelles et climatiques entre 2005 et 2019, et nous trouvons certaines différences entre ces deux produits. Finalement, une dernière étape consiste en une tentative de comprendre, en utilisant cette même base de données, comment de nouveaux facteurs et de nouvelles méthodes d'analyse de *machine learning* peuvent également être utilisés afin de mieux expliquer et de prédire le rendement des obligations d'entreprises et différencier les obligations d'entreprise traditionnelles et climatiques.

Mots clés : Finance responsable, finance verte, ISR, ESG, Taxonomie européenne, Obligations climatiques, Obligations d'entreprises, Evaluation d'actifs, Modèle factoriels, Machine Learning

This PhD dissertation is composed of five articles and a summary note. This summary note is meant to provide a general understanding of the relevance of this entire research work. It outlines its incorporation in the financial literature, its main sources of data and data treatment procedures, as well as empirical methodologies and corresponding results. Limitations of this work and required future research are also discussed throughout this thesis.

Given the nature of the summary note, and the fact that it summarizes a series of papers that have been prepared for potential publication, it is possible that some material contained in the summary note is also contained in the research papers it summarizes.

Table of contents

<u>CHAPTER 1 – SUMMARY NOTE</u>	8
1. GENERAL INTRODUCTION	8
1.1. THE CLIMATE CHANGE URGENCY AND ASSET PRICING	8
1.2. IDENTIFYING CLIMATE-ALIGNED FIRMS	11
1.3. ASSET PRICING APPLIED TO CLIMATE BONDS	15
2. DEFINING SUSTAINABLE FINANCE	21
2.1. A LACK OF COMMON DEFINITION AMONGST PRACTITIONERS	21
2.2. A LACK OF COMMON DEFINITION AMONGST ACADEMICS	24
2.3. THE IMPORTANT ROLE OF REGULATORS AND THE EUROPEAN COMMISSION	26
2.4. THE FOCUS ON LARGE FIRMS AND ASSET MANAGERS	28
3. CAPITAL MARKET FINANCING AND BOND DATA FOR TRADITIONAL AND CLIMATE FIRMS	31
3.1. DIVIDENDS, STOCK REPURCHASES AND CORPORATE BOND ISSUANCES FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	32
3.2. CORPORATE BOND DATA FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	36
4. MODERN EMPIRICAL ASSET PRICING AND CLIMATE-ALIGNED FIRMS	38
4.1. FROM PREDICTING LEVELS OF RETURNS TO PREDICTING CROSS-SECTIONAL DIFFERENCES IN RETURNS	39
4.2. WHAT ABOUT CORPORATE BONDS?	41
4.3. POSSIBLE LIMITATIONS OF TRADITIONAL ASSET PRICING METHODOLOGIES	43
4.4. ADVANTAGES OF MACHINE LEARNING AND THE “ZOO OF FACTORS”	46
5. LITERATURE REVIEW	49
5.1. LITERATURE ON THE SRI MARKET	50
5.2. THE ACADEMIC LITERATURE ON ESG/CSP AND CEP	50
5.3. CORPORATE FINANCING VIA CAPITAL MARKETS	53
5.4. DATA PROCESSING METHODOLOGY TO STUDY TRADITIONAL AND CLIMATE BOND RETURNS	57
5.5. THE NASCENT CLIMATE-ALIGNED BOND LITERATURE	60
5.6. ASSET PRICING AND TRADITIONAL AND CLIMATE BONDS	62
5.7. MACHINE LEARNING IN ASSET PRICING APPLIED IN THE CONTEXT OF SUSTAINABLE FINANCE	63
6. DATA	65
6.1. DATA TREATMENT METHODS FOR CAPITAL MARKET FINANCING FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	66
6.2. THE CORPORATE BOND DATA TREATMENT METHODOLOGY	72
6.3. DATA USED FOR TRADITIONAL AND CLIMATE BOND PRICING	78
6.4. OTHER TYPES OF DATA ON SUSTAINABLE FINANCE	81
7. METHODOLOGY	82
7.1. STRUCTURED ANALYSIS OF THE LITERATURE ON SUSTAINABLE FINANCE	82
7.2. THE CROSS-SECTION OF TRADITIONAL AND CLIMATE BOND RETURNS	86
7.3. MACHINE LEARNING AND ASSET PRICING FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	91
8. RESULTS	95
8.1. RESULTS FROM THE LITERATURE ON SRI, ESG AND CEP	95
8.2. RESULTS ON THE CAPITAL MARKET FINANCING OF TRADITIONAL AND CLIMATE-ALIGNED FIRMS	96
8.3. A BETTER UNDERSTANDING OF CLIMATE BOND AND GREEN BOND DATA AVAILABILITY	100
8.4. INSIGHTFUL RESULTS FROM THE CROSS-SECTION OF TRADITIONAL AND CLIMATE BOND RETURNS	101
8.5. THE IRRELEVANCE OF THE CLIMATE FACTOR FOR MACHINE LEARNING METHODS	109
8.6. CONCLUDING REMARKS ON RESULTS	115
9. FUTURE RESEARCH	117
10. SUMMARY CONCLUSION	121
<u>CHAPTER 2 - THE EU TAXONOMY: A NEW SUSTAINABLE FINANCE PARADIGM?</u>	122
<u>CHAPTER 3 - THE ROLE OF CAPITAL MARKETS IN FINANCING CLIMATE CHANGE MITIGATION</u>	178
<u>CHAPTER 4 - A DATA PROCESSING METHODOLOGY FOR GREEN BONDS AND CLIMATE BONDS</u>	214
<u>CHAPTER 5 - A STUDY OF THE CROSS-SECTION OF CLIMATE BOND RETURNS</u>	248
<u>CHAPTER 6 - AN INTRODUCTION TO THE PRICING OF CLIMATE BONDS VIA MACHINE LEARNING</u>	286
GENERAL CONCLUSION	321
BIBLIOGRAPHY	325
TABLE OF CONTENTS (DETAILED)	330

Chapter 1

Summary Note

1. General Introduction

1.1. The climate change urgency and asset pricing

In a special report performed in 2018, the Intergovernmental Panel on Climate Change (IPCC) published a report to inform the global community of the possible impacts of global warming in the decades to come. Based on more than 6000 scientific publications on the subject, the report warned policy makers that a 2 °C or 3 °C pathway would expose hundreds of millions of people to water stress and scarcity, heatwave events, flooding, risks related to power production, crop yield or habitat degradation, with corresponding impacts on food security, health, poverty, public unrest and political destabilization (IPCC, 2018). Less than a year later, another study focused on determining the most probable trajectory in terms of GhG emissions estimated that the most likely pathway corresponded to a 3 °C increase of temperature by 2100 (UNEP, 2019). Unsurprisingly, such a global environmental and social crisis would also be an economic one. Focusing more precisely on this aspect, the 2020 Global Risks Report performed on a yearly basis by the World Economic Forum identified five environmental risks as the five leading global risks in terms of likelihood. The most impactful risk, “climate action failure”, was estimated as being more impactful for the global community than “weapons of mass destruction” (World Economic Forum, 2020). In the beginning of 2020 when this report was performed, global warming had been clearly identified for the first as both the most probable and the most impactful threat for world economies.

But has this been understood by financial markets? If this is the case, then surely financial products that participate in financing the most sustainable or green firms, assets or projects behave differently than their traditional equivalents. Throughout this PhD dissertation, we try to answer

this central research question: are financial markets putting a price on the positive sustainable and environmental performance of firms?

In order to provide an answer to this question, our research is carried out within the theoretical framework of asset pricing. If there is indeed a growing interest from financial markets for sustainable financial products, then applying the most modern models developed in the context of this theoretical framework will lead us to obtain - at least in part - an answer.

Asset pricing theory is considered to have started with Sharp's capital asset pricing model (Sharp, 1964), which itself was developed from the modern portfolio theory (MPT) pioneered by Markowitz (1952). This single-beta pricing model then led to the development of many more renown multiple-beta models in continuous time (Merton, 1973; Breeden, 1979; Cox, Ingersoll and Ross; 1985). As these multiple-beta models were developed, the arbitrage pricing theory was put forward as an alternative to the capital asset pricing model by Stephen Ross (Ross, 1976). This in turn led to the development of a large empirical literature on the concept of factor analysis.

Originally, factor analysis is a statistical method that consists in finding a small number of unobserved variables that describe the variability amongst a larger number of observed, correlated variables. This approach was applied in empirical asset pricing with the intention of finding a small set of variables that could best explain the cross-section of stock or bond returns. The implementation of factor analysis in empirical asset pricing led to the development of a very dense literature that focused on finding the most efficient "factors" that could explain or predict stock market returns. The birth of the literature on factor models is considered to have occurred with the development of the three-factor model of Fama and French (1992).

Since 1992, many academics have tried to outperform this original model, and to find new factors that explained the cross-section of stock returns. This resulted in a very large number of

factors being included in this literature, and other important factors were discovered and added to the model, such as the liquidity (LIQ) factor (Pastor and Stambaugh, 2003), the momentum (MOM) factor (Carhart, 1997), and more recently, the investment (CMA) and profitability (RMW) factors (Fama and French, 2015). In more recent work on the subject, Harvey and Liu (2016) reference 313 papers that developed factors that explained the cross-section of stock returns since the original paper of Sharp (1964). Today, these two authors maintain a database on the factors developed in the literature, and this number had reached 514 by the end of 2018 (Harvey and Liu, 2019)¹. Our approach in this PhD dissertation was not so much to develop new factors to predict the cross-section of stock returns, but rather to understand if these latest models performed similarly on traditional firms/products and on sustainable or green firms/products.

Today, as the study of the cross-section of stock returns almost seems congested, many authors are starting to look at the cross-section of corporate bond returns. This is of interest to us, as we find, through our research, that the corporate bond market is a much more important source of financing than stock markets for both traditional large firms and climate-aligned firms in both the US and the EU. In addition, though they represent much smaller trading volumes, the total size of corporate bond markets in the US and the world is much superior to that of stock markets.

Based on this observation, Bai, Bali and Wen (2019) develop the first factors specific to the US corporate bond market to explain and predict the cross-section of corporate bond returns, when the last factors that had been put forward by the literature were the default (DEF) and term (TERM) factors of Fama and French (1993). A large part of our own research is based on the methodology and factors developed by Bai, Bali and Wen (2019).

¹ See www.docs.google.com/spreadsheets/d/1mws1bU56ZAc8aK7Dvz696LknM0Vp4Rojc3n61q2-keY/edit#gid=0

As the asset pricing literature expands towards the corporate bond market, it is also currently being impacted by the arrival of new statistical and machine learning methods. Interestingly enough, Campbell Harvey, former President of the American Finance Association and leading scholar in the asset pricing literature, states that “The first wave of quantitative innovation in finance was led by Markowitz optimization. Machine learning is the second wave, and it will touch every aspect of finance” (Lopez de Prado, 2018). These methods are starting to be applied in the context of asset pricing, and we explore in the final part of this PhD dissertation the advantages that are provided by machine learning methods compared to the more traditional asset pricing methods. In this context, one of the most impactful study is that of Gu, Kelly and Xiu (2020), in which the authors apply a series of machine learning methods to predict the cross-section of stock returns and obtain very insightful results. Much like we refer to Bai, Bali and Wen (2019) as a methodological backbone for a traditional asset pricing approach in our own study, we refer to Gu, Kelly and Xiu (2020) to explore the possibilities of applying machine learning methods to predict the cross-section of bond returns.

1.2. Identifying climate-aligned firms

With this theoretical framework and corresponding literature defined, the next step of our work consisted in identifying what are sustainable or green firms or sustainable and green financial products, and how are they identified by actors of the financial industry? In order to find an answer to this question, we inspect the nascent literature on green products, and find that most of the empirical studies that focus on sustainable products study the growing green bond market. Most of this literature focuses on trying to find if investors pay a premium to acquire green bonds. To date, this literature has inconclusive findings regarding the existence of such premium (Hachenberg and Schiereck, 2018; Bachelet et al, 2019; Zerbib, 2019). Authors that participate in

the literature inform us that the restrictions in terms of sample size, historical data and data quality on the green bond market lead their results to be quite complex to obtain and less robust than what could be expected when studying larger, more detailed datasets.

These findings leads us to take our focus away from specific green financial products such as green bonds and focus more on the concept of ESG and investing in firms that have better sustainable performance. This, however, does not simplify our approach, as we find that both practitioners and academics struggle to define the concept of ESG. A thorough inspection of both the practitioner and academic literature leaves us with the troubling conclusion that consensus on what constitutes the basic concepts of sustainability and sustainable finance has not been reached to date.

We find however that challenges faced by sustainable finance practitioners seem to be understood by policy makers. More specifically, the European Commission has started to develop a more specific framework for sustainable finance. To construct this framework, the European Commission defined environmental objectives that needed to be reached by financial actors in its member states and created a precise set of industry-specific criteria that need to be followed by all financial market participants and large firms in its geographical region. This solution, however, has only just started to be implemented by the Commission, and does not yet provide a solution to academics that wish to study this issue today.

In the context of this PhD dissertation, a solution was found in previous work performed by the Climate Bond Initiative, when the organization published a report on bonds and climate change in 2018 (CBI, 2018). In this report. The CBI defines a list of climate-aligned issuers by considering firms that “derived at least 75% of their revenue from green business lines in at least one of 6 climate themes: clean energy, low-carbon transport, water management, low-carbon

buildings, sustainable land use.” CBI provided the author with a list of these identified climate-aligned firms throughout the world. To date and to the author’s knowledge, this approach is the most robust means of identifying climate-aligned firms without further information on the specific business practices of corporations. It is interesting to acknowledge how this approach resembles in many ways the more advanced methodology being developed by the European Commission. Using this data, we can identify climate firms in the United States and in Europe on our different databases. Furthermore, we can study climate-aligned firms by applying the different asset pricing models that we previously defined and try to find existing differences between these firms and their traditional equivalents.

Having identified climate-firms in the US and in the EU, we first used this data to understand how capital markets participate in financing both traditional and climate-aligned firms in these two regions. This study was performed by first considering the massive investment needs that are progressively being identified by international agencies to face climate change. One estimation performed by the OECD in 2017 estimated global infrastructure investment needs to be consistent with a 2°C scenario to amount to USD 6.9 trillion per year in the next 15 years (OECD, 2017a). That same year, focusing on the energy sector, the international energy agency (IEA) estimated that consistency with this scenario would require 95% of the electricity to be low-carbon by 2050, that 70% of new cars would need to be electric, that the entire building stock would have to be modernized and that the industrial sector would need to produce 80% less CO₂ (IEA, 2017). Such important changes cannot be implemented without the participation of corporations that need to adapt their business models and provide new sustainable products in order to face climate change mitigation. Our results, which we explore in further detail in later sections, suggest that corporate bonds participate much more than stocks in financing both traditional and climate-

aligned firms, which in turn leads us to focus on the corporate bond market for the remainder of the studies in this PhD dissertation.

However, the study of corporate bonds comes with its own challenges. Compared to the literature addressing the study of stocks, the literature which deals with the study of corporate bonds is quite recent, and academics in the field have constantly been challenged with issues related to data availability and quality. The primary source of data on the corporate bond market, the Lehman Brothers Fixed Income database, closed in 1997, and there was no quality data available on this market until the Mergent Fixed Income Securities Database opened in 2004, and the Trade Reporting and Compliance Engine database (TRACE) was put in place starting in 2002. Furthermore, TRACE data was at first quite unreliable, and required rigorous data treatment that wasn't clearly identified by the literature until 2009. To date, many studies on the subject still do not address some essential steps of this data cleaning procedure which could bias results. In the context of this PhD dissertation, we create a clear data treatment procedure, applied to a sample of green and climate bonds, to create a framework to study these products using the thorough methodology that needs to be implemented when one studies the corporate bond market.

Once this work was completed, we were able to have a clearer comprehension of our data and understand what type of analysis could be performed on our green bond and climate bond datasets. The conclusion, much like those of our predecessors, was that even though green bond issuances have increased dramatically in the last years, there is too few green bonds in the US to be able to perform robust analysis on our green bond sample. This is the case both in terms of time-series analysis, since global green bonds issuances have only reached the \$10 billion threshold in 2013, which give us only 6 years of data, but also in the context of cross-sectional analysis: the trading volume for green bonds in the US grew from \$1.17 billion to \$13,556 billion

between 2014 and 2018, while it grew from \$44,463 billion to \$76,254 billion over the same period for climate bonds. Even in 2018, when the green bond market had reached its peak number of issuances and trading volume, this market was still almost 6 times smaller than the climate-bond market in terms of trading volume. Therefore, the last studies of this PhD dissertation focus on climate-bonds, both for reasons of data availability, but also because the methodology for to determine climate-aligned firms resembles that of the EU Taxonomy. Research approaches based on this product could very well be applied to “Taxonomy-aligned” firms in the near future.

1.3. Asset pricing applied to climate bonds

Using our dataset of climate-aligned firms, we are able to apply traditional asset pricing methodologies to our sample of climate bonds. One of the main research agenda in empirical asset pricing research focuses on studying the differences in expected returns across specific asset classes. To date, a large majority of this research has applied to stock returns given both the simplicity of these products compared to other market products and the availability of reliable data on stock prices. However, in addition to the fact the corporate bonds are a larger source of finance for large firms than stocks, the global bond market represented a total of \$102.8 trillion in securities outstanding in 2018, while the global equity market capitalization represented \$74.7 trillion. In the US, these markets amounted to \$41 trillion and \$30 trillion respectively. Furthermore, the yearly value of US corporate issuances of debt products have been in average 9 times superior to the value of equity issuances between 2004 and 2018² (SIFMA, 2019). Given the size of these corporate debt markets, empirical asset pricing research on products that compose these markets could prove important for its different actors.

² Corporate debt products include public and private, investment grade and high yield bonds, convertible debt, asset-back securities and non-agency mortgage-backed securities, while equity issuances include common stock issuances (IPOs and follow-ons) and preferred stock issuances.

Earlier research on the cross-section of corporate bond returns mostly developed factors using either stock-level data, treasury bond data and macroeconomic data. This is the case for the long-established Fama-French (1993) factors composed of the market risk factor (Mm-Rf), the size factor (SMB) and the book-to-market factor (HML) that originate from stock data and treasury bond data, and the term spread (TERM) and default spread (DEF) factors that originate from treasury bond data and government bond data³. Other factors that have complemented Fama and French's work, such as the liquidity (LIQ) factor (Pastor and Stambaugh, 2003), momentum (MOM) factor (Carhart, 1997), and more recently, the investment (CMA) and profitability (RMW) factors (Fama and French, 2015) all originate from stock-level data.

Realizing that these factors performed poorly in their ability to explain industry-sorted and size/maturity sorted portfolios of US corporate bonds, Bai, Bali and Wen (2019) introduced new bond-implied risk factors based on characteristics specific to corporate bonds. These characteristics were determined by focusing on the three most prominent differences between stocks and corporate bonds: (1) bondholders are more sensitive to downside risk given the fact that their opportunities in terms of upside are limited (2) bondholders are much more exposed to default risk and (3) as the corporate bond market is much more illiquid than stock markets, bondholders are more exposed to liquidity-related risks. As a model composed of these factors and a bond market beta outperforms other models in the literature to explain the cross-section of corporate returns for industry-sorted and size/maturity sorted portfolios, we use these factors as benchmarks to study differences between US climate bonds and the US corporate bond market.

³ In order to compute the default factor, a market portfolio of long-term corporate bonds was also required, and the data needed to compute such portfolio return could for the most part only be accessed through Ibbotson Associates, a private investment advisory firm.

Other than allowing us to develop a preliminary understanding of the differences that might exist between traditional corporate bonds and climate bonds, applying traditional asset pricing methodologies to our dataset allowed us to understand the limitations that are currently faced by these methods. The first difference resides in the fact that all traditional asset pricing models rely on multiple linear regression, which is one of the most basic forms of statistical model and is limited in the amounts of variables that can be used. This is an important limitation given the fact that hundreds of different variables have been developed by the academic and practitioner literature to explain or predict the cross-section of capital market returns. In addition, the linear aspect of the multiple linear regression could also prove to be limitative in terms of explaining or predicting returns. Furthermore, the multiple linear regression model does not consider interactions amongst predictors. Overall, statistical learning and machine learning methods bring three important strengths to the multiple linear regression model. Firstly, they are more diverse: many different models with different strengths and weaknesses can be applied to the cross-section of returns, and one can therefore choose the best performing models for their analysis. Relying exclusively on one type of model is in itself a limitation. Second, machine learning allows to construct models that are non-linear, when one could also easily expect capital markets to behave non-linearly. Third, specific methodologies applied in the context of statistical learning and machine learning methods have been developed to deal with model overfit and false discoveries, which are currently major issues faced by the asset pricing literature.

Data science and machine learning are changing the entire financial landscape, and machine learning methods and algorithms are increasingly viewed by academics in finance as the future of empirical financial research. In the first sentence of his very well received book on the subject, Marcos Lopez de Prado, one of the leading scholars on machine learning (ML) in finance,

states that “today’s machine learning (ML) algorithms have conquered the major strategy games, and are routinely used to execute tasks once only possible by a limited group of experts. Over the next few years, ML algorithms will transform finance beyond anything we know today.” (Lopez de Prado, 2018). Machine learning methods are slowly making their way into the stock pricing literature. To date, however, these methods have not been used in the context of the corporate bond pricing literature, which is a more recent and less developed.

The last study of this PhD dissertation allows for an exploration of some of these methods applied to the prediction of traditional corporate bond returns using 22 predictors. Much more work needs to be done on this subject, as machine learning methods continue to make progress in the asset pricing literature, amongst many others.

When faced with the objective of performing research on the subject of sustainable finance, academics face many different challenges. In a way, this PhD dissertation explores some of these different challenges through its five research papers. The first challenge lies in defining, quantifying and measuring sustainability in the context of empirical financial research. The fact that the most recent academic literature struggles to define and measure what constitutes ESG as a concept and that the overall quality of ESG data is questioned represents a significant issue both in academia and in practice and is an issue that is addressed in our first article.

Another central issue faced by the sustainable finance community lies in the massive financial efforts that need to be made to mitigate and adapt to climate change. It is quite clear that a large part of these financial efforts will have to be performed by large firms and their investors. Looking at the current financial relationship between large firms and capital markets allows us to understand that stock markets no longer finance large corporations. Though stock investors continue to play a central role in the governance of these firms, only corporate bond investors

provide positive capital market financing to large corporations. This is even more pronounced for climate-aligned firms, even though these are smaller in size. This subject is addressed in our second article, though much more work needs to be done to understand how corporations and their investors are responding to the massive investment needs required by climate change.

Central to the overall concept of research is that of data and data quality. Research in finance is no different, and we dedicate another article to the challenges that lie in the analysis of corporate bond data. This market, which is much larger than the stock market, attracts much less attention from both academics and practitioners. Corporate bond data has historically been scarce and most corporate bond products are complex financial products that are more difficult to price. However, we choose to focus on this market which in comparisons seems to attract more attention from the sustainable finance community given the important rise of the green bond market and other climate-aligned fixed income products. As more and more academics choose to study this market, having developed this methodology could prove useful to future academics, but also allow to enhance our general understanding of the corporate bond market.

With this methodology, we can develop a dataset of corporate bond returns using the highest quality data available to date in order to apply traditional asset pricing methods to the study of the corporate bond market. Empirical asset pricing is a central subject in financial research, one that is currently facing many controversies as academics start to question the statistical significance and underlying methodologies used in this literature. Furthermore, very few ESG factors have been developed in the “zoo of factors” that impact the cross-section of stock returns and the empirical asset pricing literature has given little attention to the corporate bond market. By developing a practical understanding of traditional asset pricing methods applied to corporate bonds and dedicating part of the study to the pricing of climate bonds, we get a foothold in what could well

soon be a new branch of the asset pricing literature: the pricing of climate-aligned (or Taxonomy-aligned) firms and assets.

This work has also led to the development of our fifth and final article, which focuses on some of the modern statistical and machine learning methods that are used by the literature in order to find solution to the methodological challenges they are currently facing. This article constitutes an attempt to understand and apply modern statistical learning methods to empirical asset pricing for corporate bonds, which is also quite a specific yet promising branch of the asset pricing literature: as the climate urgency continues to grow, so does the application of machine learning methods in asset pricing, and it is also quite probable that the upcoming pricing of climate-aligned firms and assets is performed through these methods.

In the following sections of this summary note, we expand on the specific context that lead us to develop our different research approaches and provide an outline of the data that was used, the different methodologies that were applied to these data and our main results. Section 2 provides a background on the central problematic of defining what constitutes sustainable finance within the industry. Section 3 explores how capital markets finance large and climate-aligned firms and the data issue relative to the corporate bond market. In Section 4, we explore recent developments in the asset pricing literature and look at the advantages of machine learning methods when compared to traditional asset pricing methodologies. We present the different literature that we have explored in Section 5, lay out our different data sources and treatment procedures in section 6, our methodologies in Section 7 and our results in section 8. We then explain the opportunities for future research that our work entails, before concluding.

2. Defining sustainable finance

One of the central challenges of every actor of the sustainable finance industry consists in defining what constitutes sustainable finance precisely. In this section, we explore how this issue is dealt with by practitioners of the sustainable finance industry and academics that work on defining the industry's main concepts. Finally, we look at regulators and the European Commission and show how large firms and asset managers have been selected to initiate a shift in how financial markets work.

2.1. A lack of common definition amongst practitioners

While there are many definitions of what constitutes a sustainable investment, most actors agree on the fact that this investment approach “considers environmental, social and governance (ESG) factors in portfolio selection and management” (GSIA, 2018). According to the Global Sustainable Investment Alliance (GSIA), an organization that performs a market study of the global sustainable investing industry every two years, the sustainable investment industry represented \$30.7 trillion at the start of 2018, a 34 percent increase in two years (GSIA, 2018). This would represent 38.8% of the \$79.2 trillion in total assets under management at the end of 2017 (BCG, 2018).

Much like there is common agreement that sustainable investment approaches use ESG factors in their investment decision process, there also is agreement in the industry that there are seven distinct approaches to perform sustainable investing. Ordered by the magnitude of the underlying market they represent, they are:

1. ***Negative/Exclusionary Screening***, which refers to the practice of not investing – or divesting – from certain activities that are considered as unethical or contrary to social or environmental investor principles.

2. ***ESG Integration***, which refers to the practice of integrating ESG factors into the investment decision process and investing only in projects, assets and firms that are aligned with these ESG factors.
3. ***Corporate Engagement and Shareholder Action***, which consists in using shareholder power to influence corporations and the socially or environmentally oriented actions that they take.
4. ***Norms-Based Screening***, which consists in implementing social and environmental international norms into the investment screening process.
5. ***Positive/Best-in-Class Screening***, which consists in investing exclusively in projects, assets or firms that have the best ESG performance in specific industry or sector.
6. ***Sustainability Themed Investing***, which consists in investing in specific industries and economic activities that are known to be sustainable.
7. ***Impact and Community Investing***, the former consisting in investments targeted specifically at environmental and social issues and the latter consisting in investments in identified under-served communities

As the growth of the sustainable investing market has been capturing a lot of attention by a variety of financial market actors, going into a detailed analysis of these different approaches and underlying markets provides interesting insights⁴. First and foremost, to date, the sustainable investment market is not solely composed of investments in activities with positive environmental and social performance. As an example, the negative/exclusionary screening approach was applied for \$19.77 trillion of assets under management that are counted as part of the SRI industry and considered as a form of sustainable investment in 2018 (GSIA, 2018). The industry that represents

⁴ This detailed analysis is performed in the context of the first article of this thesis, which can be found in Chapter 2.

only investments in projects or firms with positive social and environmental performance would therefore be quite smaller.

Secondly, looking at 2018 data, it seems that the environment is not yet a priority for sustainable investors. There are only two exclusion strategies that are related to the environment in Europe. The only investment approach that focuses exclusively and clearly on the environment - Sustainability Investing - is one of the least popular and has grown by only 3% in Europe between 2016 and 2018 (Eurosif, 2018). In the United States, the Environment is not the most important ESG category for both money managers and institutional investors. Climate change issues do come third in terms of shareholder proposals in the US, but sustainability reporting and other environmental issues are at the bottom of the shareholder proposal list (USSIF, 2018). As clear investments in favor of positive environmental projects that mitigate climate change become more urgent, this did not seem to be a priority for sustainable investors as of 2018.

Finally, an analysis of what to date are considered three of the most informative reports on the sustainable investing industry leaves us with the conclusion that actors still lack clarity. Even though efforts of the GSIA allow us to have a global view of the industry, we understand that European and US markets differ and provide different approaches of the same market. Sustainable investing is still quite a broad endeavor, and the terminology for sustainability or ESG criteria is very dense. For Europe, the report mentions 9 different exclusion criteria, 13 different European norms, 7 different types of sustainable investments, and in the US, reports mention 6 top ESG criteria and 10 types of shareholder proposals. Most of the terms used are not similar and do not address the same issue. Furthermore, information on the specifics of approaches like ESG Integration, which does represent \$17.54 trillion of global assets under management according to

the GSIA, are not provided. In Europe, specifics of the €4.8 trillion “Corporate Engagement and Voting” industry are also not provided.

2.2. A lack of common definition amongst academics

Similarly, to date, academic literature has not yet reached a consensus on this issue and is still trying to define what constitutes sustainable investing and ESG as a concept. The struggles in academia are somewhat similar to those concerning financial practitioners: there is no clarity on what truly constitutes ESG factors and disclosure, research approaches concerning ESG criteria vary greatly and data on ESG is not always entirely reliable.

One central issue in the academic’s literature resides in the different concepts that address the extra-financial aspect of companies and the frequently used terms of corporate social responsibility (CSR) and corporate social performance (CSP) in comparison with ESG. In practice, these three terms are almost used interchangeably, as they are used to refer to the broad category of factors that do not address financial performance or risk directly. From this analysis of the literature, it can be concluded that ESG as a concept is still not clearly defined by academics, and is still one of many terms that is used to refer to the plethora of concepts and factors that address the “extra-financial” aspect of firms, projects and assets. This is quite concerning given the fact that the ESG concept is taken as a basis for what has been identified as a 30.6\$ trillion SRI industry that focuses on sustainability issues with no clear framework to understand and demonstrate how investments are truly sustainable. Furthermore, given this lack of definition and clear framework, the true relationship between ESG as a mono-dimensional concept and financial performance and risk is quite challenging to determine.

Challenges faced by academics regarding the definition of ESG/CSP are also present in the literature that focuses on corporate environmental performance (CEP) specifically. Focusing on

the literature on the relationship between environmental performance and financial performance and risk allows for deeper understanding of the general issues faced by academics that focus on the concept of ESG/CSP. There seems to be common agreement on the terminology used in this literature – contrarily to the literature that focuses on ESG as a mono-dimensional concept – as academics do have a common usage for the main concepts linked to environmental performance. However, the problem of evaluating and measuring environmental performance remains. Unlike financial performance or financial risk, which can broadly be considered as mono dimensional concepts as their measurement rely on one metric, environmental performance has a great variety of dimensions. Environmental performance can be measured by using various metrics such as CO2 emissions, GhG emissions, tons of waste, gigajoules of energy consumption. Both the literature that focuses broadly on ESG/CSP and the literature that focuses more specifically on environmental performance suffer from a lack a common definition of what constitutes their central concepts. It is impossible to find true consensus in the literature on what “ESG” means, what are ESG factors and how they can be measured, much like it is impossible to find consensus on the evaluation and measure of corporate environmental performance.

In addition to these important challenges, recent literature has also started to look more precisely into the reliability of ESG data, and it seems that academics and practitioners also face important challenges in this specific domain as well. Overall, this recent literature seems to indicate that ESG data is not yet reliable, and that different data providers are susceptible to provide different ESG data on similar firms. This means that most studies to date that rely on ESG data and ESG data providers to determine specific relationships between ESG factors and financial performance would probably have different results if they used a different source of data, even if they kept similar samples. This constitutes yet another challenge for academics that already have

to face the lack of common definitions, concepts and metrics in this field. Interestingly, it seems that it is also a lack of common definitions, concepts and metrics that is the cause for this disparity in terms of data between data providers.

Finally, with close to 40% of global assets under management using sustainable investing strategies according to the 2018 GSIA report and the organization's underlying data, it could be expected that environmental, social and governance factors had started to impact asset pricing and the corresponding literature in financial research. Given the increased urgency relative to global warming, environmental factors specifically could also be considered to have an increasing impact on asset pricing. Historically, this has not been the case according to the literature. This could possibly mean that challenges faced by academics in the field prevent them from being able to provide robust empirical research on this subject using currently available environmental or social data. The absence of environmental or social risk factors in what has been described in the literature as a "zoo of factors" that attempt to explain the cross-section of stock returns provides further evidence that academics and practitioners lack a common framework and comparable metrics to evaluate sustainability risk precisely.

2.3. The important role of regulators and the European Commission

In an effort to find a solution to this general issue faced by the sustainable finance industry, the European Commission created its "Action Plan on Financing Sustainable Growth" in 2018 and its three underlying objectives consisting in reorienting investments to sustainable investments, in managing climate risk and promoting transparency and long-termism in finance and the economy.

For each objective, a series of specific actions were determined, adding up to a total of 10 distinct actions. Even though this is not specifically articulated by the European Commission in the official communication of the action plan, each action seems to correspond to one specific type

of actor. Objective 1 has five underlying actions which apply to actors that provide supporting functions to financial market participants: (1) The European Commission itself through the creation of an EU classification system for sustainable activities (the EU Taxonomy) (2) providers of labels and standards (3) EU financial institutions (4) financial advisors (5) benchmark administrators. Objective 2 has 3 underlying actions for more active participants of financial markets: (6) credit agencies and market research providers (7) institutional investors and asset managers and (8) banks and insurance companies. Objective 3 has 2 underlying actions for corporations through (9) corporate disclosure and (10) corporate governance. The objectives, sub-categories of objectives and corresponding actions of the official European Commission communication are summarized in Table 1, which also contains the different commitments that are taken by the Commission concerning each action.

The plan's first and most important action consists in developing the EU Taxonomy in order to create a framework to define what precisely constitutes sustainable finance for the EU Commission. In December 2019, the 'Taxonomy Regulation', which represents the legal basis for the new taxonomy framework, was approved by the European parliament and the European Council (Council of the European Union, 2019), and in March 2020, the latest report on the Taxonomy was published (European Commission, 2020d). In order to develop its classification system, six environmental objectives were identified:

- Climate change mitigation
- Climate change adaptation
- Protection of water and marine resources
- Transition to a circular economy
- Pollution prevention and control

- Protection and restoration of biodiversity and ecosystems

The approach put forward by the EU Taxonomy was developed by focusing on economic activities and identifying how each economic activity performed by a firm could be aligned with these objectives. An economic activity is then considered “Taxonomy-aligned” if it contributes to at least of these objectives, without harming any specific objective, and respective the OECD’s and the UN’s minimum social safeguards.

Given the complexity of their task and the growing urgency of climate change action, initial focus was given to climate change mitigation and climate change adaptation (Objective 1 and Objective 2). In the Taxonomy report of March 2020, the technical expert group define the technical criteria for 68 economic activities for both climate change mitigation and climate change adaptation. These 68 economic activities together represent 93.5% of total European carbon emissions and are identified by the expert group as the most likely to substantially contribute to climate change mitigation and adaptation in their respective sectors.

2.4. The focus on large firms and asset managers

Based on this work, the Taxonomy Regulation brings the EU Taxonomy into practice and enforces its use by three groups of users:

- Large companies that are already required to provide non-financial statements in the EU
- Financial market participants that offer financial products in the EU
- The European Union and Member States when setting public measures, standards or labels

As can be observed, the initial focus of the European Commission is put on only two private actors that constitute its financial markets, as larger players such as banks, insurance providers, pension funds or credit rating agencies are not yet affected. Through this regulation, the EU Taxonomy framework, unlike international frameworks set by the UN, has a legal basis and must

Table 1 – The EU Action Plan on Financing Sustainable Growth

Objective	Section	Action	Main Commitments
Objective 1: Reorienting capital flows towards a more sustainable economy	A unified classification system for sustainable activities	1. Establishing an EU Classification System for Sustainability Activities	- Progressive development of an EU taxonomy - Focus on climate mitigation and climate change adaptation
	Standards and labels for sustainable financial products	2. Creating Standards and Labels for Green Financial Products	- Report on green bond standard and prospectus of green bond issuances - Exploration of an EU Ecolabel
	Fostering investment in sustainable projects	3. Fostering Investment in Sustainable Projects	- Further measures to improve instruments aiming at sustainable investment support - Idea of establishing a single investment fund integrating all EU market-based instruments
	Sustainability considerations in financial advice	4. Incorporating Sustainability when Providing Investment Advice	- Amend the MiFID II and IDD delegated acts to ensure that sustainability preferences are incorporated in the suitability assessment - Invite ESMA to include provisions on sustainability preferences in its guidelines on the suitability assessment
	Sustainability benchmarks	5. Developing Sustainability Benchmarks	- Adoption of delegated acts on transparency of benchmark methodologies - Put forward an initiative to harmonise benchmarks comprising low-carbon issuers
Objective 2: Mainstreaming Sustainability into risk management	Sustainability in market research and credit ratings	6. Better Integrating Sustainability in Ratings and Research	- Explore the merits of amending the Credit Rating Agency Regulation - Invite ESMA to assess current sustainability-related practices in the credit rating market - Comprehensive study on sustainability ratings
	Institutional investors and asset managers' sustainability duties	7. Clarifying Institutional Investors and Asset Managers' Duties	- Proposal aiming at requiring (1) institutional investors and asset managers to integrate sustainability considerations in the investment decision-making process and (2) increase transparency to end-investors
	Prudential requirements for banks and insurance companies	8. Incorporating Sustainability in Prudential Requirements	- Explore the feasibility of including environmental risk in institutions' risk management policies and potential calibration of capital requirements of the Capital Requirement Regulation and Directive - Invite EIOPA to provide an opinion on the impact of prudential rules for insurance companies on sustainable investments
Objective 3: Fostering transparency and long-termism	Disclosure and accounting	9. Strengthening Sustainability Disclosure and Accounting Rule-Making	- Fitness check of EU legislation on public corporate reporting - Revision of guidelines on non-financial information - Creation of the European Financial Reporting Advisory Group (EFRAG) - Assess the impact of new or revised IFRSs on sustainable investments
	Corporate governance and undue capital market short-termism	10 - Fostering Sustainable Corporate Governance and Attenuating Short-termism in Capital Markets	- Assess the possible need to require corporate boards to develop and disclose sustainability strategy - Collect evidence on short-term pressure from capital markets on corporations

be respected, and puts considerable pressure on both these actors. Asset managers, which have never been legally required to report on their sustainability performance, will have to adapt rapidly to these new changes.

Through its structure, the EU Taxonomy puts particular pressure on asset managers as these actors will have to disclose the proportion of their financial products that are aligned with the taxonomy before underlying companies, but also because only large companies will have to disclose information on their alignment with the Taxonomy, in which case asset managers that provide investment products in smaller companies, assets or projects will have to gather this information themselves. This, in turn, should also put pressure on companies, as asset managers will necessarily develop a preference for companies that provide the most information related to their taxonomy alignment.

As global warming is showing no signs of deceleration, clear, impactful and efficient investments in climate change mitigation and adaption need to increase exponentially if climate risk is to be dealt with seriously. Creating a true sustainable finance industry would be an essential stepping-stone to reach that goal. In that sense, the Commission's EU Taxonomy and the surrounding work to create efficient EU sustainable financial markets could well be interpreted as a new sustainable finance paradigm for practitioners and academics alike.

Though much work is yet to be done, and implementation of disclosure criteria for the 6 objectives will not be enforced in practice before the 1st of January 2023, regulation has been adopted on this subject in order for the Commission to enforce its news vision of the European financial system. This important transformation is already taking place and is mostly centered on asset managers. These actors, more than any actors of the financial industry, will have to adapt rapidly to these new changes as EU financial markets evolve in the years to come.

As taxonomy-aligned firms start to be identified, financial market participants will also have more transparency regarding which firms participate most in climate change mitigation and adaptation, amongst other objectives. Given the important investment needs that these environmental objectives require, one might think that these firms will be more attractive to financial markets and obtain more investments and financing from these market participants. But is this the case today? In the following section, we explore this notion by analyzing how capital markets are currently financing both traditional and climate-aligned firms. Having identified the essential role of corporate bond markets in financing these firms, we also explore the corporate bond market more specifically, starting by providing a solution to the diverse data issues that are faced by academics.

3. Capital market financing and bond data for traditional and climate firms

As extra-financial issues relative to climate adaptation and mitigation – amongst others – grow more urgent with time, capital markets will inevitably have to adapt to these changes. In this section, we explore how stock markets participate in financing large firms and climate-aligned firms in the US and in the EU, by differentiating dividends and stock issuances and repurchases. We then move on to explore the corporate bond issuances. Finally, we address the specific issue relative to data availability and quality for the study of corporate bonds, which we needed to deal with rigorously in order to price both traditional and climate bonds in later sections of our work.

3.1. Dividends, stock repurchases and corporate bond issuances for traditional and climate-aligned firms

The investment needs to mitigate and adapt to climate change are increasingly being evaluated by inter-governmental agencies, and, as we've previously mentioned, these needs represent trillions of dollars of additional investments aimed specifically at global warming. Through the exact amount of investments needed remains of course uncertain, there is no doubt that there is a pressing need for investments to make the economy more aligned with global climate objectives in the next few years. These investments needs cannot be assured by public entities alone, and firms and investors will have a major role in creating and investing in the necessary projects.

Part of our work therefore focuses on how large firms and climate-aligned firms are financed by capital markets. Much like the subject of climate risk has been one of the main focuses of financial practitioners in 2020, the investment decisions of large corporations has attracted a lot attention in the last years. As we previously showed, Blackrock CEO Larry Fink chose to discuss climate risk in his 2020 letter to CEOs. However, his 2016 letter was centered on the fact that capital expenditures within firms had been cut to favor dividend distribution and share buybacks⁵. We can easily understand how the combination of these two important issues creates a challenge that could prove complex to overcome: as the need for investments in climate mitigation and adaptation continues to grow, it seems companies are investing less and less in new projects in order to increase payments to their stockholders.

Though this important challenge seems quite straightforward to identify, the literature on the subject is scarce. To the authors' knowledge, one of the most comprehensive study on the

⁵ See <https://www.blackrock.com/corporate/investor-relations/2016-larry-fink-ceo-letter>

subject is the dual study performed by Gutierrez and Philippon (2016, 2017). In this work, authors demonstrate that non-financial businesses are profitable but do not invest, that competition between firms has decreased and institutional ownership and payouts have increased. After performing a series of different tests, the authors conclude that the decrease in investments is due to a decrease in competition and changes in the governance structures of large firms that allow for an increase in shares buybacks.

The key concept in this literature is the fact that with decreased competition, firms have less incentive to invest, and that US businesses have faced decreased competition since the beginning of the century (Gutierrez and Philippon, 2016). In parallel to this trend, institutional ownership of US businesses has increased from 2% to 6% between 1980 and 2015, with a large portion of these institutional investors being quasi-indexers that favor short-term performance (Gutierrez and Philippon, 2016). Focusing on ownership of the S&P500, Fichtner et al. (2016) demonstrate that the three largest asset managers taken together – Blackrock, Vanguard and State Street - are the largest shareholder of 88% of the index and represent 82% of the index total market capitalization. Unsurprisingly, the literature also demonstrates that common ownership amongst competitors reduces competition (Salop and O'Brien, 2000). This, however, is an interesting insight from the literature, as large asset managers such as Blackrock seem to understand the economic risk that represents climate change and are asking firms in their portfolios to take action on this subject.

This leaves us with the following interrogation: are investments from corporations really decreasing, and is the distribution of cash to investors through dividends and share buybacks really increasing? Is this the case for climate-aligned firms as well?

On this specific subject, another branch of literature focuses on the concept of “maximizing share value” (MSV). In a series of different papers, William Lazonick shows that the payout ratio – the ratio of dividends to after-tax adjusted corporate profits – grew from 42% to more than 49% between 1980 and 1998, with share buybacks representing 5% of corporate profits in 1982 but more than 25% by 1985. By 1989, dividends had risen to \$134.4 billion and stock repurchases to over \$60 billion, increasing the combined payout ratio to over 81% (Lazonick et al., 2000). Lazonick performed a more recent study on the subject showing that for the period 2004-2013, companies in the S&P 500 paid \$3.4 trillion in stock buybacks, which represented 51% of these companies’ total income, when dividends already represented 35% (Lazonick, 2015).

Having clearly determined this growing trend throughout the years, Lazonick exposes the different consequences that this could have on businesses. He demonstrates that senior executives that choose to pay large yearly stock buybacks lose their capacity to understand what investments are needed for firms to remain innovative (Baldwin, 1991; Christensen et al., 2008), as stock buybacks represent a powerful incentive for executives to focus more on increasing their pay. Another consequence of large buybacks in a firm concerns skill development. Lazonick states that by spending cash on stock buybacks, the firm is not investing as much as it should in its employees’ skills and careers, which penalizes the firm’s innovative capacities. The last impact of stock buyback regards financial commitment. Lazonick argues that many companies that perform large stock buyback operations throughout the years lack cash once they enter a period of crisis and are unable to finance a restructuring process to become innovative again (Lazonick, 2015).

To date, and to the authors’ knowledge, the central subject of the role of investment and innovation of large firm to mitigate and adapt to climate change in the years and decades to come has not yet been thoroughly analyzed by academia, and many essential questions have been left

unanswered. In a general manner, it seems academics have very little information on the nature of corporations' capital expenditures. Even the general relationship between a firm's capital expenditures and its payouts to shareholders has not yet been clearly established. Studying this relationship further is necessary if one wishes to determine how large companies are financing sustainable and climate-aligned projects, and whether or not the objectives that have been set by the global community can be reached.

In that sense, the work performed by the European Commission is an essential step, and, if correctly applied and regulated, will allow much clearer visibility on the investments that are performed by these large firms. The EU taxonomy and its regulation will lead to large firms publishing the percentage of their turnover and capital expenditures that is aligned with European sustainable objectives on a yearly basis. This will mark a new step in sustainable finance, and may participate in demonstrating how the climate urgency also depends tremendously on the investments that are performed by these large firms, which today mostly finance their projects with their own cash, as we see in the literature and in our own work that stock markets now take more cash from large companies than they provide overall. Trillions of dollars are taken out of these companies to pay their shareholders while trillions of dollars are required yearly to face the devastating possible impacts of climate action failure.

However, some preliminary work can still be performed before obtaining clearer information on the internal investing processes of large corporations. Having identified climate-aligned firms, we can start looking at how large EU and US firms are financed by capital markets, but also how climate-aligned EU and US firms are financed by these same capital markets, and understand how capital market financing differs between these two types of firms. We are also able to determine the part of the long-term debt of US firms that is composed of corporate bond

financing. These figures can then be compared, and we are able to understand whether these firms mostly obtain cash from stock or bond markets, through other means of long-term financing, or simply use their own cash to finance their internal projects or perform acquisitions. As firms start to provide more information on their alignment with climate-related objectives and underlying sustainability or climate performance, the more precise relationships between capital markets and large firms on the subject of climate adaptation and mitigation and overall sustainability can be addressed more thoroughly.

3.2. Corporate bond data for traditional and climate-aligned firms

This study of capital marketing financing brings us to the interesting conclusion that overall, the corporate bond markets are the only positive source of financing for large firms and climate-aligned firms, both in the US and the EU. This leads us to explore the corporate bond literature and understand why these products are typically much less studied by academics and practitioners, when they are both larger and provide much more positive financing for large and climate-aligned firms.

The fact that corporate bonds are less studied than stocks as financial products resides mainly in the complexities of these products and the lack of available data. As we explore extensively in later sections of this PhD dissertation, corporate bonds are more complex financial products than stock, and pricing these products is challenging. Furthermore, quality data on these products is also hard to obtain. In fact, we dedicate an entire article to developing a rigorous data treatment methodology to obtain a reliable database on corporate bonds.

We explore the five distinct databases that provide data on corporate bonds and explain how the types of data differ from one database to another. We then walk through the different steps that need to be taken into account to get from the raw data of these five databases to a clean

database containing the best possible data on corporate bond prices, which in turn allows for quality corporate bond return data. Throughout the entire data treatment process three different types of databases are made available: one provides rich insight on the liquidity of the US corporate bond secondary market, another provides daily transaction-based prices of corporate bonds when the final one provides monthly returns for US corporate bonds. Different alternatives to this data treatment process can be implemented, specifically when transforming intra-day information into daily and monthly prices. We explore these specificities in detail in the data section of this PhD dissertation and in the context of our third article available in Chapter 3.

Amongst the different insights that this data treatment procedure provides us, one consists in understanding how limited is the dataset composed of green bonds, which is not large enough to apply any form of truly robust analysis at this state of the market's development. However, we do get a better understanding of the size of the climate-bond market, which is large enough to be studied, though much smaller than the entire US corporate bond market.

Perhaps the richest output that we compute from this data treatment procedure resides in obtaining monthly corporate bond returns for the entire US corporate bond market from 2002 to 2019, which we are able to use extensively in the last articles of our thesis which focus on applying traditional and modern asset pricing methodologies to study the cross-section of US corporate bonds, for both traditional and climate-aligned firms. We explore traditional empirical asset pricing techniques and more modern machine learning methods and their application to traditional and climate bond markets in the following section.

4. Modern Empirical Asset Pricing and Climate-Aligned Firms

Asset pricing is one of the central subjects of financial economics, and, as previously mentioned, within the asset pricing literature, the literature that focuses on factor models is one of the most – if not the most – prominent. Throughout this literature, four different approaches have been taken in order to find these small sets of variables, referred to as “factors”, that best explain the cross-section of asset returns.

The first approach consists in relying on theory. This was the approach taken by Sharpe (1964) for the original capital asset pricing model and which has been taken by other authors throughout the years. Though this approach was most popular when the first asset pricing models were developed, recent investment-based factors have been identified by relying on theory (Hou, Xue, and Zhang, 2015; Fama and French, 2015). A second approach is purely statistical, and mostly consists in determining factors using the principal component analysis statistical method. The third approach consists of finding factors by identifying patterns in the cross-section of asset returns. A typical example of this approach can be found in Fama and French (1996). The fourth approach has grown more popular in the last decades, and mostly consists in relying on economic or financial intuition to test factors that could influence the cross-section of returns.

Interest to find the most efficient factors grew in the last decades, since determining these factors could allow practitioners or academics some form of predictability of future asset returns. Predictability is an essential subject in asset pricing since it would provide practitioners who find these factors to have access to investment opportunities that outperform their peers’ and for academics to better understand whether or not capital markets are efficient and how asset allocation decisions are made.

Today, many different factors have been identified, and it is also important to have an idea of what these factors are. These factors are first explored by Fama (1970), in which the author differentiates weak-form return prediction - predictive factors in the literature that rely on previous price trends to predict future prices - from semi-strong form return prediction - predictive factors that rely on publicly known lagged variables.

Fama (1970) explores the weak-form predictors and finds, at that time, that evidence for weak-form return prediction to be lacking. However, later studies did demonstrate that certain types of weak-form predictors did exist. Lagged bond yields have been shown to predict bond returns (Stambaugh 1988, Campbell and Shiller, 1991, Cochrane and Piazzesi, 2005). Lo and MacKinlay (1990) find that future stock returns from small firms are correlated with past stock returns from large firms.

The first evidence of strong-form predictability also concerned bond returns (Fama and French, 1989), but many strong-form predictors were later discovered. A review of these predictors was performed recently by Chang and Ferson (2018) and later summarized in Ferson (2019). Chang and Ferson categorize these predictors into “valuation ratios”, “interest rates and spreads”, “anomaly formation spreads”, “market activity measures”, “macro-economic variables” and “miscellaneous and technical indicators”. Though exploring the precise and dense literature on all these variables is outside the scope of this work, it is important for us to understand that the literature has put forward a very large amount of semi-strong predictors in the forty eight years that separate Fama (1970) from Chang and Ferson (2018).

4.1. From predicting levels of returns to predicting cross-sectional differences in returns

Presumably the most popular form of predictability in the recent asset pricing literature is cross-sectional or relative predictability. In these forms of studies, authors try to evaluate the differences that exist between portfolio returns. Typically, these portfolios are created by using weights that depend on certain firm or asset characteristics. Generally, these portfolios are sub-categorized in deciles or quintiles, and the difference between the lowest decile or quintile portfolio return and the highest decile or quintile portfolio return is what is being studied. With this approach, authors can use these characteristics to predict the future differences in cross-sectional returns.

The literature for weak-form and strong-form cross-sectional return prediction is also very dense. Amongst the most put forward weak-form predictors are “momentum” and “reversals”. The “momentum” effect was discovered by Grinblatt and Titman (1993) and exists when past winning stocks have high future returns, and the “reversal” effect was discovered by Debondt and Thaler (1985) and exists when past winning stocks have low future returns. Though consensus on the existence of different forms of momentum exists in the literature, the precise causes of this effect have not yet been clearly identified. Momentum has been found in different asset classes, and over different horizons. The literature specific to this effect is very dense, and we integrate some momentum factors in our work. On the other hand, the literature for reversal is less dense, and this effect has mostly been found in small stocks, low-priced stocks and has mostly been identified in earlier data samples rather than most recent ones. For these reasons it is considered less robust of an effect than momentum.

The literature on semi-strong predictors is centered on the different works of Fama and French. In Fama and French (1992), the authors develop the famous three-factor model. This model is composed, in addition to the market risk factor, of the “small-minus-big” (SMB) factor,

referred to as the “small firm effect” and which represents the excess returns of small firms over large firms, and the “high-minus-low” factor (HML), which represents the excess returns of high book-to-market ratio stocks over low book-to-market ratio stocks (otherwise known as value stocks vs growth stocks). Fama and French (2016) added a “conservative-minus-aggressive” (CMA) factor based on conservative and aggressive investment policies and a “robust-minus-weak” (RMW) factor based on robust and weak profitability. Today, new factors need to resist any form of “explanation” from at least the three original Fama-French factors to be considered. Today, hundreds of factors that explain the cross-section of stock returns have been put forward in the asset pricing literature. These are referenced by Harvey, Liu and Zhu (2016) and Harvey and Liu (2019).

4.2. What about corporate bonds?

Historically, academics in asset pricing and factor analysis have focused on studying stock markets and government bonds. Our focus is on the cross-section of corporate bonds, which is much less studied by academics in financial economics and empirical asset pricing. As previously mentioned, this can be explained by the fact that corporate bonds can prove to be complex financial products and the fact that data on corporate bonds has historically been hard to come by.

The first major paper to study the cross-section of corporate bond returns was Fama and French (1993). In this paper, the authors identify five risk factors that are common to both stock markets and corporate bond markets. In addition to their original three-factor model, the authors develop two bond-market factors. These are the TERM factor, which is a proxy for the changes in interest rates, and the DEF factor, a proxy for default risk. In this study, Fama and French find that the five factors explain the average returns of both stocks and bonds.

Since then, very few papers have focused on predicting the cross-section of corporate bond returns. The most recent work on the subject is that of Bai, Bali and Wen (2019). Realising that the literature on predicting the cross-section of asset returns rarely focused on corporate bonds and that very few factors specific to the corporate bond market existed, the authors developed three bond-specific factors. Understanding the context of this study and the previous literature allows us to better understand the logic behind this study and its possible limitations.

As we've previously seen, the approach to finding factors can take multiple forms. In this work, Bai, Bali and Wen (2019) use financial intuition to develop their factors. They note that bondholders have little upside opportunities and are therefore probably more sensible to downside risk and develop a downside risk factor. They note that these actors are also more exposed to default risk and develop a default risk factor and finally develop a liquidity risk factor given the fact that the corporate bond market is much more illiquid than the stock market. The fact that Bai, Bali and Wen (2019) were able to develop these factors is also dependent on the fact that they entirely and uniquely rely on the best quality trade-based data available for corporate bond markets.

The author's methodological approach is typical of the literature that focuses on predicting cross-sectional differences in returns. An initial univariate portfolio analysis is performed by sorting portfolios in quintiles based on downside risk and measuring their average returns, as well as some other average portfolio characteristics. Going a bit further, the authors perform a bivariate portfolio analysis, in which quintile portfolios based on credit rating, maturity, size and illiquidity are formed before being further subdivided in quintiles based on downside risk. In the context of this study, authors initially refer to portfolio returns, but then focus on a ten-factor alpha as a means to demonstrate that differences between low quintiles and high quintiles portfolios cannot be

explained by the ten most popular risk factors in past literature. The authors proceed to perform a Fama-Macbeth cross-sectional regression to verify that the characteristics used to develop their risk factors do exhibit positive excess risk premiums, with and without selected control variables. The bond risk factors are then constructed following the independent sorts approach of Fama and French (2015), and the factors are tested on size/industry-sorted portfolios, where the authors show that a four-factor model composed of their three risk factors and the excess bond market return factor outperforms other models in explaining the cross-section of corporate bond expected returns.

The methodology developed by Bai, Bali and Wen (2019) can be considered as a state-of-the-art approach of traditional empirical asset pricing. We chose to imitate this methodology using our own database, while also applying, to the best of our abilities, this methodology to our dataset of climate bonds. This allowed us to understand how these traditional asset pricing methodologies are applied in practice, along with the variety of complex challenges that they entail, but also in some way to understand some of their limitations. Furthermore, we are able to produce initial work on differentiating the behavior of climate bonds compared to traditional corporate bonds in the context of the US corporate bond market.

4.3. Possible limitations of traditional asset pricing methodologies

As we made progress in applying the methodology of Bai, Bali and Wen (2019), we did notice some areas that did present, to our view, some form of limitation due to the structural process of the applied statistical methodology.

4.3.1. Control Variables

Throughout the different analyses that are performed in the paper, a series of different control variables are used by the author. In the context of their univariate analysis, the authors use

five bond factors and five stock factors as control variables in order to demonstrate the positive relationship that exists between their downside risk factor and quintile portfolio returns. These factors are namely, for the stock factors, the size (SMB) and book-to-market (HML) factors of Fama-French (1992), the excess stock market return, a momentum factor and a liquidity factor, and for the bond factors, the default (DEF) and term (TERM) of Fama-French (1993), the excess bond market return, another bond liquidity risk factor and a momentum factor. Though this is already valuable work to demonstrate the validity of the bond-specific factor, it does put forward certain logical questions: why were these factors used specifically ? Why weren't other factors used, such as, as one example amongst many others, the aforementioned investment (CMA) and profitability (RMW) factors of Fama-French (2015) ? As the authors progress in their analysis, they use other control variables in the context of the Fama-Macbeth analysis: bond exposures to the DEF and TERM factors, bond amount outstanding, years-to-maturity and the lagged excess return. Once again, we can ask ourselves why the bond exposure to other factors were not used.

4.3.2. Data mining

Before going any further, we would like to state that we are not suggesting here that the authors of the study participated in any form of data mining. However, the general subject of data mining is a central methodological concern in current asset pricing literature and needs to be addressed. A typical empirical asset pricing paper such as Bai, Bali and Wen (2019) is composed of numerous phases of data treatment and data analysis during which many different decisions are taken by the authors. The choices of control variables mentioned above is one amongst many. Another example resides in the many different steps that are taken by the authors to transform their raw TRACE data into an exploitable dataset of vanilla corporate bond returns. Bai, Bali and Wen (2019) use TRACE data from July 2002 to December 2016, when other authors only start using

TRACE data starting from 2005 given the fact that TRACE data was concerned less qualitative during the period going from 2002 to 2005 as the database was being implemented (Bessembinder et al., 2018). Bai, Bali and Wen (2019) do not account for inter-dealer transactions in TRACE data, as is mentioned in Dick-Nielsen (2014), which could impact corporate bond prices and the authors' liquidity measure. Though the validity of the authors' results is by no means questioned here, in fact a large series of additional work is performed by the paper Online Appendix to improve the robustness of their results, the actual structure of traditional asset pricing methodology does leave room for a variety of small changes that could alter final results. To supplement this fact, it has now been established by the literature that there is a publication bias by financial journals, and the fact that only statistically significant results are shown and published in academic papers has been demonstrated (Harvey, Liu and Zhu, 2015).

4.3.3. Out-of-sample data

One of the main responses to this issue provided by the literature consists in trying to determine if the predictive relations holds when applied in out-of-sample data. In the case of studies applying traditional asset pricing, tests on the predictive capacity of factors are performed on portfolios composed of stocks or bonds that had already been used to fit the different models developed by authors. In the case of Bai, Bali and Wen (2019), size and industry-sorted portfolios are used, and the final predictive power of the bond-specific factors are demonstrated through their application on these portfolios. But all of the authors' analyses, whether this regards their univariate and bivariate portfolio analyses, their Fama-Macbeth analysis or the final development of their risk factors, have been performed on the same data, though structured differently.

Different solutions to this issue have been applied to solve this issue, even though the evidence supporting these approaches is mixed. The most popular solution consists in performing

a step-ahead approach, in which only data available at the date of the forecast is used to create the model and the corresponding forecast. This only solves the issue partially, because this does not truly prohibit authors from performing data mining and publishing their best results regardless. The validity of the step-ahead approach is far from having reached consensus, as some authors have found support for this approach in their work (Fama and French, 1991; Ferson and Harvey 1993, 1999) whilst others find evidence of the contrary (Goyal and Welch, 2003, 2007).

4.4. Advantages of machine learning and the “zoo of factors”

4.4.1. Machine learning methods for prediction in asset pricing

Machine learning (ML) methods are slowly starting to be integrated into the financial economics literature, specifically in the field of asset pricing and prediction. In the context of our PhD dissertation, we choose to follow the definition provided by Gu, Kelly and Xiu (2020), who have recently published a paper named “Empirical Asset Pricing via Machine Learning” in the Review of Financial Studies. Much like we’ve chosen to take Bai, Bali and Wen (2019) as a methodological backbone for our paper focusing on traditional asset pricing methods, we chose to follow the Gu, Kelly and Xiu (2020) as a backbone to our own exploratory paper on ML asset pricing methodologies. In their paper on ML in empirical asset pricing, the authors use the term “machine learning” to describe : “(a) a diverse collection of high-dimensional models for statistical prediction, combined with (b) so-called “regularization” methods for model selection and mitigation of overfit, and (c) efficient algorithms for searching among a vast number of potential model specifications.” These three distinct elements of machine learning are mainly what differentiates ML methods from more traditional approaches to asset pricing.

As more and more data on capital markets becomes available to financial analysts and academics and as an increasing number of factors have been put forward as predictors of the cross-

section of asset returns, data on the possible behavior of capital market returns has never been so high-dimensional. Most ML models perform best in context of high-dimensionality, which is not the case of most traditional asset pricing models that rely on multiple linear regression.

These ML models have many different forms of specifications that need to be determined that allow for more flexibility and better fit with data, in particular non-linear data that can hardly be well represented with the linear approach of traditional models. When ML models are backed with efficient, power computational power, then efficient and powerful algorithms can indeed perform intense searches to find the best possible model specification to fit the model to the data. This does come, as is mentioned by Gu, Kelly and Xiu (2020), with an increased risk of overfitting the data, a concern that is central to the machine learning community, inside and outside the realm of empirical asset pricing.

Overfitting arises when ML models are fit too closely to the training data and therefore perform badly when trying to predict unseen test data. As this is a central concern for data scientist, the machine learning community has developed powerful tools against overfitting with the creation of regularization methods that allow to minimize the risks of overfit.

These characteristics make ML methods particularly well-suited to perform prediction tasks, and therefore to be applied to predict future capital market price. In addition, with more than 300 predictors of the cross-section of stock returns having been put forward by the literature, these methods can help analysts and academics better understand which of these predictors truly impacts the cross-section of returns and how. Furthermore, ML methods provide another useful advantage to empirical asset pricing by being particularly well suited to deal with correlation amongst variables, as correlation amongst predictors in the empirical asset pricing literature is a central

concern. Finally, as aforementioned, ML allows for non-linearity when it can easily be accepted that complex capital markets do not behave linearly.

In the context of this PhD dissertation, we dedicate our last article to attempt to apply some of these ML methods to the study of the cross-section of corporate bond returns, which, to the author's knowledge, has not yet been truly attempted in the literature. Though this study is only exploratory at this stage and has important limitations given the author's constrained computational power and still developing, recently acquired data science skills, it provides interesting insight to understand the potential of ML methods in the context of asset pricing.

4.4.2. New factors for corporate bond pricing

Having a robust database of corporate bond returns represents an opportunity to apply machine learning methods to the cross-section of corporate bond returns. However, contrary to the literature on asset pricing that focuses on stock return, the corporate bond literature has only put forward a few predictors other than those developed by Bai, Bali and Wen (2019).

Our approach to finding predictors is two-fold. The first step consisted in finding the most powerful predictors amongst those used in Gu, Kelly and Xiu (2020) and trying to develop similar factors for corporate bonds. The most successful factors in the authors' study are price trends factors, liquidity factors and volatility factors. We are able to develop seven price-trend predictors amongst which are six momentum factors and an adaptation of the "maximum daily return" factor developed by "Bali, Cakici & Whitelaw" and two liquidity predictors : the "dollar trading volume" of Chordia, Subrahmanyam and Anshuman (2001) and the Amihud liquidity measure (Amihud, 2002).

The second step consisted in using the bond-level risk factors of Bai, Bali and Wen (2019) as predictors, but also to adapt some of the control variables used by the authors to include them

in our work. In total, we therefore have a set of 22 potential predictors of the cross-section of corporate bond returns. Though it stands to reason that more factors can probably be computed to explain the cross-section of corporate bond returns, this number of factors is sufficient for the purpose of our study, which consists in an exploration of machine learning methods.

The application of these methods on a dataset of traditional and climate bond returns that is performed in this final article is only possible because of the preceding articles that set the framework for sustainable finance, the current relationship between capital markets and climate-aligned firms, allow the development of a quality dataset of corporate bond returns and that have explored the strengths and limitations of traditional asset pricing methods. But at the base of this work, a variety of different literature was explored, focusing not only on sustainable finance related subjects, but also on more standard issues faced by academics relative to traditional capital markets and bond markets. We present these different literatures in the following section.

5. Literature Review

Our work explores a variety of different literatures. In the context of our first article, we look at the practitioner literature on the SRI market and the academic literature on the general concept of ESG/CSP and on the more specific concept of Corporate Environmental Performance (CEP). In the context of our second article, we explore the literature on the financing of corporation via capital markets, through either share issuances and buybacks or corporate bond issuances and repayments. Finally, our last three articles focus more specifically on the general literature on corporate bonds, with the third article focusing more specifically on the methodological aspects relative to data treatment for corporate bond pricing while the fourth and fifth article take a closer look at the general asset pricing methodologies applied to corporate bonds. Our fifth and last article also studies the literature on statistical and machine learning methods applied to asset pricing in

the context of the study of the cross-section of corporate bond returns. We briefly summarize each of these literatures in the following subsections.

5.1. Literature on the SRI market

The practitioner literature on the SRI market is quite large and complex, as many practitioners try to create their own definition of what constitutes sustainability investing and sustainable finance. In the context of our first article, we summarize the different sustainable investing approaches, provide the definition and corresponding global assets under management for each approach using information for the GSIA global report for 2018. We attempt to provide further details on each approach in Europe and in the United States, using information from the Eurosif SRI Report for 2018 (Eurosif, 2018) and the US SIF Trends report from 2018 (US SIF, 2018). The GSIA, the Eurosif and the US SIF are respectively the main references in terms of market studies of the sustainable investment environment in their geographical regions, Eurosif and US SIF being underlying members that constitute the GSIA. We focus solely on these three reports to construct our first analysis, since they each represent the most reliable source of SRI data for their respective geographical regions.

5.2. The academic literature on ESG/CSP and CEP

An important focus on our first article is to understand how the literature differentiated similar concepts such as CSP, CSR and ESG. In Brooks and Oikonomou (2018), which is dedicated to a comprehensive review of literature on the relationship between ESG disclosure and performance of firm value⁶, the term “ESG” is used 19 times throughout the paper, and is

⁶The title of the authors’ paper is “*The effects of environmental, social and governance disclosures and performance on firm value: A review of the literature in accounting and finance*” (Brooks and Oikonomou, 2018)

mentioned for the first time in section 7 of the paper which directly precedes the paper's conclusion, when the term "CSR" is used 58 times and the term "CSP" is used 82 times throughout the paper. No explanation of the relationship between "CSR", "CSP" and "ESG" is provided, and none of these concepts are defined. Concluding remarks address the link between "CSP/ESG and financial performance", clearly demonstrating that the authors consider CSP and ESG to broadly address the same concepts.

In Friede et al. (2015), which is considered the largest and most recent meta-study on the relationship between ESG and financial performance⁷, 26 different concepts are explored in a total of 50 studies, going from the traditional "Corporate Social Responsibility" and "Corporate Social Performance" to the more specific "Environmentally Sustainable Supply Chain Management Practices", "Social Capital of Entrepreneurs" or "High Performance Work Practices".

The two other main meta-studies used in the literature to demonstrate the relationship between "CSP/ESG" and corporate financial performance are Orlitzky et al. (2003) and Margolis et al. (2009). Throughout the various underlying studies used in their work, these authors reference more than 100 different measures for CSP/ESG. Amongst commonly found measures in both papers are the Kinder, Lydenberg & Domini (KLD) ratings, pollution expenditure measures, the Council for Economic Priorities rankings, various types of annual report disclosures, Fortune magazine ratings, the Domini 400 Social stock index, the Toxics Release Inventory (TRI) measure, lawsuit filings, articles in the Wall Street Journal or affiliation to the Sullivan Principles/divestment from South Africa movements.

⁷Here the title of the paper is "*ESG and financial performance: aggregated evidence from more than 2000 empirical studies*" (Friede et al., 2015)

Focusing on the literature on corporate environmental performance (CEP) more specifically allows for deeper understanding of the issues faced by academics that focus on the concept of ESG/CSP. Though there seems to be common agreement on the terminology used in this literature, many issues remain in order to evaluate and measure CEP.

The first issue consists in the great variability of data sources that are used by academics, who refer to the ASSET4 ESG database, but also to the KLD database, the CDP disclosure database, the Trucost database or use data collected from surveys. Within these databases, different data is used to construct metrics to measure CEP. Depending on the studies, authors refer to carbon emissions, energy consumption, water withdrawal, waste, hazardous waste or pollutants production and a series of other types of metrics to evaluate corporate environmental performance. Finally, most studies rely on samples composed of firms in different industries and geographies. Though academics seem to have come to an agreement regarding the structural definitions and terminologies used in the literature, there seems to be strictly no consensus regarding precise metrics that need to be used in order to measure and evaluate a firm's environmental performance.

Recent literature has also started to look more precisely into the reliability of ESG data. In their paper on the subject, Kotsantonis and Serafeim (2019) hand-collected information on Employee Health and Safety data from sustainability reports of 50 large publicly listed companies across a variety of sectors and found that this information was reported in 20 different ways by these randomly selected firms, using different terminology and units of measure. Understandably, this poses important limitations in terms of aggregation as well as comparability of data. This lack of standardized data also means that ESG data providers must deal with the absence of common indicators. As an example, while a set of companies report on the "Number of accidents with fatal consequences", another might simply report on "Injury rates". Different methodologies are used

by ESG data providers to fill this gap, but it is quite clear that the provided data is then more of an approximation than a reality. Such approximations would in turn lead to less precise results.

Kotsantonis and Serafeim (2019) address another key issue relative to benchmarking methodologies to compute company ESG scores. On this issue, the authors demonstrate how the definition of peer groups for benchmarks and the lack of transparency on the choices made by ESG data provider on their selection of peer groups leads to market-wide inconsistencies in ESG metrics and data.

In turn, these issues lead to a lack of agreement between ESG data providers on the ESG scores of companies. In a study focusing on the ESG ratings of five ESG rating companies, Berg, Koelbel and Rigobon (2019) find that the correlation between ESG ratings is of 0.61 in average, and goes from 0.42 to 0.73, when correlation between credit ratings for Moody's and Standard & Poor's are correlated at 0.99. The authors conclude that given these disparities, ESG performance is therefore unlikely to be properly reflected in asset prices, discourages firms from improving their ESG ratings and poses a significant problem in empirical research as using different ESG datasets would lead to obtaining different results. This concept is verified by Diebecker, Rose and Sommer (2019) that use three distinct datasets to verify the effect of corporate sustainability performance on the cost of equity and find a negative effect when using two renown ESG databases but do not when using a third.

5.3. Corporate financing via capital markets

Part of the literature that was reviewed in the context of this PhD dissertation is not specifically focused on climate-aligned or sustainable financial products or firms. This is necessary in order to understand how the literature approaches certain subjects by focusing on

traditional products and firms. This is the case of the literature on corporate financing via capital markets, regarding either stock market or bond market financing.

Acknowledging the fact that massive investments in climate adaptation and mitigation will be needed in the years and decades to come, we study the literature on the role capital markets in financing large and climate-aligned firms in the US and the EU. In recent years companies have not invested as much as they could, regardless of whether these investments were sustainable or not. In March 2016, BlackRock's CEO Laurence Fink stated in his yearly letter to the executives of S&P 500 firms that, "in the wake of the financial crisis, many companies have shied away from investing in the future growth of their companies. Too many companies have cut capital expenditure and even increased debt to boost dividends and increase share buybacks." Even though this issue seems to have been clearly identified by financial practitioners, academic literature on the subject is scarce. As aforementioned, recent studies that address the subject most thoroughly have been performed by Gutierrez and Philippon (2016, 2017). The authors base their first study on five facts which they demonstrate : (1) non-financial businesses are profitable but do not invest, (2) investments are low even when Tobin's Q is high, (3) depreciation has remained stable since 2000, which means that the decrease in net investment is due to a decrease of gross investment, (4) firm entry has decreased, with a sharp decline of entry rates, exit rates and average number of firms by industry and finally (5) institutional ownership and payouts have increased. Using these facts as a basis for their analysis, Gutierrez and Philippon (2016) explore 8 potential explanations for firm-underinvestment : financial frictions from external finance, bank dependence and safe asset scarcity, measurement errors in intangibles or due to globalization, lack of competition due to either regulation or other factors and tighter governance. The authors' conclusion is the

following: “this investment wedge appears to be linked to decreased competition and changes in governance that encourage shares buyback instead of investment”.

These results participate in several strands of literature that are quite recent. The first is that of competition, and the link between competition, investment and innovation. On this subject, Alghion et al. (2014) find that increased competition leads to increased R&D investments by firms that operate at the same technological level, but decreased R&D investment by firms that are trailing behind. Competition also reduces the number of firms that operate on the same technological level and reduces the number of industry leaders. However, in their study, Gutierrez and Philippon (2016) state that their data suggests that firms that operate in concentrated or aging industries and growing firms that do not face any entry risks could have weak incentive to invest. This notion is supplemented with the literature that supports the hypothesis that competition is decreasing in several economic sectors (CEA 2016), which could then explain why investments are decreasing. This subject is specifically addressed in Gutierrez and Philippon (2016), in which the authors find a causal relationship between competition and investment, and state that US businesses have been under-investing since the early 2000's.

Focusing more specifically on dividends, share issuances and share repurchases, Lazonick et al. (2000) investigate the impacts of maximising shareholder value (MSV) on corporate governance in US firms. They find that pay-out ratios (the ratio of dividends to after-tax adjusted corporate profits) of US firms stayed quite stable between 1960 and 1980 at around 42 per cent, but rose between 1980 and 1998 to more than 49 per cent. The authors investigate the growth of share repurchases that occurred during the 1980s, with share buy-backs representing less than 5 per cent of corporate profits before 1982 and exceeding 25 per cent by 1985. By 1989, dividends

had risen to \$134.4 billion and stock repurchases to over \$60 billion, increasing the combined payout ratio to over 81 per cent.

The trend that had been identified by Lazonick et al. (2000) did not stop then, and in 2015 Lazonick published a study that focused specifically on the stock buyback issue and the concept of “retain-and-reinvest to downsize-and-distribute”, a term coined to illustrate the shift that occurred in US firms as they started to use their profits to pay shareholders instead of reinvesting them in their business activity (Lazonick, 2015). For the period 2004-2013, Lazonick (2015) shows that companies in the S&P 500 paid \$3.4 trillion in stock buybacks, which represented 51% of these companies’ total income, when dividends already represented 35%. For the author, the explanation for this growth in repurchases is linked to the stock-based pay of firm executives. Between 2006 and 2013, these ranged from representing 66 per cent to 84 per cent of total annual remuneration of highest-paid executives in these same firms when salaries and bonuses only ranged from 5 to 12 per cent. Furthermore, Lazonick also adds that “the vesting of stock awards is often dependent on the company hitting quarterly earnings per share (EPS) targets, for which well-timed manipulative boosts from stock buybacks can be very helpful.” In essence, highest-paid executives have a clear incentive to perform stock buybacks, which reduces the number of shares outstanding, increases EPS and helps them reach EPS targets to obtain stock awards.

Lazonick (2015) addresses two distinct flaws of the MSV approach. The first is that this approach is based on the assumption that shareholders are the sole corporate participants that bear risk, while this is not the case. The author argues that taxpayers and workers are also risk-bearers that have an economic claim on distribution of profits. On the one hand, the government invests and offers subsidies to large firms that may not generate profits that can be taxed. On the other hand, many workers’ pay does not correspond to their level of participation in the firm’s business

activities (Lazonick, 1990). The arguments provided by Lazonick (2015) are ever more interesting when considering the climate change challenge that is currently faced by US firms and which concerns all of their stakeholders. The authors' second argument is that public shareholders are not long-term investors that invest in the value-creating capabilities of the corporation, but rather traders that wait for the share value of the firm to rise before selling their shares for a profit.

Finally, Lazonick (2015) provides a summary of his research on how stock buybacks undermine a firm's innovative capabilities. One of the impacts of stock buybacks on the firms concerns strategic control. Lazonick states that senior executives that choose to pay large yearly stock buybacks lose their capacity to understand what investments are needed for firms to remain innovative (Baldwin, 1991; Christensen et al., 2008), as stock buybacks represent a powerful incentive for executives to focus more on increasing their pay. Another consequence of large buybacks in a firm concerns skill development. Lazonick states that by spending cash on stock buybacks, the firm is not investing as much as it should in its employees' skills and careers, which penalizes the firm's innovative capacities. The last impact of stock buyback regards financial commitment. Lazonick argues that many companies that perform large stock buyback operations throughout the years lack cash once they enter a period of crisis and are unable to finance a restructuring process to become innovative again (Lazonick, 2015).

5.4. Data processing methodology to study traditional and climate bond returns

We study the traditional bond literature to understand what databases are used, how data is treated and more importantly for what purpose. Research on traditional bonds relies on six financial databases, each providing different types of data, on different markets. Some of the data provided by the databases overlap, and the literature also provides a ranking on the most qualitative

data sources depending on the data types that are provided. The main databases are Mergent's Fixed Income Securities Database (FISD), the Trade Reporting and Compliance Engine (TRACE) database, the National Association of Insurance Commissioners Database (NAIC), DataStream, Bloomberg, and the Lehman Brothers Fixed Income Database. (Jostova, et al, 2013; Chordia et al., 2017). As the Lehman Brothers Fixed Income Database provides data from January 1973 to March 1998, we exclude it from our analysis, as our dataset spans from January 2013 to July 2019.

In addition to providing us with explicit details on the advantages and limitations of each financial database that provides information on traditional bonds, the literature also gives precedence to specific databases depending on the types of data they provide. Each database can provide information on the characteristics of bonds, information on trades and transactions that have been on the secondary bond market, and information on the bonds prices on the primary bond market. Regarding data on the characteristics of bonds, precedence is given to FISD over Bloomberg, these two financial databases being the only ones that offer such types of data. This is justified by the fact that FISD is recognized as the most comprehensive database on bonds, but also due to the fact that Bloomberg offers only a limited amount of data extractions. As aforementioned, regarding data on bond prices, the literature differentiates dealer-quote data provided by Datastream and Bloomberg from transaction data provided by TRACE and NAIC. Precedence is given to transaction data (Jostova et al., 2013). TRACE transaction data is preferred to NAIC transaction data given the fact that NAIC data only provides information relative to insurance companies. Datastream quote-based data is preferred to Bloomberg quote-based data since it gives no restriction on data extractions. The data selection sequence regarding bond prices

is therefore the following: TRACE, NAIC, Datastream, Bloomberg⁸. Regarding data on characteristics, Datastream only provides us with issuers identification and currency used for the issuance, and TRACE and NAIC with issuer identification and information that is specific to each transaction individually. In the context of this PhD dissertation we do not integrate NAIC data into our database given the fact that great majority of NAIC transaction data is integrated into the TRACE database.

As FISD and TRACE data only provide information on the US bond market, we understand that we can only apply the most qualitative analyses to the United States market. The study of other geographical markets will be both limited by Bloomberg data extractions for bond characteristics, and by the fact that transaction data is not available for academics. Therefore, we choose to focus specifically on the US market in this PhD dissertation.

Having developed a clear understanding of the general scope of available bond data that can be used to analyze traditional bonds on financial databases, we use the literature to differentiate data treatment processes that are applied to these datasets. We differentiate two types of research designs that require distinct data treatment processes: cross-sectional studies and event studies. Two main differences distinguish these approaches. In a majority of cases, cross-sectional studies can be performed using monthly bond prices while it has been demonstrated that event studies perform better when using daily transaction data. Furthermore, it is recommended that noninstitutional trades below \$100,000 be eliminated in the case of event studies. (Bessembinder et al., 2009). Regardless of these two approaches, much like it is the case for research on equity

⁸ Regarding data on characteristics, Datastream only provides us with issuers identification and currency used for the issuance, and TRACE and NAIC with issuer identification and information that is specific to each transaction individually.

markets, most research on traditional bonds focus on bond returns, whether monthly returns for cross-sectional studies, and daily returns for event studies.

We continue to follow the literature to develop our database. Bessenbinder et al. (2009) allows to transform transaction data into daily prices. We follow Jostova et al. (2011) to transform daily prices into monthly prices by taking the last daily price available within the last five business days of each month. In order to transform daily prices into monthly prices, we develop methodology inspired by Bao, Bali and Wen (2019). We discuss this methodology in the data section of this PhD dissertation.

5.5. The nascent climate-aligned bond literature

The growing number of green bond issuances has led academics to develop an interest in this product. Green bonds are bond products issued in order to specifically finance projects with positive environmental outcomes, and naturally a stem of literature is focused on trying to determine if green bond issuers benefit from a green bond premium given the nature of their project.

However, this initiative has met important challenges. First and foremost, the pricing of bond products is more challenging than that of equity products, given the fact that bonds have multiple specific characteristics that directly impact their pricing, such as coupon rates, credit rating, maturity or size (Zerbib, 2019). Determining if a green bond premium exists would require comparing green bonds with traditional bonds that have precisely the same characteristics, which very rarely exist (Bachelet et al, 2019). In addition to these important limitations, the green bond market represented less than 3.6% of the global bond market issuances⁹ in 2019, and issuances are

⁹ Global bond issuances in 2019 amounted to \$7.148 trillion. See <https://www.dealogic.com/insight/dcm-highlights-full-year-2019/>

still sporadic throughout the year, which leads both to issues in terms of liquidity, as well as in terms of available data for pricing. Moreover, the academic literature that focuses on studying the corporate bond market refers to one specific transaction-based database to obtain the best quality data on the pricing of corporate bonds – the TRACE database¹⁰ – which only applies to US corporate bonds. For the year 2018, Moody’s Analytics¹¹ reported that \$1.553 trillion in bonds were issued in the US, while the Climate Bond Initiative reported \$34 billion in US green bond emissions (Climate Bond Initiative, 2018). This represents less than 2.2% of US corporate bond emissions, meaning that academics that wish to study this database would have access to too little data to perform robust analyses.

Studies on the pricing of green bonds therefore generally have small samples and must refer to other databases with less precise pricing data, such as dealer quotes provided by market-makers or matrix-prices, which only provide approximations of real prices. To provide an idea of the general ranking of data on corporate bond pricing in terms of quality, Jostova, Nikolova, Philipov and Stahel (2011) built a database of bond returns using the five databases that gave information on corporate pricing and took “the first available return in the following sequence: TRACE, FISD, Lehman, Datastream and Bloomberg”, clearly giving precedence to trade-based data. Combined, these restrictions in terms of sample size, historical data and pricing data quality results in inconclusive findings concerning the existence of a green bond premium (Hachenberg and Schiereck, 2018; Bachelet et al, 2019; Zerbib, 2019). Similar studies have also focused on the relationship between ESG ratings and corporate bond performance, finding that bonds issued by firms with higher ESG ratings have tighter spreads and outperform peers with lower ESG ratings

¹⁰ As TRACE transaction data is the main source of data for this paper, more information is provided in section 4 of this paper on the specificities of this database

¹¹ See <https://www.moodysanalytics.com/-/media/article/2019/weekly-market-outlook-corporate-bond-issuance-reflects-business-activitys-heightened-to-rates.pdf>

(Polbennikov, Desclée, Dynkin and Maitra, 2016). Ge and Liu (2015) find that better CSR performance is associated with better credit ratings. In a paper focusing specifically on corporate green bonds, Flammer (2018) finds a positive reaction from stock markets to green bond issuance announcements, that green bond issuers improve their environmental performance after the issuance, and that they experience increase in ownership by long-term and green investors. To date and to the author's knowledge, there is no academic literature focusing on empirical studies of climate bonds.

5.6. Asset pricing and traditional and climate bonds

Academics have tried to explain the behaviors of corporate bond returns through variety of factors. This was initiated by Fama and French (1993) when the authors identified five risk factors that were common to the returns of stocks and corporate bonds. Since then, the study of the cross-section of corporate bond returns has created interest in the asset pricing literature. Academics that focused on this subject mostly developed factors using either stock-level data, treasury bond data and macroeconomic data. This is the case for the long-established Fama-French (1993) factors composed of the market risk factor ($Mm-Rf$), the size factor (SMB) and the book-to-market factor (HML) that originate from stock data and treasury bond data, and the term spread (TERM) and default spread (DEF) factors that originate from treasury bond data and government bond data. Other factors that have complemented Fama and French's work, such as the liquidity (LIQ) factor (Pastor and Stambaugh, 2003), momentum (MOM) factor (Carhart, 1997), and more recently, the investment (CMA) and profitability (RMW) factors (Fama and French, 2015) all originate from stock-level data. Realizing that these factors performed poorly in their ability to explain industry-sorted and size/maturity sorted portfolios of US corporate bonds, Bai, Bali and Wen (2019) introduced new bond-implied risk factors based on characteristics specific to corporate bonds.

These are to date, and to the author's knowledge, the only bond-specific factors to have been developed. While the asset pricing literature on factor models for stock markets is quite dense, the literature for corporate bonds is quite smaller. We've discussed most of the asset pricing literature in the previous section of this thesis in order to contextualize our approaches for applying traditional and machine learning asset pricing techniques in our last two articles, so we will not repeat ourselves in this section of our work.

5.7. Machine learning in asset pricing applied in the context of sustainable finance

Within the financial literature, machine learning methods have mostly emerged in the context of the asset pricing research. A lasso regression¹² method was applied to predict international stock returns and understand the role of the United States in these global market returns (Rapach, Jack, Strauss and Zhou, 2013). Regression trees were used to predict credit card risk (Butaru, Florentin, Chen, Clark, Das, Lo and Siddique, 2016) and model consumer credit-risk (Khandani, Amir, Kim, Lo, 2010). Neural networks have been used to price and hedge derivatives (Hutchinson, James, Lo and Poggio, 1994), model mortgage prepayment and delinquencies (Sirignano, Justin, Sadhwani and Giesecke, 2016), and to perform portfolio selection (Heaton, Polson and Witte, 2017).

Focusing more precisely on the cross-section of expected returns, several technical reports for the University of Chicago and Duke University have applied machine learning techniques to improve their regression models. These include using the bootstrap procedure (Harvey and Liu, 2019) as well as dimension reduction methods for factor pricing models (Giglio and Xiu, 2017;

¹² Lasso (least absolute shrinkage and selection operator) regression is a machine learning method that consists in performing both variable selection and regularization to optimize the accuracy and interpretability of the produced statistical model.

Kelly, Pruitt and Su, 2017). Tree-based models have also been used to sort portfolios in order to establish a relationship between past and future stock returns (Moritz and Zimmermann, 2016). Other reports from the University of Michigan (Kozak, Nagel and Santosh, 2017) and the University of Wisconsin-Madison (Freyberger, Neuhierl and Weber, 2017) have worked on applying selection methods and shrinkage in the context of the study of the cross-section of expected returns.

Looking at the machine learning literature, many authors have realized that machine learning methods could be applied to the consequential challenge of using times series data to predict stock market returns, and that their models could be used to develop profitable strategies regardless of the efficient market hypothesis (Henrique, Sobreiro, Kimura, 2019). The efficient market hypothesis was developed by Malkiel and Fama (1970) but was revised more than 20 years later by Fama (1991) when the author established that financial markets followed random directions and were therefore unpredictable. More recently, the efficiency of financial markets has been challenged many times, and authors both from machine learning research (Atsalakis and Valavanis, 2009) and finance research (Malkiel, 2003) have addressed this issue by summarizing different works in their field. Given the non-stationary, non-linearity, noisy and chaotic nature of price time series in financial markets, and the fact that these can be influenced by a variety of economic, financial, political or even psychological factors, research on the subject has generated much interest in the machine learning community (Henrique, Sobreiro, Kimura, 2019). Prediction models developed by the machine learning academic community are applied to financial distress estimation (Chen and Du, 2009) and stock trading prediction (Chang, Liu, Lin, Fan and Ng, 2009; de Oliveira, Nobre, and Zarate, 2013), with a series of studies successfully identifying profitable

opportunities (Doeksen, Abraham, Thomas, and Paprzycki, 2005; Huang, Yang and Chuang, 2008; Patel, Shah, Thakkar and Kotecha, 2015).

To date, there are few studies that focus on applying machine learning methods to bond markets. Bianchi, Buchner and Tamoni (2019) apply such methods to estimate bond risk premia following the work on equity risk premia by Gu, Kelly and Xiu (2019) on Treasury Bond data. Other studies focus on more specific technical subjects, such as using machine learning to identify clients that are most likely to be interested in a given bond (Wright, Capriotti, Lee, 2018), or to improve the quality of corporate bond yield data (Kirczenow, Hashemi, Fathi and Davison, 2018). Given the positive results of machine learning on market price prediction both for equity and treasury bond markets, it seems clear that machine learning methods will increasingly be used in financial research.

In the following section, we explore the essential steps that were taken in the context of the PhD dissertation in order to obtain reliable data necessary to understand the fundamentals issues of sustainable finance, obtain a dataset of climate-aligned firms, understand the relationship between capital markets and traditional and climate-aligned firms and finally obtain a rich dataset of traditional and climate-bond bond returns.

6. Data

Our data sources vary depending on the study. The only study that does not refer to any database per say is our first study, which focused on studying the academic and practitioner literature on sustainable finance and provide an explanation of how regulators, and more specifically the EU Commission, were making progress on finding solutions to challenges faced by actors of the sustainable finance industry. In the context of this initial paper, our work is solely

based on an analysis of the practitioner and academic literature on the subject, supplemented by a thorough analysis of regulatory texts on this specific subject. Our second study uses the list of climate-aligned firms provided by CBI, data from the Compustat North America and Compustat Global databases as well as the data from the FISD database. Our third study focuses specifically on developing a data treatment methodology for the US corporate bond market which relies on four databases: TRACE, FISD, Datastream and Bloomberg, and also refers to the list provided by CBI. Our fourth and fifth studies rely only on data from the TRACE dataset, which is considered as the database that provides the best quality data on corporate bond pricing. We explore the specifics of the data approach for each study individually.

6.1. Data treatment methods for capital market financing for traditional and climate-aligned firms

In the context of our study on capital market financing, four sources of data are used. We use Compustat North America to obtain information on the financial cash flows of US firms in the S&P500 index and Compustat Global to obtain information on the financial cash flows of firms in the S&P350 Europe index in countries that use the euro as a currency. We obtain similar information for US climate-aligned firms and European climate-aligned firms. Compustat North America provides fundamental and market information on US and Canadian firms such as annual and quarterly income statements, balance sheets and statement of cash flows. Annual data for most companies is available since 1950, and quarterly history and monthly market history since 1962. Compustat North America also provides information on index constituents which allows us to obtain information on companies that are part of the S&P500 index. Compustat Global provides similar data in more than 80 countries other than the US and Canada and covers 96% of European market capitalization. It provides annual and quarterly information starting from 1987.

We use the Mergent Fixed Income Securities Database (FISD) to obtain information on the US corporate bond market and to evaluate the proportion of long-term corporate debt issued by US firms in the S&P500 that is traded on these capital markets. We obtain similar information for US climate-aligned firms. The FISD database contains information on issues of more than 140,000 publicly offered US bonds. This information can be specific to both corporate bond issuers and corporate issues.

Finally, we use data provided by the Climate Bond Initiative (CBI) to identify US climate-aligned firms and European climate-aligned firms. CBI is an international NGO that focuses on the bonds as financing solution for climate change mitigation and adaptation. In 2018, CBI identified firms around the world that had more than 75% of their turnover than originated from climate aligned activities. The CBI research team were kind enough to provide this list of firms to the author.

6.1.1. Cumulative Cash Flows for the period 2009-2019

In order to obtain the required information from the Compustat database, we first identify companies that are part of the S&P500 in Compustat North America and companies that are part of the S&P350 Europe Index in Compustat Global from 2009 to 2019. The S&P 500 index in Compustat contains a total of 754 companies that have been part or currently are part of the S&P 500 and the S&P 350 Europe Index contains 364. We exclude financial services companies from our sample and keep only companies that have available data from 2009 to 2019. Our final sample contains 531 companies from the S&P 500 Index and 233 companies from the Europe S&P 350 Europe Index.

In order to have a synthetic view of firms in our sample, we focus on the main fundamental indicators from cash flow statements, differentiating, much like it is the case in firm's consolidated

accounts, net cash flow from operating activities from net cash flow from investing and financing activities. Main fundamentals for each categories of activities are net income and depreciation and amortization for operating activities that provides us with the main information on the cash that is generated by the firm's activity throughout the year, capital expenditures and acquisitions for investing activities, and information on share sales and repurchases, dividends and changes in debt for financing activities. For each section of the cash flow statement, we include an “Other” category that sums up the remaining cash flow items in the section. In total, we obtain information from 15 cash flow items from cash flow statements of both US and European firms between 2009 and 2019.

6.1.2. Average Cash Flows for US 1971-2019 and Euro 1999 - 1919

The Compustat North America and Global databases can also be used to obtain average values for cash flow items of US and Euro firms going back many years before 2009. As aforementioned, the Compustat North America database provides annual information on US firms since 1950. However, the first year for which all cash flow information is available on the database is 1971. We therefore run our sample of average values for US cash flow from 1971 to 2019. The Compustat Global database provides data from 1987 but given the currency issue we face with this dataset, we run our sample from 1999 to 2019 provided that the euro was launched on the 1st of January 1999.

6.1.3. Corporate bonds issuances and reductions 1995 - 2018

We use the FISD database to identify corporate bond issuances and reductions from US firms in the S&P500 in order to understand what proportion of US long-term debt is composed of debt securities publicly tradable on US capital markets. FISD is considered as the most

comprehensive database focusing on bonds and contains essential information on bond issuers as well as specific bond issues such as bond issue date, maturity, size, coupon, type, and any information that can be used to identify and categorize US bonds. FISD allows us to have visibility on the corporate bond market and understand operations that occur between companies that issue bonds and the market of bondholders. In addition to comprehensive information of bond issues and bond issuers, FISD provides detailed information on the amount outstanding history for every bond in its database. This information can be used to understand the different amounts that are either paid or received by companies during the lifespan of a corporate bond, which corresponds to amounts that are taken under account by firms when they report their yearly long-term debt issuances and long-term debt reductions.

We first identify corporations of our sample in the FISD database using CUSIP information. A CUSIP number is a unique identification given to each individual security. The CUSIP number is composed of nine characters. The first six characters of a CUSIP number identify the issuer, and the last three identify the specific security issued by this issuer. Compustat and FISD both provide CUSIP numbers for issuers and securities, and we are able to identify every security linked to US issuers in our sample using the first six characters of the CUSIP number. Using the FISD database, we obtain historical data on the different operations that altered the amount outstanding for each bond, from its date of issuance to its date of maturity if that date is reached, or to the corresponding date where the amount outstanding reaches zero due to another form of operation, such as a bond conversion or a call. A great number of these bonds are still active, and therefore the amount outstanding has not yet reached zero.

Some of the bond issuances of companies in our sample are performed using foreign currency. Even though this is the case for a minority of our bonds, we do need to convert these

amounts in US dollars in order to be able to consider them. Much like it was the case for our sample of EU firms, we limit our sample to 1999 and the creation of the Euro, so that this conversion is more precise. There are five distinct currencies used by US issuers in our sample other than the US dollars: the Euro, the Japanese Yen, the Swiss Franc, the United Kingdom Pound Sterling and the Canadian Dollar. Using publicly available data on monthly historical conversion rates between these currencies and the US dollar provided by the Federal Reserve on its website⁴, we convert values for every bond issued in a foreign currency to US dollars.

We then move on to categorizing every bond operation into either a bond issuance, where the company receives capital from bond holders, or a bond reduction, where the company provides capital to bondholders. This is a particularly complex operation given the different specific characteristics of bonds, which can be convertible, callable or reviewed during the life cycle of the bond.

In table 2, we provide an explanation of the different types of operations that are listed in the FISD database for our sample, and whether each operation is categorized as a bond issuance or a bond reduction. To summarize, initial offerings (type “I”) correspond to the initial bond issuance where capital is provided by bondholders to firms. Operations such as “Initial Offering Increase”, “Over-allotments” and “Reopenings” (type “II”, “OA” and “RO”) correspond to further bond issuances that occur after the initial bond offering, and also correspond to an operation where capital is provided by bondholders to firms. Different types of operations that occur during the bond’s lifecycle are linked to calls, refunds, repurchases, firm reorganizations and tender-offers (type “B”, “E”, “F”, “IRP”, “P”, “R”, “S” and “T”) and correspond to bond reductions where capital is provided by the firm to the bondholders. The amount outstanding of a bond can also be increased or decreased following a review (type “REV”), in which case this operation is either an issuance

or a reduction depending on whether the review led to a decrease or increase of funds. The bond can of course reach maturity (type “IM”) in which case the remaining amount outstanding of the bond is paid to bond holders. Other important types of operations include when issues are converted into stocks (type “C”) or exchanged for other securities (type “X”), in which case they either correspond to new share issuances or a new bond issuance. Since we already consider new bond issuances in our sample, this new bond issuance is already accounted for and type “C” and “X” operations are not taken under account.

Our available data for FISD spans to the end of 2018, therefore our sample consists in every bond operation for S&P 500 companies from 1999 to 2018. Our initial dataset for S&P 500 firms from Compustat contained 630 different issuers from 1971 to 2019, and we find information on 14,115 corporate bond operations for 456 bond issuers between 1999 and 2018 in the FISD database. The corresponding information on the different types of operations and their frequencies is available in table 2.

6.1.4. US and Euro climate-aligned firms

Finally, we use a list of climate bond issuers provided directly by the Climate Bond Initiative (CBI) research team. This list was used by CBI in their study on the climate bond market (Climate Bond Initiative, 2018). To identify climate bonds, CBI identified issuers that originated at least 75% of their revenues from green business lines in either clean energy, low-carbon buildings and transport, water and waste management and sustainable land use. Climate bonds were included if they were issued after the 1st of January 2005 and before the end of Q2 2018.

The list provided by the CBI research team did not provide any form of identification number that could be used to rapidly identify issuers in the Compustat database, and firms have to

be identified manually. We identify 28 US climate-aligned firms and 26 Euro climate-aligned firms.

Table 2 – The Corporate Bond Issuance/Reduction Methodology

Code	FISD Description	Description	Type of operation	Number of observations
B	Balance of Issue Called	Part of the issue is called by the issuer.	Reduction	281
C	Issue Converted	The issue is converted to the issuer's stocks.	Other	55
E	Entire Issue Called	Entire issue is called by issuer.	Reduction	406
F	Issue Refunded	Part or all of the issue is refunded by the issuer to bondholders.	Reduction	4
I	Initial Offering of An Issue	Initial bond issuance from to the issuer to bondholders.	Issuance	5502
II	Initial Offering Increase	The initial amount of the issue is increased.	Issuance	4
IM	Issue Matured	The issue has reached maturity and bondholders are paid the principal and final interests.	Reduction	1814
IRM	Issue Remarketed	Changes are made to specific conditions of the bond. In our sample, this does not affect bond amount outstanding.	Other	5
IRP	Issue Repurchased	Part of or all of the bond is repurchased by the issuer.	Reduction	138
OA	Over-Allotment	The initial amount of the issue is increased.	Issuance	274
P	Part of an Issue Called	Part of the issue is called by the issuer.	Reduction	71
R	Reorganization	Reorganisation of the issuing firm leads to part or all of the bond being repurchased by the issuer.	Reduction	88
REV	Review	Review of the terms of the bond issuance can lead amount outstanding to be increase or decreased.	Reduction/Issuance	3126
RO	Reopening	Issuer re-opens, and the initial amount of the issue is increased.	Issuance	158
S	Sinking Fund Payment	Part or all of the issue is paid by the issuer's sinking fund.	Reduction	262
T	Issue Tendered	Issuer offers to repurchase a specific number of bonds to bondholders.	Reduction	859
X	Issue Exchanged	Bond issue is exchanged for another security.	Other	1065

6.2. The corporate bond data treatment methodology

To date, there are five distinct sources to find financial data relative to US green bonds and climate bonds. For the purpose of our database, it is important to differentiate three types of data that are important to obtain to study these bonds. Firstly, data on bond characteristics is essential. This includes information on bond issuance dates, maturities, currency, ratings, industry, and any form of information that can help us identify bonds and group them by specific categories. Other types of data include historical prices, whether this regards daily, weekly or monthly prices provided by either quote-based databases or trade-based databases. As previously mentioned, the

literature gives precedence to trade-based databases, as these are regarded as giving higher quality data on corporate bond prices. Each of the following databases gives different types of data that we must categorize in order to develop our data processing algorithms. Table 3 summarizes the different used databases, the types of data that they provide as well as the markets they describe.

Table 3 – Corporate Bond Databases

Database	Type of Data	Advantages	Limitations
FISD	Characteristics	Most comprehensive database on US Bonds	Limited to the US
Bloomberg	Characteristics, Quote-Based Prices	Provides the list of green bonds. Provides characteristics and historical prices. Is the most comprehensive database overall.	Limited Extractions. Characteristics data on Bonds is less precise than FISD. Quote-based prices are less precise than trade-based prices.
TRACE	Transaction Data	Provides transaction data on US Bond market	Limited to the US
Datastream	Quote-Based Prices	Provides historical prices for all bonds	Quote-based prices are less precise than trade-based prices.

The Mergent Fixed Income Securities Database (FISD) is considered the most comprehensive database regarding the characteristics of publicly offered U.S. bonds. As aforementioned, It contains information on more than 140,000 corporate bonds, medium term notes, supranational or US Agency and Treasury debt products. It provides information both on issuers and on specific issues. The FISD database is composed of a series of datasets that focus on specific characteristics of bond issuers and issuances. In the context of this paper, we focus on the datasets that provide information on issues, issuers, agents, coupons, industry codes and ratings. The FISD database provides us with all necessary information relative to bond characteristics for US green bonds and climate bonds.

Since July 2002, all corporate bond transactions in the secondary market have been made available with the TRACE system through the Trade Reporting Compliance Engine. Before this

new dataset was made available, most studies that focused on the corporate bond market only used daily quotes and matrix prices for corporate bonds, which could bias results. In their paper on the subject, Sarig and Warga (1989) explore the fact that there can be liquidity-driven noise errors in daily prices for corporate bonds since daily prices are given even on days when bonds have not been traded for multiple days. When this is the case, brokers set matrix prices based on similar bonds issued by issuers with similar characteristics, which creates bias. More recently, Dick-Nielsen et al. (2009) show that this bias still exists with prices from Datastream. This gives TRACE data a considerable edge when focusing on daily prices that can be used for microstructure research such as event studies, as well as an edge for weekly and monthly prices. Important changes have been made to the TRACE database from 2002 to 2012, however since our database spans from January 2013 to June 2018, only the latest version of TRACE needs to be considered for our data processing methodology. Two distinct versions of TRACE exist: the standard TRACE data that censors trading volumes that are greater than \$5 million for investment grade bonds and greater than \$1 million for speculative grade bonds and that usually has a three-month lag for the availability of data, and the enhanced TRACE data, that has information on all transaction volumes but has an 18 month lag for the availability of data.

The last two databases are Bloomberg and Datastream. Datastream provide us with corporate bond prices, while Bloomberg offers both data on bond characteristics and on prices. However, Bloomberg has limited monthly extractions, which is quite an important constraint when considering the sizes of the samples we wish to focus on. It is interesting to note that data from Datastream is preferred to Bloomberg by academics that focus on the corporate bond market.

The data processing methodology that was developed in this research work in order to obtain green bond and climate bond data, in addition to tradition bond data for the US corporate bond market consists in the following steps:

- Obtain a list of US Green Corporate Bonds using the Bloomberg database

Amongst the different databases available, Bloomberg is the only one that clearly identifies green bonds. As information of bond characteristics is more precise when referring to the FISD database, we only extract monthly prices from Bloomberg and information relative to the identification of each bond. We can then use this information to identify green bonds on the FISD database.

- Obtain a sample of US Climate Bonds using the list provided by the Climate Bond Initiative

In order to identify US climate bonds, we use a list of climate bond issuers provided directly by the Climate Bond Initiative (CBI) research team. This list was used by CBI in their study on the climate bond market (Climate Bond Initiative, 2018). we use the list of issuers provided by the Climate Bond Initiative to identify climate bonds in the FISD database.

- Obtain Enhanced TRACE data

By using the list of CUSIP numbers we obtain from FISD, we extract 107,848 observations for our green bond dataset and 755,847 observations for our climate bond dataset from the enhanced TRACE database spanning from 01/01/2013 to 09/30/2019¹³. It is important to note that

¹³ At this study is being performed, this sample is composed only of bonds issued before the 31st of December of 2018, as the author does not have access to FISD 2019 data.

according to Dick-Nielsen (2014), the filters that need to be applied to on the enhanced dataset filter around 35% of observations.

- Develop the Enhanced TRACE data filter

The enhanced TRACE dataset contain a certain amount of reporting errors that need to be identified and deleted. These errors have been made by agents as they were reporting transactions and corrected at a later date through another report. In addition, some transactions are reported multiple times since dealers and agencies that trade for final customers have to report the same trade to TRACE. This can have implications in terms of computing both liquidity and price from the TRACE dataset, and therefore needs to be corrected.

- Verifying and applying the TRACE filter

The Trade Reporting Compliance Engine provides a TRACE fact book that gives the official number of transactions on TRACE for specific bonds. In his 2009 paper, Dick-Nielsen tests the performance of his error filter by matching the number of transactions he obtains after applying his filter to the number of transactions for the same bonds from the official TRACE fact book. We apply a similar approach and obtain very positive results for our own filter.

- Transforming TRACE clean data into daily bond prices

Once we have successfully cleaned and processes TRACE data, we can obtain daily prices for each corporate bond from this data. Following Jostova et al. (2013), the first method consists in computing daily prices as the trade-size weighted average of intraday prices, as findings in Bessembinder et al. (2009) show that a daily price based on trade-size weighted intraday prices is less noisy than the last price of the day. I apply this approach commonly used in the literature to compute daily prices for the green bond and climate samples.

- Transforming TRACE daily bond prices into monthly bond prices

In order to obtain monthly prices from daily prices, I develop a methodology similar to that of Bai, Bali and Wen (2019)¹⁴. I first identify the bond price for the last trading day for each bond-month. If the last trading day is one of the last five trading days within month t , then this daily price is used as a monthly price. If that is not the case, I identify the bond price for the first trading day of month $t+1$. If this trading day is one of the first five trading days of month $t+1$, it is used as a price for month t . This allows for a more complete dataset of bond monthly prices in order to be able to compute more monthly returns using data from months t and $t+1$, where returns are either computed using (1) end of month $t-1$ to end of month t daily prices (2) start of month t to end of month t daily prices (3) start of month t to start of month $t+1$.

- Extracting Bloomberg and Datastream Data and Merging databases

Once we have computed monthly prices using trade-based data from TRACE, we extract monthly prices from Bloomberg and Datastream. Even though TRACE provides the best quality data on corporate bond pricing, Datastream seems to have a very consequent lead regarding the quantity of available data on monthly corporate bond prices and maintains a strong correlation with TRACE trade-based data.

As we continue to follow the literature, we merge databases We keep all information regarding bond characteristics from FISD, and, when there are prices from the same bond and

¹⁴ In Bai, Bali and Wen (2019), the authors identify two scenarios to compute end-of-month returns for corporate bonds using daily data. To compute a return for month t , they either use data from the end of month $t - 1$ and the end of month t , or data from the beginning of month t and the end of month t . If a return can be computed through both methods, they apply the first method (end of month $t-1$ to end of month t).

month available from multiple sources, we take the first available price in the following sequence, willingly setting precedence to trade-based data: TRACE, Datastream, Bloomberg.

- Computing monthly returns using month-end prices

We refer to the formula developed in Jostova et al. (2011) to compute monthly returns:

$$r_{i,t} = \frac{(P_{i,t} + AI_{i,t} + Coupon_{i,t}) - (P_{i,t-1} + AI_{i,t-1})}{P_{i,t-1} + AI_{i,t-1}}$$

where $P_{i,t}$ is the transaction price, $AI_{i,t}$ is accrued interest, and $Coupon_{i,t}$ is the coupon payment, if there is any, of bond i for month t .

Computing accrued interest requires the bond's coupon size, coupon frequency, and day count convention. If the coupon frequency is missing, we assume it is semiannual. If the day count convention is missing, we assume it is 30/360.

Once again, information relative to the bond's coupon size, frequency and day count convention can be obtained from FISD. We can use this to transform our datasets of monthly prices obtained from TRACE, Datastream and Bloomberg data into monthly returns.

6.3. Data used for traditional and climate bond pricing

Article 4 and Article 5 rely on two of the same sources of data: the FISD database for essential information on bond characteristics, the TRACE database for high-quality monthly corporate bond prices and returns, and the list of climate-aligned firms provided by CBI. Though we developed a methodology that includes two additional databases in the previous article, we chose to follow the current literature that focuses on asset pricing applied to corporate bonds and

focus solely on TRACE, which offers much more precise prices for corporate bonds than Datastream and Bloomberg.

The data processing methodology used for these papers is quite similar to that developed in Article 3, specifically regarding the treatment of TRACE, FISD and CBI data. It is also developed by following some more specific measures taken by Bai, Bali and Wen (2019). Following the authors, once we have developed our TRACE database for the entire US corporate bond market intraday transactions, we remove bonds that are not traded in in the US public market, bonds that are structured notes, mortgage or asset-backed, agency backed or equity-linked, convertible bonds, bonds with floating coupon rates, bonds that trade below \$5 or above \$1000 as well as bonds that have less than one year to maturity. Regarding transaction data, much like Bai, Bali and Wen, we remove bond transactions that are labeled as “when issued”, “locked-in”, that have special sales conditions and that have more than a two-day settlement. Furthermore, we also remove transactions that have a volume inferior to \$10,000. Finally, we follow the filtering methodology developed in Dick-Nielsen (2014) to remove cancellations and corrections made on the TRACE Enhanced database, as well as inter-dealer transactions that are reported twice in the trace database. Much like the author, we apply the two distinct approaches for pre- and post-2012 Enhanced TRACE data. After applying these different filters, our dataset is composed of 91,294,517 intraday transactions spanning from June 2002 to March 2019.

Once we obtain this intra-day information on bond transactions, we compute a daily price by referring to the “trade-weighted price, all trades” approach of Bessembinder et al. (2009), which consists in performing a value-weighted average of all intraday transaction prices. This methodology puts more emphasis on trades from institutional investors that benefit from lower

execution costs, therefore providing a better accuracy for daily prices. Once we apply this methodology, our daily price dataset consists of 13,770,971 daily prices¹⁵.

In order to obtain monthly prices from daily prices, I develop a methodology similar to that of Bai, Bali and Wen (2019). I first identify the bond price for the last trading day for each bond-month. If the last trading day is one of the last five trading days within month t, then this daily price is used as a monthly price. If that is not the case, I identify the bond price for the first trading day of month t+1. If this trading day is one of the first five trading days of month t+1, it is used as a price for month t. This allows for a more complete dataset of bond monthly prices in order to be able to compute more monthly returns using data from months t and t+1, where returns are either computed using (1) end of month t-1 to end of month t daily prices (2) start of month t to end of month t daily prices (3) start of month t to start of month t+1. Using the methodology, we compute 1,761,543 monthly prices for our dataset.

Similarly to our article on the corporate bond data treatment methodology, the monthly corporate bond return at time t is computed as:

$$r_{i,t} = \frac{(P_{i,t} + AI_{i,t} + Coupon_{i,t}) - (P_{i,t-1} + AI_{i,t-1})}{P_{i,t-1} + AI_{i,t-1}}$$

where $P_{i,t}$ is the transaction price, $AI_{i,t}$ is accrued interest, and $Coupon_{i,t}$ is the coupon payment, if there is any, of bond i for month t. We obtain the necessary information to compute corporate bond returns from the FISD database¹⁶. In order to compute our returns, month prices

¹⁵ In Bessembinder et al (2009), four distinct strategies are used to compute daily prices from TRACE data: using the last price reported in TRACE for the day (“last price, all trades”), construct daily prices by weighing each trade by size (“trade-weighted price, all trades”), eliminating all trades below \$100,000 dollars and using the last available price of the day (“last price, trade \geq 100k”), and eliminating all trades below \$100,000 and weighing each trade by size (“trade-weighted price, trade \geq 100k”). Even though we do apply a “trade-weighted price, all trades” in this paper, we do exclude trades below \$10,000 from our sample following Bai, Bali and Wen (2019).

¹⁶ This includes the coupon rate, the dated date marking the start of the interest period, as well as the interest frequency.

need to be available for adjacent months. Our final dataset is composed of 1,530,745 corporate bond returns. We use this same database of corporate bond returns as a source for the analyses performed in Article 4 and Article 5.

6.4. Other types of data on sustainable finance

Given the importance of defining sustainable finance and the underlying framework in which it operates, many other forms of data were used other than purely quantitative data. In the first paper of this PhD dissertation, extensive work was performed to identify how practitioners, academics and regulators approached the notion of sustainable finance. As we demonstrate in the methodology section of this paper, we looked throughout the practitioners literature for common definitions on the general subject of sustainable investing through the most important reports on the SRI market, taking note of the different terminologies and indicators used by these practitioners. The study of the academic literature took a similar approach, looking for this terminology in academic papers on the subject of ESG and CSR, and even looking further into the more specific literature on CEP. Finally, studying regulation required analyzing what could be considered as more qualitative data on the different steps that are taken by the European Commission to implement sustainable finance throughout the different financial sectors.

Defining and understanding the current framework of sustainable finance was a necessary first step before going into further analysis in this domain. The corresponding work to develop an understanding of the industry and how it was evolving required thorough analyses of these three types of literature and rigorous accounting for the terminologies, concepts and metrics that are used by different actors. These terminologies, concepts and metrics could well be considered as another form of qualitative data that was used in the context of this PhD dissertation.

Once we have developed our different datasets, we apply different methodologies to analyze this data which we describe in the following section.

7. Methodology

In this section, we explore the different methodologies developed throughout the articles of this PhD dissertation. Overall, we can differentiate three types of methodological approaches applied throughout our thesis. The first consists in different methodologies for gathering qualitative data from practitioner, academic and regulatory literature and was applied in the context of our first article. The second consists more in data treatment methodologies applied in order to perform more precise descriptive analyses of different financial markets. These types of methodologies were applied in the context of our second article on capital market financing and our third article on structuring corporate bond data. Since these two articles focus on obtaining and structuring data in order to be able to perform descriptive analyses, we have described them in the data section of this paper (Section 6.1 and 6.2). Finally, the last type of methodology consists in more technical empirical asset pricing methodologies that are performed in the context of our last two articles.

7.1. Structured analysis of the literature on sustainable finance

7.1.1. Analysis of the literature on the SRI Market

In the context of our initial paper, various analyses were performed on the practitioner and academic literature. The first of these analyses consists in understanding how different SRI markets approached the notion of “sustainable investment”. In this context, we read through the reports of the three main actors of the SRI industry which all represented a geographical region.

Eurosif (2018) focuses on Europe, USSIF (2018) focuses on the US and GSIA (2018) on the global market. Realizing that the most present form of commonality between the report was the fact that seven distinct SRI approach had been categorized, we summarized the different information that was available for each approach in the global, European and US markets. This resulted Table 4 which contains global assets under management (AUM) for each approach, the part of these global AUM that were either in Europe or in the US, and possible information on the distribution of criteria specific to each SRI approach.

7.1.2. Analysis of the academic literature on ESG

A similar approach was taken regarding the academic literature that focuses on the concept of ESG, the results of which were previously discussed as they are the starting point for the entire work that was performed in this PhD dissertation. By thoroughly analyzing the three prominent meta-studies on the relationship between ESG and financial performance, we were able to understand how the terminology and approaches of ESG as a concept varied through the literature. Focusing on the most recent and comprehensive meta-study on the subject (Friede, Busch and Bassen, 2015), we analyze the titles of the underlying studies used by the authors to perform their analysis and show that 26 different concepts are used to refer to ESG in only 40 different studies. We expose these different terms along with the corresponding reference in Table 5.

7.1.3. Analysis of the academic literature on CEP

Finally, we push our analysis further and try to understand if more challenges are met by academics that focus on the same specific branch of ESG and study the academic literature on CEP. This allows us to go beyond issues related to terminology and to focus more specifically on

Table 4 - Global SRI approaches and corresponding assets-under-management

SRI Approach	Definition	Global AuM	Europe (% of Global AuM)	Distribution	US (% of Global AuM)	Distribution
Negative/Exclusionary Screening	"the exclusion from a fund or portfolio of certain sectors, companies or practices based on specific ESG criteria"	\$19.77 trillion	55%	Top Exclusion Criterias: - Controversial Weapons (63,6%) - Tobacco (49.1%) - All Weapons (45.7%) - Gambling (34.9%) - Pornography (34.4%) - Nuclear Energy (33.9%) - Alcohol (30.6%) - GMO (24.5%) - Animal testing (19.3%)	40%	ESG Investing ESG Categories by Money Managers 2016-2018: - Social (\$10,8 trillion) - Governance (\$10,8 trillion) - Environmental (\$10,1 trillion) - Product (\$4.5 trillion) Top specific Criteria - Climate Change / Carbon : \$3.00 trillion - Tobacco : \$2.89 trillion - Conflict Risk : \$2.26 trillion - Human Rights : \$2.22 trillion - Transparency and Anti-Corruption : \$2.22
ESG Integration	"the systematic and explicit inclusion by investment managers of environmental, social and governance factors into financial analysis"	\$17.54 trillion	54%	Indicators used by the Eurosif are: (1) the formalisation of an ESG integration policy in the investment process (19%) (2) the number of ESG analysts in the team (61% have between 1 and 5)	28%	ESG Categories by Institutional Investors 2016-2018: - Social (\$5.2 trillion) - Governance (\$3.5 trillion) - Environmental (\$3.5 trillion) - Product (\$2.9 trillion)
Corporate Engagement and Shareholder Action	"the use of shareholder power to influence corporate behavior, including through direct corporate engagement (i.e., communicating with senior management and/or boards of companies), filing or co-filing shareholder proposals, and proxy voting that is guided by comprehensive ESG guidelines."	\$9.83 trillion	56%	- No information on distribution of different criterias. - Mostly performed in the UK (€2.8 trillion), Sweden (€874 billion) and the Netherlands (around €700 billion) which together represent around 90% of the market	18%	Top specific Criteria - Conflict Risk : \$2.97 trillion - Tobacco : \$2.56 trillion - Climate Change / Carbon : \$2.24 trillion - Board Issues : \$1.73 trillion - Transparency and Anti-Corruption : \$1.69 trillion
Norms-based Screening	"screening of investments against minimum standards of business practice based on international norms, such as those issued by the OECD, ILO, UN and UNICEF"	\$4.68 trillion	77%	Applied Norms : - UN Global Compact (42%) - OECD Guidelines for MNC's (25%) - ILO Conventions (26%) - Others (7%)	na	Shareholder Action - In 2018, 165 institutional investors and 54 investment managers controlling \$1.8 trillion in assets filed or co-filed shareholder resolutions on ESG issues
Positive/Best-in-Class Screening	"investment in sectors, companies or projects selected for positive ESG performance relative to industry peers"	\$1.84 trillion	36%	- No information on distribution of different criterias. - Mostly performed in France which represents more than 50% of the market with close to €300 billion out of a total of €586 billion, with Netherlands and Italy representing close to €100 billion and 75€ billion respectively	60%	Leading ESG Issues by number of Shareholder proposals filed: - Proxy Access (≈350) - Corporate Political Activity (≈300) - Environment: Climate Change (≈270) - Labor & Equal Employment Opportunity (≈190) - Executive Pay (≈180) - Human Rights (≈150) - Independent Board Chair (≈150) - Special Meetings (≈120) - Board Diversity (≈90) - Sustainability Reporting (≈90) - Environment: Other issues (≈90)
Sustainability Themed Investing	"investment in themes or assets specifically related to sustainability (for example clean energy, green technology or sustainable agriculture)"	\$1.02 trillion	17%	Very segmented market: - Water management (17%) - Renewable Energy (12%) - Energy Efficiency (11%) - Sustainable transport (11%) - Buildings sector (11%) - Land use/Forestry/Agriculture (11%) - Waste management (9%) - Other (18%)	77%	
Impact/Community Investing	"targeted investments aimed at solving social or environmental problems, and including community investing, where capital is specifically directed to traditionally underserved individuals or communities, as well as financing that is provided to businesses with a clear social or environmental purpose"	\$ 444,26 billion	28%	Eurosif does not provide information on the specific sectors that are addressed by Impact Investing	66%	

Table 5 - ESG-related terminology

Concept used in title	Different terminology used	References
CSR	"Corporate Social Responsibility"	Cochran and Wood (1984), Aupperle, Carroll, and Hatfield (1985), Pava and Krausz (1996), Griffin and Mahon (1997), Richardson, Welker, and Hutchinson (1999), Orlitzky and Benjamin (2001), McWilliams, Siegel, and Wright (2005), Boaventura, Santos da Silva, and Bandeira-de-Mello (2012), Fifka (2013), Schröder (2014), Pavie and Filho (2008)
CSP	"Corporate Social Performance"	Wood and Jones (1995), van Beurden and Gössling (2008), Pelozo (2009), Orlitzky (2001), Orlitzky, Schmidt, and Rynes (2003), Allouche and Laroche (2005), Wu (2006), Margolis, Elfenbein, and Walsh (2009), Vishwanathan (2010)
CEP	"Corporate Environmental Performance" or "Environmental Performance"	Blanco, Rey-Maqueira, and Lozano (2009), Horváthová (2010), Günther, Hoppe, and Endrikat (2011), Schultze and Trommer (2012), Dixon-Fowler et al. (2013), Endrikat, Guenther, and Hoppe (2014)
SRI	"Socially Responsible Investing" or "Sustainable Investing"	Sjostrom (2011), Rathner (2013), Viviers and Eccles (2012), Kleine, Krautbauer, and Weller (2013), Revelli and Viviani (2013), Capelle-Blancard and Monjon (2014), Revelli and Viviani (2015)
Others - Environmental	"Being Green", "Green Management", "Environmental Characteristics", "Environmental Programs", "Environmental Management", "Environmentally Sustainable Supply Chain Management Practices"	Ambec and Lanoie (2007), Molina-Azorin et al. (2009), Rosenbusch, Bausch, and Galander (2007), Darnall and Sides (2008), Albertini (2013),
Others - Social	"Social Responsiveness", "Social Performance", "Social Initiatives", "Social Capital", "Social Investors", "Socially Irresponsible and Illegal Behaviour", "Human Capital", "Social Capital of Entrepreneurs"	Arlow and Gannon (1982), Ullmann (1985), Roman, Hayibor, and Agle (1999), Margolis and Walsh (2003), Westlund and Adam (2010), Derwall, Koedijk, and Horst (2011), Frooman (1997), Crook et al. (2011), Unger et al. (2011), del Mar Miras-Rodríguez et al. (2015), Stam, Arzlanian, and Elfring (2014)
Others	"Sustainability", "Shareholder Activism", "Governance", "High Performance Work Practices", "Knowledge Transfer", "Innovation", "Entrepreneurial Talent"	Salzmann, Ionescu-Somers, and Steger (2005), Gillan and Starks (2007), Love (2010), Clark, Feiner, and Viehs (2015), Combs et al. (2006), van Wijk, Jansen, and Lyles (2008), Rosenbusch, Brinckmann, and Bausch (2011), Rubera and Kirca (2012), Mayer-Haug et al. (2013)

metrics used by these academics. We explore this more specific literature and apply a similar methodology. We categorize the work of these academics and create specific common sections for each study. We look at the source of the data, the type of performances that are analyzed (in this literature, academics differ Environmental Management Performances (EMP) that rely more on general initiatives that are taken by firms from Environmental Operation Performance (EOP) that rely on specific metrics), the final metrics used and the studied samples. Table 6 summarizes this analysis.

7.2. The cross-section of traditional and climate bond returns

Throughout our study on the cross-section of corporate bond returns, we refer to the Newey-West adjusted t-statistic (Newey and West, 1987) to measure and indicate the statistical significance of our results. The Newey-West estimator was developed to overcome autocorrelation and heteroskedasticity in the error terms of models, in most cases for regressions that are applied to time series data. Since this study mostly relies on performing cross-sectional regressions and computing time series averages that are quite likely to have correlated error terms over time, the use of estimators such as the Newey-West estimator is required to demonstrate robust statistical significance. This study on the cross-section of corporate bond returns follows the methodology developed in Bai, Bali and Wen (2019) quite closely, since these authors are the first to explore the possibilities of developing factors specific to corporate bonds to explain the cross-section of corporate bond returns.

7.2.1. Univariate portfolio analysis

The first section of this study consists in examining the relationship that exists between the downside risk factor of Bai, Bali and Wen (2019) and corporate bond returns through portfolio analysis. For each type of portfolio analysis, we perform both a general analysis on our entire dataset of US corporate bonds and perform a more specific application on our dataset of climate-aligned bonds.

In the context of our univariate portfolio analysis, we form quintile portfolios for every month between July 2005 and March 2019 by sorting corporate bonds depending on the values of their downside risk variable for that month. Though our dataset applies from July 2002 onwards, the downside risk factor can only be computed if the corporate bond has existed for at least 36

Table 6 - Corporate Environmental Performance in the literature

Study title	Database	EMP	EOP	Final Metrics used	Sample studied
Xue, B., Zhang, Z. and Li, P., 2020. Corporate environmental performance, environmental management and firm risk. <i>Business Strategy and the Environment</i> , 29(3), pp.1074-1096.	ASSET4 ESG	Equal-weighted average of 41 Yes/No KPIs	Carbon Performance (carbon emissions per net sales)	Total CO2 and CO2 equivalents emissions in tonnes divided by net sales	UK,924 firm year in manufacturing, 708 service industry
Ren, S., He, D., Zhang, T. and Chen, X., 2019. Symbolic reactions or substantive pro-environmental behaviour? An empirical study of corporate environmental performance under the government's environmental subsidy scheme. <i>Business Strategy and the Environment</i> , 28(6), pp.1148-1165.	KLD	Dummy Variable (passed ISO14001 or have not)	Pollutants of three wastes (gas waste, water waste and solid waste)	Equal-weighted dummy variables for SO2, NOX, CO2, soot and dust, COD, ammonia nitrogen, solid waste, and energy consumption	China
Xie, S. and Hayase, K., 2007. Corporate environmental performance evaluation: a measurement model and a new concept. <i>Business Strategy and the Environment</i> , 16(2), pp.148-168.	Survey	48 measurement items for 4 management performance indicators	18 measurement items for 2 operational performance indicators	Answers from the survey of 68 measurement items	Japan, the electrical machinery and instrument manufacturing sector
Anton, W.R.Q., Deltas, G. and Khanna, M., 2004. Incentives for environmental self-regulation and implications for environmental performance. <i>Journal of environmental economics and management</i> , 48(1), pp.632-654.	IRRC from surveys	13 dummy variable EMPs			S&P 500 firms included in the Corporate Environmental Profile Directories
Post, C., Rahman, N. and McQuillen, C., 2015. From board composition to corporate environmental performance through sustainability-themed alliances. <i>Journal of Business Ethics</i> , 130(2), pp.423-435.	KLD	Year-to-year increases in 7 KLD environmental strenghts			US Oil and Gaz companies that were listed in the 2009 Forbes.com Special Report The Global 2000
Trumpp, C., Endrikat, J., Zopf, C. and Guenther, E., 2015. Definition, conceptualization, and measurement of corporate environmental performance: A critical examination of a multidimensional construct. <i>Journal of Business Ethics</i> , 126(2), pp.185-204.	ASSET4 ESG	32 Yes/No KPIs	5 continuous KPIs	Energy consumption in gigajoules, water withdrawal in cubic meters, CO2 and CO2 equivalents emissions in tons, waste produced in tons, hazardous waste produced in tons	Mostly Europe Manufacturing industry and US
Trumpp, C. and Guenther, T., 2017. Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. <i>Business Strategy and the Environment</i> , 26(1), pp.49-68.	ASSET4 for Waste, CDP for GHG emissions	-	GHG emissions and Waste	Negative GHG emissions devided by sales, negative amount of waste produced by a firm divided by sales	International, manufacturing and Service Industries
Escrig-Olmedo, E., Muñoz-Torres, M.J., Fernández-Izquierdo, M.Á. and Rivera-Lirio, J.M., 2017. Measuring corporate environmental performance: A methodology for sustainable development. <i>Business Strategy and the Environment</i> , 26(2), pp.142-162.	ASSET4	62 Indicators taken from ASSET4			Agri-food industry
Hartmann, J. and Vachon, S., 2018. Linking environmental management to environmental performance: The interactive role of industry context. <i>Business Strategy and the Environment</i> , 27(3), pp.359-374.	ASSET4	Carbon Emission Reduction	-	Absolute Carbon Emissions	Publicly traded European manufacturing firms listed in the Asset4 database
Delmas, M.A., Etzion, D. and Nairn-Birch, N., 2013. Triangulating environmental performance: What do corporate social responsibility ratings really capture?. <i>Academy of Management Perspectives</i> , 27(3), pp.255-267.	KLD, Trucost, SAM	Creation of 5 variables: KLD Total Concerns, KLD Total Strenghts, Trucost Total Damage, SAM Eco-efficiency, SAM Reporting			741 firms, no detail on industry or geography
Bhattacharyya, A. and Cummings, L., 2015. Measuring corporate environmental performance–stakeholder engagement evaluation. <i>Business Strategy and the Environment</i> , 24(5), pp.309-325.	CDP Disclosure	GHG Emissions			127 companies, various industries and geographies
Bhattacharyya, A. and Cummings, L., 2015. Measuring corporate environmental performance–stakeholder engagement evaluation. <i>Business Strategy and the Environment</i> , 24(5), pp.309-325.	Survey	36 MPis and 11 OPis			A final sample size of 150 (from chemical, industrial engineering and pharmaceutical & biotech) from Australia

months. Quintile 1 contains bonds with the lowest VaR values, while quintile 5 contains bonds with the highest VaR values. The portfolios are weighted using size as weights.

We then compute average next-month excess returns for each quintile portfolio in order to demonstrate that as the average downside risk of monthly portfolios grows, so does their average next-month excess returns. This allows us to visualize if a positive relationship exists between downside risk and next-month bond returns. To compute excess returns, we subtract the risk-free rate from corporate bond returns. We use the risk-free measure available on Kenneth French's website.

In addition to providing the next-month average excess return for every quintile, we also use control variables to understand if this relationship can be explained by other factors that have previously been developed in the literature. We regress next-month portfolio excess returns on the well-known stock market factors. These are the excess-stock market return (MKTstock), the size factor (SMB), the book-to-market factor (HML), the profitability factor (RMW), the investment factor (CMA), the momentum factor (MOM) and the liquidity risk factor (LIQ), which have been developed and computed in Fama and French (1993), Carhart (1997) and Pastor and Stambaugh (2003).

We also regress the next-month portfolios excess returns on well-known bond market factors: the bond market excess return (MKTbond), the term factor (TERM) and the default factor (DEF). We obtain the default factor (DEF) by computing the difference between the monthly returns of a market portfolio of long-term corporate bonds and monthly long-term government bond returns, and the term factor (TERM) by computing the difference between monthly long-term government bond returns and the risk-free rate.

Finally, we also compute the averages for the market beta, illiquidity, rating, maturity and size for each VaR quintile and observe that some relations might also exist between next-month excess bond returns and some of these variables. We perform similar analysis for our dataset of climate bonds, which allows us to compare our results for the market sample and the climate bond sample.

7.2.2. Bivariate portfolio analysis

Much like for the previous univariate analysis, the bivariate portfolio analysis consists in sorting our bonds monthly according to specific characteristics. However, the process here is a bit different, as we first sort bonds according to either credit rating (Panel A), maturity (Panel B), size (Panel C) or liquidity (Panel D) into five different quintiles. Once this initial process has been performed, we further sort each quintile into five quintiles based on downside risk. We therefore have 25 distinct portfolios per month, and each portfolio has very similar characteristics while downside risk maintains some dispersion. We compute value-weighted average returns for each portfolio. Much like the univariate sort, portfolios are weighted using size as weights. We then compute average returns for every VaR quintile category (portfolios with the highest VaR quintile being VaR quintile 5 and the lowest being VaR quintile 1). We therefore have five portfolios per month corresponding to each VaR category. We perform time series regressions for each of these monthly portfolios using the 10 stock and bond factors previously computed for our univariate portfolio analysis. We refer to the Newey-West adjusted t-statistic for statistical significance.

7.2.3. Fama-Macbeth regressions

Once we have tested the relationship between downside risk and corporate bond returns through portfolio analyses, we perform bond-level analysis through the application of Fama-

Macbeth (1973) regressions. The Fama-MacBeth procedure consists in running a cross-sectional regression every month in order to obtain monthly coefficients, and then perform a time series average to obtain average coefficients for the entire time period considered. Once again, in this procedure, the use of Newey-West adjusted t-statistics is essential to establish statistical significance. Following Bai, Bali and Wen (2019) we perform ten distinct Fama-French regressions, regressing each factor individually and with control variables and then regressing all four factors with and without control variables. Our control variables are bond exposure to both the term factor (β_{TERM}) and the default factor (β_{DEF}), time-to-maturity, size and lagged excess return.

We follow Bai, Bali and Wen (2019) and run the cross-sectional regressions using the following model:

$$R_{i,t+1} = \lambda_{0,t} + \lambda_{1,t}VaR_{i,t} + \lambda_{2,t}Rating_{i,t} + \lambda_{3,t}ILLIQ_{i,t} + \lambda_{4,t}\beta_{i,t}^{Bond} + \sum_{k=1}^K \lambda_{i,k,t}Control_{i,k,t} + \epsilon_{i,t+1}$$

Where $R_{i,t+1}$ is the excess return on bond i in month $t + 1$.

7.2.4. Determining bond-level risk factors

In order to construct our bond factors, we follow Fama and French (2015) and rely on independent sorts. To construct the downside risk factor for corporate bonds, for each month from July 2004 to December 2016, we form bivariate portfolios by independently sorting bonds into five quintiles based on their credit rating and five quintiles based on their downside risk (measured by 5% VaR). The downside risk factor, DRF, is the value-weighted average return difference between the highest-VaR portfolio and the lowest-VaR portfolio across the rating portfolios. The credit risk factor, CRF VaR, is the value-weighted average return difference between the lowest-

rating (i.e., highest credit risk) portfolio and the highest-rating (i.e., lowest credit risk) portfolio across the VaR portfolios. The liquidity risk and the return reversal factors are constructed similarly using independent sorts. The liquidity risk factor, LRF, is the value-weighted average return difference between the highest-illiquidity and the lowest- illiquidity portfolios across the rating portfolios. The return reversal factor, REV, is the value-weighted average return difference between the short-term loser and the short- term winner portfolios (losers-minus-winners) across the rating portfolios. These bond-level factors are then used on a set of size and industry-sorted portfolios to determine their efficiency, as well as a portfolio composed only of climate bonds.

7.3. Machine learning and asset pricing for traditional and climate-aligned firms

Throughout our study on the application of machine learning methods to the prediction of the cross-section of corporate bond returns, we follow the methodology developed in Gu, Kelly and Xiu (2020). In their work, the authors compare the performance of ML methods in predicting stock returns. This includes (1) linear models, including the traditional ordinary least squares (OLS); (2) generalized linear models that apply penalization methods such as elastic net, LASSO¹⁷ and ridge regressions; (3) methods that apply dimension reduction techniques such as principal component regressions and partial least squares; (4) regressions trees and (5) neural networks.

For each type of model, Gu, Kelly and Xiu (2020) provide a statistical model describing how the method is adapted to risk premium predictions, an objective function to estimates the parameters of the model and the computational algorithms corresponding to the model. The authors' objective for each model is to minimize the mean square predictions error (MSE). In this study, the authors describe an asset's excess return as an additive prediction error model:

¹⁷ LASSO stands for least absolute shrinkage and selection operator

$$r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1}$$

where

$$E_t(r_{i,t+1}) = g^*(z_{i,t})$$

And stocks are indexed as $i = 1, \dots, N_t$ and months by $t = 1, \dots, T$.

With this approach, the authors attempt to isolate $E_t(r_{i,t+1})$ as a function of variables that maximizes the out-of-sample explanatory power for $r_{i,t+1}$. $z_{i,t}$ represents the P -dimensional vector of variables and $g^*(\cdot)$ is a flexible function of these variables.

This approach does have important limitations. $g^*(\cdot)$ does not depend on i or t . Contrarily to standard asset pricing approaches that re-estimate a time-series model for each asset or a cross-section model for each time-period, this model maintains the same form through time and across assets, providing more stability with regards to the risk premium estimates of assets. Furthermore, $g^*(\cdot)$ depends only on z through $z_{i,t}$, meaning the model's predictions do not use any information from historical data prior to t or any other asset than the i th.

As we continue to follow the methodology of Gu, Kelly and Xiu (2020) every model description in this study can be found in the authors' work, with the following exceptions: our linear models rely on a standard least squares objective function without its robust extension and we perform our generalized linear model without performing a group lasso¹⁸.

¹⁸ In Gu, Kelly and Xiu (2020), the linear model description is performed in section 1.2, the description of the statistical model for the elastic net is performed in section 1.3, PCR and PLS methods are described in section 1.4 and the generalized linear model is explained in section 1.5.

7.3.1. Sample splitting and tuning via cross-validation

The application of ML methods requires performing specific preliminary steps. In order to be able to perform regularization - the central tool applied in the context of ML methods to minimize overfitting - a choice regarding hyperparameters needs to be made. Hyperparameters control the complexity of models, and for each ML method applied to data, a choice of hyperparameter¹⁹ will lead to the best possible result from the model in out-of-sample data. In order to determine the best value for a model's hyperparameter, the data sample is divided in three different time periods maintaining similar temporal ordering. The first sample, the "training" sample, is used to estimate the model that will be subject the hyperparameters. The second sample, the "validation" sample, is used to determine the optimal value for the hyperparameter. In order to do that, forecasted datapoints are determined in the validation sample using the model applied to the training sample, and the optimal hyperparameter is found iteratively based on forecasts errors in the validation set. It is important to note here that the model used in the training data needs to be re-estimated every time a new hyperparameter is tested. The idea of the validation sub-sample is to stimulate a form of out-of-sample test of the model and find the hyperparameter values that correspond to the optimal level of complexity of the model that produces the best results in the validation sample. Once the optimal model, hyperparameter and model complexity have been determined, the "testing" subsample, which is truly out-of-sample because it has not been used for either estimation or tuning, is used to evaluate the out-of-sample predictive performance of a method.

There are a few differences between our approach and Gu, Kelly and Xiu (2020) that we can benefit from, and which impact our approach to sample splitting and tuning via validation. First

¹⁹ Main examples of hyperparameters include penalization parameters for elastic net, the number of trees in boosting or the number of random trees in a forest

and foremost, in the context of our approach, we are more limited in terms of computational power, and therefore limit the ML methods we apply to the less computationally intensive. The methods we apply throughout this paper are linear-based : they are the standard ordinary least square (OLS) method, using either the four factor of Bai, Bali and Wen (2020) or the our entire set of factors, the Elastic Net method, the Principal Component Regression (PCR), the Partial Least Square (PLS) method and the Generalized Linear Model (GLM) method.

Furthermore, the period we study for our sample is smaller than that of Gu, Kelly and Xiu (2020). The authors study stock prices from 1956 to 2016 when our sample spans from 2005 to 2019. This provides us with another opportunity which consists in fitting our models monthly instead of yearly. This is even more interesting because our variables are computed monthly and therefore provide new monthly information that can be used by our different models.

The last central difference between our study and Gu, Kelly and Xiu (2020) consists in the fact that we apply a cross-validation technique that maintains the temporal ordering of our data. Given the fact that some of our variables require 36 months of previous monthly data to be computed, and that data from TRACE is first made available in June 2002, our models run on data from June 2005 to March 2019. Our approach then consists in the following steps:

- We use a 5 years / 60 months period for our training sample.
- We require 20 further months for our validation sample.
- We test our models on the month following the last month of the validation sample

Each month in our validation sample is used to fit a model using a different hyperparameter. This gives our ML methods 20 different samples to try to fit the best possible model. Every time

a month in our validation sample is used to fit a model using a new hyperparameter, we then integrate this month in our training sample while maintaining a training sample of 60 months. We then use the following month as our validation sample. This method allows us to perform cross-validation while maintaining the temporal ordering of the data. Since our dataset starts on June 2005 and we maintain a 60-month training sample and a 20 months validation sample, our first test month is on March 2012. We therefore have 7 years (84 months) of different test results for our different models.

Once we have applied our different methodologies, we obtain an array of different results for each type of methodology which we outline in the following section.

8. Results

Given the wide range of studies that have been performed in the context of this doctoral thesis, our results also vary and provide different forms of insight on issues relative to sustainable finance. From the analysis of the practitioner, academic, and regulatory literature we obtain insight on the current state of the definition of a common sustainable finance framework. From our study on capital market financing, we understand how both stock and corporate bond markets participate in financing firms, both large and climate aligned. Finally, our approach to data and the pricing of corporate bond data provides interesting results to better understand traditional and climate bond products, as well as the different methodologies required to obtain these results.

8.1. Results from the literature on SRI, ESG and CEP

Our results for this initial analysis is three-fold. Our analysis of the practitioner literature on the SRI market, which can be found in Table 4, leads to the conclusion that there are significant

differences in terms of approaches between the US and European market. Reports on these markets, though performed by similar entities, offers very few common approaches in the concepts and terminology used throughout the reports. Some important SRI investment strategies are clearly defined. Investors do not seem to have similar priorities. The global investor report offers very little information on other geographic regions other than the US and the EU, and no information on the Chinese market which could be expected to be quite large.

Looking into the academic literature on ESG and financial performance offers similar insights, the most troubling one being that no common definition or approach to ESG seem to exist throughout the literature. Major studies simply seem to group the entire terminology of concepts that refer to the extra-financial aspects of firms to refer to a general type of criteria referred to as “ESG”, “CSR” or “CSP”. This is even more troubling given the fact that practitioners of the SRI market refer to ESG as a central concept for the whole of the SRI market. Table 5 shows how diverse the terminology used in the main studies that focus on the concept of ESG/CSP are.

Finally, we focus more specifically on the literature environmental performance and notice that this lack of common definitions and approaches is maintained. Though some concepts such as EMP and EOP are common throughout the literature, academics face an important issue on the commonality of metrics and samples that are used to perform studies. This is an important finding, since it marks a central problem in the general sustainable finance industry: there is strictly no consensus of what defines a sustainable company and sustainable finance as a whole. Those results lead us to put more trust in the CBI’s approach to defining climate-aligned firms, which does depend on the definition on ESG and on subjectively selected ESG metrics.

8.2. Results on the capital market financing of traditional and climate-aligned firms

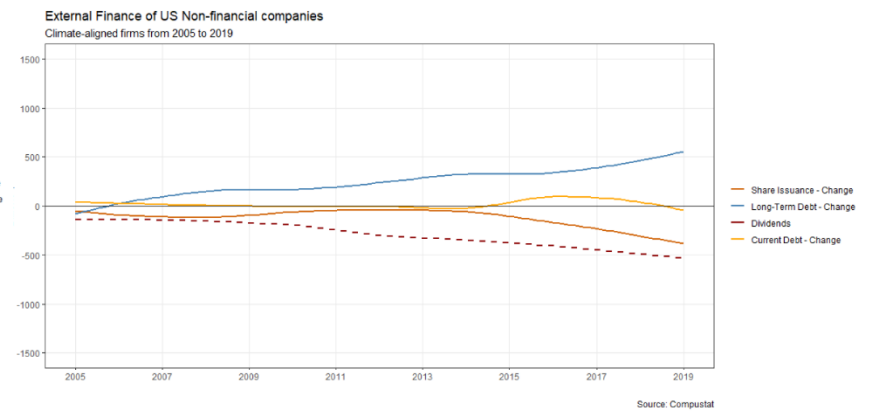
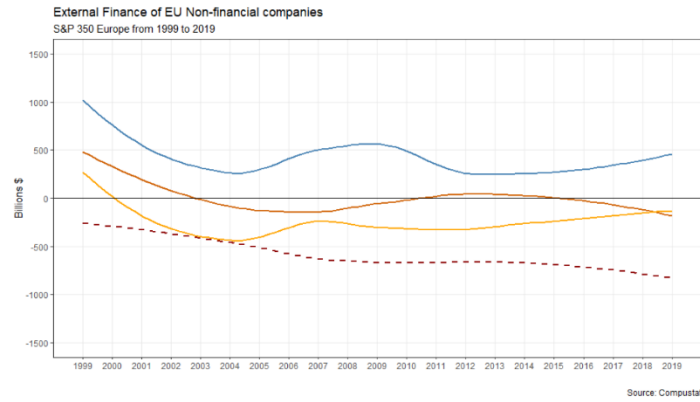
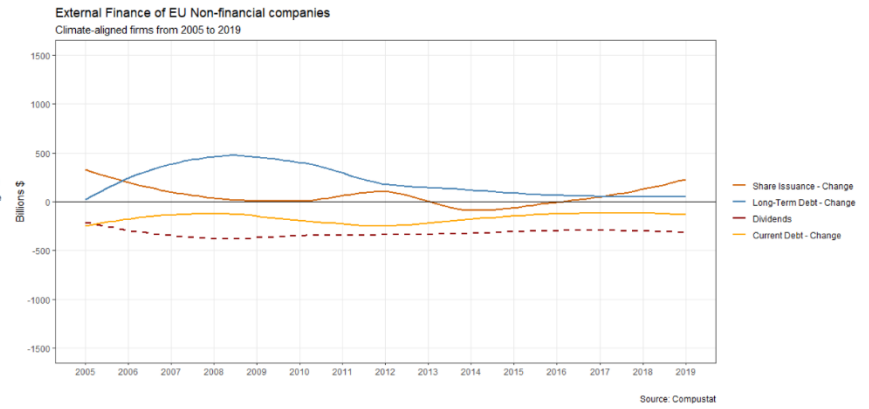
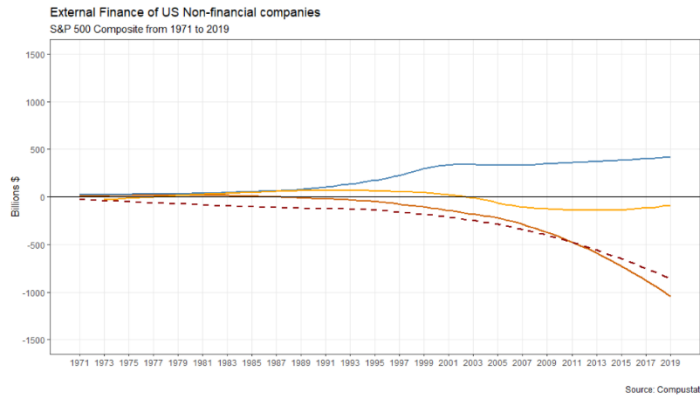
8.2.1. Negative stock market financing, positive long-term debt financing

After having applied the data structuring methodologies described in section 6.1 and 6.2, we are able to perform a thorough descriptive analysis of the relationship between capital markets and traditional and climate-aligned firms. These results are best visualized in Figure 1, though a much deeper analysis is performed in our article on the subject in Chapter 3.

Looking more precisely at trends for each sample, we observe that long-term debt has been the only positive source of financing for S&P500 firms since 2002, stabilizing at around \$400 billion per firm. Much like we have observed in previous sections, negative stock market financing for large US firms has been growing steadily since the beginning of the 1990s. The situation for large EU firms is quite different. Since the starting period of our sample in 2005, though long-term debt has been by far the largest source of financing, it has never strayed too far from changes in share issuances, which were in positive territory between 1999 and 2003 and between 2011 and 2015. Dividends, however, have always been almost proportionally negative since the beginning of our sample period.

For climate-aligned firms in the US, our graphical representation illustrates a slow but constant growth of long-term debt financing since the beginning of our sample, mirrored by a slow but constant growth of distribution of dividends. Net share issuances, though being constantly negative since the beginning of our sample, have mostly decreased recently, between 2014 and 2019. Finally, the situation for climate-aligned Euro firms is also quite specific, with a larger participation of long-term debt financing between 2007 and 2011, after which long-term debt financing starts a slow but quite steady downward stream until the end of our sample period. Net share issuances alternate between slightly positive and slightly negative territory throughout the sample period. Overall, we do find interesting patterns in these graphical representations: while

Figure 1 - External Financing of US and EU traditional and climate-aligned firms



debt financing is the only constantly positive source of financing for all our sample, dividends always closely mirror long-term debt within negative territory. We also find that the behavior of net share issuances differ mostly between our two geographical regions: net share issuances are constantly negative in the US and alter between positive and negative territory in the Eurozone.

8.2.2. Corporate bonds as a major source of financing for traditional and climate-aligned firms

Using the methodology developed in the context of section 6.3, we can extract corporate bond issuances and repayments for every US firms that has issued a corporate bond between 1995 and 2018. Using the Compustat US database, we can also extract total long-term debt issuances and repayments of these firms. This allows us to understand what part of these firms’ long-term debt originates from capital markets through the US corporate bond market. This approach also applies to US climate-aligned firms.

Using this approach, we compute the total amounts of long-term debt and corporate bond issuances and repayments for both our sample of traditional firms and climate-aligned firms. Our results are available in Table 7.

Table 7 - Long-term debt and corporate bond financing (in billions)

	All corporate bond issuers (1995-2018)		Climate-Aligned Firms (2005-2018)	
Net Corporate Bond Debt	2956.99	59%	50.24	66%
Other Net Long-term Debt	2079.31	41%	26.23	34%
Total Net Long-Term Debt	5036.30		76.47	

Both yearly long-term debt issuances and long-term debt repayments are vastly superior to yearly corporate bond issuances and repayments. However, when computing the differences between issuances and repayments for total long-term debt and total corporate bond debt, we find

that net corporate bond debt represents a majority of the total net long-term debt of US firms in our sample. With total net long-term debt on the period 1995 to 2018 reaching \$5.036 trillion for firms in the FISD database, \$2.957 trillion of this amount is corporate bond debt, which represents 59% of this total. This is quite interesting provided the fact that we observed in the previous section that long-term debt was already by far the primary source of financing for US firms in the S&P 500 by 1995. The importance of corporate bond financing for climate-aligned bonds is even pronounced, as it represents 66% of total net long-term debt between 2005 and 2018.

8.3. A better understanding of climate bond and green bond data availability

Even though our study on corporate bond data focuses on trying to obtain a database of corporate bond returns in order to be able to apply modern asset pricing models to cross-section of corporate bond returns, our analysis of this data allows to understand, with numbers, how small the US green bond market is compared to the US climate bond market, which itself only represents a fraction of the entire US corporate bond market. In Table 8, we show the trading volumes from 2014 to 2018 of the US green bond and climate bond markets.

Table 8 - Trading Volume of Green Bond and Climate Bond Markets

Year	Green Bonds			Climate Bonds		
	Trading Volume (Millions)	Corporate Bond Outstanding Amount (Millions)	Trading Volume Relative to Amount Outstanding	Trading Volume (Millions)	Corporate Bond Outstanding Amount (Millions)	Trading Volume Relative to Amount Outstanding
2014	1,172	750	1.56	44,463	14,756	3.01
2015	3,709	5,050	0.73	47,998	13,924	3.45
2016	10,057	7,075	1.42	44,484	14,732	3.02
2017	14,737	10,400	1.42	49,625	16,359	3.03
2018	13,556	3,875	3.50	76,254	21,340	3.57

This table shows how even in 2017 when trading volumes for green bonds were at their peak over the 2014-2018 period, they were still less than three times inferior to that of climate bonds. In 2018, trading volumes for green bond decreased while those for climate bonds greatly increased, and climate bond volumes were more than 5 times those of green bonds. Given the fact that our database relies mainly on trade-based data, this provides some form of evidence that a thorough study of the green bond market is not yet possible as of 2018.

8.4. Insightful results from the cross-section of traditional and climate bond returns

8.4.1. Univariate portfolio results

Results of our univariate portfolio analysis can be found in Table 4, which shows the average VaR for each quintile as well as average next-month excess returns. This allows to illustrate the relationship that exists between downside risk and future returns. As the average VaR grows from 0.82% for quintile 1 to 4.77% for quintile 5, we observe a similar phenomenon for next-month average bond excess returns which grow from 0.11% to 0.37%. The difference between high VaR portfolios and low VaR portfolios is both economically and statistically significant with an average next-month excess return difference of 0.26% and t-statistic of 5.10.

We also obtain economically and statistically significant stock alphas, which grow from an average of 0.10% for the lowest VaR quintile to 0.34% for the highest VaR quintile, with an economically and statistically significant difference of 0.25% in terms of return difference and a t-statistic of 4.64. We also regress the next-month portfolios excess returns on well-known bond market factors. Much like for our stock market factors, we obtain statistically significant alphas that grow as the downside risk grows. Finally, we also perform a regression of the next-month portfolio excess returns on all our factors. Our results are economically and statistically significant

when regressing on our bond factors and on all our factors combined, with a difference in returns of 0.23% for both type of regressions, and t-statistics of 4.43 and 4.40 respectively.

Looking at average market beta, illiquidity, rating, maturity and size for each VaR quintile, we observe that some relations might also exist between next-month excess bond returns and some of these variables.

Average bond market beta grows from an average of 0.34 for VaR 1 to 1.27 for VaR 5, illiquidity goes from an average of 0.15 to 0.78, and time-to-maturity grows from an average of 3.21 years to 15.32 years with VaR. The existence of such relation is less clear regarding size, as growth of average size for VaR quintiles is not constant, even though size does grow from 370.48 million in average for VaR 1 to 392.07 million for VaR 5. Credit rating diminishes²³ as value-at-risk grows for the market, with credit ratings going from 6.90 (which approximately corresponds to an A-/A3 rating) for the lowest VaR quintile to an 8.99 (which approximately corresponds to a BBB/Baa rating) for highest Var quintile.

Comparing our results for the market sample with the results for the climate bond sample, we do observe some differences in terms of statistical significance, as none of the differences between Var 1 and Var 5 in terms of next-month excess return or alphas have a statistical significance at the 1% level. More specifically, the bond factor alpha is strictly not statistically significant, even though the difference in average alphas for cross-sections with all factors alphas is statistically significant at the 10% level. This lower statistical significance could be attributed to the fact that there are much less bond observations in our climate bond sample, which could lead to less significant results. Economic significance in returns and alphas between average VaR 1 quintile portfolios and VaR 5 quintile portfolios is lower for climate bonds with 0.19% in average excess return difference and 0.16% when controlling for the ten stock and bond market factors. This is

not surprising given the fact that climate bonds in our sample have both lower average returns and lower downside risk than the market sample. Finally, average portfolio characteristics for our **Table 9 – Univariate Portfolio Analysis** climate bond sample behave in a similar fashion than for the market sample with regards to bond market beta, illiquidity, maturity and size, while the relationship between VaR, excess return and credit rating seems to disappear as credit rating for VaR 1 and for Var 5 are similar (8.40 and 8.38 respectively). This could in part be explained by the lack of variation in credit ratings within our climate bond sample.

Panel A - Corporate Bond market univariate portfolios										
Quintiles	Average VaR	Average return	Stock Factors alpha	Bond Factors alpha	All Factors alpha	Average Portfolio Characteristics				
						Bbond	ILLIQ	Rating	Maturity	Size
Low VaR	0.82	0.11	0.10	0.09	0.09	0.34	0.15	6.90	3.21	370.48
		3.06	3.06	3.22	2.27					
2	1.50	0.18	0.16	0.14	0.14	0.54	0.30	7.29	4.74	414.49
		3.45	2.79	3.12	3.23					
3	2.21	0.23	0.21	0.19	0.19	0.74	0.47	7.92	7.36	390.86
		3.67	3.13	3.52	3.37					
4	3.11	0.28	0.26	0.22	0.23	0.96	0.63	8.10	11.54	397.51
		3.97	3.43	3.99	3.56					
High Var	4.77	0.37	0.34	0.32	0.32	1.27	0.78	8.99	15.32	392.07
		4.91	4.28	4.73	4.43					
High-Low	3.95***	0.26***	0.25***	0.23***	0.23***					
	22.08	5.10	4.64	4.43	4.40					

Panel B - Climate Bond univariate portfolios										
Quintiles	Average VaR	Average return	Stock Factors alpha	Bond Factors alpha	All Factors alpha	Average Portfolio Characteristics				
						Bbond	ILLIQ	Rating	Maturity	Size
Low VaR	0.90	0.15	0.13	0.12	0.12	0.32	0.18	8.40	5.32	367.44
		2.85	3.10	2.59	2.85					
2	1.60	0.21	0.18	0.15	0.15	0.55	0.41	8.78	6.26	404.60
		3.68	3.22	2.52	2.19					
3	2.20	0.24	0.22	0.21	0.21	0.72	0.48	8.75	8.43	506.50
		3.78	3.46	2.74	2.79					
4	2.93	0.28	0.25	0.21	0.21	0.92	0.61	8.54	14.89	482.19
		2.80	2.65	2.61	2.52					
High Var	4.41	0.34	0.31	0.25	0.28	1.26	0.81	8.38	21.00	429.19
		3.66	3.04	2.29	2.49					
High-Low	3.52***	0.19**	0.18**	0.13	0.16*					
	15.69	2.53	2.23	1.52	1.86					

Following Bai, Bali and Wen (2019), these apparent interactions between downside risk and other variables leads us to test whether the relationship between downside risk and returns holds when controlling for market beta, credit rating, maturity, liquidity and size using both a bivariate portfolio analysis and Fama-Macbeth (1973) bond-level regressions.

8.4.2. Bivariate portfolio results

We expose our result for our bivariate portfolio analysis in Table 10. As we can see from our results, the relationship between downside risk and next month excess returns holds for every characteristic, both for our market sample and our climate bond sample after controlling for the ten stock and bond market factors. Looking at the market sample, difference between VaR 5 portfolios and VaR 1 portfolios are both economically and statistically significant, with differences of 0.19% when controlling for credit rating, 0.15% when controlling for maturity, 0.18% when controlling for size and 0.15% when controlling for liquidity, with each result having a statistical significance at the 1% level. Results for the climate bond sample only vary slightly, with lower economic significance for differences in excess returns when controlling for credit rating (0.13%) and maturity (0.07%) and overall lower statistical significance, as the difference in excess returns when controlling for size is the only one that has a statistical significance at the 1% level.

8.4.3. Bond-level Fama-MacBeth regressions

In table 11, we expose our results for the Fama-Macbeth regression perform in the context of our study of the cross-section of corporate bonds.

In regression (1), the average slope coefficients of monthly regressions of excess bond returns on VaR without control variables is 0.078 with a t-statistic of 4.90. The economic significance of this effect resembles that of the univariate portfolio analysis in which we found a spread of 3.95

for our VaR values between low Var and high Var portfolios. Multiplying this spread by the slope coefficient, we obtain a monthly downside risk premium of 30 basis points. Slope coefficients for regressions (3), (5) and (7), which correspond to the regressions of excess bond returns on credit rating, illiquidity and market beta individually and without control variables, are also positive and statistically significant. This also corresponds to our results for the univariate portfolio analysis, in which we can observe a positive difference between low Var quintiles and high Var quintiles in terms of average credit rating, illiquidity and bond market beta.

Table 10 – Bivariate Portfolio Analysis

VaR quintile	Panel A: Controlling for credit rating		Panel B: Controlling for maturity		Panel C: Controlling for size		Panel D: Controlling for liquidity	
	Market	Climate Bonds	Market	Climate Bonds	Market	Climate Bonds	Market	Climate Bonds
1	0.10	0.11	0.13	0.14	0.13	0.11	0.13	0.11
	<i>2.56</i>	<i>2.32</i>	<i>3.62</i>	<i>2.63</i>	<i>3.80</i>	<i>2.90</i>	<i>3.41</i>	<i>2.99</i>
2	0.14	0.14	0.15	0.14	0.16	0.16	0.15	0.15
	<i>3.41</i>	<i>2.78</i>	<i>2.91</i>	<i>2.30</i>	<i>4.43</i>	<i>3.11</i>	<i>3.61</i>	<i>2.60</i>
3	0.19	0.22	0.19	0.22	0.20	0.14	0.18	0.21
	<i>3.39</i>	<i>2.88</i>	<i>3.28</i>	<i>3.39</i>	<i>4.73</i>	<i>2.42</i>	<i>3.81</i>	<i>2.85</i>
4	0.24	0.20	0.21	0.19	0.24	0.22	0.22	0.19
	<i>3.54</i>	<i>2.34</i>	<i>3.80</i>	<i>3.07</i>	<i>4.68</i>	<i>2.53</i>	<i>3.66</i>	<i>2.45</i>
5	0.29	0.24	0.29	0.20	0.31	0.27	0.28	0.26
	<i>3.87</i>	<i>2.63</i>	<i>4.83</i>	<i>2.87</i>	<i>5.23</i>	<i>3.01</i>	<i>4.42</i>	<i>2.60</i>
VaR5 - VaR1	0.19	0.13	0.15	0.07	0.18	0.17	0.15	0.16
	<i>3.53</i>	<i>1.96</i>	<i>6.28</i>	<i>1.85</i>	<i>4.37</i>	<i>2.68</i>	<i>4.41</i>	<i>1.87</i>

Performing these same regressions but controlling for beta term, beta default, maturity, size and lagged excess returns does not change our results. Controlling for these characteristics maintains significantly positive and significant results for all our risk variables and does not affect

their positive relation with next-month bond returns. Surprisingly, however, controlling for these factors even enhances our results for regressions (2), (4) and (8).

Our results diverge slightly for Bai, Bali and Wen (2019) when all our variables are regressed together. In regression (9), though our VaR measure is still positive and statistically significant, our liquidity measure is neither economically nor statistically significant, but our credit rating measure is. However, when controlling for our different bond characteristics in regression (10), economical and statistical significance for credit rating disappears and illiquidity becomes positive and statistically significant. Much like Bai, Bali and Wen (2019), statistical significance for bond market beta disappears for both regression (9) and (10), and the return reversal control variable is very significant statistically and has a negative relation with next-month excess return.

Finally, we observe that many of our control variables are statistically significant when regressing variables independently. More specifically, bond exposure to the default factor is always both positive and statistically significant at the 1% level. This is also the case for bond exposure to the term factor for regressions (4) and (6). Regarding our climate bond sample (Panel B), our bond-level Fama-Macbeth regression analysis is limited by the lack of monthly data available for our climate bond sample²⁴, and we are only able to perform individual regressions without control variables.

This is sufficient however to detect positive and statistically significant slope coefficients for regression of one-month ahead excess return on VaR (1), illiquidity (3) and bond market beta (4) and that no such results are obtained for the regression of one-month ahead excess returns on credit rating (3). These results correspond to results obtained in our univariate portfolio analysis for the climate bond sample, where we established a positive statistically significant difference in terms of VaR between VaR 5 portfolios and VaR 1 portfolios, but also observed such positive

differences for average quintile liquidity and average bond market beta, while this relation did not exist for our average credit ratings.

Table 11 – Bond-level Fama-MacBeth Analysis

Panel A - Bond-level Fama-MacBeth cross-sectional regressions (Market)											
	Intercept	5% VaR	Rating	ILLIQ	BBond	BDEF	BTERM	Maturity	Size	REV	Adj. R2
(1)	0.446	0.078									0.025
	<i>60.516</i>	<i>4.899</i>									
(2)	0.363	0.080				0.322	0.079	0.016	-0.017	-0.185	0.158
	<i>2.408</i>	<i>6.592</i>				<i>4.621</i>	<i>0.524</i>	<i>0.647</i>	<i>-1.630</i>	<i>-28.493</i>	
(3)	0.484		0.034								0.022
	<i>61.602</i>		<i>3.535</i>								
(4)	-0.026		0.071			0.594	0.558	0.087	-0.037	-0.191	0.165
	<i>-0.161</i>		<i>5.488</i>			<i>4.657</i>	<i>2.733</i>	<i>3.057</i>	<i>-3.143</i>	<i>-28.118</i>	
(5)	0.484			0.026							0.018
	<i>81.291</i>			<i>3.747</i>							
(6)	-0.239			0.021		1.032	0.542	0.082	0.017	-0.164	0.176
	<i>-1.282</i>			<i>3.868</i>		<i>5.438</i>	<i>2.259</i>	<i>2.604</i>	<i>1.131</i>	<i>-30.171</i>	
(7)	0.472				0.018						0.035
	<i>92.110</i>				<i>5.808</i>						
(8)	0.399				0.068	0.289	0.064	0.067	-0.033	-0.188	0.158
	<i>2.794</i>				<i>4.782</i>	<i>2.776</i>	<i>0.309</i>	<i>2.356</i>	<i>-2.900</i>	<i>-28.801</i>	
(9)	0.424	0.057	0.045	0.009	0.029						0.112
	<i>56.756</i>	<i>6.299</i>	<i>2.747</i>	<i>1.471</i>	<i>1.303</i>						
(10)	0.252	0.062	0.010	0.056	0.025	0.223	0.221	0.000	0.000	-0.174	0.214
	<i>1.678</i>	<i>7.908</i>	<i>1.842</i>	<i>3.830</i>	<i>0.958</i>	<i>1.467</i>	<i>1.423</i>	<i>0.004</i>	<i>0.025</i>	<i>-46.298</i>	

Panel B - Bond-level Fama-MacBeth cross-sectional regressions (Climate Bonds)											
	Intercept	5% VaR	Rating	ILLIQ	BBond	BDEF	BTERM	Maturity	Size	REV	Adj. R2
(1)	0.448	0.072									0.083
	<i>56.660</i>	<i>2.404</i>									
(2)	0.478		0.039								0.020
	<i>40.930</i>		<i>1.329</i>								
(3)	0.484			0.026							0.046
	<i>71.040</i>			<i>2.386</i>							
(4)	0.464				0.026						0.094
	<i>21.020</i>				<i>2.495</i>						

Much like it was the case for the univariate portfolio analysis, this result could well be due to the lack of variation in terms of credit ratings in our climate bond sample. Furthermore, by multiplying our slope coefficient of 0.072 for regression (1) by the difference in downside risk in our univariate

portfolio analysis of 3.52%, we obtain a monthly down-side risk premium for climate bonds of 27,5 basis points, slightly lower than our results for the market sample.

8.4.4. Testing the factors on Size/maturity-sorted and industry-sorted portfolios for both traditional and climate bonds

Continuing to follow Bai, Bali and Wen (2019), we apply the authors' most performant model – the four-factor model composed of the excess bond market return (MKTBond), the downside risk factor (DRF), the credit risk factor (CRF) and the bond specific liquidity factor (LRF) – to both size/maturity-sorted portfolios and industry-sorted portfolio. Though our results differ from the authors', the four-factor model, when applied to our dataset, results in an average adjusted R-square of 0.49 for size/maturity-sorted portfolios and 0.50 for industry-sorted portfolios, and outperforms other models proposed by Bai, Bali and Wen (2019).

However, results for the four-factor model do differ in our case, since we obtain a positive and statistically significant average alpha of 0.15 with a t-statistic of 3.99 for size/maturity-sorted portfolios of the market sample. In addition, model performance, when using adjusted R-square as a measure, seems to increase as maturity increases. This leads the four-factor model to obtain an adjusted R-square of 0.91 for the portfolio containing the largest and most mature bonds (size quintile 5 and maturity quintile 5). This pattern is also recognizable in our size/maturity-sorted portfolio constructed from the climate bond sample. Looking at industry sorted portfolios, we notice some industries have much higher adjusted R-squared than others. As an example, the “Electric” and “Transportation” industries having adjusted R-squared of 0.76 and 0.74 respectively, while industries such as “Water” and “Savings and Loans” have adjusted R-squared of 0.26. Similar differences can be observed in our climate bond sample. Results for our

size/maturity-sorted portfolios and industry-sorted portfolios are available in our original article in Chapter 5.

8.5. The irrelevance of the climate factor for machine learning methods

8.5.1. Models without Climate dummy variables

We test each of our machine learning methods for every month from March 2012 to February 2019. Our results vary between months and ML methods and are summarized in Table 7. Monthly out-of-sample R-squared are represented in Figure 1.

Table 12 – Machine Learning Models Results without Climate dummy variable

Month	OLS4 - Roos	OLS - Roos	Elastic Net - Roos	PCR - Roos	PLS - Roos	Glm - Roos
Min	-0.517	-0.782	-0.187	-0.110	-0.130	-0.189
Median	-0.009	-0.065	-0.001	-0.001	-0.004	-0.065
Mean	-0.017	-0.085	-0.003	0.003	-0.003	-0.085
Max	0.290	0.419	0.164	0.087	0.093	0.419
Standard Deviation	0.097	0.134	0.064	0.032	0.065	0.134

Our results demonstrate that our models are unable to constantly obtain positive out-of-sample R-squared using the developed predictors. Though all models obtain both negative mean and median values over the sample period, some perform better than others. Both traditional linear models that used all variables – the Ordinary Least Square approach and generalized linear model approach - have the worst performance, with a mean out-of-sample R-square of -8.5% and a medium out-of-sample R-squared of -6.5%. These models also have the highest standard deviation for their performance with standard deviation of 0.134 each.

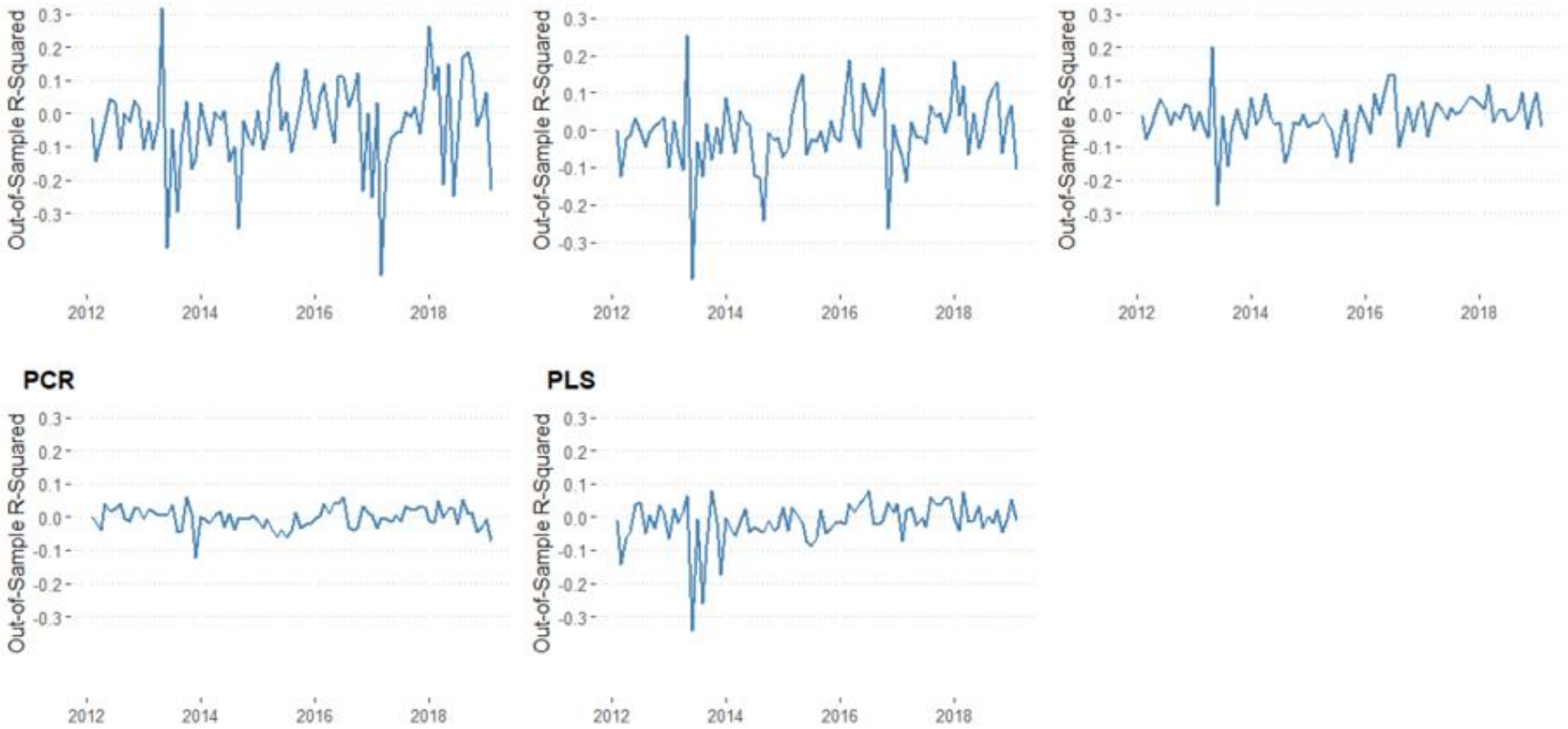
The OLS4 model, which represents the Ordinary Least Square approach using only the four factors of Bai, Bali and Wen (2019), has a much better performance, with a mean out-of-sample R-squared of -1.7% and a median out-of-sample R-squared of -0.9%. Though Bai, Bali and Wen (2019) do

argue that their four- factor model outperform all other models to explain the cross-section of corporate returns, it seems this is still not sufficient to create a model that performs positively in predicting next-month bond returns.

Finally, as it could be expected, ML methods that use shrinkage methods – such as the elastic net – and dimension reduction techniques – such as the principal component regression (PCR) or partial least square (PLS) method – perform better than standard linear models. The elastic net method has a mean out-of-sample R-Square of -0.3% and median out-of-sample R-squared of -0.1%, while the PLS approach has a similar mean out-of-sample R-Square of -0.3% and a slightly lower median out-of-sample R-squared of -0.4%. The PCR method is the only method to have positive mean out-of-sample R-squared of 0.3%, though median out-of-sample R-squared values are slightly in negative territory with -0.1%. The PCR method is also the one with the lowest standard deviation (0.032) compared to the other two methods (0.064 for elastic net and 0.065 for PLS). The fact that the PCR approach outperforms other approach could be explained by the fact that many of our price-trend variables are correlated and dimension reduction technique are best for dealing with correlated variables.

The volatility of monthly out-of-sample R-squared for each method can be visualized in Figure 1. Through these figures, we understand how each test-month showcases different performances for each type of method, and that some test months have quite high yet isolated out-of-sample R-squared. As an interesting example, we can observe that for the test month of May 2013, the standard OLS regression using every variable results in an out-of-sample R-squared of 30%, but then results in a very low out-of-sample R-squared of -40.6% for the next test-month of June 2013. From these representations, we can also notice how standard deviation differs between

Figure 2 - Variation of out-of-sample R-Squared through time



our models, with Elastic Net, PCR and PLS methods demonstrating much less variation and having out-of-sample R-squared performances varying at a much closer distance from 0.

8.5.2. Variable importance for models without climate dummy variable

While the traditional linear models that use none of the machine learning methods such as cross-validation, shrinkage or dimension reduction use all of the variables they are provided, Elastic Net, PCR and PLS machine learning methods use specific sets of predictors to obtain the best performing models. As previously mentioned, the Elastic Net method will in most cases use only a subset of provided predictors to create its best models and use the predictors they find to be the most impactful. Machine learning methods that use dimension reduction techniques perform a different type of operation and learning methods that use dimension reduction techniques perform a different type of operation and construct new components that are composed of different predictors and use these constructed components as new variables for their models. Much like it is the case for the elastic net approach, both PCR and PLS methods use the most impactful predictors to construct these components. In table 6, we show the importance of each predictor in the construction of each methods' final model for every test month.

These results in terms of variable importance are insightful, both in the context of the work of Bai, Bali and Wen (2019) and of Gu, Kelly and Xiu (2020). In Bai, Bali and Wen (2019), the authors establish that the best model available to date to explain the cross-section of corporate bond returns is composed of 4 variables that they develop in their paper : the Value-at-risk variable (VaR) the credit rating variable (meangrade), and the illiquidity measure of Bao, Pan and Wang (2011) (ILLIQ) as well as the bond market risk (MKTbond). Table 8 confirms that throughout the hundreds of different model fits performed though these different machine learning methods,

although three of these predictors are in the top six most important predictors overall, a bond's time to maturity (timetomat) and a bond's beta with the profitability factor (betaRMW) are the most impactful predictors to determine next-month corporate bond returns. Interestingly, a bond's beta with the stock market risk factor (betaMktRF) is more important than its beta with the corporate bond market risk return (MKTBond).

Table 12 – Variable Importance for models without Climate dummy variable

Variables	Elastic Net	PCR	PLS	Mean
timetomat	0.362	1.000	1.000	0.787
betaRMW	0.550	0.376	0.893	0.606
meangrade	0.473	0.412	0.805	0.563
betaMktRF	0.154	0.418	0.974	0.515
VaR	0.212	0.343	0.751	0.435
MKTbond	1.000	0.000	0.304	0.435
betaMom	0.100	0.349	0.809	0.419
betaSMB	0.056	0.289	0.752	0.366
betaterm	0.032	0.325	0.609	0.322
betaHML	0.063	0.230	0.610	0.301
betadef	0.067	0.237	0.548	0.284
mom12m	0.332	0.142	0.314	0.263
maxday	0.317	0.106	0.361	0.261
realindmom	0.381	0.078	0.242	0.234
mom36m	0.113	0.146	0.367	0.209
ILLIQ	0.108	0.121	0.338	0.189
mom1m	0.395	0.029	0.114	0.180
betaCMA	0.073	0.109	0.299	0.160
size	0.040	0.194	0.171	0.135
momt6t1	0.018	0.125	0.220	0.121
betaAgg.Liq.	0.155	0.061	0.123	0.113
chmom	0.160	0.000	0.118	0.093
sd_turn	0.000	0.244	0.000	0.081

Regarding Gu, Kelly and Xiu (2020), to date our linear machine learning models are unable to determine future corporate bond returns continually. Furthermore, it seems like predictors based on price trends have little role to play in determining future corporate bond returns compared to stock returns, as the first price trend predictor – twelve-month momentum (mom12m) – is only the twelfth most significant predictor for our different models. According to this study, and unlike Gu, Kellu and Xiu's (2020) on the stock market, the most impactful predictors in our sample are not

the traditional price-trend, volatility and liquidity-based predictors but rather bond specific predictors such as time-to-maturity, credit rating, and betas with the literature’s most important stock and bond-level risk factors.

It is also interesting to notice that for our elastic net models, the leading predictor is the bond market risk factor when our PCR models do not even use this predictor in constructing their models, which, in turn, seem slightly more efficient.

8.5.3. Models with a Climate dummy variable

Once we have obtained our results, we run our models once again, this time integrating the climate dummy variable. This can help us understand whether or not the fact that a bond is issued by climate-aligned firms has an impact on its pricing and can be used by the machine to better approximate how the corporate bond’s future price. Descriptive statistics of our new dataset are available in Table 9.

Table 13 - Machine Learning Models Results with Climate dummy variable

Month	OLS4 - Roos	OLS - Roos	Elastic Net - Roos	PCR - Roos	PLS - Roos	Glm - Roos
Min	-0.397	-0.486	-0.277	-0.157	-0.345	-0.489
Median	-0.002	-0.022	-0.010	-0.003	-0.013	-0.023
Mean	-0.004	-0.030	-0.011	-0.004	-0.016	-0.030
Max	0.256	0.318	0.200	0.061	0.081	0.318
Standard Deviation	0.097	0.134	0.064	0.035	0.065	0.134

As we can see, adding a climate dummy variable changes some of our results. Interestingly, mean and median out-of-sample R-squared for our classic linear models are higher, when they are reduced for our Elastic Net, PCR and PLS methods. However, standard deviation for each method does not change. Overall, since our Elastic Net, PCR and PLS methods were providing us with the best results for our different predictors, adding a climate bond dummy does reduce our best mean

out-of-sample R-squared from 0.3% from our PCR method to -0.4% from both our OLS4 approach and our PCR approach with climate dummy variable.

Table 14 - Variable Importance for models with Climate dummy variable

Variables	Elastic Net	PCR	PLS	Mean
timetomat	0.37	1.00	1.00	0.79
betaRMW	0.56	0.45	0.89	0.63
meangrade	0.48	0.44	0.80	0.57
betaMktRF	0.17	0.50	0.96	0.54
MKTbond	1.00	0.03	0.36	0.46
VaR	0.23	0.40	0.76	0.46
betaMom	0.11	0.41	0.82	0.45
betaSMB	0.07	0.36	0.77	0.40
betaterm	0.04	0.37	0.64	0.35
betaHML	0.08	0.29	0.64	0.34
betadef	0.08	0.29	0.58	0.32
maxday	0.33	0.14	0.42	0.30
mom12m	0.34	0.17	0.37	0.30
realindmom	0.39	0.12	0.30	0.27
mom36m	0.13	0.18	0.42	0.24
size	0.11	0.35	0.23	0.23
ILLIQ	0.12	0.16	0.39	0.22
mom1m	0.40	0.05	0.19	0.21
betaCMA	0.09	0.16	0.36	0.20
betaAgg.Liq.	0.17	0.09	0.21	0.15
momt6t1	0.03	0.15	0.27	0.15
sd_turn	0.00	0.30	0.10	0.13
chmom	0.17	0.02	0.20	0.13
climate	0.00	0.00	0.00	0.00

We perform a new variable importance analysis on the machine learning methods and models that include a climate dummy variable, and find that the climate variable is the least used variable for all three machine learning methods that are elastic net, PCR and PLS. This means that even though hundreds of models are fitted through the 165 month period, and taking under account that our cross-validation methodology fits hundreds of models to fit its best fit for every month, the climate dummy variable contains no information used by machine learning algorithm to

determine a bond's pricing. This allows us to conclude that climate bonds are not priced differently from traditional bonds. Variable importance for our new model permutations are in table 14.

8.6. Concluding remarks on results

Our analyses allows us to obtain some insights on the cross-section of corporate bond and climate bond returns, through more needs to be done. There is room to transform the work performed in our article on the pricing of corporate and climate bonds into two distinct studies : a more precise replication study of the work performed by Bai, Bali and Wen (2019) since we do not find precisely the same results than the authors throughout the analysis. Some of our results have lower statistical significance and our Fama-Macbeth analysis leads us to develop dissimilar conclusion than the authors. Since could also prove beneficial since this work is so recent and has not yet been tested thoroughly by other academics. Furthermore, the bond-level factor model could also be tested on more recent, out-of-sample data. In our study, we build our models on the entire dataset from 2002 to 2019, but we could also fit our model to precisely the same historic data than Bai, Bali and Wen (2019) going from 2002 to 2016 and observe how models perform on “out-of-sample” data from 2017 onwards.

Another study could consist in applying more precise and specialized comparison methods to better differentiate traditional bonds from climate bonds focusing more precisely on the specific impact of a climate dummy variable applied in a traditional asset pricing context.

Regarding our work on machine learning methods, more variables can be developed, the machine learning methods, algorithms and results could benefit from solid robustness checks, and more computer power could allow authors to run more consuming and powerful machine learning methods such as random forests and neural networks.

Nevertheless, this work represents to the authors' knowledge one of the first attempts to apply traditional asset pricing to climate-aligned assets, as well as one of the first attempts to develop bond-specific factors in addition to those developed by Bai, Bali and Wen (2019) in order to run machine learning methods to predict corporate bond returns. For the author, these studies represent important steppingstones for future research in sustainable finance, and more specifically research in modern sustainable asset pricing.

9. Future Research

The five distinct research papers that were developed in the context of this PhD dissertation can all be taken as a basis for future work. The first article opens the way to the what we chose to describe as a “new paradigm” of sustainable finance, one in which every major European firm and asset manager will be more or less aligned with the environmental and social goals of the EU Taxonomy. This represents a very diverse set of interesting research opportunities not only for practitioners but for academics as well. New relationships will emerge between traditional financial and risk indicators and environmental and social EU Taxonomy indicators. Capital markets and their actors will react to this new wave of information, and it is our belief that many new interesting discoveries will be made on the general and specific relationships between environmental and social performance and financial performance and risk. As previously discussed in the context of this thesis, this could also impact in turn the asset pricing literature, which has seen only very few ESG-related factors enter its “zoo of factors”. While it might have been a bit too early to obtain reliable ESG-related data to perform truly robust analysis on the performance of what can be considered sustainable assets, the EU Taxonomy might well mark the start of a sustainable finance empirical literature.

The second paper focuses on the role of capital markets in financing large US and EU firms. It is perhaps the less developed subject in the PhD dissertation because its focus is so wide and dispersed. It studies four different datasets - non-financial US firms in the S&P 500, non-financial firms in Euro countries in the S&P 350, US climate-aligned firms and EU climate-aligned firms – and the two capital markets – stock markets and bond markets. Though it does analyze large sets of data, these analyses are purely descriptive, and lead to many further underlying questions, which all belong in different specific underlying literature: what is the relationship between dividend distribution or share repurchases and performance in these firms? Combined with the arrival of the taxonomy-alignment indicators, one could also study the relationship between taxonomy-alignment and capital market financing. The methodology developed in this second paper could lead to interesting studies on the relationship between corporate bond financing and performance, risk or investment. To the author, this second article is considered as a basis for much more work on the subject, specifically given the fact that recent literature on these different subjects has been quite difficult to come by.

Realizing both the importance that US corporate bond markets had on financing large and climate-aligned companies in the US and the interest the sustainable finance practitioners had taken in corporate bond markets given the growth of the green bond market since 2013, the third paper focuses on the issue of corporate data. This paper, much like previous paper, is a gateway to a large literature that has just started to emerge on the corporate bond primary and secondary markets in the US as well as the pricing of corporate bonds. Once again, this study has no inferential or predictive dimension to it but focuses on a thorough data treatment procedure for corporate bond prices, also offering a preview of what the most developed corporate bond databases – FISD and TRACE – can offer in terms of research. Though we briefly explore volume

and pricing of the green bond and climate bond markets, this is purely exploratory, and paves the way for much more thorough and in-depth studies. A thorough understanding of these databases could lead to interesting work on liquidity, momentum or volatility on the growing green bond and climate bond markets.

Amongst the diverse possibilities that are associated with developing such a database on corporate bonds is the possibility of moving into the asset pricing literature. Asset pricing can be considered of fundamental importance for the economy as it provides key information for economic and financial decision makers in the context of physical investment and consumption.

Looking at asset pricing from a sustainable finance point of view, and considering the growing climate change urgency, we can understand the pressing need for sustainable concerns to enter this economic and financial decision-making process. In addition, while financial asset prices are considered to reflect underlying fundamental values, financial crises have proven the contrary. Today, an interesting issue resides in the idea that the current environmental crisis might well be the result of the fact that environmental factors have historically not been part of these fundamental values that are supposedly reflected in asset prices. To the author, this concept is perhaps what is most promising about what the PhD dissertation could possibly entail in terms of future research. Before this occurs, however, much progress needs to be made regarding the development, communication, quality and comparability of ESG data. In some sense, this is the EU Taxonomy's purpose.

Developing an understanding of some of the traditional asset pricing methodologies allows for further research on the general subject of sustainable asset pricing. While studying climate-aligned firms today could well be considered one of the most reliable ways of studying both time-series and cross-sections of sustainable assets, it is the author's belief that the study of Taxonomy-

aligned – or related – products will soon enter the empirical asset pricing literature. Understanding these traditional asset pricing methodologies also allows for an understanding of their possible limitations, which we have discussed in previous sections. Big data, data science and machine learning represent a new wave of quantitative research tools that are progressively making their way into the academic literature and are seen by some of the most influential financial academics as revolutionary. In some way, the fourth paper of this PhD dissertation represents in itself an introduction to its fifth and final paper and the corresponding exploration of statistical and machine learning methods for asset pricing.

Traditional asset pricing methodologies were invented and developed well before modern computers were made readily available to consumers. It stands to reason that these methodologies evolve with rising computational power and data availability. Today, statistical learning and machine learning are progressively making their way into the financial and asset pricing literature. They allow to rapidly fit hundreds if not thousands of diverse models – linear, quasilinear and non-linear – to financial time-series and cross-sectional data each with various numbers of specifications, all the while implementing algorithms preventing risks of overfit and false discovery. They can work more efficiently with more data, and will necessarily provide opportunities for better results, specifically as an exponential amount of data is made available to financial analysts and academics, and an increasing number of factors now participate in explaining or predicting the cross-section of stock returns.

Though the author's capacities in terms of constructing reliable statistical and machine learning algorithms are of course limited given the fact that these algorithms have been constructed and run on his personal computer, this work does suffice, in his sense, in demonstrating some of the obvious advantage of machine learning methods. Once again, as is the case for every paper in

this PhD dissertation, this last paper can also be considered as an introduction to the possibilities that lie ahead for future research work. Of course, as a sustainable finance researcher, the main reflection behind this work is that ESG factors explaining the cross-section of asset returns exist but have not yet been identified. By the time quality comparable ESG data becomes available to analysts and academics, it is the author's belief that statistical and machine learning methods the like of which are explored in this final paper will be the ones best fit to find and understand the role of such factors in explaining and predicting the cross-section of asset returns.

10. Summary conclusion

This summary note allows for a better understanding of the overall research process that was performed in the context of this PhD thesis. Central to the general subject of sustainable finance is the issue regarding the definition of what constitutes sustainability for firms, financial actors and regulators, and work on this subject allows us to understand that the in which a clear regulatory framework to create a common language for sustainable finance has only just begun. Sustainable actors will progressively be identified and much more transparency regarding the possibilities of sustainable investments will be available for financial market participants.

As the need for investments in climate change mitigation and adaptation will inevitably grow in the following years, identifying these firms and corresponding investment could allow firms, investors and regulators to better understand how capital markets can better participate in financing sustainable firms, assets and projects. Furthermore, the importance of the corporate bond market in providing finance for these firms, assets and projects will inevitably grow clearer to these actors. It could well be assumed that given this possible evolution, much more attention will be given to corporate bonds, particularly in the asset pricing literature, as ESG-related and climate-

related factors start to be identified as relevant to explain or predict the cross-section of returns. This phenomenon could well be amplified given the powerful innovative machine learning tools that are increasingly being used to understand financial market returns. We explore each of these subjects individually in our five research articles.

The chapters that follow this summary note contain each of these articles developed in the context of the PhD dissertation, going into further detail on each research topic covered. Chapter 2 contains our research article that focuses on understanding the framework for sustainable finance through the analysis of the practitioner, academic and regulatory literature. Chapter 3 contains our work on capital market financing of traditional and climate-aligned firms. Chapter 4 presents our detailed data treatment methodology for corporate bond data for the green bond and climate bond markets. Chapter 5 contains our work on the analysis of the cross-section of corporate bonds for traditional and climate-aligned firms and Chapter 6 our work on the application of Machine Learning methods to the cross-section of corporate bond returns. Finally, we perform a general conclusion of our work.

Chapter 2

The EU taxonomy: a new sustainable finance paradigm?

Abstract

As the scientific community continues its progress on evaluating the likeliness and possible impact of global environmental risks, focus on the role of financial market participants to mitigate these risks increases. However, these actors lack clarity on how capital can be properly channeled to firms, assets and projects that positively benefit the environment. This article reviews the latest literature on challenges faced by practitioners and academics to evaluate the environmental performance of investments and to establish relationships between environmental performance, financial performance and risk. Considering this current state of play, I perform a detailed analysis of how the EU taxonomy on sustainable activities and related measures of the European Commission could provide the necessary framework for a new sustainable finance paradigm where investments and environmental performance will be directly linked.

Keywords: SRI; Corporate Environmental Performance; ESG; Sustainable Finance; Climate Change; EU Taxonomy

1. Introduction

In October 2018, the Intergovernmental Panel on Climate Change (IPCC) - the United Nations body for assessing the science related to climate change - published a special report on the impacts of global warming for various global warming pathways. This report built on the work of more than 6000 recent publications to provide a summary of the situation to policy makers, and its conclusion is best described by the following excerpt:

“(...) recent trends in emissions and the level of international ambition indicated by nationally determined contributions, within the Paris Agreement, deviate from a track consistent with limiting warming to well below 2°C. Without increased and urgent mitigation ambition in the coming years, leading to a sharp decline in greenhouse gas emissions by 2030, global warming will surpass 1.5°C in the following decades, leading to irreversible loss of the most fragile ecosystems, and crisis after crisis for the most vulnerable people and societies.”

Throughout the report, it is made quite clear that a 2 °C or 3 °C pathway would expose hundreds of millions of people to water stress and scarcity, heatwave events, flooding, risks related to power production, crop yield or habitat degradation, with corresponding impacts on food security, health, poverty, public unrest and political destabilization. The report also addresses the irreversible economic damages that could result from these different pathways (IPCC, 2018). More recently, it was estimated that the world’s current trajectory in terms of GhG emissions will lead global warming to rise by at least 3 °C by the end of the century (UNEP, 2019). Unsurprisingly, the 2020 Global Risks Report performed on a yearly basis by the World Economic Forum identified five environmental risks as the five leading global risks in terms of likelihood, all five

of these risks also being part of the top ten risks in terms of impact, with “climate action failure” being top of the ranking, in front of “weapons of mass destruction” (World Economic Forum, 2020). For the first time, global warming has been clearly identified as the most important threat to world economies, both in terms of likelihood and potential impact.

In his March 2020 letter, which represents one of the most awaited statements of each year for financial market participants, Blackrock CEO Larry Fink calls for a “Fundamental Reshaping of Finance”, strongly advocating for its readers to consider climate risk as investment risk, and calling for governments, companies and shareholders to confront climate change. More practically, Larry Fink asks CEO’s to start to publish industry-specific sustainability disclosure as well as climate-related risk disclosure aligned with the recommendations of specialized organizations. The letter contains a clear warning : “Given the groundwork we have already laid engaging on disclosure, and the growing investment risks surrounding sustainability, we will be increasingly disposed to vote against management and board directors when companies are not making sufficient progress on sustainability-related disclosures and the business practices and plans underlying them²⁰”. This clear warning from the largest asset manager in the world marks new progress for sustainable finance and sustainability disclosures. Much like it was the case in 2015 for COP 21, the Paris Agreements and the SDGs, particular focus in this letter is given to the possible partnerships between governments, companies and financial market participants that need to be implemented in order to face the climate urgency. Furthermore, the global asset manager puts an important emphasis on the need for all actors to provide the most optimal and comprehensive information using comparable frameworks that have been developed. The

²⁰ See <https://www.blackrock.com/corporate/investor-relations/larry-fink-ceo-letter>

underlying message that is given through this letter is that more data is needed from companies and market participants in order for actors to understand and evaluate sustainability risks, and that these systemic risks need to be taken under account by every type of financial market participant before they can be properly managed and possibly mitigated. In essence, what the industry needs is to clearly evaluate the relationships that exist between economic activities performed by companies and their impacts in terms of sustainability.

In this paper, we explore this notion by investigating how practitioners and academics are evaluating the extra-financial performance of companies and financial market participants, focusing primarily on the environmental aspect of this overall extra-financial corporate performance. I first study the global sustainable investment industry (SRI), as well as the more specific US and European markets for SRI. From this analysis, and using the most recent available data on these markets, we conclude that even though actors from this industry seem to have common approaches and concepts regarding sustainable investing, there is a lack of common definition and metrics in order for extra-financial performance to be easily measured and compared across actors. Furthermore, even actors that specialize in this market are yet unable to provide sufficient information on each approach, and most of the available data comes from the United States or Europe, while very large actors such as China have not provided any form of reliable data on the state of the Chinese SRI market. To date, it seems that a clear transparent link between investments and environmental performance does not exist, as important work still needs to be done regarding the development of common terminology, definitions and concepts that can be used across industries, geographies and type of actors.

I then investigate how academics define the concept of corporate environmental performance by first studying the literature focusing on the relationship between environmental, social and

governance (ESG) factors and financial performance before focusing more specifically on the concept corporate financial performance as both a unidimensional and multi-dimensional concept. My conclusion on this study resembles that of actors of the SRI market in the sense that there seems to lack common terminology, definitions and concepts within this academic literature. Studies on the relationship between ESG and financial performance lack a common definition for what truly constitutes ESG as a concept, with most of the work to date using a diverse set of terms to refer to the overall concept of extra-financial performance such as ‘corporate social responsibility’ (CSR), ‘corporate social performance’ (CSP), SRI, or terms more specific to social and environmental factors such as ‘environmental management’, ‘green management, social performance’ or ‘social capital’. When focusing more specifically on the academic concept of corporate environmental performance (CEP), authors disagree on whether CEP is a unidimensional or multidimensional concept composed of a process-based environmental management performance (EMP) and a outcomes-based environmental operational performance. An investigation of the different metrics used to define these concepts leads to the conclusion that most authors develop their own measure for these concepts, and that most studies focus on samples from different industries and geographies. Finally, there also seems to lack comparability between data provided by different ESG data providers, which further emphasizes the fact that overall, it seems that both practitioners and academics are in need of a common framework to allow for a common and comparable evaluation of what truly constitutes sustainability performance. I also note that ESG factors are almost entirely absent from the asset pricing literature even though hundreds of different factors have been put forward by academics in the last decades.

Following this finding, I study the recent European Union framework for sustainable finance to try to understand how the European Commission is tackling this challenge specifically. My

analysis of available documentation on the different measures that are taken on this subject provides insight on how the European Commission has created a framework in which each different type of actor of the financial industry is being targeted individually in order to start integrating sustainability into business practices and regulation. At the center of this framework is the EU taxonomy, a detailed, industry-specific classification system which categorizes different economic activities according to their extra-financial impact in the context of 6 environmental objectives. I understand from my analysis that this classification system is intended to create common terminology, concepts and definitions to all actors the European Union's financial system, therefore providing a solution to the aforementioned challenges faced by both practitioners and academics of the sustainable finance industry.

In the following section, I go into a more detailed analysis of the EU taxonomy and the corresponding regulation that has been implemented to date, and show how the European Commission is planning on creating a financial system where every actor has to clearly demonstrate its participation in terms of sustainability with an emphasis on climate mitigation and adaptation given the current state of the climate urgency. I also show how companies and asset managers are put at the center of the transition towards sustainable EU financial markets by being both the first to have to demonstrate their alignment with the EU taxonomy but also have to be the most precise regarding this alignment, with both types of actors having to provide precise percentage of their economic activity that is taxonomy-aligned.

In the final section of this paper, I then explore the current state of plan regarding the integration of sustainability into the business practices and regulation of other types of actors such as banks, pension providers, insurance providers, credit agencies, benchmark and label providers as well as corporate governance. This final section provides a general view of the current shift that

is occurring in the European Union financial system regulation regarding the creation of what was can be considered as a new sustainable finance paradigm.

Overall, this paper is structured as follows. Section 2 addresses the study of the SRI industry and the corresponding ESG approach. Section 3 investigates the literature related to the definition of ESG and Corporate environmental performance and their relationships with financial performance. Section 4 provides insight on the European Union's new framework for sustainable finance, while Section 5 focuses on the specific study of the EU taxonomy and corresponding regulation and Section 6 on the state of play of sustainability-linked measures and regulations for other actors of the EU financial markets. Section 7 concludes the paper.

2. The Sustainable Investment Industry and the ESG approach

When referring to investments made by private investors for sustainability, actors of financial markets generally refer to the sustainable investment industry, which is also known as the socially responsible investing (SRI) industry. While there are many definitions of what constitutes a sustainable investment, most actors agree on the fact that this investment approach “considers environmental, social and governance (ESG) factors in portfolio selection and management” (GSIA, 2018). According to the Global Sustainable Investment Alliance (GSIA), an organization that performs a market study of the global sustainable investing industry every two years, the sustainable investment industry represented \$30.7 trillion at the start of 2018, a 34 percent increase in two years (GSIA, 2018). This would represent 38.8% of the \$79.2 trillion in total assets under management at the end of 2017 (BCG, 2018).

Much like there is common agreement that sustainable investment approaches use ESG factors in their investment decision process, there also is agreement in the industry that there are seven

distinct approaches to perform sustainable investing. In Table 1, I summarize the different sustainable investing approaches, provide the definition and corresponding global assets under management for each approach using information for the GSIA global report for 2018. Furthermore, I attempt to provide further details on each approach in Europe and in the United States, using information from the Eurosif SRI Report for 2018 (Eurosif, 2018) and the US SIF Trends report from 2018 (US SIF, 2018). The GSIA, the Eurosif and the US SIF are respectively the main references in terms of market studies of the sustainable investment environment in their geographical regions, Eurosif and US SIF being underlying members that constitute the GSIA. To date, Europe and the United States constitute by far the largest SRI markets globally, considering that the Chinese market on sustainable investing has not yet been identified given a lack of available data. When looking at Table 1, a few observations can be made regarding each investment approach.

2.1. Negative/Exclusionary Screening

We notice the largest investment approach is the “Negative/Exclusionary Screening Approach”, which in practice does not represent concrete investments in sustainable projects, but rather non-investment - and in some cases divestment - in industries that are considered unethical or contrary to social or environmental investor principles. As this paper primarily focuses on investments in climate change and the environment, I observe that only two environmental criteria are included in the top European criteria for exclusion. Eurosif does address the divestment from fossil fuel industries separately in their report, but this approach does not seem to be a priority for investors as of 2018. Specifications regarding the preferences of US investors in terms of negative screening are not provided by US SIF.

2.2. ESG Integration

The second largest approach is “ESG integration”, for which the GSIA provides a definition that corresponds quite closely to their definition of the SRI industry as a whole²¹. This subject is addressed specifically by Eurosif in their study, which states clearly that there is a “lack of clarity in the parameters governing the integration of ESG factors”. In some sense, this is addressed by the US SIF which provides an analysis of the market that solely focuses on ESG categories and specific ESG criteria but does not address each investment approach separately. Looking at US SIF data, we understand that, as of 2018, sustainable investors in the US focused quite equally between social, environmental and governance categories, even though “Environment” is the third category for both Money Managers and Institutional Investors. However, in the case of the US market, the specific “Climate Change / Carbon” criteria does make the ranking of top ESG criteria used by both Money Managers and Institutional Investors.

2.3. Corporate Engagement and Shareholder Action

Another largely implemented sustainable investment approach resides in “Corporate Engagement and Shareholder Action”, which consists in using shareholder power to influence corporation action. In Europe, this approach is mostly performed in the United Kingdom, which together with Sweden and the Netherlands represent 90% of the European market. Though Eurosif does not provide information on specific criteria used for this approach, US SIF provides information on leading ESG issues in the US. While the top 2 issues are governance issues, climate change does stand in third place with 270 shareholder proposals in the 2016-2018 period.

²¹ In the GSIA 2018 report, the overall definition for a sustainable investment is, as aforementioned, an investment approach that “considers environmental, social and governance (ESG) factors in portfolio selection and management” while the definition for the specific “ESG Integration” investment approach refers to “the systematic and explicit inclusion by investment managers of environmental, social and governance factors into financial analysis”

2.4. Norms-Based Screening

According to the GSIA, norms-based screening is not performed in the US and consists in implementing international norms into the investment screening process. The three main norms used in Europe are the UN Global Compact (42%), the OECD Guidelines for Multinational Companies (25%) and the International Labour Organisation conventions (26%). These mostly focus on human, labour rights and anti-corruption, though the environment has also been integrated. The Eurosif also includes more than 10 other international norms that are used by the 7% of remaining investors that compose this market

2.5. Positive/Best in Class Screening

Amongst the smallest investment approaches in terms of corresponding assets under management, “Positive/Best-in-Class Screening” consists in selecting investment opportunities that have the best ESG performance in an industry or sector. According to Eurosif, in Europe, investors in France mostly contribute to this approach, but it is most popular in the US, which represents 60% of the global market for this approach. Neither the GSIA, Eurosif or the US SIF provide information on which ESG criteria are most used to construct best-in-class ESG investment approaches.

2.6. Sustainability Themed Investing

This investment approach focuses on specific industries and economic activities that are known to be beneficial for climate and the environment. This approach is however one of the least popular approaches in the sustainable investment industry, as it represents only \$1,02 trillion globally, with 77% of assets under management managed in the US. Though the US SIF does not provide information on the specific market fragments that are targeted by this approach, data from

Eurosif make us understand that this market is very fragmented between different activities (7 distinct activities composing respectively 17% to 9% of the market).

2.7. Impact and Community Investing

The least popular sustainable investment approaches in terms of assets under management are the “Impact Investing” and “Community Investing” approaches. Impact Investing involves investments targeted specifically at environmental and social issues. Impact investors invest with the intention of generating measurable positive social and environmental impact while also maintaining positive financial results. Community Investing consists in investing in identified under-served communities and is mostly popular in the United States.

2.8. General remarks

As the growth of the sustainable investing market has been capturing a lot of attention by a variety of financial market actors, going into a detailed analysis of this market provides interesting insight. First and foremost, to date, the sustainable investment market is not solely composed of investments in activities with positive environmental and social performance, but also of non-investments in some specific activities that are subjectively assessed as non-ethical by investors. An interesting example is the exclusion of nuclear energy by some investors from their portfolios, while the negative or positive social and environmental impact of nuclear energy is still quite a debate²². Furthermore, there is debate between actors of the sustainable finance industry of whether exclusion can be considered a sustainable investment approach, as one investor that chooses not to invest or to divest from an activity or firm is in most cases replaced by another. As the need for

²² Greenhouse gas emissions from nuclear generation is quite low, but nuclear energy generation has other environmental impacts linked to uranium mining, radioactive emissions and waste heat. Though this debate is clearly outside the scope of this study, nuclear energy constitutes a good example of the subjective aspect of some SRI strategies.

investments in sustainable projects grows with the climate urgency, it seems quite clear that non-investment or divestment in “brown” activities will not be sufficient. The exclusion approach, however, is applied for \$19.77 trillion of assets under management that are counted as sustainable investments. The industry that represents only investments in projects or firms with positive social and environmental performance would therefore be quite smaller.

Table 1 - Global SRI approaches and corresponding assets-under-management

SRI Approach	Definition	Global AuM	Europe (% of Global AuM)	Distribution	US (% of Global AuM)	Distribution
Negative/Exclusionary Screening	"the exclusion from a fund or portfolio of certain sectors, companies or practices based on specific ESG criteria"	\$19.77 trillion	55%	<p>Top Exclusion Criteria:</p> <ul style="list-style-type: none"> - Controversial Weapons (63.6%) - Tobacco (49.1%) - All Weapons (45.7%) - Gambling (34.9%) - Pornography (34.4%) - Nuclear Energy (33.9%) - Alcohol (30.6%) - GMO (24.5%) - Animal testing (19.3%) 	40%	<p>ESG Investing</p> <p>ESG Categories by Money Managers 2016-2018:</p> <ul style="list-style-type: none"> - Social (\$10.8 trillion) - Governance (\$10.8 trillion) - Environmental (\$10.1 trillion) - Product (\$4.5 trillion) <p>Top specific Criteria</p> <ul style="list-style-type: none"> - Climate Change / Carbon : \$3.00 trillion - Tobacco : \$2.89 trillion - Conflict Risk : \$2.26 trillion - Human Rights : \$2.22 trillion - Transparency and Anti-Corruption : \$2.22
ESG Integration	"the systematic and explicit inclusion by investment managers of environmental, social and governance factors into financial analysis"	\$17.54 trillion	54%	<p>Indicators used by the Eurosif are:</p> <p>(1) the formalisation of an ESG integration policy in the investment process (19%)</p> <p>(2) the number of ESG analysts in the team (61% have between 1 and 5)</p>	28%	<p>ESG Categories by Institutional Investors 2016-2018:</p> <ul style="list-style-type: none"> - Social (\$5.2 trillion) - Governance (\$3.5 trillion) - Environmental (\$3.5 trillion) - Product (\$2.9 trillion) <p>Top specific Criteria</p> <ul style="list-style-type: none"> - Conflict Risk : \$2.97 trillion - Tobacco : \$2.56 trillion - Climate Change / Carbon : \$2.24 trillion - Board Issues : \$1.73 trillion - Transparency and Anti-Corruption : \$1.69 trillion
Corporate Engagement and Shareholder Action	"the use of shareholder power to influence corporate behavior, including through direct corporate engagement (i.e., communicating with senior management and/or boards of companies), filing or co-filing shareholder proposals, and proxy voting that is guided by comprehensive ESG guidelines."	\$9.83 trillion	56%	<ul style="list-style-type: none"> - No information on distribution of different criterias. - Mostly performed in the UK (€2.8 trillion), Sweden (€874 billion) and the Netherlands (around €700 billion) which together represent around 90% of the market 	18%	<p>Shareholder Action</p> <ul style="list-style-type: none"> - In 2018, 165 institutional investors and 54 investment managers controlling \$1.8 trillion in assets filed or co-filed shareholder resolutions on ESG issues
Norms-based Screening	"screening of investments against minimum standards of business practice based on international norms, such as those issued by the OECD, ILO, UN and UNICEF"	\$4.68 trillion	77%	<p>Applied Norms :</p> <ul style="list-style-type: none"> - UN Global Compact (42%) - OECD Guidelines for MNC's (25%) - ILO Conventions (26%) - Others (7%) 	na	<p>Leading ESG Issues by number of Shareholder proposals filed:</p> <ul style="list-style-type: none"> - Proxy Access (≈350) - Corporate Political Activity (≈300) - Environment: Climate Change (≈270) - Labor & Equal Employment Opportunity (≈190) - Executive Pay (≈180) - Human Rights (≈150) - Independant Board Chair (≈150) - Special Meetings (≈120) - Board Diversity (≈90) - Sustainability Reporting (≈90) - Environment: Other issues (≈90)
Positive/Best-in-Class Screening	"investment in sectors, companies or projects selected for positive ESG performance relative to industry peers"	\$1.84 trillion	36%	<ul style="list-style-type: none"> - No information on distribution of different criterias. - Mostly performed in France which represents more than 50% of the market with close to €300 billion out of a total of €586 billion, with Netherlands and Italy representing close to €100 billion and 75€ billion respectively 	60%	
Sustainability Themed Investing	"investment in themes or assets specifically related to sustainability (for example clean energy, green technology or sustainable agriculture)"	\$1.02 trillion	17%	<p>Very segmented market:</p> <ul style="list-style-type: none"> - Water management (17%) - Renewable Energy (12%) - Energy Efficiency (11%) - Sustainable transport (11%) - Buildings sector (11%) - Land use/Forestry/Agriculture (11%) - Waste management (9%) - Other (18%) 	77%	
Impact/Community Investing	"targeted investments aimed at solving social or environmental problems, and including community investing, where capital is specifically directed to traditionally underserved individuals or communities, as well as financing that is provided to businesses with a clear social or environmental purpose"	\$ 444,26 billion	28%	<p>Eurosif does not provide information on the specific sectors that are addressed by Impact Investing</p>	66%	

Secondly, looking at 2018 data, it seems that the environment is not yet a priority for sustainable investors. There are only two exclusion strategies that are related to the environment in Europe using Eurosif data, the only investment approach that focuses exclusively and clearly on the environment - Sustainability Investing - is one of the least popular and has grown by only 3% in Europe between 2016 and 2018. In the United States, the Environment is not the most important ESG category for both money managers and institutional investors. Climate change issues do come third in terms of shareholder proposals in the US, but sustainability reporting and other environmental issues are at the bottom of the shareholder proposal list. As clear investments in favor of positive environmental projects that mitigate climate change become more urgent, this did not seem to be a priority for sustainable investors as of 2018.

Finally, an analysis of what to date are considered three of the most informative reports on the sustainable investing industry leaves us with the conclusion that actors still lack clarity. Even though efforts of the GSIA allow us to have a global view of the industry, we understand that European and US markets differ and provide different approaches of the same market. Sustainable investing is still quite a broad endeavor, and the terminology for sustainability or ESG criteria is very dense. Table 1 only contains criteria that is provided in the reports, which generally consists in a list of the most used criteria. For Europe, the report mentions 9 different exclusion criteria, 13 different European norms, 7 different types of sustainable investments, and in the US, reports mention 6 top ESG criteria and 10 types of shareholder proposals. Most of the terms used are not similar and do not address the same issue. Furthermore, information on the specifics of approaches like ESG Integration, which does represent \$17.54 trillion of global assets under management according to the GSIA, are not provided. In Europe, specifics of the €4.8 trillion “Corporate Engagement and Voting” industry are also not provided.

As it seems clear today that interest for sustainability is growing amongst investors, we understand from our analysis of the industry that there still lacks visibility on what precise investments are made in projects and activities that have positive environmental outcomes, specifically with regards to climate change. Another insight from this analysis resides in the fact that while sustainability investing is described as the application of ESG factors in portfolio selection and management, there seems to lack a precise delimitation of what constitutes an ESG factor. Furthermore, the descriptions given by the GSIA for every approach mention “ESG criteria”, “ESG factors”, “ESG guidelines” and “ESG performance” without further precision of what differentiates these terms and what they refer to specifically.

3. Corporate Environmental Performance

3.1. ESG and Financial Performance

In order for companies and financial market participants to truly take action to mitigate global warming, these actors need better understanding of the relationships that exist between environmental factors and factors linked to financial performance and risk. However, to date, academic literature has not yet reached a consensus on this issue and is still trying to understand the relationship of ESG disclosure as a one-dimensional concept with financial performance. We observe similar struggles in academia than that which were identified in the previous section concerning practitioners: there is no clarity on what truly constitutes ESG factors and disclosure, few studies focus on the specific relationship between environmental performance and financial performance and risk, research approaches concerning ESG criteria vary greatly and data on ESG is not always entirely reliable.

Looking at the literature, we do get a sense that a positive relationship exists between ESG and financial performance. Brooks and Oikonomou (2018) provide a comprehensive review of the effects of ESG disclosure on firm financial performance. According to the authors, a consensus has been reached regarding the statistically significant and economically modest positive relationship between ESG criteria and corporate financial performance, with only 6% to 8% of the literature finding a negative relationship between the two concepts. Furthermore, the authors point out that the three main meta-analyses on the subject found very similar correlation between ESG and corporate financial performance, going from 0.12 (Friede et al., 2015), to 0.13 (Margolis et al., 2009) to 0.15 (Orlitzky, 2003). However, further analysis of this work provides interesting insight on some underlying issues that are faced by the literature.

The first of these issues consists in the different concepts that address the extra-financial aspect of companies and the frequently used terms of corporate social responsibility (CSR) and corporate social performance (CSP) in comparison with ESG. In practice, these three terms are almost used interchangeably, as they are used to refer to the broad category of factors that do not address financial performance or risk directly. In the aforementioned work of Brooks and Oikonomou (2018), which is dedicated to a comprehensive review of literature of the relationship between ESG disclosure and performance of firm value²³, the term “ESG” is used 19 times throughout the paper, and is mentioned for the first time in section 7 of the paper which directly precedes the paper’s conclusion, when the term “CSR” is used 58 times and the term “CSP” is used 82 times throughout the paper. No explanation of the relationship between “CSR”, “CSP” and “ESG” is provided, and none of these concepts are defined. Concluding remarks address the

²³ The title of the authors’ paper is “*The effects of environmental, social and governance disclosures and performance on firm value: A review of the literature in accounting and finance*”

link between “CSP/ESG and financial performance”, clearly demonstrating that the authors consider CSP and ESG to broadly address the same concepts. Further analysis of the three main meta-studies on the relationship between ESG/CSP and corporate financial performance gives further insight on the challenges faced by the literature.

In Friede, Busch and Bassen (2015), which is considered the largest and most recent meta-study on the relationship between ESG and financial performance²⁴, authors identified 35 vote-count studies and 25 meta-analyses which together analyzed 3718 primary studies. In this work, authors broadly determine whether underlying studies focus on Environmental, Social or Governance criteria by applying “a definition of ESG that reflects the exemplary list of variables of Clarkson (1995), Wood (2010) and the investment approaches of GSIA (2013)”. However, this definition is not provided to the reader, Clarkson (1995) and Wood (2010) refer to work on the evaluation and measure of corporate social performance (CSP) and GSIA (2013) refers to the Global Sustainable Investment Review 2012, which, much like the 2018 report that was reviewed in the previous section, does not provide a definition of what constitutes ESG criteria or factors. Looking at the terminology of the titles of the underlying studies that are used in Friede, Busch and Bassen’s meta-study, it can be observed that 26 different concepts are explored in a total of 40 studies, going from the traditional “Corporate Social Responsibility” and “Corporate Social Performance” to the more specific "Environmentally Sustainable Supply Chain Management Practices", "Social Capital of Entrepreneurs" or "High Performance Work Practices". Table 2 contains the different terminology used in the titles of the 40 underlying studies used in Friede, Busch and Bassen (2015) as well as corresponding citations, where we can observe the variety of

²⁴ Here the title of the paper is “*ESG and financial performance: aggregated evidence from more than 2000 empirical studies*”

different concepts that are used in studies that focus on the relationship between ESG/CSP and financial performance.

The two other main meta-studies used in the literature to demonstrate the relationship between “CSP/ESG” and corporate financial performance are Orlitzky et al. (2003) and Margolis et al. (2009). In their paper, Orlitzky et al. (2003) provide an overview of the studies included in their meta-studies as well as the measures of CSP and corporate financial performance that were used. More than 30 different measures of CSP and more than 20 measures of corporate financial performance are used in the 52 underlying studies. In Margolis et al. (2009) though the similar 20+ corporate financial performance measures are used, more than 100 different measures of CSP are identified in the 182 underlying studies. Amongst commonly found measures in both papers are the Kinder, Lydenberg & Domini (KLD) ratings, pollution expenditure measures, the Council for Economic Priorities (CEP) rankings, various types of annual report disclosures, Fortune magazine ratings, the Domini 400 Social stock index, the Toxics Release Inventory (TRI) measure, lawsuit filings, articles in the Wall Street Journal or affiliation to the Sullivan Principles/divestment from South Africa.

From this analysis of the literature, it can be concluded that ESG as a concept is still not clearly defined by academics, and is still one of many terms that is used to refer to the plethora of concepts and factors that address the “extra-financial” aspect of firms, projects and assets. This is quite concerning given the fact that the ESG concept is taken as a basis for what has been identified as a 30.6\$ trillion SRI industry that focuses on sustainability issues with no clear framework to understand and demonstrate how investments are truly sustainable. Furthermore, given this lack of definition and clear framework, the true relationship between ESG as a mono-dimensional concept and financial performance and risk is quite challenging to determine.

While authors like Brooks and Oikonomou (2018) provide important work to understand the current state of the literature on the concept of ESG/CSP, and do provide rich insight by putting forward the fact that there seems to be a general positive correlation between ESG/CSP and financial performance, it seems clear from the analysis of meta-studies of the field that ESG/CSP is far from a mono-dimensional concept, as these meta-studies do refer to hundreds of different underlying concepts that are categorized as ESG criteria or factors. In addition, the variables used in these studies to measure financial performance also differ greatly, further challenging the robustness of these results.

A literature specific to the study of corporate environmental performance – which is broadly categorized as the environmental branch of ESG/CSP - has also been developing as focus on the climate change urgency grew. We focus on this literature to try to further understand the specific challenges of academics trying to measure corporate environmental performance and evaluate the relationship between environmental performance and financial performance.

Table 2 - ESG-related terminology

Concept used in title	Different terminology used	References
CSR	"Corporate Social Responsibility"	Cochran and Wood (1984), Aupperle, Carroll, and Hatfield (1985), Pava and Krausz (1996), Griffin and Mahon (1997), Richardson, Welker, and Hutchinson (1999), Orlitzky and Benjamin (2001), McWilliams, Siegel, and Wright (2005), Boaventura, Santos da Silva, and Bandeira-de-Mello (2012), Fiska (2013), Schröder (2014), Pavie and Filho (2008)
CSP	"Corporate Social Performance"	Wood and Jones (1995), van Beurden and Gössling (2008), Pelozo (2009), Orlitzky (2001), Orlitzky, Schmidt, and Rynes (2003), Allouche and Laroche (2005), Wu (2006), Margolis, Eifenbein, and Walsh (2009), Vishwanathan (2010)
CEP	"Corporate Environmental Performance" or "Environmental Performance"	Blanco, Rey-Maqueira, and Lozano (2009), Horváthová (2010), Günther, Hoppe, and Endrikat (2011), Schultze and Trommer (2012), Dixon-Fowler et al. (2013), Endrikat, Guenther, and Hoppe (2014)
SRI	"Socially Responsible Investing" or "Sustainable Investing"	Sjostrom (2011), Rathner (2013), Viviers and Eccles (2012), Kleine, Krautbauer, and Weller (2013), Revelli and Viviani (2013), Capelle-Blancard and Monjon (2014), Revelli and Viviani (2015)
Others - Environmental	"Being Green", "Green Management", "Environmental Characteristics", "Environmental Programs", "Environmental Management", "Environmentally Sustainable Supply Chain Management Practices"	Ambec and Lanoie (2007), Molina-Azorin et al. (2009), Rosenbusch, Bausch, and Galander (2007), Darnall and Sides (2008), Albertini (2013),
Others - Social	"Social Responsiveness", "Social Performance", "Social Initiatives", "Social Capital", "Social Investors", "Socially Irresponsible and Illegal Behaviour", "Human Capital", "Social Capital of Entrepreneurs"	Arlow and Gannon (1982), Ullmann (1985), Roman, Hayibor, and Agle (1999), Margolis and Walsh (2003), Westlund and Adam (2010), Derwall, Koedijk, and Horst (2011), Frooman (1997), Crook et al. (2011), Unger et al. (2011), del Mar Miras-Rodríguez et al. (2015), Stam, Arzlanian, and Elfving (2014)
Others	"Sustainability", "Shareholder Activism", "Governance", "High Performance Work Practices", "Knowledge Transfer", "Innovation", "Entrepreneurial Talent"	Salzmann, Ionescu-Somers, and Steger (2005), Gillan and Starks (2007), Love (2010), Clark, Feiner, and Viehs (2015), Combs et al. (2006), van Wijk, Jansen, and Lyles (2008), Rosenbusch, Brinckmann, and Bausch (2011), Rubera and Kirca (2012), Mayer-Haug et al. (2013)

3.2. Environmental Performance and Financial Performance

Challenges faced by academics regarding the definition of ESG/CSP are also present in the literature that focuses on corporate environmental performance (CEP) specifically. The first dichotomy in terms of approaches in the literature concerns whether (CEP) is a mono-dimensional construct or a multi-dimensional one. As some academics directly construct unique measures to represent firm CEP, some differentiate CEP into two distinct types of performances: an environmental management performance (EMP) that focuses on process-based environmental performance such as firm strategies or policies and an environmental operational performance (EOP) that focuses on measurable environmental outcomes. This is part of a current challenge faced by academics that seems central in the literature: the lack of common measures for CEP. In table 3, I breakdown different approaches used in the CEP literature by authors that constructed a measure for CEP in their work. This table does not intend to provide a comprehensive view of the

entire literature but does help us understand the complexity and variety of measuring CEP for academics, and the absence of common measures and comparability in results.

Looking at Table 3, we observe great variability in the databases that are used, as academics refer to the ASSET4 ESG database, but also to the KLD database, the CDP disclosure database, the Trucost database or use data collected from surveys. Within this databases, different data is used to measure CEP. Xue, Zhang and Li (2019) use ASSET4 ESG to create 41 EMP KPIs and get data on carbon emissions, Trumpp, Endrikat, Zopf and Guenther (2015) to create 32 EMP KPIs and get data on energy consumption, water withdrawal, waste and hazardous waste produced in addition to carbon emissions data, Trumpp & Guenther (2017) only for waste production data, Escrig-Olmedo, Muñoz-Torres, Fernández-Izquierdo, & Rivera-Lirio (2017) to create 62 environmental indicators for the Agri-food industry and Hartmann & Vachon for carbon emissions. The KLD database is used to create variables such as KLR environmental strengths indicators (Post, Rahman and McQuillen, 2015), the KLD Total Strengths and KLD Total Concerns (Delmas, Etzion, and Nairn-Birch, 2013) or to get metrics on pollutants (Ren, He, Zhang & Chen, 2019). Given the additional flexibility that is enabled through the use of survey in terms of variable construction, customized variables are constructed using survey data (Xie & Hayase, 2007; Anton, Deltas, and Khanna, 2004; Bhattacharyya & Cummings, 2015). In addition, samples in these studies mostly consist of firms in different industries and geographies.

Focusing on the literature on the relationship between environmental performance and financial performance and risk allows for deeper understanding of the general issues faced by academics that focus on the concept of ESG/CSP. Their seems to be common agreement on the terminology used in this literature – contrarily to the literature that focuses on ESG as a mono-dimensional concept – as academics do have a common usage for the main concepts linked to

environmental performance such as CEP, EMP and EOP. However, the problem of evaluating and measuring these performances remains. Unlike financial performance or financial risk, which can broadly be considered as mono dimensional concepts as their measurement rely on one metric, environmental performance has a great variety of dimensions. Much like we can observe in Table 3, environmental performance can be measured by using various metrics such as CO2 emissions, GhG emissions, tons of waste, gigajoules of energy consumption. In that sense, we understand the importance of a general framework to define what truly constitutes environmental performance.

3.3.Environmental and Social Data

As we understand from the analysis performed in the previous sections, both the literature that focuses broadly on ESG/CSP and the literature that focuses more specifically on environmental performance suffer from a lack a common definition of what constitutes their central concepts. It is impossible to find true consensus in the literature on what “ESG” means, what are ESG factors and how they can be measured, much like it is impossible to find consensus on the evaluation and measure of corporate environmental performance. In addition to these important challenges, recent literature has also started to look more precisely into the reliability of ESG data, and it seems that academics also face important challenges in this specific domain as well.

Table 3 - Corporate Environmental Performance in the literature

Study title	Database	EMP	EOP	Final Metrics used	Sample studied
Xue, H., Zhang, Z. and Li, P., 2020. Corporate environmental performance, environmental management and firm risk. <i>Business Strategy and the Environment</i> , 29(3), pp.1074-1096.	ASSET4 ESG	Equal-weighted average of 41 Yes/No KPIs	Carbon Performance (carbon emissions per net sales)	Total CO2 and CO2 equivalents emissions in tonnes divided by net sales	UK,924 firm year in manufacturing, 708 service industry
Ren, S., He, D., Zhang, T. and Chen, X., 2019. Symbolic reactions or substantive pro-environmental behaviour? An empirical study of corporate environmental performance under the government's environmental subsidy scheme. <i>Business Strategy and the Environment</i> , 28(6), pp.1148-1165.	KLD	Dummy Variable (passed ISO14001 or have not)	Pollutants of three wastes (gas waste, water waste and solid waste)	Equal-weighted dummy variables for SO2, NOX, CO2, soot and dust, COD, ammonia nitrogen, solid waste, and energy consumption	China
Xie, S. and Hayase, K., 2007. Corporate environmental performance evaluation: a measurement model and a new concept. <i>Business Strategy and the Environment</i> , 16(2), pp.148-168.	Survey	48 measurement items for 4 management performance indicators	18 measurement items for 2 operational performance indicators	Answers from the survey of 68 measurement items	Japan, the electrical machinery and instrument manufacturing sector
Anton, W.R.Q., Deltas, G. and Khanna, M., 2004. Incentives for environmental self-regulation and implications for environmental performance. <i>Journal of environmental economics and management</i> , 48(1), pp.632-654.	IRRC from surveys	13 dummy variable EMPs			S&P 500 firms included in the Corporate Environmental Profile Directories
Post, C., Rahman, N. and McQuillen, C., 2015. From board composition to corporate environmental performance through sustainability-themed alliances. <i>Journal of Business Ethics</i> , 130(2), pp.423-435.	KLD	Year-to-year increases in 7 KLD environmental strengths			US Oil and Gaz companies that were listed in the 2009 Forbes.com Special Report The Global 2000
Trumpp, C., Endrikat, J., Zopf, C. and Guenther, E., 2015. Definition, conceptualization, and measurement of corporate environmental performance: A critical examination of a multidimensional construct. <i>Journal of Business Ethics</i> , 126(2), pp.185-204.	ASSET4 ESG	32 Yes/No KPIs	5 continuous KPIs	Energy consumption in gigajoules, water withdrawal in cubic meters, CO2 and CO2 equivalents emissions in tons, waste produced in tons, hazardous waste produced in tons	Mostly Europe Manufacturing industry and US
Trumpp, C. and Guenther, T., 2017. Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. <i>Business Strategy and the Environment</i> , 26(1), pp.49-68.	ASSET4 for Waste, CDP for GHG emissions	-	GHG emissions and Waste	Negative GHG emissions (divided by sales, negative amount of waste produced by a firm divided by sales)	International, manufacturing and Service Industries
Escrig-Olmedo, E., Muñoz-Torres, M.J., Fernández-Izquierdo, M.Á. and Rivera-Lirio, J.M., 2017. Measuring corporate environmental performance: A methodology for sustainable development. <i>Business Strategy and the Environment</i> , 26(2), pp.142-162.	ASSET4	62 Indicators taken from ASSET4			Agri-food industry
Hartmann, J. and Vachon, S., 2018. Linking environmental management to environmental performance: The interactive role of industry context. <i>Business Strategy and the Environment</i> , 27(3), pp.359-374.	ASSET4	Carbon Emission Reduction	-	Absolute Carbon Emissions	Publicly traded European manufacturing firms listed in the Asset4 database
Delmas, M.A., Etzion, D. and Naim-Birch, N., 2013. Triangulating environmental performance: What do corporate social responsibility ratings really capture?. <i>Academy of Management Perspectives</i> , 27(3), pp.255-267.	KLD, Trucost, SAM	Creation of 5 variables: KLD Total Concerns, KLD Total Strengths, Trucost Total Damage, SAM Eco-efficiency, SAM Reporting			741 firms, no detail on industry or geography
Bhattacharyya, A. and Cummings, L., 2015. Measuring corporate environmental performance stakeholder engagement evaluation. <i>Business Strategy and the Environment</i> , 24(5), pp.309-325.	CDP Disclosure	GHG Emissions			127 companies, various industries and geographies
Bhattacharyya, A. and Cummings, L., 2015. Measuring corporate environmental performance stakeholder engagement evaluation. <i>Business Strategy and the Environment</i> , 24(5), pp.309-325.	Survey	36 MPIs and 11 OPIs			A final sample size of 150 (from chemical, industrial engineering and pharmaceutical & biotech) from Australia

In a study on this subject, S. Kotsantonis and G. Serafeim (2019) hand-collected information on Employee Health and Safety data from the latest sustainability reports of 50 large publicly listed companies across a variety of sectors and found that this information was reported in 20 different ways by these randomly selected firms, using different terminology and units of measure (Kotsantonis and Serafeim, 2019). This is a significant challenge both in terms of aggregation of data as well as comparability, as it poses both a challenge within a firm to create an aggregate measure of their “Employee Health and Safety” performance and between firms to understand how and to what extent their performance might be compared to their peers. It seems that the definition challenge that can be identified in the academic literature on the subject of ESG/CSP is common to practitioners as well, as we can understand that “Health and Safety concerns” constitute only one of the many dimensions of ESG/CSP.

This in turn causes a second limitation of ESG data relative to lack of data. The lack of standardized data means that even when regulation requires a certain type of disclosure, ESG data providers must deal with the absence of common indicators. As an example, referring to measures reported in Kotsantonis and Serafeim (2019), while a set of companies report on the “Number of accidents with fatal consequences”, another might simply report on “Injury rates”. Different methodologies are used by ESG data providers to fill this gap, but it is quite clear that the provided data is then more of an approximation than what the data would indicate in reality. Such approximations would in turn lead to less precise results. Kotsantonis and Serafeim (2019) address another key issue relative to ESG scores provided by ESG data providers on the subject of benchmarking. On this issue, the authors demonstrate how the definition of peer groups for benchmarks and the lack of transparency on the choices made by ESG data provider on their selection of peer groups leads to market-wide inconsistencies in ESG metrics and data.

Finally, another field of study addresses the lack of agreement between ESG data providers on the ESG scores of companies. In a study focusing on the ESG ratings of five ESG rating companies, Berg, Koelbel and Rigobon (2019) find that the correlation between ESG ratings is of 0.61 in average, and goes from 0.42 to 0.73, when correlation between credit ratings for Moody's and Standard & Poor's are correlated at 0.99. The authors conclude that given these disparities, ESG performance is therefore unlikely to be properly reflected in asset prices, discourages firms from improving their ESG ratings and poses a significant problem in empirical research as using different ESG datasets would lead to obtaining different results. This concept is verified by Diebecker, Rose and Sommer (2019) that use three distinct datasets to verify the effect of corporate sustainability performance on the cost of equity and find a negative effect when using the MSCI ESG KLD STATS dataset and the original ASSET4 dataset but do not when using the Thomson Reuters ESG dataset²⁵. In his study on the subject, Doyle (2018) found that larger companies obtained higher ESG ratings, that there was a geographical bias toward companies in regions with high reporting requirements and that ESG agencies oversimplify industry weighting and company alignment, providing multiple concise examples of companies receiving vastly different ratings from different rating agencies. Yang (2019), even finds evidence of a link between green washing and ratings inflation. First controlling for firm size, the author finds that environmental and social ratings are largely uninformative about future corporate bad behavior. However, without controlling for size, the author finds that better environmental ratings predict more future bad behavior. Finally, findings of Christensen, Serafeim and Sikochi (2019) suggest that firms that

²⁵ MSCI ESG KLD STATS and the Thomson Reuters ESG dataset are two of the leading ESG data providers to date. Thomson Reuters purchased the Asset 4 dataset and integrated it in some of its data products in 2009. Interestingly enough, even though it could be considered that the Asset 4 data was not modifiable, it provided different results in the context of this study, demonstrating how ESG data could still have contain a subjective aspect to date.

provided greater ESG disclosure were faced with greater disagreement across ESG rating agencies.

Overall, this recent literature seems to indicate that ESG data is not yet reliable, and that different data providers are susceptible to provide different ESG data on similar firms. This means that most studies to date that rely on ESG data and ESG data providers to determine specific relationships between ESG factors and financial performance would probably have different results if they used a different source of data, even if they kept similar samples. This constitutes yet another challenge for academics of the field that already have to face the lack of common definitions, concepts and metrics in this field. Interestingly, it seems that it also a lack of common definitions, concepts and metrics that is the cause for this disparity in terms of data between data providers. Once again, we understand how it is becoming increasingly urgent to develop a common approach to ESG and sustainability in order for the industry – both practitioners and academics - to be able to reliably evaluate the extra-financial aspects of firms, assets and projects.

3.4. The absence of ESG factors in asset pricing

With close to 40% of global assets under management using sustainable investing strategies according to the 2018 GSIA report and the organization's underlying data, it could be expected that environmental, social and governance factors had started to impact asset pricing and the corresponding literature in financial research. Furthermore, given the increase urgency relative to global warming, environmental factors specifically could also be considered to have an increasing impact on asset pricing.

Looking at firm environmental performance and criteria from an asset pricing perspective does make sense, as it could be reasonable to expect firms with higher environmental performance

not only to have better financial performance or to be less exposed to risk, but most importantly, such highly performing environmental firms would inevitably create interest from market participants that wish to use their investment to help mitigate climate change.

Historically, this has not been the case. In a renowned study in the asset pricing literature, Harvey, Liu and Zhu (2015) identify more than 300 papers in the literature that had been put forward new factors that explained the cross-section of expected stock returns. The considered period for the identification of these factors starts from the original market return factor of Sharp (1964) and stop at the year 2012. Out of more than 300 papers, only 1 paper mentions an environmental factor (Brammer, Brooks, and Pavelin, 2006) out of only 6 papers in total that mention environmental or social factors with factors such as labor income (Campbell, 1996; Jagannathan and Wang, 1996; Gómez, Priestley, and Zapatero, 2016²⁶), local unemployment (Korniotis and Kumar, 2009), labor unions (Chen, Kacperczyk and Ortiz-Molina, 2011) and employment and community indicators (Brammer, Brooks, and Pavelin, 2006). The environmental factor developed by Brammer, Brooks and Pavelin (2006) is built using the EIRIS database²⁷. Authors create 3 categories of environmental performance (environmental policies, environmental management systems and management reporting) which are all given a grade between 0 to 4 for a total environmental responsibility score out of 12. There is no reason not to think this approach to estimating environmental performance does not suffer from the same issues mentioned in the previous sections.

The sample used in Harvey, Liu and Zhu (2015) stops in 2012, and represents a period where studying the extra-financial performance of the firm was less popular, and there was much less

²⁶ This paper is cited in Harvey, Liu and Zhu (2015) as a working paper in 2012, but the paper was later published in the *Journal of Financial and Quantitative Analysis* in 2016.

²⁷ EIRIS was an ESG research provider that also add a database of ESG data available

focus on the climate urgency. However, Harvey and Liu (2019) provide an update of these factors and find 159 additional factors between 2012 and 2018. Once again, environmental and social factors are almost absent from this list of new factors, with one paper putting forward labor income as a risk factor (Gomez, Priestley and Zapatero, 2016) which was already mentioned as a working paper in 2012, and another creating a boycott risk factor based on the value-weighted return of the tobacco, alcohol, fossil, weapons and gaming firms (Luo and Balvers, 2017). Since 2019, the author has not found any additional literature in top ranking financial journals that put forward one or multiple environmental factors as possible risk factors for asset pricing. This could possibly mean that challenges faced by academics in the field prevent them from being able to provide robust empirical research on this subject using currently available environmental or social data. The absence of environmental or social risk factors in what has been described in the literature as a “zoo of factors” that attempt to explain the cross-section of stock returns provides further evidence that academics and practitioners lack a common framework and comparable metrics to evaluate sustainability risk precisely.

4. The EU’s Framework

The aforementioned challenges faced by actors of the sustainable finance industry seem to be understood by public entities and policy makers. Governments and intergovernmental entities have increasingly been focusing on creating a framework of incentives, guidelines and regulations to encourage sustainable finance. In 2015, the global community represented by the United Nations, has developed both a framework for global sustainable development through the Sustainable Development Goals (SDGs) and a framework specific to the climate change challenge through the Paris Agreements and the objective of limiting global warming to “well-below” 2 °C (UN, 2015; UNFCCC 2015). In the context of the Paris Agreement, every nation had to put forward “nationally

determined contributions” (NDCs) and strengthen their efforts in the years to come. Even though these agreements provided a useful framework for governments to make progress, they were not legally binding, and did not go into detail on specific necessary measures that had to be implemented to face climate change.

In 2016, the European Commission started to focus on implementing a more specific framework for sustainable finance with the creation of a High-Level Expert Group (HLEG) on sustainable finance. The group provided advice to the Commission on steering public and private capital towards sustainable investments, on protecting the stability of the financial system from environmental risks and on deploying related policies throughout the European Union. Its final report was published in January 2018 (HLEG, 2018). This work led to the creation of the “Action Plan on Financing Sustainable Growth”, an action plan focused predominantly on the EU private sector and each category of private actors. The action plan has three objectives:

- reorient capital flows towards sustainable investment, in order to achieve sustainable and inclusive growth
- manage financial risks stemming from climate change, environmental degradation and social issues
- foster transparency and long-termism in financial and economic activity

For each objective, a series of specific actions were determined, adding up to a total of 10 distinct actions. Even though this is not specifically articulated by the European Commission in the official communication of the action plan, each action seems to correspond to one specific type of actor. Objective 1 has five underlying actions which apply to actors that provide supporting functions to financial market participants: (1) The European Commission itself through the creation of an EU classification system for sustainable activities (2) providers of labels and

standards (3) EU financial institutions (4) financial advisors (5) benchmark administrators. Objective 2 has 3 underlying actions for (6) credit agencies and market research providers (7) institutional investors and asset managers and (8) banks and insurance companies. Objective 3 has 2 underlying actions for corporations through (9) corporate disclosure and (10) corporate governance. The objectives, sub-categories of objectives and corresponding actions of the official European Commission communication are summarized in Table 4, which also contains the different commitments that are taken by the Commission concerning each action.

Amongst these different actions, the European Commission emphasized the essential role of the EU classification system for sustainable activities, which it refers to as the EU Taxonomy. The official communication refers to this taxonomy as “the most important and urgent action of this Action Plan” and provides a diagram to illustrate how the action plan depends on the development of the EU taxonomy and its integration into EU legislation (European Commission, 2018). This diagram is available in Figure 1.

The EU Taxonomy tackles a challenge that has been central for actors of sustainable finance which we have explored in the previous section: to create a regulatory framework to define and delineate what constitutes a sustainable investment. Today, the sustainable finance industry is composed of many different actors that have developed their own definition of what sustainable investments are, and there still is to date an absence of common definitions and approaches on the subject. Creating a clear framework developed in collaboration with expert members of this industry that would progressively become enforceable by the Commission is a way of aggregating the different approaches of sustainable finance in a single comprehensive environment and creating a common language as well as comparability between actors and sustainable financial products.

As the new European Commission presidency started in 2019, the European Green Deal - a growth strategy to make Europe the first climate neutral continent by 2050 - was put forward, followed by a corresponding European Green Deal Financial Plan in January 2020²⁸. This new financial plan focuses on mobilizing at least € 1 trillion of sustainable investments over the next decade, on creating a framework to enable such investments for both private and public actors and on supporting project creators that could benefit from such investments. Supplementing these efforts, the European Commission's approach also consists in putting forward both a climate tracking methodology and a method to assess the environmental, climate and social impact of private investors. Financial market participants will have to implement such methodologies if they wish to benefit from European funds. These methodologies will be developed by relying on the EU Taxonomy (European Commission, 2020a). Throughout the different EU communications, it is made quite clear that the EU Green Deal Investment Plan supplements the EU Action plan on Sustainable Finance but does not replace it. Even though the plan mentions that the European Commission will renew its sustainable finance strategy in Autumn 2020, the taxonomy will keep a central role in this strategy and the progress made in the context of the Action Plan on Sustainable Finance will be used as a foundation for the new strategy²⁹. The EU Taxonomy therefore represents a new approach to sustainable finance that provides both a clear, material, metrics-based definition - developed by a large group of both private and public experts – of what constitutes sustainable finance as well as a legally-backed regulatory framework. We investigate the EU Taxonomy in more detail in the following section.

²⁸ See https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_24

²⁹ This is also detailed on the commission's website. See https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_24

5. The EU Taxonomy

The beginning of the new European Commission presidency was also an important milestone for the EU taxonomy. In December 2019, the proposal for a regulation on the establishment of a framework to facilitate sustainable investment (what is referred as the ‘Taxonomy Regulation’) was approved by the European parliament and the European Council (Council of the European Union, 2019a), and in March 2020, the final report on the EU Taxonomy was published by the TEG (European Commission, 2020c). In order to develop its classification system, the TEG first identified six environmental objectives:

1. Climate change mitigation
2. Climate change adaptation
3. Protection of water and marine resources
4. Transition to a circular economy
5. Pollution prevention and control
6. Protection and restoration of biodiversity and ecosystems

The approach put forward by the EU Taxonomy was then developed by focusing on economic activities and identifying how each economic activity performed by a firm was or could be aligned with these objectives. In order for this to be the case, a defined set of criteria, both general and specific to the activity, are to be respected. An economic activity is then considered as “Taxonomy-aligned” if it:

- Makes a substantive contribution to at least one of the 6 objectives according to activity-specific technical criteria
- Does no significant harm (DNSH) to other objectives according to the taxonomy’s framework

- Complies with the minimum social safeguards of the OECD Guidelines on Multinational Enterprises (OECD,2011) and the UN Guiding Principles on Business and Human Rights (United Nations, 2011)

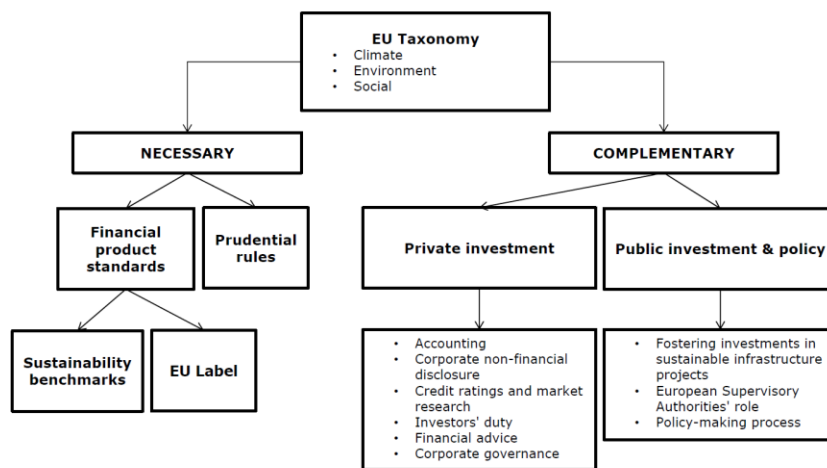


Figure 3 - The EU Taxonomy and the EU Action Plan on Sustainable Finance

Given the complexity of their task and the growing urgency of climate change action, the TEG first focused on what it considered to be the most urgent objectives of climate change mitigation and climate change adaptation (referred as Objective 1 and Objective 2).

In the Taxonomy report of March 2020, the technical expert group defined the technical criteria for 68 economic activities for both climate change mitigation and climate change adaptation. These 68 economic activities were selected amongst 8 economic sectors that together represented 93.5% of total European carbon emissions and were identified by the expert group as the most likely to substantially contribute to climate change mitigation in their respective sectors. These activities were specifically selected using the NACE classification system³⁰.

³⁰ The Statistical classification of economic activities in the European Community (abbreviated as NACE as the original name of the classification system was in French) is an industry standard classification system that uses four levels to categories economic activities in 21 sections (level 1), 88 divisions (level 2), 272 groups (level 3) and 615 classes (level 4).

Table 4 - The European Union Action Plan

Objective	Section	Action	Main Commitments
Objective 1: Reorienting capital flows towards a more sustainable economy	A unified classification system for sustainable activities	1. Establishing an EU Classification System for Sustainability Activities	<ul style="list-style-type: none"> - Progressive development of an EU taxonomy - Focus on climate mitigation and climate change adaptation
	Standards and labels for sustainable financial products	2. Creating Standards and Labels for Green Financial Products	<ul style="list-style-type: none"> - Report on green bond standard and prospectus of green bond issuances - Exploration of an EU Ecolabel
	Fostering investment in sustainable projects	3. Fostering Investment in Sustainable Projects	<ul style="list-style-type: none"> - Further measures to improve instruments aiming at sustainable investment support - Idea of establishing a single investment fund integrating all EU market-based instruments
	Sustainability considerations in financial advice	4. Incorporating Sustainability when Providing Investment Advice	<ul style="list-style-type: none"> - Amend the MiFID II and IDD delegated acts to ensure that sustainability preferences are incorporated in the suitability assessment - Invite ESMA to include provisions on sustainability preferences in its guidelines on the suitability assessment
	Sustainability benchmarks	5. Developing Sustainability Benchmarks	<ul style="list-style-type: none"> - Adoption of delegated acts on transparency of benchmark methodologies - Put forward an initiative to harmonise benchmarks comprising low-carbon issuers
Objective 2: Mainstreaming Sustainability into risk management	Sustainability in market research and credit ratings	6. Better Integrating Sustainability in Ratings and Research	<ul style="list-style-type: none"> - Explore the merits of amending the Credit Rating Agency Regulation - Invite ESMA to assess current sustainability-related practices in the credit rating market - Comprehensive study on sustainability ratings
	Institutional investors and asset managers' sustainability duties	7. Clarifying Institutional Investors and Asset Managers' Duties	<ul style="list-style-type: none"> - Proposal aiming at requiring (1) institutional investors and asset managers to integrate sustainability considerations in the investment decision-making process and (2) increase transparency to end-investors
	Prudential requirements for banks and insurance companies	8. Incorporating Sustainability in Prudential Requirements	<ul style="list-style-type: none"> - Explore the feasibility of including environmental risk in institutions' risk management policies and potential calibration of capital requirements of the Capital Requirement Regulation and Directive - Invite EIOPA to provide an opinion on the impact of prudential rules for insurance companies on sustainable investments
Objective 3: Fostering transparency and long-termism	Disclosure and accounting	9. Strengthening Sustainability Disclosure and Accounting Rule-Making	<ul style="list-style-type: none"> - Fitness check of EU legislation on public corporate reporting - Revision of guidelines on non-financial information - Creation of the European Financial Reporting Advisory Group (EFRAG) - Assess the impact of new or revised IFRSs on sustainable investments
	Corporate governance and undue capital market short-termism	10 - Fostering Sustainable Corporate Governance and Attenuating Short-termism in Capital Markets	<ul style="list-style-type: none"> - Assess the possible need to require corporate boards to develop and disclose sustainability strategy - Collect evidence on short-term pressure from capital markets on corporations

5.1. Activity-specific Technical Criteria

For each of these identified economic activities, a specific set of technical criteria was defined by the expert group. These criteria have two distinct formats. (1) They can take the form of principles that have to be respected. The expert group's definition of principles is the following:

“Principles: The underlying rationale for how the activity will result in a substantial contribution and/or avoidance of significant harm to the environmental objective in question.”

(2) They can also take the form of metrics and thresholds. The expert group's definition of metrics and thresholds is the following:

“Criteria: including both metrics and thresholds: The method(s) by which the environmental performance of the economic activity will be measured, including defining the boundary for this measurement and the qualitative or quantitative conditions which must be met to enable the performance of the activity in a way that is considered environmentally sustainable.”

If one of the firm's economic activity is listed amongst these 68 economic activities, the company will need to disclose whether or not it is aligned with these different activity-specific criteria. If it is not able to demonstrate that it is, the firm's economic activity is considered not to be Taxonomy-aligned. If the activity meets the expert groups criteria, the firm needs to further demonstrate that the activity does no significant harm (DNSH) to other objectives.

5.2. Activity-specific DNSH Criteria

Once the technical criteria for an economic activity has been met, the firm still needs to demonstrate that its economic activity is not detrimental to other objectives. By analyzing the technical annex of the taxonomy, we understand that these criteria are also mostly activity-specific. Using the March 2020 taxonomy report as the source of our analysis, we also find that these are

still mostly process-based approaches where no specific metric or threshold needs to be specifically respected. However, the expert group does mention that these may vary as further work is done on developing the technical criteria for the last 4 objectives of the Taxonomy.

5.3. Meeting minimal social safeguards

The final required step for the activity to be Taxonomy-aligned is for it to comply minimum social safeguards. More precisely, these activities will have to be “in alignment with the OECD Guidelines for Multinational Enterprises and UN Guiding Principles on Business and Human Rights, including the International Labour Organisation’s (‘ILO’) declaration on Fundamental Rights and Principles at Work, the eight ILO core conventions and the International Bill of Human Rights”. The OECD Guidelines for Multinational Enterprises corresponds to a set of legally nonbinding recommendations providing a series of principles and standards encouraging responsible business conduct for multinational corporations operating in and from countries in the OECD (OECD, 2011). The UN Guiding Principles also represent a set of legally nonbinding principles encouraging better business conduct. It focuses more specifically on human rights and is addressed to businesses of all sizes and any sector throughout the world. Any firm that wishes to have one of its economic activities be taxonomy-aligned needs to conduct due diligence to show alignment with these principles.

5.4. General Remarks

Interestingly, the distinction between these two types of approaches resembles that which currently exists in the literature on corporate environmental performance between the process-based environmental performance EMP approach that focuses on general firm policies and the outcome-based environmental performance EOP approach that focuses on measurable environmental outcomes.

This parallel is even more striking when looking at the difficulties that were faced by academics concerning the differences in industries and geographies of their studied samples. Looking at the literature and Table 3, it would seem that each industry and geography required different measures for environmental performance. This notion is confirmed by the approach taken by the European Commission and the technical expert group. Focusing on the geographical region it represents, the European Commission developed its 6 environmental objectives in accordance with the environmental priorities of its Member States, but also by taking the Sustainable Development Goals (SDGs) in consideration. Composed mostly of developed economies that had less stringent issues regarding purely social SDGs such as poverty or hunger, the European Commission chose to take the lead on environmental objectives which can more easily be addressed by developed economies. Its approach of environmental issues was then decomposed into different economic activities that all had a specific role in the climate and environmental urgency, and developed measures at the scale of these activities. Analyzing the March 2020 taxonomy-report is sufficient to understand the underlying objective of the European Commission: to create a clear classification system in which the participation of each individual economic activity in each of the six environmental objectives is identifiable and to create a transparent and precise link between economic performance and measurable, comparable corporate environmental performance (European Commission, 2020c). The European Commission is also progressively constructing a legal basis in order to enforce this new approach to finance. We explore this legal basis in the following section.

6. The Taxonomy Regulation

Based on the technical expert group's previous work, the Taxonomy Regulation brings the EU Taxonomy into practice and enforces its use by three groups of users:

- Financial market participants offering financial products in the EU
- Large companies who are already required to provide non-financial statement under the Non-Financial Reporting Directive (Official Journal of the European Union, 2014)

31

- EU and Member States when setting public measures, standards or labels

6.1. Financial Market Participants

The regulation states that specified financial market participants will have to disclose, for each relevant product:

- how and to what extent they have used the Taxonomy in determining the sustainability of the underlying investments;
- to what environmental objective(s) the investments contribute; and
- the proportion of underlying investments that are Taxonomy-aligned, expressed as a percentage of the investment, fund or portfolio. This disclosure should include details on the respective proportions of enabling and transition activities, as defined under the Regulation.

As it is mentioned in the TEG's report, Taxonomy-related disclosure will be included in the Regulation on Sustainability-Related Disclosures in the Financial Services Sector (commonly referred as the 'Disclosure Regulation'³²). In this context, financial market participants will have to provide information on their alignment with the EU Taxonomy in pre-contractual disclosures,

³¹ The Non-Financial Reporting Directive lays down rules on necessary non-financial disclosure that need to be performed by large companies in the European Union, and states that companies have to include non-financial statements in their annual reports starting from 2018.

³² The Disclosure Regulation was adopted by the European Union in 2019 and "lays down harmonized rules for financial market participants and financial advisers on transparency with regard to the integration of sustainability risks and the consideration of adverse sustainability impacts in their processes and the provision of sustainability - related information with respect to financial products (Official Journal of the European Union, 2019a).

through their websites and in periodic reports. Information disclosed on websites are the most detailed: “Description of the environmental or social characteristics or objectives of the fund, information on the methodologies used to assess, measure and monitor the characteristics or impact of the underlying investments, data sources and screening criteria”. The disclosure regulation goes even further by requiring that any financial product that has sustainable investment as an objective or promotes any form of environmental or social characteristic of an investment must complete Taxonomy disclosures. For any other type of financial product, financial market participants must either complete the Taxonomy disclosure or carry the following disclaimer “the investment(s) underlying this financial product do not take into account the EU criteria for environmentally sustainable investments”. Any market participant that offers a financial product that has an environmental or social aspect to it will have to disclose to what environmental objective(s) the investment contributes and the proportion of the underlying investments that are Taxonomy-aligned, expressed as a percentage.

By December 31st, 2021, every specified financial product in the EU that addresses climate change mitigation and adaptation - objectives 1 and 2 of the Taxonomy - will have to disclose the specific percentage of the investment, fund or portfolio that is taxonomy-aligned. By December 31st, 2022, every specified financial product in the EU that addresses any of the taxonomy’s objectives will have to disclose the specific percentage of the investment, fund or portfolio that is taxonomy-aligned.

6.2. Companies

The Taxonomy will also require specified companies to report on the proportion of their activity that is taxonomy-aligned. This covers at the minimum large public-interest companies with more than 500 employees, including listed companies, banks and insurance companies that

already had to disclose extra-financial information in the context the Non-Financial Reporting Directive. More specifically, these companies will have to disclose the percentage of their turnover that is aligned with the Taxonomy as well as the percentage of its capital expenditures and, in some cases, even operational expenditures, that is aligned with the Taxonomy. Disclosure on financial year 2021 will have to be performed in the course of financial year 2022.

The Taxonomy Regulation and the Disclosure Regulation are an attempt to create a transparent financial market in which each major actor is attributed a corresponding form of sustainability rating. This is done through the creation of a common framework for all these actors, that provides a common definition of sustainable investing, with common, industry-specific processes and metrics to respect. In the European Commission's framework, sustainability is transformed into binary concept in which activities are either aligned or not. There is no more room for grey areas and actor specific, subjective definition of what constitutes sustainable business practices.

This, in turn, could create a new market for sustainability in which companies can compare their taxonomy-alignment with that of their peers, within or outside their own industry. As the focus on climate mitigation and adaptation grows as global warming becomes an increasingly present and central systemic risk, there will undoubtedly be increasing pressure for actors to optimize their taxonomy-alignment. In Table 5, I show an example in which two companies perform similar economic activities but originate different proportions of their turnover from these activities. Both companies have similar taxonomy-alignment for their economic activities. However, one company is 11% taxonomy-aligned but is below the overall average alignment a company with similar structure should have, when the other is 9% taxonomy-aligned, which corresponds to the average alignment a company with similar structure should have. If both

companies are equally profitable, would an investor prefer company A or company B? From this simple exercise, we can already understand how some investors would choose company A over Company B by favoring overall taxonomy-alignment when others would use industry averages to select their investments. Furthermore, we also understand how Company A would focus on improving alignment for Economic Activity 1 when Company would choose Economic Activity 3. Going one step further, we also understand how a financial market participant that offers a financial product consisting in an investment portfolio composed of these two companies could then weight its portfolio in order to optimize its taxonomy-alignment in different economic activities.

6.3. Pressure on financial-market participants

Through its structure, the EU taxonomy puts pressure on financial market participants that offer financial products in the EU in two essential ways:

- Financial market participants will have to disclose the proportion of their financial products that are aligned with the taxonomy before underlying companies
- Not all companies will have to disclose information on their alignment with the Taxonomy, in which case financial market participants will have to gather this information themselves

This, in turn, should also put pressure on companies, as financial market participants will necessarily develop a preference for companies that provide the most information related to their taxonomy alignment to them. Furthermore, given the central role the EU Taxonomy will play for EU institutions and other public-entities, this will provide further pressure for companies to start

working on their taxonomy-related disclosures, specifically given the fact that the EU plans to invest more than EUR 1 trillion in sustainable investment in the next 10 years.

Table 5

Company A

	Economic Activity 1	Economic Activity 2	Economic Activity 3
Proportion of turnover	65%	20%	15%
Alignment with Taxonomy	10%	15%	7%
Industry Average	12%	22%	5%
Overall Taxonomy-alignment		11%	
Average Taxonomy-alignment		13%	

Company B

	Economic Activity 1	Economic Activity 2	Economic Activity 3
Proportion of turnover	10%	20%	65%
Alignment with Taxonomy	10%	15%	7%
Industry Average	12%	22%	5%
Overall Taxonomy-alignment		9%	
Average Taxonomy-alignment		9%	

7. Implementation of sustainability measures in the rest of the market

The relationship between financial market participants and companies will be central to the development of the sustainable finance framework set out by the European Commission, and these actors will be the first to implement measures to align their activities with the Taxonomy. However, looking back at the Action Plan on Sustainable Finance and its 10 actions, it seems that all actors of the market will have to start adapting as well. As the European Commission adopted the ‘Taxonomy Regulation’ and the ‘Disclosure Regulation’ oriented at (1) establishing an EU Classification System for Sustainability Activities (4) Incorporating Sustainability when Providing Investment Advice (7) Clarifying Institutional Investors and Asset Managers’ Duties and (9)

Strengthening Sustainability Disclosure and Accounting Rule-Making, other progress was made with other actions and actors.

7.1. The Benchmark Regulation

The Benchmark Regulation refers to another Regulation approved by the European Parliament and the Council in November 2019 to incorporate sustainability in EU Benchmarks. Two new benchmarks were created:

- EU Climate Transition Benchmarks that will consist in portfolios that are on a decarbonization trajectory.
- Paris-aligned Benchmarks that consist in portfolios that have a carbon emission trajectory aligned with the objectives of the Paris Agreement.

In addition, the regulation also states that:

- By the 30th of April 2020, benchmark administrators will have to provide ‘an explanation of how the key elements of the methodology (...) reflect ESG factors for each benchmark or family of benchmarks.
- ‘By 1 January 2022, administrators which are located in the Union and which provide significant benchmarks (...) shall endeavour to provide one or more EU Climate Transition Benchmarks.’
- ‘By 31 December 2021, benchmark administrators shall (...) include in their benchmark statement an explanation of how their methodology aligns with the target of carbon emission reductions or attains the objectives of the Paris Agreement.’

This means that, much like it will be the case for large firms in the European Union, benchmarks providers will have to provide financial market participants with information on the

alignment of their products with climate objectives, and will create a new market for climate-aligned benchmarks that will also require firms to provide information regarding their carbon emissions and alignment with the EU taxonomy.

7.2. EU Labels

The European Commission has been working on the creation of an EU Green Bond Standard (European Commission, 2019c), and the TEG has published a usability guide for the EU Green Bond Standard in March 2020 (European Commission, 2020c) to start providing some guidance on the use of the standard. As this article is being written, the Commission is exploring the possibility of a legislative initiative for an EU Green Bond Standard. The European Commission has also started to work on an EU Ecolabel³³ scheme for retail funds, savings and deposits.

7.3. EU initiatives to mobilize private investments

In the context of the European Green Deal Investment Plan, the European Commission will dedicate 25% of the EU budget to Climate and Environment, which corresponds to €503 billion. Other EU funds will also participate in triggering investments, such as the EU Emissions Trading System (ETS) funds (€25 billion), national co-financing structural funds (€114 billion) and the Just Transition Fund (€143 billion over 10 years).

In addition, the EUInvest program will help mobilize private investments. The EUInvest program is the realization of the European Commission's 'Idea of establishing a single investment fund integrating all EU market-based instruments', which it had communicated in the EU Action plan on Sustainable Finance in 2018. It brings together the European Fund for Strategic Investments and 13 other EU financial instruments and allows private investors that wish to

³³ See <https://ec.europa.eu/environment/ecolabel/>

perform sustainable investments in the EU to benefit from the InvestEU Guarantee. Through this mechanism, the European Commission plans to mobilize €279 billion in private sustainable investment.

7.4. Credit Rating Agencies

To date, no strict regulation on the sustainability disclosure of credit rating agencies has been discussed by the European Commission. The latest progress on the subject was made through guidelines published by ESMA in July 2019, and applicable starting April 2020. These guidelines require increased transparency on whether ESG factors were a key driver in the credit rating process and provide some guidance regarding what should be disclosed when a credit rating is issued (ESMA, 2019). The European Commission will perform a report on progress on this subject in 2021.

7.5. Insurances

The European Commission has also addressed the subject of integrating sustainability in the European insurance industry. The European Insurance and Occupation Pensions Authority (EIOPA) has been mandated to provide its technical advice on the integration of sustainability risk and factors in the Solvency II directive in the context of the 2020 Review of the directive. In September 2019, the EIOPA had already identified additional practices that should be adopted by insurance companies to guarantee that companies considered sustainability risk in their risk management approaches. As this paper is being written, the 2020 Review of Solvency II is meant to occur at the end of December 2020.

7.6. *Banking Prudential Framework*

Some actions were taken by the European Commission related to the integration of ESG risks in banks' capital requirements regulations. However, the European Commission stated in its consultation on the renewed sustainable finance strategy in April 2020 that “given the new objectives under the European Green Deal, it can be argued that the efforts in this area need to be scaled up in order to support a faster transition to a sustainable economy and increase the resilience of physical assets to climate and environmental risks. Integrating sustainability considerations in banks' business models requires a change in culture which their governance structure needs to effectively reflect and support.” (European Commission, 2020b)

7.7. Pension Providers

Some measures relative to sustainability reporting and ESG integration by EU pension providers have been applied in 2016 and 2017. However, according to a stress test on Institutions for Occupational Retirement Provision (IORPs) ran by the EIOPA in 2019 only about 30% of IORPs in the EU manage ESG-related risks relative to their investments. Moreover, while most of these EU pension providers claimed to have taken appropriate steps to identify ESG risks relative to their investments, only 19% assessed the impact of ESG factors on investments' risks and returns. The European Commission plans to review the IORP II Directive by January 2023. In 2019, EIOPA published an opinion on the subject of ESG risks faced by the IORPs. The Commission has not communicated on any specific regulatory action on the subject.

7.8. Corporate Governance

In the context of its action plan on sustainable finance, the European Commission also focused on corporate governance and short-termism. In June 2019, it requested that the European Supervisory Authorities – composed of the European Banking Authority (EBA), the European Securities and Markets Authority (ESMA) and the European Insurance and Occupational Pensions

Authority (EIOPA) each publish advice on “undue short-term pressure from the financial sector on corporations”.

The EBA did not find evidence of short-termism. It committed to assess whether a dedicated prudential treatment would be required for assets or activities that were aligned with the new European objectives. Its main recommendations included adding sustainability considerations into the European banking sector legislation, continue to develop a framework for disclosure on long-term risk that includes ESG-related information for both banks and corporations, and encouraged transparency and data availability between banks and corporations on sustainability-related issues (EBA, 2019). ESMA finds that there is short-term focus from financial analysts. ESMA’s recommendations addressed improving issuers’ ESG disclosures to allow for comparability and reliability and creating an international standard for ESG disclosures (ESMA, 2019b). Finally, EIOPA did not find evidence of undue short-termism for insurance and institutions for occupational retirement provision and recommended that a framework for long-term investments be created, along with the creation of long-term performance benchmarks that focus on long-term value creation (EIOPA, 2019).

Adding to these inputs from the ESA’s on short-termism, the European Commission has also performed a study on due diligence requirements to assess the “possible need to require corporate boards to develop and disclose a sustainability strategy” and “to clarify the rules according to which directors are expected to act in the company’s long-term interest.”³⁴ This study could possibly lead to the development of regulations by the European Commission on this issue, and provides different options that can be chosen by the commission. Overall, the options include “no

³⁴ These citations originate from Action 10 of the European Action Plan on Financing Sustainable Growth

policy change”, “new voluntary guidelines”, “new regulation requiring due diligence reporting” or “new regulation requiring mandatory due diligence as a legal duty of care”. The study does conclude that social, human rights and environmental impacts from this last option are expected to be most significant (European Commission, 2020d).

7.9. General remarks

Much like it has communicated in its initial action plan on financial sustainable growth in 2018, it seems that the European Commission has focused its initial actions on two actors – large firms and financial market participants - it considered as central to its financial system. However, it seems that progress is also gradually being made to create a sustainable environment for other essential actors of the financial system to make their transition into a more sustainable approach to finance and growth. The EU taxonomy is still work in progress, and the Commission still has to develop criteria for 4 other objectives. It also mentions implementing social objectives as well as creating a “brown” Taxonomy in the context of which activities with negative environmental or social impact will also be identified and have to be disclosed. Though much work is yet to be done, and implementation of disclosure criteria for the 6 objectives will not be enforced in practice before the 1st of January 2023, regulation has been adopted on this subject in order for the Commission to enforce its new vision of the European financial system. There is no reason not to believe that, given the growing urgency to mitigate and adapt to climate change and other environmental issues, this will not continue in the years to come.

8. Conclusion

An analysis of the professional and academic literature on the subject of environmental performance of investments and firms provides interesting insight on the lack of commonly agreed

definition of what constitutes environmental performance and sustainable investments. The professional industry represented by actors throughout the world that perform sustainable investing in its different forms and terminologies shows us the disparities and subjective aspects of the different approaches that are taken in this sector to date. As intergovernmental entities identified a growing need for large-scale investments in climate change mitigation and adaptation, the professional literature boasts the magnitude of an SRI industry that is still to a great majority composed of investment approaches that either do not strictly consist in investing in climate-oriented assets or projects that positively benefit the environment, or still lack clarity and transparency regarding the sustainable aspect of their investment process.

A study of the academic literature on the subject provides further insight on the puzzle that currently exists on the general subject of the extra-financial performance of investments. The terminology of authors varies greatly, as even central terms such as environmental, social and governance (ESG) factors, corporate environmental responsibility (CSR) and corporate environmental performance (CEP) are used interchangeably in some of the literature's most influential papers. As we investigate further into the specific study of corporate environmental performance, further challenges can be identified. As academics seem to agree on the overall terminology of this more specific literature, issues on the selected metrics as well as on the industries and geographies of corporations in sample arise. More recently, the validity of the different sources of ESG data is even questioned. The conclusion of this analysis is quite straightforward: it seems both practitioners and academics need a common framework to be able to have a common language on the different concepts, definitions and metrics that constitute sustainable finance.

With the clear intention of putting sustainability at the heart of its financial system, the European Commission has started to deal with these issues. Through its action plan on sustainable finance and the different measures that it is composed of, the Commission is making process in incorporating sustainability into the business practices of every type of actor in its financial system. At the center of its action plan, it has started to develop a classification system that clearly defines its central concepts for sustainability, as well as industry-specific metrics that need to be respected if any financial market participant wishes to offer sustainable financial products and any large firms wished to put forward its sustainability performance. Through this classification system, the Commission aims to answer the markets' need for a clear, common framework for sustainable finance in which actors can truly evaluate their extra-financial performance and be accountable for their actions in terms of sustainability.

The Commission is also creating a legal basis for this new framework, with strict regulation to enforce its implementation having been approved at the end of 2019. As this regulation - the Taxonomy regulation, the Disclosure regulation and Benchmark regulation - mostly focus on large firms and financial market participants that offer financial products in the EU, it is made quite clear from the analysis of other actions currently taken by the European Commission and its different agencies that the entire financial system will in some way have to adapt to include sustainability in their decision-making process.

The Commission's work is providing one can be considered as the last key element needed by the sustainable finance industry in order to be recognized as a separate financial industry, the first with an extra-financial purpose. As global warming is showing no signs of deceleration, clear, impactful and efficient investments in climate change mitigation and adaption need to increase exponentially if the climate risk is to be dealt with seriously. In this matter, creating a true

sustainable finance industry would be an essential stepping-stone towards true climate change mitigation. In that sense, the innovation work on creating a classification system for sustainable economic activities oriented towards financial market participants, along with the surrounding work to create a efficient environment for a strong sustainable finance to develop could well be interpreted as a new sustainable finance paradigm for both practitioners and academics alike.

References

Anton, W.R.Q., Deltas, G. and Khanna, M., 2004. Incentives for environmental self-regulation and implications for environmental performance. *Journal of environmental economics and management*, 48(1), pp.632-654. BCG, 2018. "Global Asset Management, 2018. The Digital Metamorphosis."

Berg, F., Koelbel, J.F. and Rigobon, R., 2020. Aggregate confusion: the divergence of ESG ratings. *Available at SSRN 3438533*.

Bhattacharyya, A. and Cummings, L., 2015. Measuring corporate environmental performance–stakeholder engagement evaluation. *Business Strategy and the Environment*, 24(5), pp.309-325.

Brammer, S., Brooks, C. and Pavelin, S., 2006. Corporate social performance and stock returns: UK evidence from disaggregate measures. *Financial management*, 35(3), pp.97-116.

Brooks, C. and Oikonomou, I., 2018. The effects of environmental, social and governance disclosures and performance on firm value: A review of the literature in accounting and finance. *The British Accounting Review*, 50(1), pp.1-15.

Campbell, J.Y., 1996. Understanding risk and return. *Journal of Political economy*, 104(2), pp.298-345.

Christensen, D., Serafeim, G. and Sikochi, A., 2019. Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings.

Clarkson, Max E. "A stakeholder framework for analyzing and evaluating corporate social performance." *Academy of management review* 20.1 (1995): 92-117.

Council of the European Union, 2019. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the establishment of a framework to facilitate sustainable investment - Approval of the final compromise text

Delmas, M.A., Etzion, D. and Nairn-Birch, N., 2013. Triangulating environmental performance: What do corporate social responsibility ratings really capture?. *Academy of Management Perspectives*, 27(3), pp.255-267.

Diebecker, J., Rose, C. and Sommer, F., 2019. Spoiled for Choice: Does the Selection of Sustainability Datasets Matter?. *Available at SSRN 3359508*.

Doyle, T., 2018. Ratings that Don't Rate: The Subjective World of ESG Ratings Agencies. *Report, American Council for Capital Formation*.

EBA, 2019. "EBA Report on Undue Short-term pressure from the financial sector on corporations".

EIOPA, 2019. “Potential undue short-term pressure from financial markets on corporates: Investigation on European insurance and occupational pension sectors.”

Escrig-Olmedo, E., Muñoz-Torres, M.J., Fernández-Izquierdo, M.Á. and Rivera-Lirio, J.M., 2017. Measuring corporate environmental performance: A methodology for sustainable development. *Business Strategy and the Environment*, 26(2), pp.142-162.

ESMA, 2019. “Final Report. Guidelines on Disclosure Requirements Applicable to Credit Ratings.”

ESMA, 2019b. “Report. Undue short-term pressure on corporations.”

Eurosif, 2018. “European SRI Study 2018.”

European Commission, 2018. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN CENTRAL BANK, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. Action Plan: Financing Sustainable Growth

European Commission, 2019a. “United in delivering the Energy Union and Climate Action - Setting the foundations for a successful clean energy transition”

European Commission, 2019b. “Communication on the European Green Deal”

European Commission, 2019c. “TEG Report. Proposal for an EU Green Bond Standard”

European Commission, 2020a. “Commission Communication on the Sustainable Europe Investment Plan”

European Commission, 2020b. “Consultation Document. Consultation on the Renewed Sustainable Finance Strategy.”

European Commission, 2020c. “Taxonomy: Final report of the Technical Expert Group on Sustainable Finance”. March 2020.

European Commission 2020d. “Study on due diligence requirements through the supply chain. Final Report.”

European Commission, 2020d. “Usability Guide, TEG Proposal for a EU Green Bond Standard”.

Friede, G., Busch, T. and Bassen, A., 2015. ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), pp.210-233.

Gómez, J.P., Priestley, R. and Zapatero, F., 2016. Labor income, relative wealth concerns, and the cross section of stock returns. *Journal of Financial and Quantitative Analysis*, 51(4), pp.1111-1133.

GSIA, 2013. “Global Sustainable Investment Review 2012.”

GSIA, 2018. “Global Sustainable Investment Review 2018.”

Hartmann, J. and Vachon, S., 2018. Linking environmental management to environmental performance: The interactive role of industry context. *Business Strategy and the Environment*, 27(3), pp.359-374.

Harvey, C.R., Liu, Y. and Zhu, H., 2016. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), pp.5-68.

HLEG, 2018. Financing a Sustainable European Economy, Final Report 2018 by the High-Level Expert Group on Sustainable Finance.

IPCC, 2018: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skeea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)].

Jagannathan, R. and Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. *The Journal of finance*, 51(1), pp.3-53.

Kotsantonis, S. and Serafeim, G., 2019. Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, 31(2), pp.50-58.

Luo, H.A. and Balvers, R.J., 2017. Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis*, 52(1), pp.365-399.

Margolis, J.D., Elfenbein, H.A. and Walsh, J.P., 2007. Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance. *Ann Arbor*, 1001, pp.48109-1234.

Misani, N. and Pogutz, S., 2015. Unraveling the effects of environmental outcomes and processes on financial performance: A non-linear approach. *Ecological economics*, 109, pp.150-160.

OECD (2011), OECD Guidelines for Multinational Enterprises, OECD Publishing. <http://dx.doi.org/10.1787/9789264115415-en>

Official Journal of the European Union, 2014. DIRECTIVE 2014/95/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 October 2014 amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups

Official Journal of the European Union, 2019a. REGULATION (EU) 2019/2088 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 November 2019 on sustainability-related disclosures in the financial services sector

Official Journal of the European Union, 2019b. REGULATION (EU) 2019/2089 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 November 2019 amending Regulation (EU) 2016/1011 as regards EU Climate Transition Benchmarks, EU Paris-aligned Benchmarks and sustainability-related disclosures for benchmarks

Orlitzky, M., Schmidt, F.L. and Rynes, S.L., 2003. Corporate social and financial performance: A meta-analysis. *Organization studies*, 24(3), pp.403-441.

OECD, 2011. OECD Guidelines for Multinational Enterprises.

Post, C., Rahman, N. and McQuillen, C., 2015. From board composition to corporate environmental performance through sustainability-themed alliances. *Journal of Business Ethics*, 130(2), pp.423-435.

Ren, S., He, D., Zhang, T. and Chen, X., 2019. Symbolic reactions or substantive pro-environmental behaviour? An empirical study of corporate environmental performance under the government's environmental subsidy scheme. *Business Strategy and the Environment*, 28(6), pp.1148-1165.

Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3), pp.425-442.

- Trumpp, C., Endrikat, J., Zopf, C. and Guenther, E., 2015. Definition, conceptualization, and measurement of corporate environmental performance: A critical examination of a multidimensional construct. *Journal of Business Ethics*, 126(2), pp.185-204.
- Trumpp, C. and Guenther, T., 2017. Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. *Business Strategy and the Environment*, 26(1), pp.49-68. UNEP, 2019. "Emissions Gap Report 2019."
- United Nations, 2015. Transforming our world: The 2030 agenda for sustainable development. *General Assembly 70 session*.
- UNFCCC, V., 2015. Adoption of the Paris agreement. *I: Proposal by the President (Draft Decision)*, United Nations Office, Geneva (Switzerland), (s 32).
- United Nations, 2011. UN Guiding Principles on Business and Human Rights.
- US SIF, 2018. "Report on US Sustainable, Responsible and Impact Investing Trends."
- Wood, D.J., 2010. Measuring corporate social performance: A review. *International journal of management reviews*, 12(1), pp.50-84.
- World Economic Forum, 2020. "Global Risks Report 2020."
- Xie, S. and Hayase, K., 2007. Corporate environmental performance evaluation: a measurement model and a new concept. *Business Strategy and the Environment*, 16(2), pp.148-168.
- Xue, B., Zhang, Z. and Li, P., 2020. Corporate environmental performance, environmental management and firm risk. *Business Strategy and the Environment*, 29(3), pp.1074-1096.
- Yang, R., 2019. What Do We Learn From Ratings About Corporate Social Responsibility (CSR)?. *Columbia Business School Research Paper*, (18-37).

Chapter 3

The role of capital markets in financing climate change mitigation

Abstract

As an investment gap of \$6.9 trillion per year has been identified by the OECD to meet the Paris Agreement goals by 2030, capital markets are mostly taking capital out of large firms rather than providing it for investment purposes. We find that through corporate stock repurchases and distribution of dividends, stock markets have been divesting from large US firms since 1994 and from large Euro firms since 2001. Long-term debt is the only positive source of external financing for these firms. However, for large US firms, corporate bonds represent less than 20% of long-term debt financing since 1995. We obtain a sample of climate-aligned firms and find similar results for these companies, even though our sample contains some less mature firms that still choose to call on equity financing.

Keywords: Climate Change; Sustainable Finance; Capital Markets; Stock Markets; Corporate Bonds; Climate-aligned Firms

JEL classification: G23, M14, D22, Q2, G14, G10, G12

1. Introduction

In a report performed in 2017 on the subject of public financing of climate-aligned growth, the OECD estimated global infrastructure investment needs to be consistent with a 2°C 66% scenario to amount to USD 6.9 trillion per year in the next 15 years (OECD, 2017a). This mostly covers energy (34%) and transport (43%) sectors, with between 60% to 70% required in emerging economies. Looking at the energy sector specifically, the international energy agency (IEA) estimates that consistency with this scenario would require 95% of the electricity to be low-carbon by 2050, that 70% of new cars would need to be electric, that the entire building stock would have to be modernized and that the industrial sector would need to produce 80% less CO₂ (IEA, 2017). Through the exact amount of investments needed remains of course uncertain, there is no doubt that there is a pressing need for investments to make the economy more energy-efficient in the next few years.

Such important changes cannot be implemented without the participation of corporations that need to adapt their business models and provide new sustainable products in order to face climate change mitigation. However, one of the main challenges faced by these corporations concerns a lack of understanding and information on which activities and products are aligned with this 2°C 66% scenario, and which are not. This issue has been clearly identified by regulators such as the European Commission, which has developed a classification system for sustainable activities (EU Taxonomy) in which technical screening criteria for 70 climate change mitigation and 68 climate change adaptation activities have been created based on sector-specific metrics. (European Commission, 2020). This classification system is enforced by regulation which will require EU firms to declare the percentage of their turnover and capital expenditures that are “taxonomy-aligned”. Financial market participants are also central to this regulation and will also be required

to disclose the percentage of each financial product they offer in the EU that is “taxonomy-aligned” (Council of the European Union, 2019). In the near future, information of the precise participation in the climate change mitigation of every large EU firm and financial market participant concerned by the regulation will be publicly available.

A similar approach has been used by the Climate Bond Initiative (CBI) when the organization published a report on bonds and climate change in 2018 (CBI, 2018). In this report, the CBI identifies a \$1.45 trillion climate-aligned bond universe composed of \$389 billion of labelled green bonds, \$811 billion of corporate bonds from climate-aligned issuers and \$250 billion of municipal bonds from climate-aligned municipal issuers. In order to identify climate-aligned issuers, the CBI considered firms “that derived at least 75% of their revenue from green business lines in at least one of 6 climate themes: clean energy, low-carbon transport, water management, low-carbon buildings, sustainable land use.” CBI then screened issuers who had no debt outstanding, had been acquired or had insufficient revenue information to determine the share of green revenues. To date and to the author’s knowledge, this approach is the most robust means of identifying climate-aligned firms without further information on the specific business practices of corporations. A large literature also refers to specific environmental metrics to measure the environmental performance of corporations, but strong debate resides on the quality and comparability of this data and the resulting measures for environmental performance (Kotsantonis and Serafaim, 2019). By clearly identifying 6 economic activities that are structurally aligned with the Paris climate goals, the CBI has more certainty on the underlying positive environmental impact of corporations in its sample than it probably would with currently available data that is provided by corporations and not verified in a great majority of cases (Kotsantonis and Serafaim, 2019). The Climate Bond

Initiative provided the author with a list of these identified climate-aligned firms throughout the world.

Before looking at climate-aligned firms specifically, we study cash-flow fundamentals of US non-financial firms in the S&P500 index and non-financial firms in the S&P 350 Europe index in the euro area since 2009 in order to understand the dynamics of these firms' operating activities, investing activities and financing activities. We find that between 2009 and 2019, an amount corresponding to \$7.72 trillion has been paid by non-financial S&P500 firms to stockholders in the form of net stock repurchases (stock issuances subtracted from stock repurchases) and dividends, a sum that represents 80% of total net income over the period. Over the same period, for our sample of S&P 350 Europe firms in the euro area, this amount was of \$3.75 trillion and 67% of total income, with dividends representing 60% of total income. Overall, for both US firms and euro firms in our sample, financing activities have constantly been negative since 2009, and the operating cash flow generated by firms in both regions has financed both investing and financing activities. This confirms the fact that once companies have reached a certain level of maturity and size, they finance their own operations and investments, even providing a generous amount of capital to stockholders. This first finding is quite alarming, specifically given the fact that the current climate mitigation urgency requires massive investments in climate-aligned activities in which we would expect financial capital markets and financing institutions to participate.

Following these initial findings, we study these firms' financing activities more specifically. We first focus on financing originating from stock markets. By studying average stock issuances, repurchases and dividends instead of total dollar amounts, our sample of US firms can be studied on a longer time period. We find that in the US, stock repurchases have been superior to stock

issuances since 1994 at the exception of 2008 following the subprime crisis. Similarly, the only year that dividends have been inferior to stock issuances since 1971 for US firms was also 2008. We also find that the difference between average stock repurchases and dividends and stock issuances has been increasing steadily since the subprime crisis. For euro firms, the difference between stock issuances and repurchases is less pronounced, but dividends have consistently been much larger than stock issuances since 2001. These findings further illustrate how capital flow has been going from large US and euro firms to stock markets, which have been divesting from these firms instead of investing in them.

Long-term debt, however, has been a constant source of positive financing for both US and euro firms. We find that large US firms have been issuing an increasing amount of long-term debt since 1980, going from a market of \$10 billions to more than \$3 trillion of yearly long-term debt issuance starting in 2015. Even though long-term term reduction has been increasing accordingly, the change in long-term issuance has – at the exception of the short period following the subprime crisis – been positive since 1971 for US firms and since 1999 for euro firms since. But long-term debt is not solely composed of securities that are publicly traded on capital markets, but also of credit lines and bank loans. Using the FISD database, a database that provides information on the US bond market, I investigate the US corporate bond market and develop a methodology to determine the part of US long-term debt that is composed of corporate bonds in our sample of S&P 500 firms, in order to have a better understanding of the role played by capital markets in this positive long-term debt.

We find that on the period going from 1995 to 2018, corporate bond have represented on average 13.6% of newly issued long-term debt for US S&P500 non-financial firms, and 5.3% of yearly long-term debt reduction. These findings indicate that the participation of bond markets in

financing large US firms is minimal and does not outweigh the negative capital flow stemming from large corporations to stockholders. Our results for the first part of this study indicate that capital markets are not financing large US corporations, but that on the contrary, it seems that large US corporations are providing capital stemming from their operating activities to capital market participants.

In this context, I use data provided by the Climate Bond Initiative to identify a set of US climate-aligned firms and investigate whether the participation of capital markets in financing these firms is similar to our findings for large US and euro firms.

2. Literature Review

The necessary investments needed to mitigate the worst impacts of climate change have been evaluated by the OECD in a special report in 2017 on the subject. In this work, the OECD focuses on infrastructure needs in energy, transport, water and telecommunications, and identifies large underinvestments in these areas. The organisation states that important investments are required to both maintain or upgrade current infrastructure and allow for basic service in middle-income countries, all the while preparing for a world population of 9 billion by 2050, with two thirds of this population being urban and with growth mostly concentrated in developing economies (OECD, 2017).

The OECD's estimate is that USD 95 trillion would be needed between 2016 and 2030, which represents USD 6.3 trillion per year, when current infrastructure spending ranges between USD 3.4 and USD 4.4 (IEA, 2017; IEA, 2016; Woetzel et al., 2016). This would have to be oriented in priority towards energy production and transport. According to the International Energy Agency (IEA), energy production represents around two-thirds of global GhG emissions, mostly due to the

combustion of fossil fuels. The energy sector faces the key challenge of having to decarbonise its electricity production. On the one hand is the necessary phasing out of coal, and on the other further deployment of nuclear power, widespread deployment of renewable energy sources, as well as investments in negative emissions technologies (IEA, 2017). Transport is also a central industry, not only because the sector represents 23% of global CO₂ emissions, but also because it is the fastest growing source globally. According to the OECD's International Transport Forum, CO₂ emissions from transport could double by 2050 (OECD 2017). Necessary investments in transport focus mostly in multiplying low-emission or zero-emission transport as well as developing energy efficient technologies. Other sectors are considered by the OECD in its report, including the building sector, which could already benefit from the availability of energy efficient technologies that are not yet widespread amongst actors, and the agriculture, forestry and land use sector (AFOLU). The AFOLU sector represents 25% of GhG emissions, mostly through deforestation (9-10%) and agriculture (10-12% mostly methane and nitrous dioxide).

Part of the OECD's report focuses on the role that can be played by private finance in this sector by investing in this infrastructure, but also in innovative technologies and divestment from carbon-intensive assets. The role of private finance is all the more important in advanced economies where public actors play a smaller role in funding infrastructure (NCE, 2016; Ahmad, 2015). Amongst the different private actors that have been identified by the OECD, corporations represent a key source of finance in the transition. These actors can participate in many ways by directly investing in renewable projects, performing power purchase agreements (PPAs) and creating partnerships with active players in the energy or transport sectors. Most importantly, corporations can also invest in their own business activities to lower their emissions profile or promote renewable energy and transport. Large global companies are being held accountable for

emissions since the industrial revolution. In a study using compiled data of firm carbon emissions, the Carbon Disclosure Project (CDP) identified companies that together represented 72% of annual global industrial GHG emissions in 2015 (Griffin, 2017). Within these companies, 100 fossil fuel producers represented 52% of GHG emissions in 2017. Given this new data, it is clear that major global companies are the central playing field where most sustainable investments need to be performed.

However, it seems that in recent years companies have not invested as much as they could, regardless of whether these investments were sustainable or not. In March 2016, BlackRock's CEO Laurence Fink stated in his yearly letter to the executives of S&P 500 firms that, "in the wake of the financial crisis, many companies have shied away from investing in the future growth of their companies. Too many companies have cut capital expenditure and even increased debt to boost dividends and increase share buybacks." Even though this issue seems to have been clearly identified by financial practitioners, academic literature on the subject is scarce. To the author's knowledge, recent studies that address the subject most thoroughly have been performed by Gutierrez and Philippon (2016, 2017). The authors base their first study on five facts which they demonstrate : (1) non-financial businesses are profitable but do not invest, (2) investments are low even when Tobin's Q is high, (3) depreciation has remained stable since 2000, which means that the decrease in net investment is due to a decrease of gross investment, (4) firm entry has decreased, with a sharp decline of entry rates, exit rates and average number of firms by industry and finally (5) institutional ownership and payouts have increased. Using these facts as a basis for their analysis, Gutierrez and Philippon (2016) explore 8 potential explanations for firm-underinvestment : financial frictions from external finance, bank dependence and safe asset scarcity, measurement errors in intangibles or due to globalization, lack of competition due to

either regulation or other factors and tighter governance. The authors' conclusion is the following: "this investment wedge appears to be linked to decreased competition and changes in governance that encourage shares buyback instead of investment".

These results participate in several strands of literature that are quite recent. The first is that of competition, and the link between competition, investment and innovation. On this subject, Alghion et al. (2014) find that increased competition leads to increased R&D investments by firms that operate at the same technological level, but decreased R&D investment by firms that are trailing behind. Competition also reduces the number of firms that operate on the same technological level and reduces the number of industry leaders. However, in their study, Gutierrez and Philippon (2016) state that their data suggests that firms that operate in concentrated or aging industries and growing firms that do not face any entry risks could have weak incentive to invest. This notion is supplemented with the literature that supports the hypothesis that competition is decreasing in several economic sectors (CEA 2016), which could then explain why investments are decreasing. This subject is specifically addressed in Gutierrez and Philippon (2016), in which the authors find a causal relationship between competition and investment, and state that US businesses have been under-investing since the early 2000's.

Another strand of literature focuses on the recent trends and impact of institutional ownership. In Gutierrez and Philippon (2016), the authors showed that institutional share of ownership of US businesses had risen from 2% to around 6% between 1980 and 2015, a majority of these institutional investors being quasi-indexer which favor short-termism over dedicated investments. Noticing this trend, Fichtner et al. (2016) go further into this analysis and find that, taken together, the three largest asset managers - Blackrock, Vanguard and State Street - are the largest shareholder of 88% of the S&P 500 and 82% of the index's market capitalization. Adding to this subject, a

pre-existing literature had already focused on the concept that common ownership amongst competitors reduces incentives for competition (Salop and O'Brien, 2000). The literature on ownership is closely related to that of corporate governance, and centers on the agency problem between managers and shareholders. For academics in the field, weak governance is synonymous to higher agency conflicts, and signifies that managers are less constrained by shareholder rights and ownership and therefore more independent from shareholders. In cases of weak governance, it is shown that managers spend more cash in capital expenditures and acquisitions (Harford et al., 2008; Richardson, 2006). This is not necessarily a good sign given the fact that other academics have also demonstrated that cash-rich firms tend to make more value-decreasing acquisitions (Harford, 1999).

Changes in ownership can also have an influence in terms of short-termism, but the literature on the existence of short-termism and its possibly negative impact on investments and R&D is strangely scarce. However, one particular author has been focusing on this subject quite extensively. Focusing on the possible negative impacts of the concept of maximising shareholder value (MSV) on corporations and their use of funds, William Lazonick's position is that equity markets and stock-based compensation have led corporations to focus on short term share price instead of investment and innovation (Lazonick, 2015).

In their paper on the concept of maximising shareholder value (MSV), Lazonick et al. (2000) investigate the impacts of this ideology on corporate governance in US firms. They find that payout ratios (the ratio of dividends to after-tax adjusted corporate profits) of US firms stayed quite stable between 1960 and 1980 at around 42 per cent, but rose between 1980 and 1998 to more than 49 per cent. The authors investigate the growth of share repurchases that occurred during the 1980s, with share buy-backs representing less than 5 per cent of corporate profits before 1982 and

exceeding 25 per cent by 1985. By 1989, dividends had risen to \$134.4 billion and stock repurchases to over \$60 billion, increasing the combined payout ratio to over 81 per cent.

The trend that had been identified by Lazonick et al. (2000) did not stop then, and in 2015 Lazonick published a study that focused specifically on the stock buyback issue and the concept of “retain-and-reinvest to downsize-and-distribute”, a term coined to illustrate the shift that occurred in US firms as they started to use their profits to pay shareholders instead of reinvesting them in their business activity (Lazonick, 2015). For the period 2004-2013, Lazonick (2015) shows that companies in the S&P 500 paid \$3.4 trillion in stock buybacks, which represented 51% of these companies’ total income, when dividends already represented 35%. For the author, the explanation for this growth in repurchases is linked to the stock-based pay of firm executives. Between 2006 and 2013, these ranged from representing 66 per cent to 84 per cent of total annual remuneration of highest-paid executives in these same firms when salaries and bonuses only ranged from 5 to 12 per cent. Furthermore, Lazonick also adds that “the vesting of stock awards is often dependent on the company hitting quarterly earnings per share (EPS) targets, for which well-timed manipulative boosts from stock buybacks can be very helpful.” In essence, highest-paid executives have a clear incentive to perform stock buybacks, which reduces the number of shares outstanding, increases EPS and helps them reach EPS targets to obtain stock awards.

Lazonick (2015) addresses two distinct flaws of the MSV approach. The first is that this approach is based on the assumption that shareholders are the sole corporate participants that bear risk, while this is not the case. The author argues that taxpayers and workers are also risk-bearers that have an economic claim on distribution of profits. On the one hand, the government invests and offers subsidies to large firms that may not generate profits that can be taxed. On the other hand, many workers’ pay does not correspond to their level of participation in the firm’s business

activities (Lazonick, 1990). The arguments provided by Lazonick (2015) are ever more interesting when considering the climate change challenge that is currently faced by US firms and which concerns all of their stakeholders. The authors' second argument is that public shareholders are not long-term investors that invest in the value-creating capabilities of the corporation, but rather traders that wait for the share value of the firm to rise before selling their shares for a profit.

Finally, Lazonick (2015) provides a summary of his research on how stock buybacks undermine a firm's innovative capabilities. One of the impacts of stock buybacks on the firms concerns strategic control. Lazonick states that senior executives that choose to pay large yearly stock buybacks lose their capacity to understand what investments are needed for firms to remain innovative (Baldwin, 1991; Christensen et al., 2008), as stock buybacks represent a powerful incentive for executives to focus more on increasing their pay. Another consequence of large buybacks in a firm concerns skill development. Lazonick states that by spending cash on stock buybacks, the firm is not investing as much as it should in its employees' skills and careers, which penalizes the firm's innovative capacities. The last impact of stock buyback regards financial commitment. Lazonick argues that many companies that perform large stock buyback operations throughout the years lack cash once they enter a period of crisis and are unable to finance a restructuring process to become innovative again (Lazonick, 2015).

3. Data

In the context of this study, four sources of data are used. We use Compustat North America to obtain information on the financial cash flows of US firms in the S&P500 index and Compustat Global to obtain information on the financial cash flows of firms in the S&P350 Europe index in countries that use the euro as a currency. We obtain similar information for US climate-aligned

firms and European climate-aligned firms. Compustat North America provides fundamental and market information on US and Canadian firms such as annual and quarterly income statements, balance sheets and statement of cash flows. Annual data for most companies is available since 1950, and quarterly history and monthly market history since 1962. Compustat North America also provides information on index constituents which allows us to obtain information on companies that are part of the S&P500 index. Compustat Global provides similar data in more than 80 countries other than the US and Canada and covers 96% of European market capitalization. It provides annual and quarterly information starting from 1987.

We use the Mergent Fixed Income Securities Database (FISD) to obtain information on the US corporate bond market and to evaluate the proportion of long-term corporate debt issued by US firms in the S&P500 that is traded on these capital markets. We obtain similar information for US climate-aligned firms. The FISD database contains information on issues of more than 140,000 publicly offered US bonds. This information can be specific to both corporate bond issuers and corporate issues.

Finally, we use data provided by the Climate Bond Initiative (CBI) to identify US climate-aligned firms and European climate-aligned firms. CBI is an international NGO that focuses on the bonds as financing solution for climate change mitigation and adaptation. In 2018, CBI identified firms around the world that had more than 75% of their turnover than originated from climate aligned activities. The CBI research team were kind enough to provide this list of firms to the author.

3.1. Cumulative Cash Flows for the period 2009-2019

In order to obtain the required information from the Compustat database, we first identify companies that are part of the S&P500 in Compustat North America and companies that are part of the S&P350 Europe Index in Compustat Global from 2009 to 2019. The S&P 500 index in Compustat contains a total of 754 companies that have been part or currently are part of the S&P 500 and the S&P 350 Europe Index contains 364. We exclude financial services companies from our sample and keep only companies that have available data from 2009 to 2019. Our final sample contains 531 companies from the S&P 500 Index and 233 companies from the Europe S&P 350 Europe Index.

In order to have a synthetic view of firms in our sample, we focus on the main fundamental indicators from cash flow statements, differentiating, much like it is the case in firm's consolidated accounts, net cash flow from operating activities from net cash flow from investing and financing activities. Main fundamentals for each categories of activities are net income and depreciation and amortization for operating activities that provides us with the main information on the cash that is generated by the firm's activity throughout the year, capital expenditures and acquisitions for investing activities, and information on share sales and repurchases, dividends and changes in debt for financing activities. For each section of the cash flow statement, we include an "Other" category that sums up the remaining cash flow items in the section. In total, we obtain information from 15 cash flow items from cash flow statements of both US and European firms between 2009 and 2019.

3.2. Average Cash Flows for US 1971-2019 and Euro 1999 - 1919

The Compustat North America and Global databases can also be used to obtain average values for cash flow items of US and Euro firms going back many years before 2009. As aforementioned,

the Compustat North America database provides annual information on US firms since 1950. However, the first year for which all cash flow information is available on the database is 1971. We therefore run our sample of average values for US cash flow from 1971 to 2019. The Compustat Global database provides data from 1987 but given the currency issue we face with this dataset, we run our sample from 1999 to 2019 provided that the euro was launched on the 1st of January 1999.

3.3. Corporate bonds issuances and reductions 1995 - 2018

We use the FISD database to identify corporate bond issuances and reductions from US firms in the S&P500 in order to understand what proportion of US long-term debt is composed of debt securities publicly tradable on US capital markets. FISD is considered as the most comprehensive database focusing on bonds and contains essential information on bond issuers as well as specific bond issues such as bond issue date, maturity, size, coupon, type, and any information that can be used to identify and categorize US bonds. FISD allows us to have visibility on the corporate bond market and understand operations that occur between companies that issue bonds and the market of bondholders. In addition to comprehensive information of bond issues and bond issuers, FISD provides detailed information on the amount outstanding history for every bond in its database. This information can be used to understand the different amounts that are either paid or received by companies during the lifespan of a corporate bond, which corresponds to amounts that are taken under account by firms when they report their yearly long-term debt issuances and long-term debt reductions.

We first identify corporations of our sample in the FISD database using CUSIP information. A CUSIP number is a unique identification given to each individual security. The CUSIP number is composed of nine characters. The first six characters of a CUSIP number identify the issuer, and

the last three identify the specific security issued by this issuer. Compustat and FISD both provide CUSIP numbers for issuers and securities, and we are able to identify every security linked to US issuers in our sample using the first six characters of the CUSIP number. Using the FISD database, we obtain historical data on the different operations that altered the amount outstanding for each bond, from its date of issuance to its date of maturity if that date is reached, or to the corresponding date where the amount outstanding reaches zero due to another form of operation, such as a bond conversion or a call. A great number of these bonds are still active, and therefore the amount outstanding has not yet reached zero.

Some of the bond issuances of companies in our sample are performed using foreign currency. Even though this is the case for a minority of our bonds, we do need to convert these amounts in US dollars in order to be able to consider them. Much like it was the case for our sample of EU firms, we limit our sample to 1999 and the creation of the Euro, so that this conversion is more precise. There are five distinct currencies used by US issuers in our sample other than the US dollars: the Euro, the Japanese Yen, the Swiss Franc, the United Kingdom Pound Sterling and the Canadian Dollar. Using publicly available data on monthly historical conversion rates between these currencies and the US dollar provided by the Federal Reserve on its website³⁵, we convert values for every bond issued in a foreign currency to US dollars.

We then move on to categorizing every bond operation into either a bond issuance, where the company receives capital from bond holders, or a bond reduction, where the company provides capital to bondholders. This is a particularly complex operation given the different specific characteristics of bonds, which can be convertible, callable or reviewed during the life cycle of the bond.

³⁵ See <https://www.federalreserve.gov/releases/h10/Hist/>

In table 1, we provide an explanation of the different types of operations that are listed in the FISD database for our sample, and whether each operation is categorized as a bond issuance or a bond reduction. To summarize, initial offerings (type “I”) correspond to the initial bond issuance where capital is provided by bondholders to firms. Operations such as “Initial Offering Increase”, “Over-allotments” and “Reopenings” (type “II”, “OA” and “RO”) correspond to further bond issuances that occur after the initial bond offering, and also correspond to an operation where capital is provided by bondholders to firms. Different types of operations that occur during the bond’s lifecycle are linked to calls, refunds, repurchases, firm reorganizations and tender-offers (type “B”, “E”, “F”, “IRP”, “P”, “R”, “S” and “T”) and correspond to bond reductions where capital is provided by the firm to the bondholders. The amount outstanding of a bond can also be increased or decreased following a review (type “REV”), in which case this operation is either an issuance or a reduction depending on whether the review led to a decrease or increase of funds. The bond can of course reach maturity (type “IM”) in which case the remaining amount outstanding of the bond is paid to bond holders. Other important types of operations include when issues are converted into stocks (type “C”) or exchanged for other securities (type “X”), in which case they either correspond to new share issuances or a new bond issuance. Since we already consider new bond issuances in our sample, this new bond issuance is already accounted for and type “C” and “X” operations are not taken under account.

Our available data for FISD spans to the end of 2018, therefore our sample consists in every bond operation for S&P 500 companies from 1999 to 2018. Our initial dataset for S&P 500 firms from Compustat contained 630 different issuers from 1971 to 2019, and we find information on 14,115 corporate bond operations for 456 bond issuers between 1999 and 2018 in the FISD

database. The corresponding information on the different types of operations and their frequencies is available in table 1.

4. Cumulative Cash Flows of US and Euro firms

4.1. The S&P500 and Euro firms in the S&P 350 Europe

Once we have performed our data treatment on cash flow items from Compustat and have kept only firms that provide data for the entire 2009-2019 period, we add these values across firms to get a general idea of the cash that has been coming in and out of US and Euro firms. Our results are available in Table 2. Looking at operating activities, US firms in our sample have had a steady increase in income since 2009, while European firms seem to have had an immediate increase in income between 2009 and 2010 which has not been surpassed until 2017.

In total, this has led EU firms to have generated little net income compared to the total value of depreciation and amortization of their assets for the same period. For Euro firms in the S&P 350 Europe index, depreciation and amortization represents 125% of total income over the period, when it only represents 64% for US firms. For these firms, depreciation and amortization represents 56% of total net cash flow obtained from operating activities, while these represent only 35% for US firms.

However, looking at investing activities, European firms are spending more in capital expenditures as a percentage of their net income, with total capital expenditures on the period reaching 131% of net income for European firms, compared to 69% for US firms. Acquisitions for both geographical regions are similar, though slightly higher for Euro firms, representing 27.5% of net income for US firms and 36% for European firms.

Table 1

Code	FISD Description	Description	Type of operation	Number of observations
B	Balance of Issue Called	Part of the issue is called by the issuer.	Reduction	281
C	Issue Converted	The issue is converted to the issuer's stocks.	Other	55
E	Entire Issue Called	Entire issue is called by issuer.	Reduction	406
F	Issue Refunded	Part or all of the issue is refunded by the issuer to bondholders.	Reduction	4
I	Initial Offering of An Issue	Initial bond issuance from to the issuer to bondholders.	Issuance	5502
II	Initial Offering Increase	The initial amount of the issue is increased.	Issuance	4
IM	Issue Matured	The issue has reached maturity and bondholders are paid the principal and final interests.	Reduction	1814
IRM	Issue Remarketed	Changes are made to specific conditions of the bond. In our sample, this does not affect bond amount outstanding.	Other	5
IRP	Issue Repurchased	Part of or all of the bond is repurchased by the issuer.	Reduction	138
OA	Over-Allotment	The initial amount of the issue is increased.	Issuance	274
P	Part of an Issue Called	Part of the issue is called by the issuer.	Reduction	71
R	Reorganization	Reorganization of the issuing firm leads to part or all of the bond being repurchased by the issuer.	Reduction	88
REV	Review	Review of the terms of the bond issuance can lead amount outstanding to be increased or decreased.	Reduction/Issuance	3126
RO	Reopening	Issuer re-opens, and the initial amount of the issue is increased.	Issuance	158
S	Sinking Fund Payment	Part or all of the issue is paid by the issuer's sinking fund.	Reduction	262
T	Issue Tendered	Issuer offers to repurchase a specific number of bonds to bondholders.	Reduction	859
X	Issue Exchanged	Bond issue is exchanged for another security.	Other	1065

However, US firms dedicate a significant amount of their investing cash flow to other investing activities such as the purchase of marketable and non-marketable securities, while EU firms have mostly generated cash from selling such securities. For this reason, a further 35% of income is spent in investing activities by US firms, when this pole generates 5% of net income in cash for EU firms.

Perhaps the most insightful information provided in table 2 regards financing activities of US and Euro firms. Over the course of the period going from 2009 to 2019, US firms in our sample

have repurchased \$5.226 trillion of their own stocks on the stock market, while they have issued only \$1.379 trillion. Furthermore, \$3.876 trillion in dividends have been distributed to shareholders. In comparison, the difference between proceeds and repayment for long-term debt totaled \$1.318 trillion and -857 billion for current debt. This means that the only positive source of external financing for industrial US corporations in our sample between 2009 and 2019 has been long-term debt, as these corporation have been mostly repaying current debt and using most of the cash generated from their activity to pay their shareholders, either through repurchases, as the difference between sales and purchases of common and preferred stock correspond to \$3.847 trillion and 39.4% of net income, or through dividends, which correspond to \$3.876 trillion and 39.7% of net income.

Interestingly, Euro firm stock sales are superior to stock purchases by \$34.6 billion, which represents 2% of total net income over the period. However, \$965 trillion in dividends have been distributed, which corresponds to 55% of net income over the period. The difference between proceeds and repayments for long-term debt is also positive for European firms by \$469 billion, and negative for current debt at -\$133 billion. Thought financing activities represent a similar percentage of net income for US (-44%) and Euro firms (-46%), there are important differences between the two regions. Euro firms perform much less stock repurchases than US firms but spend more of cash dividends and obtain almost twice as much financing compared to their net income through long term debt (26% for Euro firms against 14% for US firms).

We understand from these results that both Euro and US firms have been providing cash steadily to their shareholders over the period 2009-2019. When looking exclusively at share issuances, repurchases and cash dividends, US firms have paid a net sum of \$7.722 trillion to their shareholders and EU firms have paid a net sum of \$930 trillion to their shareholders. Our findings

are similar to that of the literature which states that stock markets are increasingly being financed by large firms instead of providing them with financing.

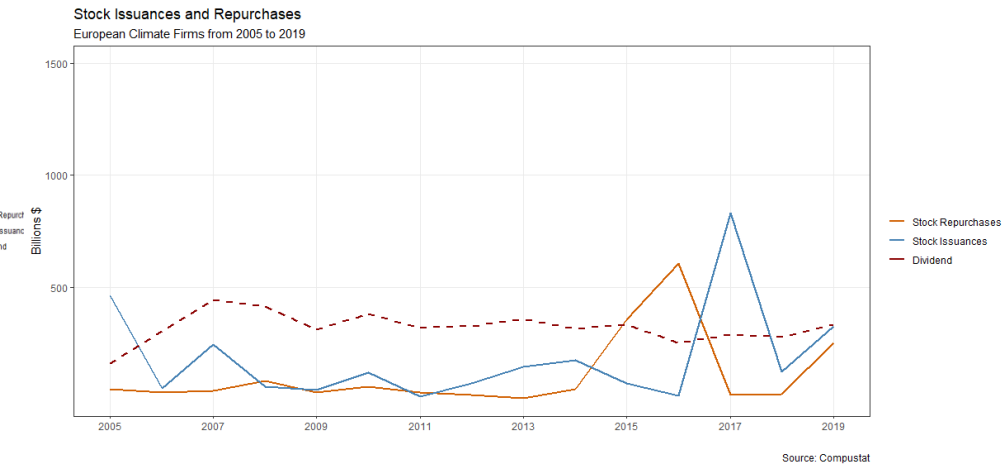
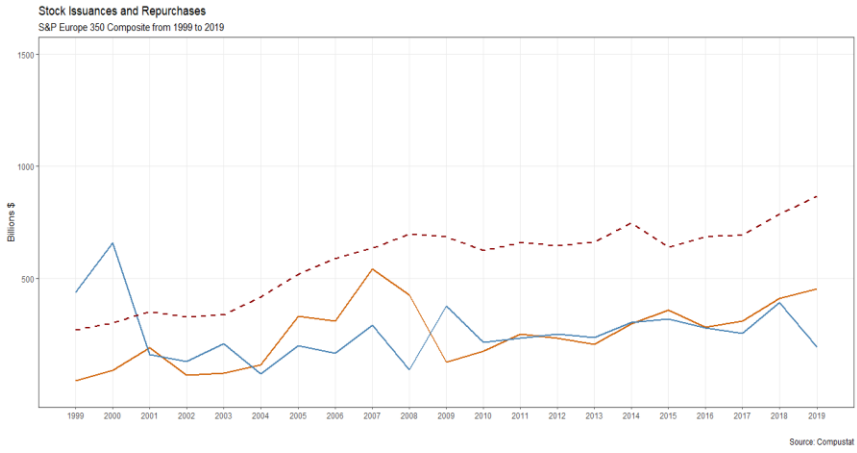
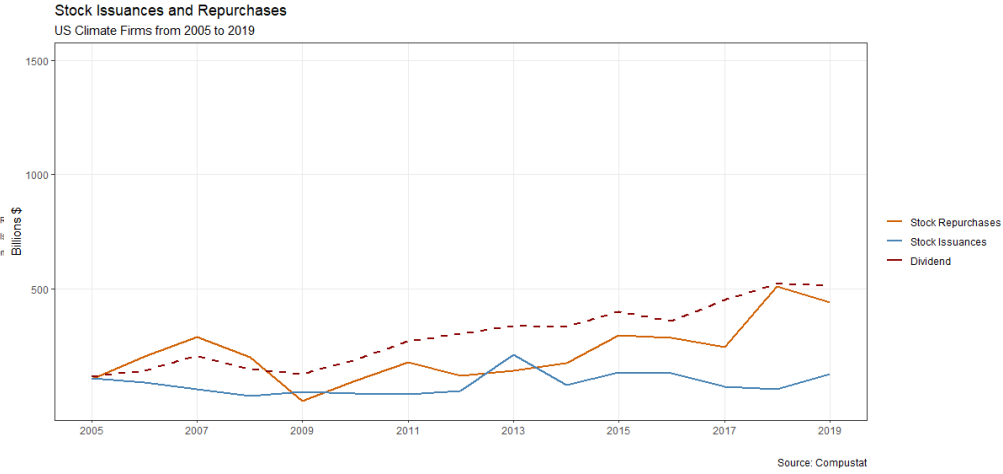
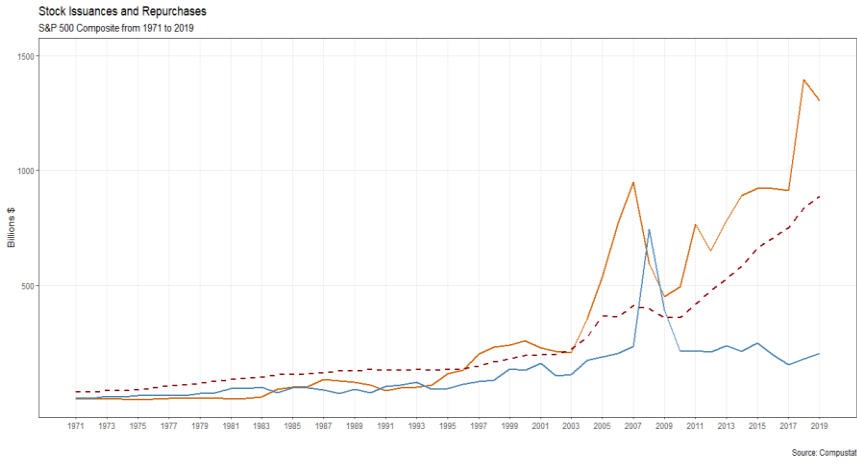
4.2. US and Euro Climate-aligned firms

Cash flow items for US and Euro climate-aligned firms differ from their traditional equivalents. US climate-aligned firms have slightly higher depreciation and amortization and significantly higher capital expenditures than US S&P 500 firms: depreciation and amortization represent 77% of net income for climate-aligned firms against 64% of S&P500 firms and capital expenditures represent 118% of net income for climate aligned-firms against 69% for S&P500 firms. Acquisitions for both samples are similar. This could very well indicate that US climate firms invest more in their business activities than S&P 500 firms.

Looking at financing activities, we find that US climate firms also spend a large amount of cash on paying shareholders through either share repurchases or dividends. While 13% of net income is spent on share issuances, 32% of this same net income is spent on repurchasing shares. A further 52% of net income is spent on cash dividends, bringing the total amount of cash flow dedicated to shareholders to 71% of net income and \$137 billion. US climate firms have a large part of their financing that is provided by long-term debt which represents 43% of their net income (against 14% of S&P 500 firms). Overall, US climate firms spend 20% more of their cash in investing activities than S&P500 firms and 20% less in financing activities.

While Euro firms in the S&P350 Europe index already had a large proportion of their net income dedicated to depreciation and amortization (125%), this is even more the case for EU climate firms for which this figure reaches 214%.

Figure 1



Overall, more than two thirds of their operating cash flow for EU climate firms originates from depreciation and amortization of their assets. Furthermore, 382% of net income for these firms is dedicated to capital expenditures, with a further 33% in acquisitions, which indicates the EU climate firms are investing large amounts of cash in their business activities.

Interestingly, much like Euro firms in the S&P 350 index, Euro climate firms are issuing more stocks than they are purchasing, with 16% of net income spent on share issuances against 9% in repurchases. However, once again quite similarly to Euro firms in the S&P 350 index, cash dividends for Euro climate firms are also very high and represent 64% of net income. Long term debt for Euro climate firms is very high at 50% of net income. This resembles more the profile of US climate firms, which also issued a lot of long-term debt as a proportion of net income (43%). Overall, this causes financing activities for Euro climate firms to be slightly positive at 3% of net income, which is quite far from the -23% of US climate firms, and the -44% and -46% of S&P 500 and S&P 350 Europe firms. Looking at these cash flow items, it seems both US and Euro climate firms are spending most of their cash in reinvesting in their business activities.

5. Average Historical Cash Flows of US and Euro firms

5.1. Share issuances, repurchases and dividends

Looking at average stock issuances, stock repurchases and dividends of firms that have been or currently are part of the S&P 500 since 1971, we notice that stock issuances have only occasionally been superior to stock repurchases: before 1984, between 1991 and 1993, and in 2008 following the subprime crisis. Furthermore, at the exception of 2008, dividends have always been higher than stock issuances for S&P500 firms since 1971. In Europe, we focus on firms that have been or currently are members of the S&P Europe 350 from countries that use the euro as a

currency since 1999. We observe that since 1999, in average, firms in our sample have had relatively similar levels of stock issuances and stock repurchases, with higher levels of stock issuances in 1999 and 2000 and higher levels of stock repurchases in the few years before the subprime crisis. In Europe, dividends have been much higher than both stock issuances and stock repurchases since 2001. Looking at this data, we understand that stock markets have not participated in financing large EU and US firms in the last decades, but have rather been receiving cash from these large firms, either in the form of dividends in Europe, or both in the form of dividends in stock repurchases in the United States since the middle of the 1990s.

Overall, the situation for climate firms is quite similar, although the large differences between stock issuances and repurchases that exist for S&P500 firms do not exist either for US climate-aligned or Euro climate-aligned firms. For US climate-aligned firms, we do notice that stock repurchases have to a great majority been superior to stock repurchases, while dividends have constantly been higher to both stock issuances and repurchases since 2009. For Euro climate-aligned firms, it seems like stock issuances are slightly higher than stock repurchases. Stock repurchases peak in 2016, quickly followed by a peak of stock issuances in 2017. Much like US climate-aligned firms, dividends are constantly higher than stock issuances and stock repurchases except for these two peaks. However, we do observe that stock issuance, repurchases and dividends have been growing for US climate firms between 2005 and 2019 while they have been steadier in the case of Euro firms.

The main conclusion of these results is quite similar to that of the previous section: stock markets have not been a positive source of financing for either large US and Euro firms or their climate-aligned equivalents. This time-series analysis does allow us to observe the current tendencies that are occurring on each of these markets. US firms – either large traditional firms or

climate-aligned firms – tend to have important differences between stock repurchases and issuances. This trend is quite pronounced for S&P500 companies since the mid-1990s, but a similar trend can be observed for US climate firms – though much smaller in magnitude – since 2014. For Euro firms, again both large Euro firms in the S&P 350 Europe and Euro climate-aligned firms, there is more balance between stock issuances and stock repurchases, though dividends have been high since the creation of the Euro. There is a growing trend for Euro dividend distribution, though less pronounced than that of large US firms.

5.2. Long-term debt and overall external financing

The situation for long-term and current debt for these same firms is not similar. At the exception of the post-crisis period corresponding to 2009-2010, US firms have always issued more long-term debt than they have reduced their existing long-term debt³⁶, even though both the amount of issued long-term debt and reduction of long-term debt have grown extensively in the last decades, growing from a little more than \$100 billion in issuances in 1980 to \$3 trillion and more starting in 2015³⁷. Whether it regards large US firms, large Euro firms or Euro and US climate-aligned firms, long-term debt has been a positive source of financing for the studied periods³⁸. Figure 3 illustrates how different sources of external finance have helped finance both US firms and EU firms in recent decades. In order to have better visibility and understanding of the dynamics of each source of external funding for US and EU firms, we apply a smoothed

³⁶ In the Compustat database, long-term debt issuance includes increase in long-term and short-term debt when combined, long-term debt issued for or assumed in an acquisition, proceeds from bonds, capitalized lease obligations, or note obligations, and reclassification of current debt to long-term debt. Long-term debt reduction includes conversion of debt to common stock, change in debt not classified into current and long-term debt in the cash-flow statement, change in long-term debt combined with change in current debt, current maturities of long-term debt for companies reporting a working capital statement, reclassification of long-term debt due to Chapter 11, transfers or reclassifications of long-term debt to current liabilities, decreases to long-term debt accounts, cash statements and LOC or revolving loan agreement

³⁷ Specific information on long-term debt issuance and long-term debt reduction is not available in Compustat Global, and therefore a similar analysis could not be performed for EU firms.

³⁸ Studied periods are 1971 to 2019 for S&P500 firms, 1999 to 2019 for Euro firms in the S&P350 Europe index, and 2005 to 2019 for both Euro and US climate aligned firms.

conditional mean using a local polynomial regression fitting. This smoothing method is commonly used to detect patterns in datasets and corresponding graphical visualization and consists in creating a smoother representation of the data by fitting data points in accordance with points in the direct and indirect neighborhood of these data points. We provide these modified graphical representations in Figure 2. The original graphical representations of the data is available in Appendix 3 and further detail on this approach and corresponding literature in appendix 4.

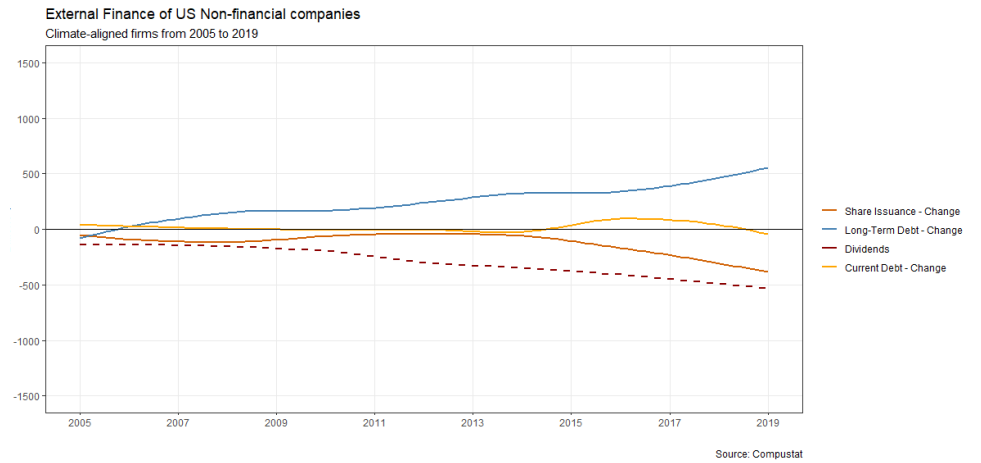
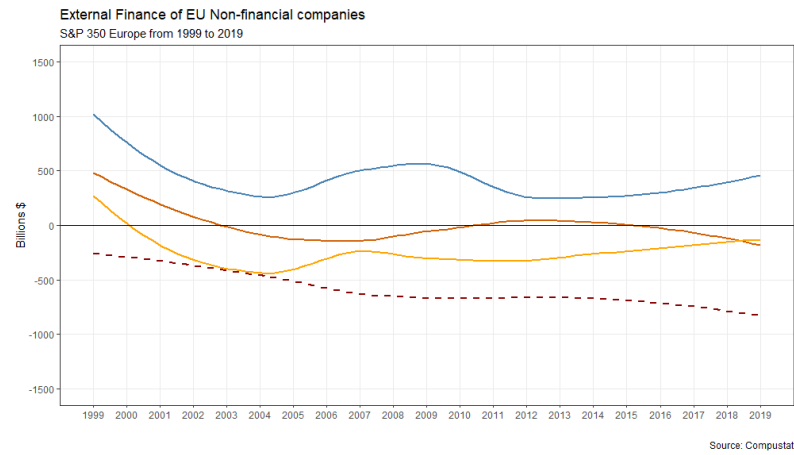
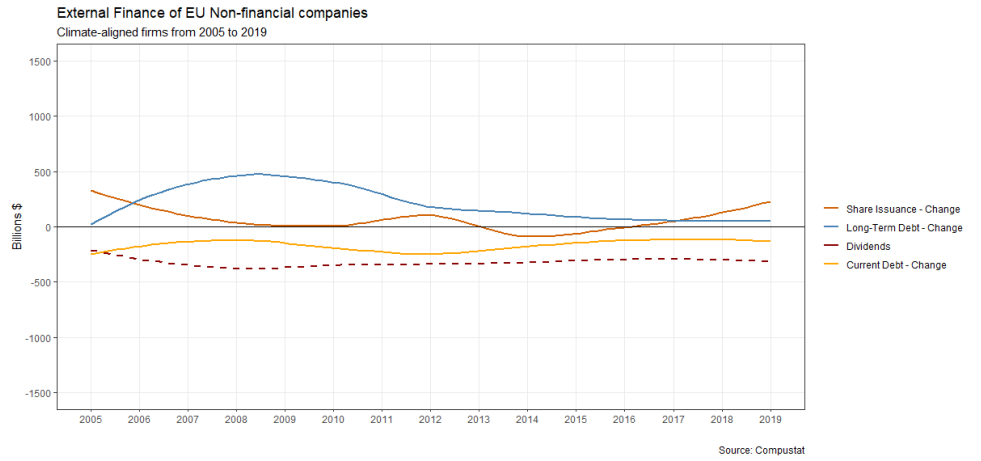
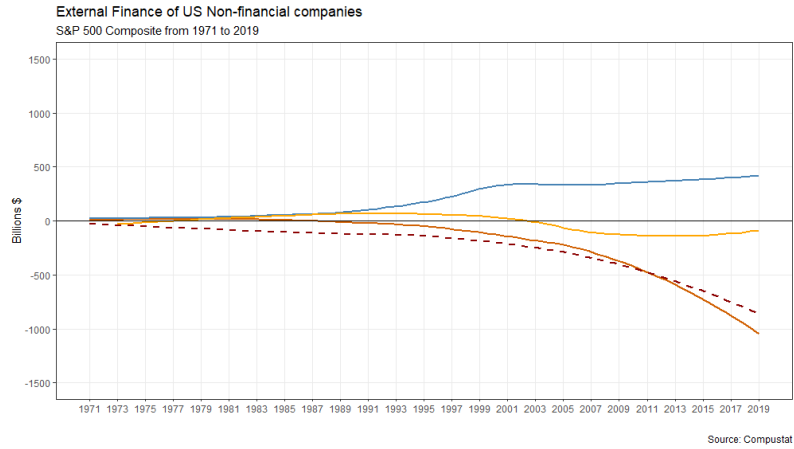
Using this method, the constant presence of long-term debt in positive financing territory is made quite clear. This allows us to observe that long-term debt has been the main source of external finance for S&P 350 Europe firms in the Eurozone for the entire 20 years period represented in our sample, and since the beginning of the 90s for US companies in the S&P 500. This is also quite clear regarding both US and Euro climate-aligned firms.

Looking more precisely at trends for each sample, we observe that long-term debt has been the only positive source of financing for S&P500 firms since 2002, stabilizing at around \$400 billion per firm. Much like we have observed in previous sections, negative stock market financing for large US firms has been growing steadily since the beginning of the 1990s. The situation for large EU firms is quite different. Since the starting period of our sample in 2005, though long-term debt has been by far the largest source of financing, it has never strayed too far from changes in share issuances, which were in positive territory between 1999 and 2003 and between 2011 and 2015. Dividends, however, have always been almost proportionally negative since the beginning of our sample period.

For climate-aligned firms in the US, our graphical representation illustrates a slow but constant growth of long-term debt financing since the beginning of our sample, mirrored by a slow but constant growth of distribution of dividends. Net share issuances, though being constantly

negative since the beginning of our sample, have mostly decreased recently, between 2014 and 2019. Finally, the situation for climate-aligned Euro firms is also quite specific, with a larger participation of long-term debt financing between 2007 and 2011, after which long-term debt financing starts a slow but quite steady downward stream until the end of our sample period. Net share issuances alternate between slightly positive and slightly negative territory throughout the sample period. Overall, we do find interesting patterns in these graphical representations:

Figure 2



while debt financing is the only constantly positive source of financing for all our sample, dividends always closely mirror long-term debt within negative territory. We also find that the behavior of net share issuances differ mostly between our two geographical regions: net share issuances are constantly negative in the US and alter between positive and negative territory in the Eurozone.

6. Corporate Bonds and Long-term debt for US firms

As the purpose of this paper is to understand the role of capital markets in financing firms, we can investigate further and try to understand what part of this long-term debt is constituted of publicly tradable securities. As we've explained in section 3 of this paper, we can extract corporate bond issuances and repayments for every US firms that has issued a corporate bond between 1995 and 2018. Using the Compustat US database, we can also extract total long-term debt issuances and repayments of these firms. This allows us to understand what part of these firms' long-term debt originates from capital markets through the US corporate bond market. This approach also applies to US climate-aligned firms.

Using this approach, we compute the total amounts of long-term debt and corporate bond issuances and repayments for both our sample of traditional firms and climate-aligned firms. Our main results are available in Table 2. Details on the total yearly amounts of long-term debt and corporate bond issuances and repayments for both traditional firms and climate-aligned firms are available in Appendix 4.

Both yearly long-term debt issuances and long-term debt repayments are vastly superior to yearly corporate bond issuances and repayments.

Table 2 - Long-term debt and corporate bond financing

	All corporate bond issuers (1995-2018)		Climate-Aligned Firms (2005-2018)	
Net Corporate Bond Debt	2956.99	59%	50.24	66%
Other Net Long-term Debt	2079.31	41%	26.23	34%
Total Net Long-Term Debt	5036.30		76.47	

However, when computing the differences between issuances and repayments for total long-term debt and total corporate bond debt, we find that net corporate bond debt represents a majority of the total net long-term debt of US firms in our sample. With total net long-term debt on the period 1995 to 2018 reaching \$5.036 trillion for firms in the FISD database, \$2.957 trillion of this amount is corporate bond debt, which represents 59% of this total. This is quite interesting provided the fact that we observed in the previous section that long-term debt was already by far the primary source of financing for US firms in the S&P 500 by 1995. The importance of corporate bond financing for climate-aligned bonds is even pronounced, as it represents 66% of total net long-term debt between 2005 and 2018.

7. Conclusion

Our study of cash flow items of S&P500 firms and Euro firms in the S&P350 Europe Index confirms what has been displayed in the literature focusing on stock-buybacks and dividends: corporations are not getting new funding from stock markets, but are instead channeling cash generated by their business activities towards their shareholders, and in some sense financing stock markets instead of being financed by them. We find that a net sum of more than \$7.7 trillion has been paid to shareholders by large US firms since 2009, and more than €3.7 trillion for large European firms in the Eurozone. By computing average cash flow items instead of cumulative

cash flow items for our samples, we are able to look extend our period of analysis as far as 1971 for large US firms and 1999 for large European firms in the Eurozone. Our results confirm that stock markets have never been a true source of positive net financing for either large US or large Euro firms for these periods of time.

We then observe that long-term debt has been the only source of positive net financing for these firms during these large periods of time. Using the FISD database on US corporate bonds, we focus on the US debt market, and find that corporate bond financing composes a majority of this long-term debt financing. We understand from our data that the US corporate bond market, contrarily to the US stock market, does represent a positive source of financing for US firms in our sample, but also for US climate firms. This provides interesting insight on the possible positive role that could be played by corporate bond markets for sustainable finance and in financing solutions for climate change adaptation and climate change mitigation in the context of growing climate urgency.

As a recent literature that focuses on the relationship between sources of external finance for corporations and their capacity to innovate and invest is starting to develop, academics will undoubtedly have to look more closely and specifically at both stock markets and bond markets to understand their role and impact on the investment decisions of large firms throughout the world. It is becoming increasingly clear that a large part of the necessary efforts against climate change will have to be performed by large international corporations that constitute financial markets, and further research needs to be performed on these specific subjects. In this context, this study represents a first attempt in understanding the roles that are played and could be played by financial markets in providing capital to finance climate change mitigation and adaptation.

References

- Aghion, P., Bechtold, S., Cassar, L. and Herz, H., 2014. *The causal effects of competition on innovation: Experimental evidence* (No. w19987). National Bureau of Economic Research.
- Ahmad, E., 2014. Public finance underpinnings for infrastructure financing in developing countries. *Paper for the G, 24*.
- Baldwin, C.Y., 1991. How capital budgeting deters innovation—and what to do about it. *Research-Technology Management*, 34(6), pp.39-45.
- CEA, 2016. Benefits of competition and indicators of market power.
- Christensen, C.M., Kaufman, S.P. and Shih, W.C., 2008. Innovation killers: how financial tools destroy your capacity to do new things. *Harvard business review*, 86(1), pp.98-105.
- Climate Bonds Initiative, 2018. Bonds and Climate Change: The State of the Market.
- Council of the European Union, 2019. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the establishment of a framework to facilitate sustainable investment - Approval of the final compromise text
- European Commission, 2020. “Taxonomy: Final report of the Technical Expert Group on Sustainable Finance”. March 2020.
- Fichtner, J., Heemskerk, E.M. and Garcia-Bernardo, J., 2017. Hidden power of the Big Three? Passive index funds, re-concentration of corporate ownership, and new financial risk. *Business and Politics*, 19(2), pp.298-326.
- Griffin, P., 2017. CDP carbon majors report 2017. *Carbon Majors Database*.
- Gutiérrez, G. and Philippon, T., 2016. *Investment-less growth: An empirical investigation* (No. w22897). National Bureau of Economic Research.
- Gutiérrez, G. and Philippon, T., 2017. *Declining Competition and Investment in the US* (No. w23583). National Bureau of Economic Research.
- Harford, J., 1999. Corporate cash reserves and acquisitions. *The Journal of Finance*, 54(6), pp.1969-1997.
- Harford, J., Mansi, S.A. and Maxwell, W.F., 2008. Corporate governance and firm cash holdings in the US. *Journal of financial economics*, 87(3), pp.535-555.
- IEA, 2016. Renewable Energy Medium-Term Market Report, 2016: Market Analysis and Forecasts to 2021.
- IEA, 2017. Perspectives for the Energy Transition: Investment Needs for a Low-Carbon Energy System. International Energy Agency. *International Renewable Energy Agency, Paris, Bonn*.
- Kotsantonis, S. and Serafeim, G., 2019. Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, 31(2), pp.50-58.
- Lazonick, W., 1990. Organizational capabilities in American industry: The rise and decline of managerial capitalism. *Business and Economic History*, pp.35-54.
- Lazonick, W. and O'sullivan, M., 2000. Maximizing shareholder value: a new ideology for corporate governance. *Economy and society*, 29(1), pp.13-35.
- Lazonick, W., 2015. Stock buybacks: From retain-and-reinvest to downsize-and-distribute. *Center for Effective Public Management at Brookings*, 17.

Appendix 1

Panel A: Firms in S&P 500 from 2009 to 2019

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total	Percentage of Net Income
Income Before Extraordinary Items	619.37	757.19	848.65	803.01	928.26	914.88	790.38	834.09	976.74	1087.94	1199.58	9760.10	
Depreciation and Amortization	429.65	439.64	466.46	493.23	519.72	542.32	598.18	608.76	647.92	687.86	765.38	6199.11	64%
Other operating activities	186.58	144.28	251.40	191.76	330.47	148.71	263.76	124.40	-50.96	195.78	-2.35	1783.83	18%
Operating Activities - Net Cash Flow	1235.60	1341.11	1566.51	1488.00	1778.46	1605.90	1652.32	1567.25	1573.70	1971.57	1962.61	17743.04	182%
Capital Expenditures	-445.12	-474.46	-566.44	-625.13	-633.36	-689.41	-645.69	-608.70	-621.39	-712.97	-736.16	-6758.83	-69%
Acquisitions	-104.63	-171.78	-198.71	-208.65	-139.54	-171.36	-345.81	-400.77	-265.08	-417.65	-260.91	-2684.88	-28%
Other investing activities	351.62	15.47	-373.06	-515.26	-592.67	-548.70	-311.69	-478.11	-375.77	-372.14	-245.54	-3445.85	-35%
Investing Activities	-198.14	-630.78	-1138.21	-1349.03	-1365.58	-1409.46	-1303.19	-1487.57	-1262.24	-1502.75	-1242.61	-12889.55	-132%
Sale of Common and Preferred Stock	242.75	126.38	127.91	114.82	111.59	116.96	144.00	107.33	82.75	95.98	109.44	1379.94	14%
Purchase of Common and Preferred Stock	-272.86	-274.74	-431.23	-361.99	-424.21	-498.75	-512.87	-511.84	-499.13	-735.08	-703.77	-5226.47	-54%
Cash Dividends	-212.87	-219.94	-250.64	-293.36	-319.38	-355.35	-397.52	-413.20	-439.28	-479.44	-494.64	-3875.63	-40%
Long-Term Debt - Change	-107.74	-325.17	-194.13	-79.96	210.89	322.76	487.90	379.74	316.52	174.52	132.99	1318.32	14%
Current Debt Changes	-478.86	-64.77	-139.93	83.12	-154.62	31.96	-164.55	-12.63	88.92	9.67	-55.66	-857.37	-9%
Other financing activities	-49.75	85.50	589.31	437.95	252.45	184.60	178.98	349.25	268.65	266.80	386.15	2949.88	30%
Financing Activities	-879.34	-672.74	-298.70	-99.43	-323.28	-197.81	-264.06	-101.35	-181.56	-667.54	-625.50	-4311.33	-44%

Panel B: Firms in S&P Europe 350 from 2009 to 2019

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total	Percentage of Net Income
Income Before Extraordinary Items	94.48	181.89	160.49	151.75	136.70	153.20	116.54	158.26	226.50	204.93	184.33	1769.06	
Depreciation and Amortization	182.91	189.38	183.45	192.78	188.18	185.40	201.53	206.02	206.62	212.73	259.82	2208.81	125%
Other operating activities	75.57	-0.92	-17.35	-18.35	-0.62	-19.31	41.86	-5.79	-80.31	-33.00	-4.63	-62.85	-4%
Operating Activities - Net Cash Flow	352.96	370.35	326.59	326.18	324.25	319.29	359.92	358.49	352.82	384.66	439.52	3915.02	221%
Capital Expenditures	-210.10	-205.42	-205.12	-214.94	-205.49	-197.89	-211.44	-206.99	-217.27	-215.49	-230.54	-2320.69	-131%
Acquisitions	-80.05	-30.49	-45.79	-35.71	-20.79	-51.43	-79.01	-51.31	-67.86	-115.96	-53.68	-632.08	-36%
Other investing activities	32.13	-14.77	13.98	18.61	2.50	36.08	5.60	-19.55	13.91	9.31	-13.14	84.65	5%
Investing Activities	-258.02	-250.68	-236.94	-232.04	-223.77	-213.24	-284.85	-277.85	-271.22	-322.15	-297.36	-2868.11	-162%
Sale of Common and Preferred Stock	35.21	17.66	17.60	14.68	19.20	23.73	22.55	20.51	19.22	26.42	13.50	230.30	13%
Purchase of Common and Preferred Stock	-4.60	-9.09	-16.07	-9.28	-12.17	-20.57	-20.21	-19.56	-21.58	-28.75	-33.83	-195.71	-11%
Cash Dividends	-79.64	-71.82	-84.29	-81.09	-79.34	-95.82	-83.76	-86.33	-89.98	-100.89	-112.45	-965.40	-55%
Long-Term Debt - Change	84.56	-24.70	35.12	77.51	22.63	25.66	23.18	55.96	52.25	88.27	28.29	468.72	26%
Current Debt Changes	-29.69	-10.42	-4.78	-19.78	-13.85	-9.57	-0.47	-13.63	-17.74	-3.73	-8.94	-132.60	-7%
Other financing activities	-21.22	-23.00	-26.36	-35.94	-17.11	-27.79	-10.27	-25.71	-0.51	-20.34	-16.33	-224.59	-13%
Financing Activities	-15.37	-121.38	-78.79	-53.90	-80.63	-104.35	-68.98	-68.75	-58.34	-39.03	-129.76	-819.29	-46%

Appendix 2

Panel A: US Climate Firms from 2009 to 2019

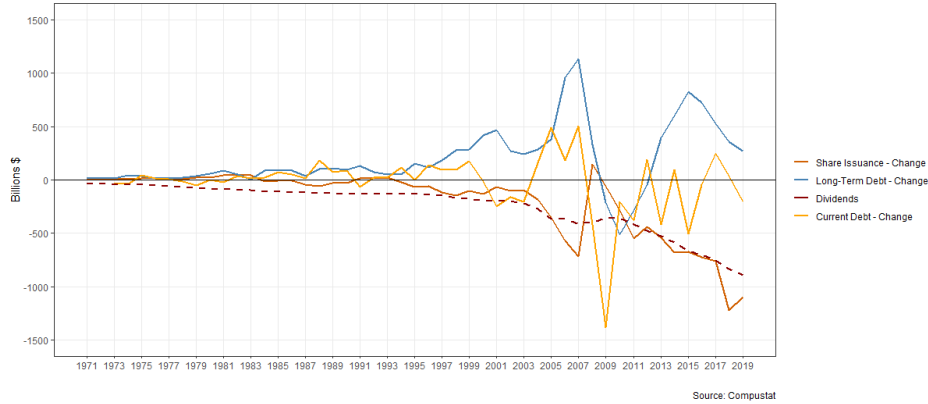
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total	Percentage of Net Income
Income Before Extraordinary Items	7.32	12.51	12.72	13.48	16.85	17.99	16.55	15.45	35.86	24.93	21.34	195.00	
Depreciation and Amortization	9.62	10.07	10.19	10.87	12.02	13.40	14.13	15.70	16.37	18.44	19.88	150.69	77%
Other operating activities	4.14	-0.79	3.35	2.66	1.84	1.72	2.25	3.92	-17.71	-1.13	3.29	3.55	2%
Operating Activities - Net Cash Flow	21.08	21.78	26.27	27.01	30.71	33.11	32.93	35.08	34.52	42.24	44.51	349.24	179%
Capital Expenditures	-12.50	-13.01	-17.89	-20.90	-18.81	-23.27	-25.96	-22.48	-25.28	-24.82	-25.00	-229.92	-118%
Acquisitions	-0.33	-2.18	-2.66	-4.75	-3.97	-4.30	-3.08	-3.41	-3.35	-1.83	-12.63	-42.48	-22%
Other investing activities	-2.59	-1.76	-0.08	-2.33	-1.16	-1.34	-3.51	-1.64	-1.21	-4.17	-3.59	-23.36	-12%
Investing Activities	-15.41	-16.95	-20.63	-27.98	-23.94	-28.91	-32.55	-27.54	-29.84	-30.81	-41.22	-295.77	-152%
Sale of Common and Preferred Stock	1.16	1.03	0.91	1.31	5.31	2.12	3.57	3.49	1.96	1.59	3.46	25.91	13%
Purchase of Common and Preferred Stock	-0.18	-2.31	-4.65	-3.09	-3.53	-4.58	-7.37	-7.14	-6.09	-13.31	-10.56	-62.81	-32%
Cash Dividends	-2.89	-4.51	-7.05	-8.20	-8.71	-9.03	-10.78	-9.66	-12.21	-14.10	-13.88	-101.03	-52%
Long-Term Debt - Change	-0.92	2.01	4.97	7.66	6.86	8.66	12.24	7.39	7.96	10.49	17.50	84.82	43%
Current Debt Changes	-0.40	-1.06	0.73	0.04	-0.62	0.63	-0.27	0.00	1.93	6.92	-4.97	2.92	1%
Other financing activities	-0.09	0.24	0.53	0.67	-1.10	-1.02	0.14	1.32	0.94	1.54	0.59	3.77	2%
Financing Activities	-3.32	-4.60	-4.55	-1.62	-1.78	-3.21	-2.48	-4.60	-5.51	-6.88	-7.87	-46.41	-24%

Panel B: Euro Climate Firms from 2009 to 2019

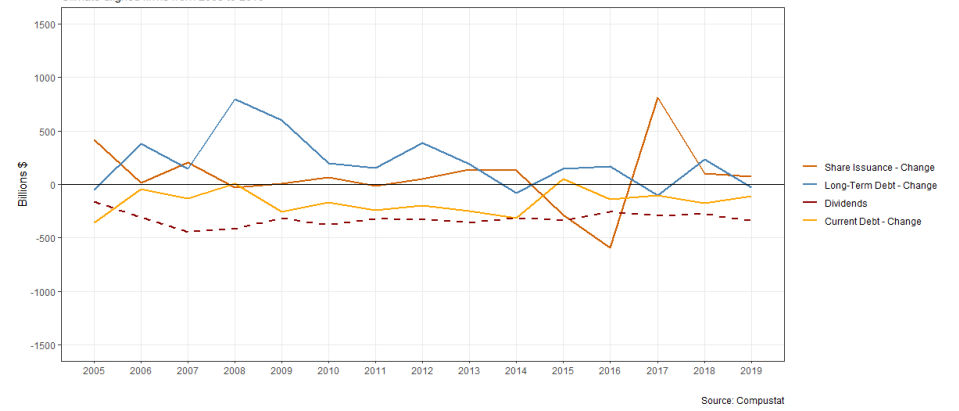
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total	Percentage of Net Income
Income Before Extraordinary Items	9.02	7.25	8.59	9.90	6.14	8.77	1.87	7.71	8.42	6.95	12.16	86.78	
Depreciation and Amortization	15.83	16.20	15.49	15.76	16.44	16.14	17.29	16.39	17.21	17.70	21.62	186.07	214%
Other operating activities	2.81	2.83	-2.26	-0.05	2.15	-0.40	4.74	0.85	-1.65	1.86	-2.94	7.94	9%
Operating Activities - Net Cash Flow	27.66	26.28	21.82	25.61	24.73	24.51	23.90	24.95	23.98	26.51	30.84	280.79	324%
Capital Expenditures	-30.74	-27.48	-26.89	-29.59	-27.99	-28.21	-30.27	-29.42	-31.39	-33.66	-35.99	-331.63	-382%
Acquisitions	-15.71	-0.98	-0.57	-0.64	-0.33	-0.20	-2.21	-1.28	-5.44	-0.45	-0.95	-28.76	-33%
Other investing activities	18.48	-0.95	11.82	6.73	7.50	13.21	20.26	4.97	13.13	7.28	11.26	113.69	131%
Investing Activities	-27.97	-29.41	-15.64	-23.50	-20.82	-15.20	-12.22	-25.73	-23.70	-26.83	-25.68	-246.70	-284%
Sale of Common and Preferred Stock	0.37	1.27	0.10	0.35	0.13	0.74	0.68	0.06	6.65	0.12	3.23	13.70	16%
Purchase of Common and Preferred Stock	-0.29	-0.33	-0.31	-0.20	-0.03	-0.45	-3.73	-0.08	-0.10	-0.30	-1.76	-7.58	-9%
Cash Dividends	-4.06	-5.36	-5.09	-5.88	-5.48	-5.21	-5.32	-3.83	-4.46	-4.81	-5.77	-55.27	-64%
Long-Term Debt - Change	13.33	4.44	3.32	8.66	5.17	-1.23	3.91	2.81	-2.66	4.04	1.61	43.40	50%
Current Debt Changes	-1.53	-0.85	-1.20	-0.98	-1.25	-1.59	0.26	-0.95	-0.53	-0.88	-0.53	-10.03	-12%
Other financing activities	-8.22	-0.23	3.91	-1.91	0.84	8.47	3.62	2.17	3.21	8.59	-1.71	18.74	22%
Financing Activities	-0.40	-1.06	0.73	0.04	-0.62	0.73	-0.58	0.18	2.11	6.76	-4.93	2.96	3%

Appendix 3

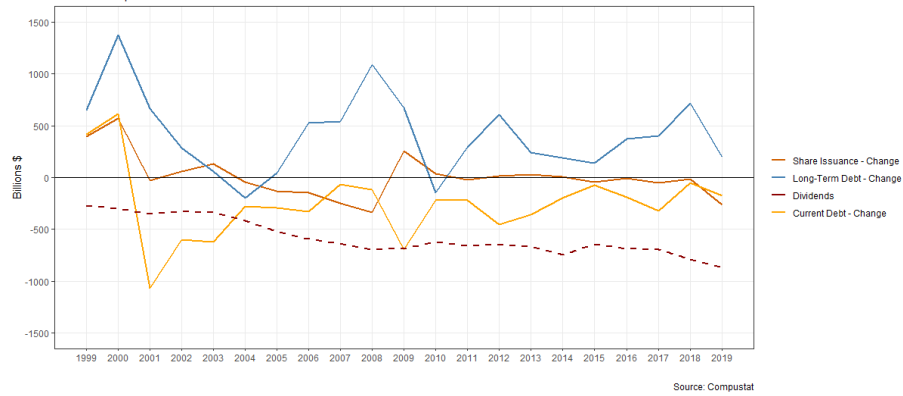
External Finance of US Non-financial companies
S&P 500 Composite from 1971 to 2019



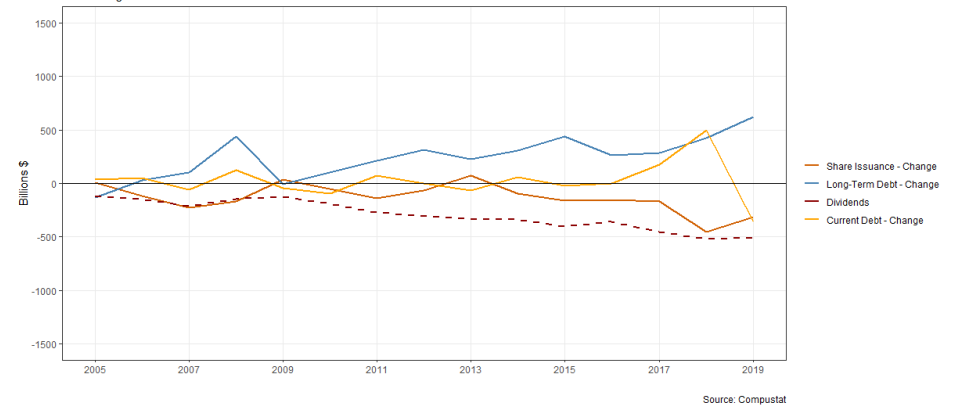
External Finance of EU Non-financial companies
Climate-aligned firms from 2005 to 2019



External Finance of EU Non-financial companies
S&P 350 Europe from 1999 to 2019



External Finance of US Non-financial companies
Climate-aligned firms from 2005 to 2019



Appendix 4

Panel A: Corporate Bond and Long-term Debt Financing for all firms in the FISD dataset from 1995 to 2018

	1995-2000	2001-2005	2005-2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Corporate Bonds Issuances	263.18	463.95	899.75	219.02	263.13	336.04	309.87	488.99	415.62	502.51	300.88	4462.94
Long-Term Debt - Issuance	2772.88	4517.24	8727.86	1218.78	1452.25	1561.73	1755.39	1999.94	2041.81	2057.18	2051.06	30156.10
Corporate Bonds Repayments	69.70	173.63	293.49	64.80	76.27	97.66	110.67	98.70	164.24	178.02	178.76	1505.95
Long-Term Debt - Repayments	1975.81	3556.81	7473.27	1429.15	1494.50	1304.85	1333.66	1437.51	1582.91	1725.46	1805.86	25119.81
Net Corporate Bond Financing	193.48	290.32	606.26	154.22	186.86	238.38	199.20	390.29	251.37	324.50	122.12	2956.99
Net Long-Term Debt Financing	797.06	960.43	1254.59	-210.37	-42.25	256.88	421.73	562.43	458.91	331.72	245.19	5036.30

Panel B: Corporate Bond and Long-term Debt Financing for US Climate Firms from 2005 to 2018

	2005-2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Corporate Bonds Issuances	19.28	3.63	4.90	6.78	8.44	8.12	6.68	9.20	10.80	77.84
Long-Term Debt - Issuance	85.04	18.51	25.51	24.81	30.39	36.37	31.16	37.22	40.06	329.08
Corporate Bonds Repayments	12.21	1.22	1.51	0.72	2.12	2.91	1.79	3.18	1.95	27.60
Long-Term Debt - Repayments	76.09	13.75	17.42	17.89	21.63	23.65	23.43	29.09	29.66	252.61
Net Corporate Bond Financing	7.08	2.41	3.39	6.05	6.33	5.21	4.89	6.03	8.85	50.24
Net Long-Term Debt Financing	8.95	4.77	8.10	6.92	8.76	12.72	7.73	8.13	10.40	76.47

Chapter 4

A data processing methodology for green bond and climate bond primary and secondary market

Abstract

The aim of this paper is to explore challenges for developing a detailed, comprehensive and up-to-date database focusing on the US primary and secondary markets for green bonds and climate bonds. In September 2018, the green and climate bond universe represented USD1.2tn, and included 869 issuers, giving research an opportunity to thoroughly analyze the first financial products to have an extra-financial purpose. This paper first focuses on identifying, acquiring and storing various types of green and climate bond data using available financial databases. It then gives a detailed walk-through on structuring and processing this data. Different data process methodologies result in obtaining various databases on green bonds and climate bonds that can be used to develop a variety of factors that can provide insight on the green bond and climate bond primary and secondary market, as well as obtaining monthly returns for these markets in order to explore the cross-section or times series of green and climate corporate bond returns.

Keywords: Green Bonds; Climate Bonds; Sustainable Finance; Climate Change; Enhanced TRACE Data

1. Introduction

With the first green bond issuance dating back to 2007, the green bond market is still young. However, since the year 2013, when total issuances grew close to \$10 billion worldwide, offerings for green bonds have grown at an increasing rate. In 2019, issuances for the year reached \$257.5 billion. This presents not only an opportunity for market practitioners, but for academics as well, as data relative to the primary green bond market starts to accumulate, and the first forms of transaction data relative to the secondary market for these products are starting to emerge.

As one of the first financial products with an extra-financial purpose to reach such a market size, the study of green bonds offers an opportunity to gather interesting and useful insights on the specificities of financial products that do not have a purely financial objective, as well as on the investors that trade these specific securities. However, green bonds are not the only products that have an underlying environmental purpose. Bonds issued by corporations that stream more than 75% of their revenues from climate-aligned activities have been categorized as a specific type of bond. These bonds, referred to as climate bonds, though not officially labeled, can undoubtedly provide further insight on the specificities of environment-related financial products.

However, creating a clean database that has the most detailed, comprehensive and up-to-date data on green and climate bonds presents a series of challenges. This paper focuses on identifying these challenges, and on constructing a robust and practical methodology in order to address them. Setting-up such a database and data treatment methodologies as the green and climate bond markets continue to grow could be useful for academics or practitioners that wish to analyze it, but also to review already existing work and data treatment procedures for the study of corporate bond markets.

Compared to the literature addressing the study of stocks, the literature which deals with the study of corporate bonds is quite recent, and academics in the field have constantly been challenged with issues related to data availability and quality. The primary source of data on the corporate bond market, the Lehman Brothers Fixed Income database closed in 1997, and there was no quality data available on this market until the Mergent Fixed Income Securities Database opened in 2004, and the Trade Reporting and Compliance Engine database (TRACE) was put in place starting in 2002. Furthermore, TRACE data was at first quite unreliable, and required rigorous data treatment that wasn't clearly identified by the literature until 2009. To date, many studies on the subject still do not address some essential steps of this data cleaning procedure which could bias results. In this paper, we create a clear data treatment procedure, applied to a sample of green and climate bonds, to create a framework to study these products using the thorough methodology that needs to be implemented when one studies the corporate bond market.

After explaining the different approaches that are taken on the subject of treatment of corporate bond data in the literature, we develop a methodology specific to green corporate bonds and climate bonds using a variety of different data sources. We use data from Bloomberg to identify green bonds and data provided by the Climate Bond Initiative (CBI)³⁹ to identify climate bonds. We then extract raw trade-based data from the Enhanced TRACE dataset and develop a data treatment procedure closely following Dick-Nielsen (2009) and Dick-Nielsen (2014), two studies in the literature that address the data treatment procedure for TRACE data in the most detail. Through various steps, this data treatment methodology transforms raw intra-day transaction data into monthly prices, creating different types of databases along the way that can be used for various type of studies on the corporate bond market. Once we obtain

³⁹The Climate Bond Initiative is an international, investor-focused not-for-profit that focuses explosively on the bond market for climate change solutions. The amount of global green bond issuances is monitored on the organization's website. See <https://www.climatebonds.net/>.

monthly prices from the TRACE database, these are merged with monthly prices from Datastream and Bloomberg. Following the literature, we first use TRACE data for our monthly prices, then, if TRACE data is unavailable, Datastream data is used, followed by Bloomberg. This choice that is put forward by academics in the field, is better understood when we compute correlations between datasets. Datastream data is strongly correlated to TRACE data, with a correlation of 0.93 for green bond data and 0.96 for climate bond data, but Bloomberg has a surprisingly low correlation with TRACE, with a correlation of only 0.28 for our green bond sample and 0.75 for our climate bond sample. This does not only provide insight regarding the quality of data provided by Bloomberg, but also on previous work performed by academics that used Bloomberg data to study the corporate bond market. We also find that Bloomberg data provides very little monthly bond prices to our datasets that is not already provided by either TRACE and Datastream.

We conclude this paper by visualizing monthly returns for our samples of green bonds and climate bonds and find that though climate bond monthly returns and green bonds climate bond returns are strongly correlated overall, these returns vary greatly during the last months of 2016, and correlation is not perfect. From this visualization, we understand that differences exist between these two types of products. Though the precise analysis of these differences is outside the scope of this paper, this provides interesting insight that could prove useful for future research.

The remainder of the paper is structured as follows. Section 2 presents the existing literature related to the study of green and climate bonds as well as the literature on the study of traditional corporate bonds that has led the author to develop such methodology. Section 3 presents the different financial databases that are used as well as their advantages and limitations and the different steps of this data processing methodology. Section 4 describes the

resulting datasets we obtain from these methodologies and their applications and the results we obtain for green bonds and climate bonds. Section 5 concludes the study.

2. Literature Review

2.1. The nascent climate-aligned bond literature

The growing number of green bond issuances has led academics to develop an interest in this product. Green bonds are bond products issued in order to specifically finance projects with positive environmental outcomes, and naturally a stem of literature is focused on trying to determine if green bond issuers benefit from a green bond premium given the nature of their project.

However, this initiative has met important challenges. First and foremost, the pricing of bond products is more challenging than that of equity products, given the fact that bonds have multiple specific characteristics that directly impact their pricing, such as coupon rates, credit rating, maturity or size (Zerbib, 2019). Determining if a green bond premium exists would require comparing green bonds with traditional bonds that have precisely the same characteristics, which very rarely exist (Bachelet et al, 2019). In addition to these important limitations, the green bond market represented less than 3.6% of the global bond market issuances⁴⁰ in 2019, and issuances are still sporadic throughout the year, which leads both to issues in terms of liquidity, as well as in terms of available data for pricing. Moreover, the academic literature that focuses on studying the corporate bond market refers to one specific transaction-based database to obtain the best quality data on the pricing of corporate bonds – the TRACE database⁴¹ – which only applies to US corporate bonds. For the year 2018,

⁴⁰ Global bond issuances in 2019 amounted to \$7.148 trillion. See <https://www.dealogic.com/insight/dcm-highlights-full-year-2019/>

⁴¹ As TRACE transaction data is the main source of data for this paper, more information is provided in section 4 of this paper on the specificities of this database

Moody's Analytics⁴² reported that \$1.553 trillion in bonds were issued in the US, while the Climate Bond Initiative reported \$34 billion in US green bond emissions (Climate Bond Initiative, 2018). This represents less than 2.2% of US corporate bond emissions, meaning that academics that wish to study this database would have access to too little data to perform robust analyses.

Studies on the pricing of green bonds therefore generally have small samples and must refer to other databases with less precise pricing data, such as dealer quotes provided by market-makers or matrix-prices, which only provide approximations of real prices. To provide an idea of the general ranking of data on corporate bond pricing in terms of quality, Jostova, Nikolova, Philipov and Stahel (2011) built a database of bond returns using the five databases that gave information on corporate pricing and took "the first available return in the following sequence: TRACE, FISD⁴³, Lehman, Datastream and Bloomberg", clearly giving precedence to trade-based data. Combined, these restrictions in terms of sample size, historical data and pricing data quality results in inconclusive findings concerning the existence of a green bond premium (Hachenberg and Schiereck, 2018; Bachelet et al, 2019; Zerbib, 2019). Similar studies have also focused on the relationship between ESG ratings and corporate bond performance, finding that bonds issued by firms with higher ESG ratings have tighter spreads and outperform peers with lower ESG ratings (Polbennikov, Desclée, Dynkin and Maitra, 2016). Ge and Liu (2015) find that better CSR performance is associated with better credit ratings. In a paper focusing specifically on corporate green bonds, Flammer (2018) finds a positive reaction from stock markets to green bond issuance announcements, that green bond issuers improve their environmental performance after the issuance, and that they experience increase in ownership

⁴² See <https://www.moodyanalytics.com/-/media/article/2019/weekly-market-outlook-corporate-bond-issuance-reflects-business-activity-heightened-to-rates.pdf>

⁴³ When referring to FISD, the authors referred to the NAIC databases which complements the FISD database, which itself only provides information on characteristic data of US bonds. The NAIC database provides transaction data on corporate bonds issued and traded by US insurance companies.

by long-term and green investors. To date and to the author's knowledge, there is no academic literature focusing on empirical studies of climate bonds.

2.2. Recent studies on the corporate bonds market

Looking at the progress that has been made relative to the study of traditional bonds can help us understand what research questions and possible results can be applied and obtained when focusing more specifically on green and climate bonds. This is quite understandable, as green bonds only differ from traditional bonds in the fact that they have been labelled as being green and are financial tools used to finance environmental projects, and climate bonds, which have only recently been identified by the Climate Bonds Initiative (CBI, 2018), are traditional bonds that finance firms that are considered to positively - or at least not negatively - affect climate change. In addition, looking at the young literature that focuses specifically on green bonds and climate bonds can help us identify the first research questions that have been applied, and results that have been obtained, on these specific bond markets, and understand how our work can benefit to this literature.

There is an important literature focusing on liquidity-related issues on the traditional corporate bond market. Acharya et al. (2013) show that the pricing of liquidity risk in the bond market is conditional to the state of the economy, and that liquidity risk is more important in times of financial and economic distress. Focusing on bond-specific liquidity measures, Chen et al. (2007) find that liquidity is priced in corporate bond yield spreads and Lin et al (2011) investigate corporate bond expected returns and find that these are partly explained by liquidity risk. Using transaction data similar to ours, Bao et al (2011) shows that illiquidity in corporate bonds is substantial and significantly greater than what can be explained by bid-ask spreads. The authors establish a strong link between illiquidity and bond prices. Two years later, Bao et al (2013) find that empirical volatilities of corporate bond returns are higher than implied by equity return volatilities and the Merton model due to illiquidity.

Focusing more specifically on traditional bond returns, DeCosta (2017) find that investment-grade bonds with short-maturity perform better than similar bonds with longer maturities. These results are attributable in part to the insurance companies' trading behavior, as insurance-company purchases create a strong demand for long-term bonds. By examining underpricing of initial public offerings (IPOs) and seasoned offerings in the corporate bond market, Cai et al. (2007) investigate whether underpricing results from an information problem or a liquidity problem and find that issues related to information cause underpricing. On a similar note, Liu et al. (2014) try to understand the relationship between information risk and the underpricing of newly issued corporate bonds and find that information risk is associated with higher underpricing for these products.

Studies on the effects of increased transparency on corporate bond markets - mostly due to improved trade reporting - have also emerged. On this subject, Bessembinder et al. (2006) find that improved trade reporting in the corporate bond market lowered trade execution costs, also showing that better pricing information regarding some bonds also improves valuation and execution cost monitoring for related bonds and find no evidence that market quality deteriorated in other dimensions. Adding to this work, Edwards et al. (2007), using a record of US over-the-counter (OTC) secondary trades in corporate bonds, find that transactions costs decrease significantly with trade size, and that costs lower for bonds with transparent trade prices, suggesting that public traders benefit significantly from price transparency. Finally, Goldstein et al. (2007) find that adding transparency has either a neutral or a positive effect on liquidity, and that transparency is not associated with greater trading volume. The authors conclude that observed decreases in transaction costs illustrate the investors' ability to negotiate better terms once they have access to better data.

The study of the relationships between equity characteristics and corporate bond characteristics has also stimulated the interest of some academics. In their paper, Chordia et al.

(2017) tell us “although it stands to reason that corporate bonds are not as sensitive to firm outcomes as equities, corporate bond return volatility is still material, at about a third of that of equities for junk bonds and about a fifth for investment-grade bonds”. The authors estimate that uncertainty in cash flow resulting from credit risk could have similarities with equities. Risk-based factors and possible investor biases that apply in equity markets might also apply to the credit risk sector. In a similar manner, De Jong et al. (2007) study the liquidity risk premia in corporate bonds and equity markets and find that corporate bond returns are sensitive to fluctuations in liquidity of the Treasury and equity markets.

In the asset pricing literature, the behavior of bond returns has also tried to be identified using a variety of factors. This was initiated by Fama and French (1993) when the authors identified five risk factors that were common to the returns of stocks and corporate bonds. Since then, the study of the cross-section of corporate bond returns has created interest in the asset pricing literature. Academics that focused on this subject mostly developed factors using either stock-level data, treasury bond data and macroeconomic data. This is the case for the long-established Fama-French (1993) factors composed of the market risk factor (Mm-Rf), the size factor (SMB) and the book-to-market factor (HML) that originate from stock data and treasury bond data, and the term spread (TERM) and default spread (DEF) factors that originate from treasury bond data and government bond data⁴⁴. Other factors that have complemented Fama and French’s work, such as the liquidity (LIQ) factor (Pastor and Stambaugh, 2003), momentum (MOM) factor (Carhart, 1997), and more recently, the investment (CMA) and profitability (RMW) factors (Fama and French, 2015) all originate from stock-level data. Realizing that these factors performed poorly in their ability to explain industry-sorted and

⁴⁴ In order to compute the default factor, a market portfolio of long-term corporate bonds was also required, and the data needed to compute such portfolio return could for the most part only be accessed through Ibbotson Associates, a private investment advisory firm.

size/maturity sorted portfolios of US corporate bonds, Bai, Bali and Wen (2019) introduced new bond-implied risk factors based on characteristics specific to corporate bonds.

The study of the dense corporate bond literature allows us to understand the wide array of possible studies that could be applied to a dataset of green bonds and climate bonds. Whether we chose to focus on liquidity, performance, increased transparency, the equity-bond relationships that can exist within firms or markets, or the more general asset pricing literature that is starting to focus more specifically on the study of the corporate bond market, the fact that these climate-aligned products are just traditional corporate bonds with an underlying climate objective is a an important opportunity to discover whether these products behave differently. However, before being able to perform such analyses, academics need to face an important challenge related to the quality of available data, and the different data treatment procedures that need to be taken care of before being able to obtain robust results.

2.3. The corporate bond data challenge

As this paper focuses on a data processing methodology for green and climate bonds, we study the traditional bond literature to understand what databases are used, how data is treated and more importantly for what purpose. Research on traditional bonds relies on six financial databases, each providing different types of data, on different markets. Some of the data provided by the databases overlap, and the literature also provides a ranking on the most qualitative data sources depending on the data types that are provided. The main databases are Mergent's Fixed Income Securities Database (FISD), the Trade Reporting and Compliance Engine (TRACE) database, the National Association of Insurance Commissioners Database (NAIC), DataStream, Bloomberg, and the Lehman Brothers Fixed Income Database. (Jostova, et al, 2013; Chordia et al., 2017). As the Lehman Brothers Fixed Income Database provides data from January 1973 to March 1998, we exclude it from our analysis, as our dataset spans from January 2013 to July 2019. We explore these databases in detail in Section 3.

In addition to providing us with explicit details on the advantages and limitations of each financial database that provides information on traditional bonds, the literature also gives precedence to specific databases depending on the types of data they provide. Each database can provide information on the characteristics of bonds, information on trades and transactions that have been on the secondary bond market, and information on the bonds prices on the primary bond market. Regarding data on the characteristics of bonds, precedence is given to FISD over Bloomberg, these two financial databases being the only ones that offer such types of data⁴⁵. This is justified by the fact that FISD is recognized as the most comprehensive database on bonds, but also due to the fact that Bloomberg offers only a limited amount of data extractions. As aforementioned, regarding data on bond prices, the literature differentiates dealer-quote data provided by Datastream and Bloomberg from transaction data provided by TRACE and NAIC. Precedence is given to transaction data (Jostova et al., 2013). TRACE transaction data is preferred to NAIC transaction data given the fact that NAIC data only provides information relative to insurance companies. Datastream quote-based data is preferred to Bloomberg quote-based data since it gives no restriction on data extractions. The data selection sequence regarding bond prices is therefore the following: TRACE, NAIC, Datastream, Bloomberg.

As FISD, NAIC and TRACE data only provide information on the US bond market, we understand that we can only apply the most qualitative analyses to the United States market. The study of other geographical markets will be both limited by Bloomberg data extractions for bond characteristics, and by the fact that transaction data is not available for academics. Therefore, in this paper, we choose to focus specifically on the US market.

⁴⁵ Regarding data on characteristics, Datastream only provides us with issuers identification and currency used for the issuance, and TRACE and NAIC with issuer identification and information that is specific to each transaction individually.

Having developed a clear understanding of the general scope of available bond data that can be used to analyze traditional bonds on financial databases, we use the literature to differentiate data treatment processes that are applied to these datasets. In this paper, we differentiate two types of research designs that require distinct data treatment processes: cross-sectional studies and event studies. Two main differences distinguish these approaches. In a majority of cases, cross-sectional studies can be performed using monthly bond prices while it has been demonstrated that event studies perform better when using daily transaction data. Furthermore, it is recommended that noninstitutional trades below \$100,000 be eliminated in the case of event studies. (Bessembinder et al., 2009). Regardless of these two approaches, much like it is the case for research on equity markets, most research on traditional bonds focus on bond returns, whether monthly returns for cross-sectional studies, and daily returns for event studies. We give further detail on the specific data processing procedures to obtain daily and monthly green bond and climate bond returns for both approaches in the following section.

3. Data and Methodology

3.1. Databases of the corporate bond market

To date, there are five distinct sources to find financial data relative to US green bonds and climate bonds. For the purpose of our database, it is important to differentiate three types of data that are important to obtain to study these bonds. Firstly, data on bond characteristics is essential. This includes information on bond issuance dates, maturities, currency, ratings, industry, and any form of information that can help us identify bonds and group them by specific categories. Other types of data include historical prices, whether this regards daily, weekly or monthly prices provided by either quote-based databases or trade-based databases. As previously mentioned, the literature gives precedence to trade-based databases, as these are regarded as giving higher quality data on corporate bond prices. Each of the following

databases gives different types of data that we must categorize in order to develop our data processing algorithms. Table 1 summarizes the different used databases, the types of data that they provide as well as the markets they describe.

The Mergent Fixed Income Securities Database (FISD) is considered the most comprehensive database regarding the characteristics of publicly offered U.S. bonds. It contains information on more than 140,000 corporate bonds, medium term notes, supranational or US Agency and Treasury debt products. It provides information both on issuers and on specific issues. The FISD database is composed of a series of datasets that focus on specific characteristics of bond issuers and issuances. In the context of this paper, we focus on the datasets that provide information on issues, issuers, agents, coupons, industry codes and ratings. The FISD database provides us with all necessary information relative to bond characteristics for US green bonds and climate bonds.

Table 1

Database Comparison

Database	Type of Data	Advantages	Limitations
FISD	Characteristics	Most comprehensive database on US Bonds	Limited to the US
Bloomberg	Characteristics, Quote-Based Prices	Provides the list of green bonds. Provides characteristics and historical prices. Is the most comprehensive database overall.	Limited Extractions. Characteristics data on Bonds is less precise than FISD. Quote-based prices are less precise than trade-based prices.
TRACE	Transaction Data	Provides transaction data on US Bond market	Limited to the US
Datastream	Quote-Based Prices	Provides historical prices for all bonds	Quote-based prices are less precise than trade-based prices.

Considered as a dataset contained within the FISD database, we consider the National Association of Insurance Commissioners Database (NAIC) independently from the FISD database, as it provides a different type of data, in a different form, for a different purpose.

NAIC data represents bond sales and purchases by US insurance companies and contains bond transactions in more than 79,000 unique issues for almost 8,000 issuers from 1994 onwards.

Since July 2002, all corporate bond transactions in the secondary market have been made available with the TRACE system through the Trade Reporting Compliance Engine. Before this new dataset was made available, most studies that focused on the corporate bond market only used daily quotes and matrix prices for corporate bonds, which could bias results. In their paper on the subject, Sarig and Warga (1989) explore the fact that there can be liquidity-driven noise errors in daily prices for corporate bonds since daily prices are given even on days when bonds have not been traded for multiple days. When this is the case, brokers set matrix prices based on similar bonds issued by issuers with similar characteristics, which creates bias. More recently, Dick-Nielsen et al. (2009) show that this bias still exists with prices from Datastream. This gives TRACE data a considerable edge when focusing on daily prices that can be used for microstructure research such as event studies, as well as an edge for weekly and monthly prices. Important changes have been made to the TRACE database from 2002 to 2012, however since our database spans from January 2013 to June 2018, only the latest version of TRACE needs to be considered for our data processing methodology. Two distinct versions of TRACE exist: the standard TRACE data that censors trading volumes that are greater than \$5 million for investment grade bonds and greater than \$1 million for speculative grade bonds and that usually has a three-month lag for the availability of data, and the enhanced TRACE data, that has information on all transaction volumes but has an 18 month lag for the availability of data. We use TRACE data conjointly with NAIC data to compute trade-based data for our database.

The last two databases are Bloomberg and Datastream. Datastream provide us with corporate bond prices, while Bloomberg offers both data on bond characteristics and on prices. However, Bloomberg has limited monthly extractions, which is quite an important constraint when considering the sizes of the samples we wish to focus on. It is interesting to note that data

from Datastream is preferred to Bloomberg by academics that focus on the corporate bond market.

3.2. Data Processing Methodology

3.2.1. Obtaining a sample of US Green Corporate Bonds

Amongst the different databases available, Bloomberg is the only one that clearly identifies green bonds. As aforementioned, Bloomberg provides two types of information on green bonds: information on bond characteristics, and monthly quote-based prices for each bond since its month of issuance. As information of bond characteristics is more precise when referring to the FISD database, we only extract monthly prices from Bloomberg and information relative to the identification of each bond. We can then use this information to identify green bonds on the FISD database.

Bloomberg provides two types of identification information on bonds: a CUSIP number, a unique identification number assigned to US and Canada stock and bonds, and an ISIN number, which similarly identifies any specific securities issue throughout the world. Either can be used to obtain a list of green bonds in the FISD database. We identify 2015 green bond instruments on Bloomberg from January 2013 to June 2018. This number represents the total number of bonds, regardless of the types of issuers, geographies or the types of bonds issued. Using bond CUSIPs, we identify 253 US green bonds on FISD.

We then select only U.S. Corporate Debentures (bond type = CDEB) and U.S. Corporate Bank Notes (bond type = USBN), filtering out of our dataset bonds issued by government agencies, medium-term notes or bonds issued in a foreign currency following the literature (Bessembinder et al, 2018). This further reduces our sample to 49 US corporate bonds. Finally, we follow Jostova et al (2011) and exclude non-U.S. dollar denominated bonds, bonds with unusual coupons (e.g., step-up, increasing-rate, pay-in-kind, and split-coupons), mortgage

backed or asset-backed bonds, convertible bonds, bonds with warrants, and bonds part of unit deals from our sample. This only reduces our sample by two additional bonds. Our final sample for green bonds is composed of 47 US green corporate bonds, out of the global sample of 2015 green bond instruments for the period 2013 to December 2018.

3.2.2. Obtaining a sample of US Climate Bonds

In order to identify US climate bonds, we use a list of climate bond issuers provided directly by the Climate Bond Initiative (CBI) research team. This list was used by CBI in their study on the climate bond market (Climate Bond Initiative, 2018). To identify climate bonds, CBI identified issuers that originated at least 75% of their revenues from green business lines in either clean energy, low-carbon buildings and transport, water and waste management and sustainable land use. Climate bonds were included if they were issued after the 1st of January 2005 and before the end of Q2 2018. Therefore, we use the list of issuers provided by the Climate Bond Initiative to identify climate bonds in the FISD database.

Using the same approach that was used on green bond data, we identify 332 US climate bond issuances in the FISD database between 2005 and December 2018 out of a total of 424 global climate bond issuers globally.

3.2.3. Obtaining Enhanced TRACE data

As aforementioned, there are two distinct disseminated TRACE datasets. The original dataset has a 6-month lag compared to the market and does not offer specific information regarding the volumes of transactions that are superior to \$1 million (marked as 1MM+) and transactions that are superior to \$5 million (marked as 5MM+). The enhanced TRACE dataset has an 18-month lag compared to the market but gives information on volumes that are superior to \$1 million.

By using the list of CUSIP numbers we obtain from FISD, we extract 107,848 observations for our green bond dataset and 755,847 observations for our climate bond dataset from the enhanced TRACE database spanning from 01/01/2013 to 09/30/2019⁴⁶. It is important to note that according to Dick-Nielsen (2014), the filters that need to be applied to on the enhanced dataset filter around 35% of observations. Enhanced TRACE data contains information relative to each transaction, providing identification data (CUSIP, TRACE Bond Symbol and Company Symbol), data relative to the time and date of each transaction, as well as data relative to volume and price of the transaction. Furthermore, various data on both the buyer and seller in each transaction is provided.

3.2.4. Developing the Enhanced TRACE data filter

The enhanced TRACE dataset contain a certain amount of reporting errors that need to be identified and deleted. These errors have been made by agents as they were reporting transactions and corrected at a later date through another report. In addition, some transactions are reported multiple times since dealers and agencies that trade for final customers have to report the same trade to TRACE. This can have implications in terms of computing both liquidity and price from the TRACE dataset, and therefore needs to be corrected.

In the literature, two different approaches are used to clean TRACE data. One follows the Dick-Nielsen (2009) procedure, and the other one refers to Bessembinder et al (2009). However, the enhanced TRACE dataset did not yet exist as these papers were being written, and Dick-Nielsen later wrote a paper in which he adapts his filter to enhanced TRACE data more specifically (Dick-Nielsen, 2014). I chose to refer to this most recent approach in this work. The cleaning procedure for reporting errors goes as follows:

⁴⁶ At this study is being performed, this sample is composed only of bonds issued before the 31st of December of 2018, as the author does not have access to FISD 2019 data.

1. Delete cancelled reports. These are reports that are later marked as cancelled by the reporting agent and need to be taken out of the dataset. In order to do this, the reporting agent files a report that consists in canceling the original report that needs to be canceled. This report (marked with an “X” in the Trading Status) and the corresponding original report (that have a similar Sequence Number and that are marked with a regular “T” in the Trading Status) both need to be taken out of the dataset.
2. Delete corrected reports. These are reports that contain a mistake that needs to be corrected. In order to do this, the reporting agent files two additional reports. One consists in indicating that the original report contains a mistake and needs to be corrected, and the other consists in providing the new information for the report containing the corrected data. The report that indicates a correction needs to be made (marked with a “C” in the Trading Status) and the corresponding original report (that have a similar Sequence Number and that are marked with a regular “T” in the Trading Status) are taken out of the dataset. The last report that contains the corrected information needs to be kept in the dataset (marked with an “R” in the Trading Status and has a similar sequence number).
3. Delete reversals. These are reports that have been marked as being cancelled at a much later date. Overall, the procedure is similar than for cancelled reports. Reversal reports (that are marked with an “R” in the “As Of” column and a “Y” in the “Trading Status” column) as well as their corresponding original reports (that have a corresponding Sequence Number and that are marked with a regular “T” in the Trading Status) need to be taken out of the dataset.

Once our error filter is applied, we also need to create an agency filter. Still following Dick-Nielsen (2014), we perform the following steps:

1. Delete agency transactions. Agency transactions occur when a broker acts on behalf of a customer and transacts with an executive broker. When this occurs, three reports are made in TRACE which correspond to the same transaction. In most cases, the broker acting on behalf of a client charges a commission, and therefore the price reported by the customer does not reflect the real market price of the bond. In these cases, we must take the original transaction reported by the customer (market with a “C” in Contra Party) out of the dataset.
2. Delete one of the reports of each inter-dealer pair. Once the original customer transaction has been taken out of the sample, we must choose to keep one of the remaining two reports on the same transaction. We follow Dick-Nielsen (2014) and keep information from the buyer (marked with a B in Buy/Sell).
3. Once we execute the agency filter, we are left with the correct number of transactions. We keep this first database in order to have the best quality data for research on volume and liquidity.

3.2.5. Verifying and applying the TRACE filter

The Trade Reporting Compliance Engine provides a TRACE fact book that gives the official number of transactions on TRACE for specific bonds. In his 2009 paper, Dick-Nielsen tests the performance of his error filter by matching the number of transactions he obtains after

applying his filter to the number of transactions for the same bonds from the official TRACE fact book.

Applying a similar approach for Green Corporate Bonds presents a challenge as the TRACE Fact Book only gives information on the 50 most traded investment grade and high-yield bonds as well as the top 25 convertible bonds, none of which are in our sample.

We resolve this issue by extracting all TRACE data for the 50 most traded investment grade bonds of Q1 and Q2 2017. We then develop the filter using this dataset and verify the filter's accuracy using the TRACE Fact Book. For the 40 Investment Grade bonds that have been most issued for the first semester of 2017, our error rate is less than 0.1%. The results of our filter are shown in appendix 1.

Once our filter has been validated, we apply it to our green bond and climate bond datasets. We obtain a dataset of 76,204 observations for green bonds, 533,418 observations for climate bonds. Table 3 describes the resulting post-filter TRACE datasets that are obtained.

3.2.6. Transforming TRACE clean data into daily bond prices

Once we have successfully cleaned and processes TRACE data, we can obtain daily prices for each corporate bond from this data. The literature applies two distinct methods.

Following Jostova et al. (2013), the first method consists in computing daily prices as the trade-size weighted average of intraday prices, as findings in Bessembinder et al. (2009) show that a daily price based on trade-size weighted intraday prices is less noisy than the last price of the day. This method puts more weight on institutional trades as these have lower execution costs and should reflect the underlying price of the bond more accurately. However, choosing this price leads to it reflecting market conditions during the day rather than at the end of the day. This is referred to by Bessembinder et al. (2009) as the "trade-weighted price, all trades" approach.

Another method consists in following Harris and Piwowar (2006) and eliminating all trades under \$100,000, which tend to be non-institutional trades, and then relying on the last trade price in the remaining sample (the "last price, trade > 100k" approach). A problem with this approach is the loss of daily observations for bonds that only have small trades during the day; further, the trade selected may not reflect end-of-day market conditions. The problem of losing observations is somewhat mitigated if the firm has multiple bonds, and this approach will tend to reflect the price changes on the most liquid bonds of a particular firm.

Our approach for this step will vary depending on the type of analysis we wish to pursue. For corporate event studies, we will tend to drop trades below \$100,000. When our focus will be on studying the cross section of bond returns, keeping all trades will be more relevant. This gives us two distinct databases. As the objective set in this paper is to obtain corporate bond returns for our sample of green bond and climate bonds, I apply the "trade-weighted price, all trades" approach commonly used in the literature to compute daily prices for the green bond and climate samples. I obtain 11,798 daily prices for the green bonds sample and 125,374 daily prices for the climate bond sample.

3.2.7. Transforming TRACE clean data into monthly bond prices

In order to obtain monthly prices from daily prices, I develop a methodology similar to that of Bai, Bali and Wen (2019)⁴⁷. I first identify the bond price for the last trading day for each bond-month. If the last trading day is one of the last five trading days within month t , then this daily price is used as a monthly price. If that is not the case, I identify the bond price for the first trading day of month $t+1$. If this trading day is one of the first five trading days of month $t+1$, it is used as a price for month t . This allows for a more complete dataset of bond

⁴⁷ In Bai, Bali and Wen (2019), the authors identify two scenarios to compute end-of-month returns for corporate bonds using daily data. To compute a return for month t , they either use data from the end of month $t - 1$ and the end of month t , or data from the beginning of month t and the end of month t . If a return can be computed through both methods, they apply the first method (end of month $t-1$ to end of month t).

monthly prices in order to be able to compute more monthly returns using data from months t and $t+1$, where returns are either computed using (1) end of month $t-1$ to end of month t daily prices (2) start of month t to end of month t daily prices (3) start of month t to start of month $t+1$. Using this methodology, we compute monthly prices for our dataset. We first compute end-of-month prices for each month and complete this database with beginning-of-month $t+1$ prices where end-of-month prices were not available. We obtain 896 monthly prices for our green bond sample and 10,190 monthly prices for our climate bond sample.

3.2.8. Extracting Bloomberg and Datastream Data

Once we have computed monthly prices using trade-based data from TRACE, we extract monthly prices from Bloomberg and Datastream. For our green bond sample, we extract 1318 bond-month observations from Bloomberg and 1595 bond-month observations from Datastream. For our climate bond sample, we extract 7826 bond-month observations from Bloomberg and 15906 from bond-month observations from Datastream. In Table 2, we show commonalities and differences amongst the different datasets in terms of bond-month observations.

For our green bond sample, the correlation between TRACE and Datastream data is of 0.93, showing that the relationship between these two data sources is strong but not perfect, and that the quote-based and matrix-based approach that is applied by Datastream is not entirely precise. This correlation stands at 0.96 for the climate bond sample, which further adds to the strength of this relationship though it does not quite reach a perfect correlation.

Correlation with data originating from Bloomberg is lower, as it corresponds to 0.28 between TRACE and Bloomberg data for green bonds and 0.41 between Datastream and Bloomberg. These values are higher for the larger climate bond sample but are still lower for Bloomberg data which only has a correlation of 0.75 with TRACE data and 0.72 with Datastream data. This analysis of correlation between bond-month observations originating

form our different datasets allows for a better understanding of the literature, which considers Datastream as a better source of information for corporate bond prices.

Going further into this analysis, we also understand that Datastream has much more data on bond-month prices available on both green bonds and climate bonds than the other two databases. For green bonds, the TRACE dataset only has 50 bond-month observations that are not available in Datastream and Bloomberg only has 43, while Datastream has 749 observations that are not available in TRACE and 320 that are not available in Bloomberg. For climate bonds, TRACE provides 203 bond-month observations that are not available in Datastream when Datastream has 5919 observations not available on TRACE and Bloomberg provides 65 observations not available on Datastream when Datastream has 8145. Even though TRACE provides the best quality data on corporate bond pricing, Datastream seems to have a very consequent lead regarding the quantity of available data on monthly corporate bond prices and maintains a strong correlation with TRACE trade-based data.

3.2.9. Merging databases

As we continue to follow the literature, we merge databases. We keep all information regarding bond characteristics from FISD, and, when there are prices from the same bond and month available from multiple sources, we take the first available price in the following sequence, willingly setting precedence to trade-based data: TRACE, Datastream, Bloomberg. Our final dataset for monthly prices green bonds is composed of the 896 bond-month observations from the TRACE dataset, of 749 bond-month observations from the Datastream dataset and 20 observations from the Bloomberg dataset. Our final dataset for climate bonds is composed of 10,190 bond-month observations from the TRACE dataset, of 5,919 bond-month observations from the Datastream dataset and 65 bond-month observations from Bloomberg. We notice that even though Datastream has the most data in terms of corporate bond monthly

prices, most of these observations are also available on TRACE, which makes the TRACE dataset our primary source of data, for both green bonds and climate bonds.

3.2.10. Computing monthly returns using month-end prices

We refer to the formula developed in Jostova et al. (2011) to compute monthly returns:

$$r_{i,t} = \frac{(P_{i,t} + AI_{i,t} + Coupon_{i,t}) - (P_{i,t-1} + AI_{i,t-1})}{P_{i,t-1} + AI_{i,t-1}}, \quad (1)$$

where $r_{i,t}$ is bond i 's month- t return, $P_{i,t}$ is its price at month-end t , $AI_{i,t}$ is its accrued interest at month-end t , and $Coupon_{i,t}$ is any coupon paid between month-ends $t-1$ and t .

Computing accrued interest requires the bond's coupon size, coupon frequency, and day count convention. If the coupon frequency is missing, we assume it is semiannual. If the day count convention is missing, we assume it is 30/360.

Once again, information relative to the bond's coupon size, frequency and day count convention can be obtained from FISD. We can use this to transform our datasets of monthly prices obtained from TRACE, Datastream and Bloomberg data into monthly returns. Our final dataset for green bond returns contains 1,636 observations and our final dataset for climate bond returns contains 16,062 observations.

Table 2
Data Treatment Process

	Green Bonds	Climate Bonds
FISD	253	332
Enhanced TRACE data (raw)	107,848	755,847
Filter 1 - Corrections and Cancellations	103,469	706,013
Filter 2 - Interdeal Transactions	76,204	533,418
Daily TRACE Prices	11,798	125,374
Monthly Datastream Prices	749	5,919
Monthly Bloomberg Prices	20	65
Total Monthly Prices	769	5,984
Total Monthly Returns	1,636	16,062

4. Resulting Datasets and Possible Applications

Advancing through the different phases of the data processing methodology results in obtaining a variety of different datasets and data types that can be used for different purposes. We explore these different datasets and the resulting insights they can provide in the following section. As an example, in their work on capital commitment and illiquidity in the corporate bond markets, Bessembinder et al. (2018) keep inter-dealer transactions

4.1. The TRACE Dataset without Reporting Errors

This dataset gives us a preview of every transaction made on TRACE, including inter-dealer transactions and agency transactions. This type of dataset is used in the literature by academics that wish to study dealer behavior and trading costs in the corporate bond market, as well as total trading volume. As an example, in their work on capital commitment and illiquidity on the corporate bond markets, Bessembinder et al.(2018) keep inter-dealer transactions in their sample in order to determine the size of total yearly trading volume for the US corporate bond market.

I apply the same approach for our sample of green bonds and climate bonds to have a better understanding of the trading volume on these markets since 2014.

4.2. The TRACE Dataset without Reporting Errors and without Agency Transactions

This dataset gives us more precise data on the amount of transactions performed on the secondary market. Interdealer transactions are a form of double counting and can have both an impact on certain liquidity measures that are developed using intraday data as well as on the computation of daily prices. In Bai, Bali and Wan (2011), the authors develop a liquidity measure that can be applied to trade-by-trade data and develop a methodology that relies on difference in prices between two trades to develop their measure. In this context, we can

understand how having a trade be counted multiple times can distort the calculation of this measure, especially if we consider the fact that inter-dealer transactions represent around 30% of raw enhanced TRACE data.

In addition, when considered the different size-weighted approaches used by Bessembinder et al. (2009) to compute daily prices using intraday transaction data on TRACE, we can also understand how counting some intraday trade multiple times could have an important impact on daily prices. However, we compute daily prices for green bonds and climate bonds with and without treating interdealer transactions and find correlations that are superior to 0.99, indicating that the treatment of interdealer data has little impact on corporate bond daily pricing.

Table 3

Trading Volume

Year	Green Bonds			Climate Bonds		
	Trading Volume (Millions)	Corporate Bond Outstanding Amount (Millions)	Trading Volume Relative to Amount Outstanding	Trading Volume (Millions)	Corporate Bond Outstanding Amount (Millions)	Trading Volume Relative to Amount Outstanding
2014	1,172	750	1.56	44,463	14,756	3.01
2015	3,709	5,050	0.73	47,998	13,924	3.45
2016	10,057	7,075	1.42	44,484	14,732	3.02
2017	14,737	10,400	1.42	49,625	16,359	3.03
2018	13,556	3,875	3.50	76,254	21,340	3.57

4.3. The TRACE Dataset without Reporting Errors and without Agency Transactions

This dataset gives us more precise data on the amount of transactions performed on the secondary market. Interdealer transactions are a form of double counting and can have both an impact on certain liquidity measures that are developed using intraday data as well as on the

computation of daily prices. In Bai, Bali and Wan (2011), the authors develop a liquidity measure that can be applied to trade-by-trade data and develop a methodology that relies on difference in prices between two trades to develop their measure. In this context, we can understand how having a trade be counted multiple times can distort the calculation of this measure, especially if we consider the fact that inter-dealer transactions represent around 30% of raw enhanced TRACE data.

In addition, when considered the different size-weighted approaches used by Bessembinder et al. (2009) to compute daily prices using intraday transaction data on TRACE, we can also understand how counting some intraday trade multiple times could have an important impact on daily prices. However, we compute daily prices for green bonds and climate bonds with and without treating interdealer transactions and find correlations that are superior to 0.99, indicating that the treatment of interdealer data has little impact on corporate bond daily pricing.

4.4. The TRACE Daily Prices Dataset

TRACE daily prices can prove useful for various research designs. As aforementioned, using the trade-weighted price for all trades is best for research focusing on the cross-section of corporate bond returns and using the trade-weighted price for all trades superior to \$100 000 is best for research focusing on corporate event studies. Daily prices can also be used to compute daily returns that are used for computing liquidity measures such as the Amihud liquidity measure (Amihud, 2002) or the Pastor-Stambaugh liquidity measure (Pastor and Stambaugh, 2003), common liquidity measure to evaluate liquidity on corporate bond markets.

4.5. The Monthly Price and Monthly Return Dataset

Depending of the different methodologies that we choose to apply to our daily price dataset, we obtain a dataset of monthly prices which we convert to monthly returns for our

green bond and climate bond datasets. The monthly return dataset is the most useful dataset to obtain, as it can not only be used to construct a variety of factors that are based on return trends, such as momentum and reversals, or volatility factors, such as market beta, return volatility or idiosyncratic volatility, but also most importantly to study the cross-section or times series of corporate bond returns.

Figure 1 shows the behavior of corporate bond returns for our sample of green bonds and climate bonds. This visualization helps us understand how the green bond market and the climate bond market have an overall similar behavior but shows a vast difference in terms of volatility in the last months before 2016. Though this graphical visualization is simply illustrative in the context of this paper, it provides interesting insight on the behavior of these two markets and is an interesting starting point for more specific research on this subject.

Figure 1 - Monthly Corporate Bond Returns
Green Bonds and Climate Bonds



5. Conclusion

As the market for climate-oriented products continues to develop, large amounts of rich data on the green corporate bond market and the climate corporate bond market is emerging. The methodology developed in this paper can be used as a guideline to develop intelligent algorithms that can clean and process the different data types as they are provided and apply various data processing approaches depending on the type, scope and aim of research. The resulting datasets could provide academics with interesting opportunities to better understand

green corporate bonds and climate bonds and their impact on corporations, other financial products and markets as well as their various stakeholders.

As the secondary market for green corporate bonds develops and the TRACE dataset grows larger, transaction data provides an opportunity to better analyze green and climate-aligned bond products in ways that were previously impossible. Precise data from TRACE on intra-day transactions can lead to more precise measures of trading volumes and liquidity, but most important provided more precise data on daily bond prices, which can open the way for research work on event studies for green corporate bonds and climate-aligned bonds. Finally, transaction data, once transformed into monthly returns and combined with Datastream and Bloomberg data allows for optimal cross-section computation for green corporate bond prices and returns. As the literature that focuses on the cross-section of corporate bond returns is still to date quite new, such methodology will undoubtedly be needed by academics as research on traditional, green and climate-aligned corporate bonds develops.

References

- Acharya, V.V., Amihud, Y. and Bharath, S.T., 2013. Liquidity risk of corporate bond returns: conditional approach. *Journal of Financial Economics*, 110(2), pp.358-386.
- Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1), pp.31-56.
- Bachelet, M.J., Becchetti, L. and Manfredonia, S., 2019. The green bonds premium puzzle: The role of issuer characteristics and third-party verification. *Sustainability*, 11(4), p.1098.
- Bai, J., Bali, T.G. and Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3), pp.619-642.
- Bao, J. and Hou, K., 2013. Comovement of corporate bonds and equities. *Charles A. Dice Center Working Paper*, (2013-11).
- Bao, J., Pan, J. and Wang, J., 2011. The illiquidity of corporate bonds. *The Journal of Finance*, 66(3), pp.911-946.
- Bessembinder, H., Kahle, K.M., Maxwell, W.F. and Xu, D., 2008. Measuring abnormal bond performance. *The Review of Financial Studies*, 22(10), pp.4219-4258.
- Bessembinder, H., Jacobsen, S., Maxwell, W. and Venkataraman, K., 2018. Capital commitment and illiquidity in corporate bonds. *The Journal of Finance*, 73(4), pp.1615-1661.
- Cai, N., Helwege, J. and Warga, A., 2007. Underpricing in the corporate bond market. *The Review of Financial Studies*, 20(6), pp.2021-2046.
- Carhart, M., 1997. On persistence in mutual fund performance. *The Journal of finance*, 52(1), pp.57-82.
- Chen, L., Lesmond, D.A. and Wei, J., 2007. Corporate yield spreads and bond liquidity. *The Journal of Finance*, 62(1), pp.119-149.
- Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A. and Tong, Q., 2017. Are capital market anomalies common to equity and corporate bond markets? An empirical investigation. *Journal of Financial and Quantitative Analysis*, 52(4), pp.1301-1342.
- Climate Bond Initiative, 2017. Bonds and climate change. *The state of the market*.
- DeCosta, D., Leng, F. and Noronha, G., 2013. Minimum maturity rules: The cost of selling bonds before their time. *Financial Analysts Journal*, 69(3), pp.45-56.
- De Jong, F. and Driessen, J., 2004. Liquidity risk premia in corporate bond and equity markets. Working paper, University of Amsterdam.
- Dick-Nielsen, J., 2009. Liquidity biases in TRACE. *The Journal of Fixed Income*, 19(2), pp.43-55.
- Dick-Nielsen, J., 2014. How to clean enhanced TRACE data. *Available at SSRN 2337908*.
- Edwards, A.K., Harris, L.E. and Piwowar, M.S., 2007. Corporate bond market transaction costs and transparency. *The Journal of Finance*, 62(3), pp.1421-1451.
- Fama, E. F., and K. R. French, 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33, 23-49.
- Fama, E.F. and French, K.R., 2015. A five-factor asset pricing model. *Journal of financial economics*, 116(1), pp.1-22.
- Flammer, C., 2018. Corporate green bonds.
- Ge, W. and Liu, M., 2015. Corporate social responsibility and the cost of corporate bonds. *Journal of Accounting and Public Policy*, 34(6), pp.597-624.
- Goldstein, M.A., Hotchkiss, E.S. and Sirri, E.R., 2007. Transparency and liquidity: A controlled experiment on corporate bonds. *The review of financial studies*, 20(2), pp.235-273.

- Hachenberg, B. and Schiereck, D., 2018. Are green bonds priced differently from conventional bonds?. *Journal of Asset Management*, 19(6), pp.371-383.
- Harris, L.E. and Piwowar, M.S., 2006. Secondary trading costs in the municipal bond market. *The Journal of Finance*, 61(3), pp.1361-1397.
- Jostova, G., Nikolova, S., Philipov, A. and Stahel, C.W., 2013. Momentum in corporate bond returns. *The Review of Financial Studies*, 26(7), pp.1649-1693.
- Liu, M. and Magnan, M., 2014. Conditional conservatism and underpricing in US corporate bond market. *Applied Financial Economics*, 24(20), pp.1323-1334.
- Pástor, Ľ. and Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3), pp.642-685.
- Polbennikov, S., Desclée, A., Dynkin, L. and Maitra, A., 2016. ESG ratings and performance of corporate bonds. *The journal of fixed income*, 26(1), pp.21-41.
- Sarig, O. and Warga, A., 1989. Some empirical estimates of the risk structure of interest rates. *The Journal of Finance*, 44(5), pp.1351-1360.
- Zerbib, O.D., 2019. The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of Banking & Finance*, 98, pp.39-60.

Appendix

Appendix 1

Post 2012 TRACE Filter - S1 2017

Issue Symbol	Cancellations & Reversals	Reversals	Pre-Filter	Agency	Post-Filter	Official Observations	Error Rate
AAPL4001809	142	0	12280	3121	8875	8875	0.00
AAPL4336441	75	0	8098	2115	5832	5834	0.03
BAC3953004	101	0	8837	2396	6239	6239	0.00
BUD4091519	110	1	10739	1806	8712	8710	-0.02
BUD4327481	148	0	10334	2518	7520	7520	0.00
BUD4327594	90	0	7744	2069	5495	5495	0.00
CBL4434050	163	0	12457	4030	8101	8101	0.00
ESRX4379433	104	0	9101	2169	6724	6726	0.03
F4433681	114	0	9425	2630	6567	6567	0.00
GE4329014	692	1	25401	4772	19239	19245	0.03
GS.AEH	104	3	8712	2460	6041	6038	-0.05
GS.YW	85	0	8232	2238	5824	5824	0.00
GS3956630	111	0	8201	2014	5965	5965	0.00
GS4030214	95	0	8525	2170	6165	6166	0.02
HPQ.AI	131	0	9850	2836	6752	6750	-0.03
HPQ4431601	152	4	8191	2108	5774	5770	-0.07
JPM3999853	71	0	9456	2812	6502	6502	0.00
JPM4132024	338	11	9461	2345	6429	6419	-0.16
JPM4135537	125	0	9931	2518	7162	7163	0.01
JPM4135538	175	0	12096	2957	8789	8785	-0.05
JPM4135539	104	0	8598	2087	6303	6303	0.00
JPM4234071	111	2	10226	2228	7774	7772	-0.03
STX4152326	194	1	9117	2374	6354	6353	-0.02
STX4269186	147	0	9247	2713	6240	6238	-0.03
STX4337814	277	2	18235	5473	12206	12204	-0.02
STX4380193	156	0	10737	3220	7205	7204	-0.01
T3818484	116	0	14653	3806	10614	10615	0.01
T4237446	83	2	8736	2241	6327	6325	-0.03
T4237447	133	1	11004	3048	7689	7688	-0.01
T4237448	160	0	13692	3639	9733	9732	-0.01
T4332470	69	1	8812	2704	5969	5968	-0.02
T4332471	100	0	9676	2521	6955	6955	0.00
T4451560	123	0	9938	2506	7186	6934	-3.63
VZ4050437	199	0	13236	3478	9360	9360	0.00
VZ4176696	95	0	11363	2904	8269	8270	0.01
WFC.LG	58	0	8182	2127	5939	5940	0.02
WFC.NW	157	0	7371	1688	5369	5371	0.04
WFC3827183	75	0	8695	2387	6158	6158	0.00
WFC4130435	72	0	8777	2466	6167	6167	0.00
WFM4411360	194	0	13024	3039	9597	9597	0.00

Chapter 5

A study of the cross-section of climate bond returns

Abstract

In this study, I investigate common predictors in the cross-section of corporate bonds returns and evaluate to what extent these predictors apply to the returns of corporate bonds issued by climate-aligned firms, commonly referred to as climate bonds. I focus on the recently proposed bond-specific risk factors – downside risk, credit risk and liquidity risk – and use traditional bond-level and portfolio-level approaches to examine the differences between the general US corporate bond market and climate bonds. Our results differ from that of previous work on the subject, but overall the bond-specific risk factors perform well in explaining the returns of industry- and size/maturity sorted portfolios of corporate bonds for both traditional and climate-aligned firms.

Keywords: Corporate Bonds; Climate Bonds; Risk Factors

1. Introduction

In recent years, there has been increasing focus on the role of the private sector regarding the necessary investments that need to be made for climate change adaptation and mitigation. Corporations, both financial and non-financial, have been identified by the global community as key actors in the efforts to reach both the United Nations' sustainable development goals and to limit global warming to a 2°C threshold by 2100 (UN, 201, 2014; Climate Policy Initiative, 2015). With the intention of developing a European framework for sustainable finance, the European Commission has started to work on an action plan with clear objectives of reorienting capital flows towards sustainable investment and managing financial risks stemming from climate change (European Commission, 2018). The first measure taken by the commission is to develop a taxonomy to help actors of the private sector, as well as European Union member states, to differentiate economic activities based on their alignment with sustainability objectives. By developing a more transparent financial market where the corporations' different economic activities are well delineated, the commission wants to help financial actors understand as clearly as possible how their investments participate in different sustainability-related issues (TEG, 2020). This initiative was backed by regulation focusing on the notion of creating a more sustainable financial system (Council of the European Union, 2019) and is part of a general momentum that is already taking place for financial actors today. The Sustainable and Responsible Investing (SRI) industry has grown from \$23 trillion to \$30 trillion between 2016 and 2018 (GSIA, 2018), the green bond market reached the \$10 billion threshold in 2013 and represented more than \$257 billion in 2019 according to the Climate Bond Initiative (CBI)⁴⁸ and a nascent green loan market is developing, which represented close

⁴⁸The Climate Bond Initiative is an international, investor-focused not-for-profit that focuses explosively on the bond market for climate change solutions. The amount of global green bond issuances is monitored on the organization's website. See <https://www.climatebonds.net/>.

to \$100 billion in 2018 (Linklaters, 2019). Furthermore, what has been identified by the CBI as the “climate-aligned bonds universe” represented a total of \$1.45 trillion in 2018⁴⁹. Though definitions of what constitutes sustainable investments and assets vary greatly between actors and geographical regions, and much work still needs to be done regarding the effective material investments that are made in sustainable projects, it is clear that business-as-usual is in the process of a major transformation to which financial and non-financial corporations will have to adapt, and financial markets with them.

In this regard, this paper constitutes an attempt to understand changes that might already be occurring on these financial markets due to the growing climate change urgency. Using transaction-based data spanning from June 2002 to March 2019, I analyze the cross-section of corporate bond returns of the US corporate bond market. By referring to a list of climate-aligned US firms provided by the Climate Bond Initiative, I obtain a dataset of climate bond returns spanning from June 2002 to March 2019. I then use these returns to determine the relationship between risk factors and the cross-section of climate bond returns and understand the differences that might exist between corporate bonds issued by traditional firms and corporate bonds issued by climate-aligned firms.

One of the main research agenda in empirical asset pricing research focuses on studying the differences in expected returns across specific asset classes. To date, a large majority of this research has applied to stock returns given both the simplicity of these products compared to other market products and the availability of reliable data on stock prices. However, the global bond market represented a total of \$102.8 trillion in securities outstanding in 2018, while the global equity market capitalization represented \$74.7 trillion. In the US, these markets

⁴⁹ According to the CBI, this climate-aligned bond universe is composed of the \$389 billion labelled green bonds market, the \$314 billion market of corporate bonds issued by firms that originate more than 75% of their revenues from climate-aligned activities, the \$497 billion market of corporate bonds issued by firms that originate more than 95% of their revenues from climate-aligned activities, and the \$250 billion market of municipal bonds from issuers that originate more than 95% of their revenues from climate-aligned activities. See (CBI, 2018).

amounted to \$41 trillion and \$30 trillion respectively. Furthermore, the yearly value of US corporate issuances of debt products have been in average 9 times superior to the value of equity issuances between 2004 and 2018⁵⁰ (SIFMA, 2019). Given the size of these corporate debt markets, empirical asset pricing research on products that compose these markets could prove important for its different actors. However, corporate bond markets have historically been less accessible than stock markets for research. Quality data on bond markets had disappeared between 1997 and 2004⁵¹ and corporate bonds are considered much more complex financial products given both the wide array of specific features these can have and the diverse risk exposures to both financial and macroeconomic factors they face. Fortunately, the Trade Reporting and Compliance Engine (TRACE) database and the Mergent Fixed Income Securities (FISD) database opened in 2002 and 2004 respectively, allowing authors to start working on the cross-section of corporate bond returns using bond-specific data.

Earlier research on the cross-section of corporate bond returns mostly developed factors using either stock-level data, treasury bond data and macroeconomic data. This is the case for the long-established Fama-French (1993) factors composed of the market risk factor (Mm-Rf), the size factor (SMB) and the book-to-market factor (HML) that originate from stock data and treasury bond data, and the term spread (TERM) and default spread (DEF) factors that originate from treasury bond data and government bond data⁵². Other factors that have complemented Fama and French's work, such as the liquidity (LIQ) factor (Pastor and Stambaugh, 2003), momentum (MOM) factor (Carhart, 1997), and more recently, the investment (CMA) and profitability (RMW) factors (Fama and French, 2015) all originate from stock-level data.

⁵⁰ Corporate debt products include public and private, investment grade and high yield bonds, convertible debt, asset-back securities and non-agency mortgage-backed securities, while equity issuances include common stock issuances (IPOs and follow-ons) and preferred stock issuances.

⁵¹ The Lehman Brothers Fixed Income Database closed in December 1997 and the Mergent Fixed Income Securities Database (FISD) opened in 2004

⁵² In order to compute the default factor, a market portfolio of long-term corporate bonds was also required, and the data needed to compute such portfolio return could for the most part only be accessed through Ibbotson Associates, a private investment advisory firm.

Realizing that these factors performed poorly in their ability to explain industry-sorted and size/maturity sorted portfolios of US corporate bonds, Bai, Bali and Wen (2019) introduced new bond-implied risk factors based on characteristics specific to corporate bonds. These characteristics were determined by focusing on the three most prominent differences between stocks and corporate bonds: (1) bondholders are more sensitive to downside risk given the fact that their opportunities in terms of upside are limited (2) bondholders are much more exposed to default risk and (3) as the corporate bond market is much more illiquid than stock markets, bondholders are more exposed to liquidity related risks. As a model composed of these factors and a market beta outperforms other models in the literature to explain the cross-section of corporate returns for industry-sorted and size/maturity sorted portfolios, we use these factors as benchmarks to study differences between US climate bonds and the US corporate bond market. Following Bai, Bali and Wen (2019), we calculate bond returns using intraday transaction records from Enhanced TRACE data from June 2002 to March 2019. Much like the authors, our proxy for downside risk is the 5% value at risk (VaR) estimated from the second lowest monthly return observation in the past 36 months, our proxy for credit risk is the bond-level credit rating, and our proxy for liquidity is that developed in Bao, Pan and Wang (2011). We also develop a market beta measure in accordance with the literature.

Though we initially strictly follow the mythology of Bai, Bali and Wen (2019), our results vary. Across the entire dataset, we obtain lower bond returns, as well as lower downside risk. However, we obtain slightly higher illiquidity and lower market beta. Looking more specifically at US climate bonds, we observe even lower returns, much less variation in terms of credit ratings, much higher time to maturity, greater size, lower downside risk and average illiquidity, and similar market beta.

Our univariate and bivariate portfolio analysis still however illustrates a strong relationship between downside risk and next-month excess-bond returns. Regarding the general US market,

bonds in the highest downside risk quintiles generate 3.12% higher yearly returns than bonds in the lowest downside risk quintiles. In the case of climate bonds, the yearly difference is of 2.28%. After controlling for the ten previously cited stock and bond factors, the risk-adjusted return difference between bonds in the highest and lowest downside risk quintiles is economically and statistically significant for the entire market with 2.76% return difference and a t-statistic of 4.40. Economic and statistical significance levels are just slightly inferior for climate bonds, with a yearly return difference of 1.92% and a t-statistic of 1.86.

Similarly, we study average portfolio characteristics and find that at the scale of the entire market, there seems to be a positive relationship⁵³ between downside risk and market risk, illiquidity, credit rating and maturity, while no such relationship seems to exist between downside risk and size, although in average bonds in the lowest downside risk quintiles are smaller than bonds in the highest downside risk quintiles. Similar relationships seem to exist for the climate bond sample as well, at the exception of credit rating which stays quite constant throughout the different levels of downside risk.

Given this information, we therefore also perform bivariate portfolio analysis and, continuing to follow the approach of Bai, Bali and Wen (2019), test if the relationship between downside risk and next-month bond excess returns is maintained when controlling for these characteristics. Our results are also both economically and statistically significant, with yearly differences in returns for the entire market of 2.28% after controlling for credit rating, 1.8% after controlling for maturity, 2.16% after controlling for size, and 1.8% after controlling for liquidity. Our results for climate bonds are also both statistically and economically significant, although to a slightly lower degree, with lower statistical significance for credit rating-, maturity- and liquidity-controlled portfolios, and lower economic significance for credit rating- and maturity-controlled portfolios.

⁵³ In the sense that as quintile portfolios represent bonds with higher downside risk, these portfolios constantly contain higher levels of each of the corresponding characteristics

I then perform bond-level Fama-Macbeth regressions in order to study the cross-sectional relationship between the three bond-specific factors, the bond market beta and next-month excess bond returns more specifically. Following the original methodology of Fama-Macbeth (1973) and continuing to follow the specific approach of Bai, Bali and Wen (2019), I perform regressions of one-month ahead excess returns on downside risk, credit risk, liquidity risk and bond market beta, while controlling for bond exposures to default and term factors, maturity, size and lagged return. While our results regarding the statistically significant positive relationships between these factors and expected returns, as well as the statistically significant negative relationship between return reversal and expected return are quite similar to that of Bai, Bali and Wen (2019), we do find important differences, as we obtain some very statistically significant alphas as well as some statistically significant control variables.

We also compute factors following the methodology in Fama and French (2015) by relying on independent sorts. Once again, our results differ from results in Bai, Bali and Wen (2019), since our downside risk factor is neither – considering the time series averages - our largest factor nor is it larger than the bond market risk premium of 0.58. Considering ten-factors alphas for the bond factors, the highest alpha is that of the return reversal factor (REV), which is more than twice that of our downside risk factor. Even though all our results are statistically significant, they do vary from Bai, Bali and Wen's previous work.

In order to verify our bond factors, we test these factors on industry- and size/maturity-sorted portfolios. Though our results also outperform previous models if we refer to adjusted R squared as measure for model performance, our results differ. Regarding tests on size/maturity portfolios, we obtain an average adjusted R squared of 0.49, but our results seem to greatly improve as maturity increases for portfolios. We obtain an average alpha of 0.15 across portfolios, with a great majority of alphas being statistically significant with an average t-statistic of 3.99. Looking at industry portfolios, our average adjusted R squared are of 0.50

across industries, with a slightly lower average alpha of 0.10 and a much lower average t-statistic of 1.86. Results are quite similar for climate bonds, which average a 0.46 adjusted R-squared for both size/maturity portfolios and industry-sorted portfolios.

2. Literature Review

Literature on the general subject of the relationship between finance and environment-related issues is still nascent. To date, two main subjects have been studied: the relationship between corporate environmental, social and governance (ESG) performances and corporate financial performance and the pricing of green bonds and the existence of a green bond premium.

A majority of studies and meta-analyses that focus on the relationship between ESG factors and corporate financial performance (CFP) reveal a positive association exists, even though work still needs to be done on establishing clearer causality (Schiereck, Friede and Bassen, 2019). This subject draws a lot of attention given the fact that it is at the cross-roads of different deeply rooted theories in literature. On the one hand, a positive relationship between ESG factors and financial performance is supported in the management literature by stakeholder theory (Donaldson and Preston, 1995), the resourced-based theory of the firm (Barney, Ketchen, and Wright, 2011) as well as in the literature that focuses on competitive advantage more specifically (Porter and Kramer, 2006). On the other, a negative relationship corresponds to more traditional financial and economic theories, such as Milton Friedman's view of the firm (Friedman, 1970) and his claim that the only social responsibility of the company is to increase its profits, and portfolio theory (Markowitz, 1959) which would consider the exclusion of certain stocks from portfolios for ESG-related matters suboptimal. Finally, we can consider that supporters of the efficient market hypothesis consider that all

information – and therefore information included in ESG factors – are constantly considered by market actors (Fama,1991), and therefore that no specific relationship exists.

As a globally positive relationship seems to be established between ESG and financial performance, it is yet unclear what are the precise relationships between the different dimensions of ESG and corporate financial performance. Two important meta-analyses have emerged in the literature that have focused on gathering a wide range of studies in order to have a more precise understanding of the different possible ESG and CFP relationships (Busch and Friede, 2018; Friede, Busch, and Bassen, 2015). Friede, Busch and Bassen (2015) find that approximately 90% of studies find non-negative ESG-CFP relations, with 47.9% in vote-count studies and 62.6% of meta-analyses resulting in positive findings, while only 6.9% and 8.0% find negative relationships. These results hold across asset-classes, geographical regions and environmental, social and governance categories, and this positive relationship is even more pronounced in North America and Emerging markets, and for non-equity products. In their paper, Busch and Friede (2018) focus on the relationship between corporate social/environmental performance (CSP) and corporate financial performance (CFP) through a second-order meta-analysis and find a “highly significant, positive, robust, and bilateral CSP-CFP relation” and conclude that “based on the extant literature, the business case for being a good firm is undeniable.”

A branch of this literature focuses specifically on the relationship between environmental performance and financial performance of corporations. However, the definition of what constitutes corporate environmental performance (CEP) varies amongst authors. Authors still do not agree on whether environmental performance is a one-dimensional concept or a multi-dimensional one. Some authors directly construct one CEP measure while others differentiate between process-based environmental performance (Environmental Management Performance – EMP) and outcomes based environmental performance (Environmental Operational

Performance – EOP). Furthermore, databases, metrics and samples vary. Using the ASSET4 ESG database Xue, Zhang and Li (2019) create a measure for EMP using 41 KPIs and measure EOP using Total CO₂ and CO₂ equivalents emissions in tonnes divided by net sales on a sample of UK firms while Trumpp, Endrikat, Zopf and Guenther (2015) use 32 KPIs for EMP and 5 continuous KPIs for EOP focusing on energy consumption, CO₂, water withdrawal and produced waste on a sample of European and US firms. Using the same source of data, Escrig-Olmedo, Muñoz-Torres, Fernández-Izquierdo, & Rivera-Lirio (2017) use 61 KPIs to create one measure for CEP for firms in the agri-food industry and Hartmann & Vachon (2018) focus only on carbon emission reduction for EU manufacturing firms. This variety in terms of approaches and samples applies to other sources of data, such as for users of the KLD database (Ren, He, Zhang, & Chen, 2019; Post, Rahman, and McQuillen, 2015; Delmas, Etzion, and Nairn-Birch, 2013), the CDP database (Misani & Pogutz, 2015; Trumpp & Guenther, 2017) or data originating from surveys (Xie & Hayase, 2007; Anton, Deltas, and Khanna, 2004; Bhattacharyya & Cummings, 2015). In addition to the challenge faced by the literature regarding this diversity of possible approaches, another challenge resides in the quality and comparability of environmental data that is provided by ESG databases (Kotsantonis and Serafaim, 2019).

As the evidence for a strictly positive relationship between corporate environmental, social and governance performance and corporate financial performance is being gathered in the literature, initiatives have emerged on the pricing of assets with underlying extra-financial purposes. Specifically, the growing number of issuances of green bonds since 2013⁵⁴ has led academics to develop an interest in this product. Green bonds are bonds issued in order to specifically finance projects with positive environmental outcomes, and naturally a stem of

⁵⁴ With the first issuance of a green bond product dating to 2007, the green bond market reached the \$10 billion milestone in 2013 and the \$100 billion milestone in 2017. In 2019, the market represented \$257.4 billion in global issuances.

literature is focused on trying to determine if green bond issuers benefit from a green bond premium given the nature of their project.

However, this initiative has met important challenges. First and foremost, the pricing of bond products is more challenging than that of equity products, given the fact that bonds have multiple specific characteristics that directly impact their pricing, such as coupon rates, credit rating, maturity or size (Zerbib, 2019). Determining if a green bond premium exists would require comparing green bonds with traditional bonds that have precisely the same characteristics, which very rarely exist (Bachelet et al, 2019). In addition to these important limitations, the green bond market represented less than 3.6% of the global bond market issuances⁵⁵ in 2019, and issuances are still sporadic throughout the year, which leads both to issues in terms of liquidity, as well as in terms of available data for pricing. Moreover, the academic literature that focuses on studying the corporate bond market refers to one specific transaction-based database to obtain the best quality data on the pricing of corporate bonds – the TRACE database⁵⁶ – which only applies to US corporate bonds. For the year 2018, Moody’s Analytics⁵⁷ reported that \$1.553 trillion in bonds were issued in the US, while the Climate Bond Initiative reported \$34 billion in US green bond emissions (Climate Bond Initiative, 2018). This represents less than 2.2% of US corporate bond emissions, meaning that academics that wish to study this database would have access to too little data to perform robust analyses.

Studies on the pricing of green bonds therefore generally have small samples and must refer to other databases with less precise pricing data, such as dealer quotes provided by market-makers or matrix-prices, which only provide approximations of real prices. To provide an idea

⁵⁵ Global bond issuances in 2019 amounted to \$7.148 trillion. See <https://www.dealogic.com/insight/dcm-highlights-full-year-2019/>

⁵⁶ As TRACE transaction data is the main source of data for this paper, more information is provided in section 4 of this paper on the specificities of this database

⁵⁷ See <https://www.moodysanalytics.com/-/media/article/2019/weekly-market-outlook-corporate-bond-issuance-reflects-business-activities-heightened-to-rates.pdf>

of the general ranking of data on corporate bond pricing in terms of quality, Jostova, Nikolova, Philipov and Stahel (2011) built a database of bond returns using the five databases that gave information on corporate pricing and took “the first available return in the following sequence: TRACE, FISD⁵⁸, Lehman, Datastream and Bloomberg”, clearly giving precedence to trade-based data. Combined, these restrictions in terms of sample size, historical data and pricing data quality results in inconclusive findings concerning the existence of a green bond premium (Hachenberg and Schiereck, 2018; Bachelet et al, 2019; Kapraun and Sheins, 2019, Zerbib, 2019). Similar studies have also focused on the relationship between ESG ratings and corporate bond performance, finding that bonds issued by firms with higher ESG ratings have tighter spreads and outperform peers with lower ESG ratings (Polbennikov, Desclée, Dynkin and Maitra, 2016). Ge and Liu (2015) find that better CSR performance is associated with better credit ratings. In a paper focusing specifically on corporate green bonds, Flammer (2020) finds a positive reaction from stock markets to green bond issuance announcements, that green bond issuers improve their environmental performance after the issuance, and that they experience increase in ownership by long-term and green investors.

While most of the aforementioned literature focuses either on corporations’ ESG performance and ratings or the performance of green bonds more specifically, this study focuses on traditional corporate bonds issued by corporations with climate-aligned activities. This approach, which consists in identifying climate-aligned business activities and differentiating them from non-climate-aligned business activities, though initially developed by the Climate Bond Initiative in order to identify and analyze a market of “climate-aligned bonds” or “climate bonds”, is being integrated by the European Commission in an effort to help investors better identify sustainable firms and corresponding financial products. It is the

⁵⁸ When referring to FISD, the authors referred to the NAIC databases which complements the FISD database, which itself only provides information on characteristic data of US bonds. The NAIC database provides transaction data on corporate bonds issued and traded by US insurance companies.

author's belief that the European Commission is taking the lead in implementing a major shift in how corporations are viewed by investors through its EU classification system for sustainable activities. In that sense, this paper investigates if significant differences already exist between "traditional" firms and "climate-aligned" firms. As previous work by the Climate Bond Initiative has identified climate-aligned bonds starting from January 1st 2005, and that there are 55 US climate-aligned firms, the dataset of US climate-aligned bonds is sufficient to obtain economically and statistically significant results both through cross-sectional analysis and time-series analysis. To the author's knowledge, this study is the first to examine the cross-sectional determinants of climate bonds returns.

3. Data

3.1. Required databases for corporate bond and climate bond pricing

For this paper, we use three distinct sources of data. We obtain transaction data for the US corporate bond market through the TRACE database, which we use to compute corporate bond returns. We use the FISD database to obtain information on bond characteristics. Finally, we obtain a list of climate bond issuers from the Climate Bond Initiative, which we use to identify climate bonds through the FISD database.

Since June 2002, all corporate bond transactions in the secondary market have been made available with the TRACE system through the Trade Reporting Compliance Engine⁵⁹. Before this new dataset was made available, most studies that focused on the corporate bond market only used daily quotes and matrix prices for corporate bonds, which could bias results. In their paper on the subject, Sarig and Warga (1989) explore the fact that there can be liquidity-driven noise errors in daily prices for corporate bonds since daily prices are given even on days when

⁵⁹ The Trade Reporting and Compliance Engine (TRACE) was developed by the National Association of Securities Dealers (NASD) and is used for over-the-counter (OTC) transactions for fixed-income securities. This trade-based data is made available to academics through the TRACE database.

bonds have not been traded for multiple days. When this is the case, brokers set matrix prices based on similar bonds issued by issuers with similar characteristics, which creates bias. More recently, Dick-Nielsen (2009) shows that this bias still exists with prices from Datastream⁶⁰. This gives TRACE data a considerable edge when focusing on daily prices that can be used for microstructure research such as event studies, as well as an edge for weekly and monthly prices.

Two distinct versions of TRACE exist: the standard TRACE data that censors trading volumes that are greater than \$5 million for investment grade bonds and greater than \$1 million for speculative grade bonds and that usually has a three-month lag for the availability of data, and the enhanced TRACE data, that has information on all transaction volumes but has an 18 month lag for the availability of data. In this paper, we refer to enhanced TRACE data, which spans from June 2002 to March 31st 2019.

Considering bond characteristic, FISD is considered as the most comprehensive database focusing on bonds and contains detailed information on the characteristics of more than 140,000 US bonds, either corporate bonds, supranational bonds, U.S. Agency Bonds and U.S. Treasury Bonds. The FISD database contains essential information on bond issuers as well as specific bond issues such as bond issue date, maturity, size, coupon, type, and any information that can be used to identify and categorize US bonds.

Finally, we use a list of climate bond issuers provided directly by the Climate Bond Initiative (CBI) research team. This list was used by CBI in their study on the climate bond market (Climate Bond Initiative, 2018). To identify climate bonds, CBI identified issuers that originated at least 75% of their revenues from green business lines in either clean energy, low-carbon buildings and transport, water and waste management and sustainable land use. Climate bonds were included if they were issued after the 1st of January 2005 and before the end of Q2

⁶⁰ As it is explained in Jostova, Nikolova, Philipov, Stahel (2013): “Extensive discussions with the DataStream support team about the source of their data confirmed that most U.S. corporate bond prices are dealer quotes reported by market-makers. These data are further augmented with trading prices for traded bonds. Like Lehman, DataStream provides no indication of whether a price is based on a quote or a trade.”

2018. Therefore, we use the list of issuers provided by the Climate Bond Initiative to identify all climate bond transactions available on the Enhanced TRACE database since June 2002 to the 31st of March 2019. Our database for traditional bonds uses all available Enhanced TRACE data on the same period.

3.2. From raw intraday transaction data to monthly corporate bond returns

As I continue to follow Bai, Bali and Wen (2019), I remove bonds that are not traded in in the US public market, bonds that are structured notes, mortgage or asset-backed, agency backed or equity-linked, convertible bonds, bonds with floating coupon rates, bonds that trade below \$5 or above \$1000 as well as bonds that have less than one year to maturity. Regarding transaction data, much like Bai, Bali and Wen, I remove bond transactions that are labeled as “when issued”, “locked-in”, that have special sales conditions and that have more than a two-day settlement. Furthermore, I also remove transactions that have a volume inferior to \$10,000. Finally, I follow the filtering methodology developed in Dick-Nielsen (2014) to remove cancellations and corrections made on the TRACE Enhanced database, as well as inter-dealer transactions that are reported twice in the trace database. Much like the author, I apply the two distinct approaches for pre- and post-2012 Enhanced TRACE data. After applying these different filters, my dataset is composed of 91,294,517 intraday transactions spanning from June 2002 to March 2019.

Once we obtain this intra-day information on bond transactions, we compute a daily price by referring to the “trade-weighted price, all trades” approach of Bessembinder et al. (2009), which consists in performing a value-weighted average of all intraday transaction prices. This methodology puts more emphasis on trades from institutional investors that benefit from lower

execution costs, therefore providing a better accuracy for daily prices. Once we apply this methodology, our daily price dataset consists of 13,770,971 daily prices⁶¹.

In order to obtain monthly prices from daily prices, I develop a methodology similar to that of Bai, Bali and Wen (2019). I first identify the bond price for the last trading day for each bond-month. If the last trading day is one of the last five trading days within month t , then this daily price is used as a monthly price. If that is not the case, I identify the bond price for the first trading day of month $t+1$. If this trading day is one of the first five trading days of month $t+1$, it is used as a price for month t . This allows for a more complete dataset of bond monthly prices in order to be able to compute more monthly returns using data from months t and $t+1$, where returns are either computed using (1) end of month $t-1$ to end of month t daily prices (2) start of month t to end of month t daily prices (3) start of month t to start of month $t+1$. Using the methodology, we compute 1,761,543 monthly prices for our dataset.

The monthly corporate bond return at time t is computed as:

$$r_{i,t} = \frac{(P_{i,t} + AI_{i,t} + Coupon_{i,t}) - (P_{i,t-1} + AI_{i,t-1})}{P_{i,t-1} + AI_{i,t-1}}$$

where $P_{i,t}$ is the transaction price, $AI_{i,t}$ is accrued interest, and $Coupon_{i,t}$ is the coupon payment, if there is any, of bond i for month t . We obtain the necessary information to compute corporate bond returns from the FISD database.⁶² In order to compute our returns, month prices need to be available for adjacent months. Our final dataset is composed of 1,530,745 corporate bond returns.

61 In Bessembinder et al (2009), four distinct strategies are used to compute daily prices from TRACE data: using the last price reported in TRACE for the day (“last price, all trades”), construct daily prices by weighing each trade by size (“trade-weighted price, all trades”), eliminating all trades below \$100,000 dollars and using the last available price of the day (“last price, trade $\geq 100k$ ”), and eliminating all trades below \$100,000 and weighing each trade by size (“trade-weighted price, trade $\geq 100k$ ”). Even though we do apply a “trade-weighted price, all trades” in this paper, we do exclude trades below \$10,000 from our sample following Bai, Bali and Wen (2019).

62 This includes the coupon rate, the dated date marking the start of the interest period, as well as the interest frequency.

4. Variable Selection

In the context of this study, we apply three bond-level factors developed by Bai, Bali and Wen (2019). Through their research, the authors observed that the literature that focus on the cross-section of corporate bond returns relied on stock market factors, based on the rational that the stock and bond markets are integrated, since both bonds and stocks value similar underlying assets. Their work goes against this more traditional approach by focusing on risk factors that are specific to the bond market.

4.1. Downside risk

One of the main concerns for bond holders regards extraordinary events that can lead to crashes on either the stock or bond markets. Both private companies and public regulators focus on understanding what are the minimal capital requirements in order to face such risks if they were to occur. Various methodologies have been developed by actors of the bond markets in order to measure such risks. Both practitioners and academics have recently focus on the concept of Value at Risk (VaR).

VaR measures how the proportion by which the price of an asset could decline over a certain period of time as a result of variations in market rates or prices. As an example, if the studied period is 10 days, and the probability of such an event is 5%, the VaR estimate would consist of evaluating the decline of an asset's value that could occur with 5% probability over the next 10 days. In order to determine a measure for VaR, we refer to the work of Bai, Bali and Wen (2019), in which the proxy for downside risk is a 5% VaR computed by selecting the second lowest monthly return distribution in a 36-month period. The result is then multiplied by -1 for convenience.

4.2. Credit quality

Credit ratings have been developed in order to assess bond default probabilities. Bond-specific ratings synthesize information relative to the issuer's financial condition, operating performance as well as risk management strategies and information relative to bond characteristics such as coupon rate and seniority. We collect credit ratings for our sample, as well as historical credit rating changes through the FISD database. We refer to credit ratings by S&P, Moody's and Fitch rating agencies. All credits ratings are assigned a number from 1 (AAA or Aaa) to 10 (BBB- or Baa3) for investment-grade bonds and numbers above 10 for non-investment grade bonds. Bond ratings in our datasets correspond to the average available credit ratings for each bond. We follow the numbering methodology provided by the FISD database and number bonds that have not been rated by either of the three rating agencies with a value of 27.

4.3. Bond Liquidity

The relationship between liquidity and corporate bond returns is well established. Empirical results have established the specific correlation between corporate bond yield spreads and bond liquidity (Chen et al., 2007; Dick-Nielsen et al., 2012). The bond-level liquidity measure developed by Bao et al. (2011) explains a significant proportion of the variations in bond yield spreads, and the market level liquidity measure developed by Lin et al. (2011) show that market liquidity beta is priced in corporate bond returns. Following Bai, Bali and Wan (2019), we chose to construct the bond-level liquidity measure developed by Bao et al. (2011) using our dataset through the following model:

Table 1

Panel A: Traditional Bonds - Cross-sectional statistics over the sample period June 2002 - March 2019									
	N	Mean	Median	Percentiles					
				5th	10th	25th	75th	90th	95th
Bond return (%)	1530745	0.51	0.34	-3.17	-1.75	-0.37	1.38	2.91	4.36
Rating	1530745	8.86	8.00	2.67	4.00	5.67	10.33	15.00	18.33
Time to maturity (maturity, year)	1530745	9.68	6.52	1.47	1.95	3.49	13.18	23.67	27.13
Size (\$million)	1530745	380.10	250.00	1.99	4.48	19.57	500.00	1000.00	1250.00
Downside risk (5% VaR)	713984	3.73	2.52	0.53	0.84	1.46	4.24	7.58	11.79
Illiquidity (ILLIQ)	975447	2.36	0.58	0.03	0.05	0.17	1.65	4.02	7.46
Bond market beta (β_{bond})	908834	0.95	0.71	0.00	0.06	0.25	1.32	2.11	2.75

Panel B: Climate Bonds - Cross-sectional statistics over the sample period January June 2002 - March 2019									
	N	Mean	Median	Percentiles					
				5th	10th	25th	75th	90th	95th
Bond return (%)	17356	0.40	0.30	-2.59	-1.53	-0.40	1.22	2.47	3.45
Rating	17356	8.54	8.50	6.00	6.50	7.50	9.00	10.33	12.00
Time to maturity (maturity, year)	17356	14.52	9.02	1.87	2.78	5.11	23.55	28.58	29.82
Size (\$million)	17356	447.60	350.00	147.50	200.00	250.00	600.00	750.00	1000.00
Downside risk (5% VaR)	9492	2.59	2.16	0.66	0.93	1.39	3.49	4.68	5.51
Illiquidity (ILLIQ)	13199	1.66	0.55	0.04	0.07	0.20	1.46	3.26	5.33
Bond market beta (β_{bond})	12162	0.95	0.75	0.05	0.11	0.31	1.36	2.21	2.58

The authors define P_t as a bond's clean price (meaning the price of a bond without accounting for accrued interests and coupon payment) of the bond at time t and p_t as the log price, and assume that p_t consists of the following components :

$$p_t = f_t + u_t$$

Where f_t represents the fundamental value of the log price in the absence of friction, and u_t is generated from the impact of illiquidity. Therefore, in the authors' framework, u_t characterizes illiquidity on the market.

Finally, the authors define their measure for illiquidity, which is aimed at extracting this component in the observed clean log price p_t , as follows:

$$\gamma = - \text{Cov}(\Delta p_t, \Delta p_{t+1})$$

With $\Delta p_t = p_t - p_{t-1}$. Bao et al. (2011) develop this model under the assumption that f_t follows a random walk, so that γ only depends on u_t and therefore increases when u_t increases.

3.1. Bond market β

As we continue to refer to Bai, Bali and Wen (2019) to develop our bond risk factors, we compute the bond market excess return (MKTBond) by subtracting the one-month Treasury bill rate from the value-weighted average returns of each corporate bond in our dataset. Therefore, for each bond-month in our dataset, we estimate the bond market beta (β_{Bond}) using a time-series regression of each bond's excess return on MKTBond using a 36-month rolling window.

4.4. Summary Statistics

Table 1 reports the cross-sectional statistics of both the entire corporate bond market sample and the climate bond market sample. Our total sample is composed of 1,530,745 bond returns. Our sample contains the same amount of data for credit rating, time-to-maturity and size. However, given the fact that our risk measures are obtained using specific computation processes, we have less observations for downside risk, illiquidity and bond market beta. Computing downside risk requires that a bond be issued at least 36 months before obtaining the first possible measure, illiquidity risk needs a minimal number of transactions or daily prices every month to be computed, and we compute the bond market beta using a 36-month rolling regression, which also require 36 month of data availability. Our total number of climate

bond monthly returns consists in 17,356 observations, and observations in terms of downside risk, illiquidity risk and bond market beta are also sparser, specifically for downside risk for which we only have 9492 observations.

A few differences can be noticed between the two samples. Mean returns for the market are slightly higher in average than the climate bond market (0.51 for the market sample against 0.40 for climate bonds), even though median returns are almost similar (0.34 and 0.30 respectively). Climate bonds have higher time-to-maturity (14.52 years in average) and size (\$447.62 million in average) than the corporate bond market (9.68 years to maturity and \$380.10 million in size). Even though credit ratings for the market and for climate bonds are quite similar when looking at both mean and median ratings, climate bond rating experience much less variation across the sample : at the 5th percentile, climate bond credit ratings correspond to an A rating while at the 95th percentile, climate bond credit ratings correspond to BB/Ba2, while the market goes from between AA+/Aa1 and AA/Aa2 at the 5th percentile to between CCC/Caa and CCC-/Caa3 at the 95th percentile. Average illiquidity for the market is higher than for climate bonds, even though median liquidity is similar for both samples. Quite logically, climate bonds have lower value-at-risk (2.59 in average) than the market (3.73 in average) since they also have lower average returns. Finally, bond market betas are quite similar for climate bonds and the market (0.95 in average for both samples).

5. Downside risk and next-month excess returns

In this section we examine the specific relationship between downside risk and corporate bond returns through portfolio analysis. Throughout the rest of this study, we refer to the Newey-West adjusted t-statistic (Newey and West, 1987) to measure and indicate the statistical significance of our results. The Newey-West estimator was developed to overcome autocorrelation and heteroskedasticity in the error terms of models, in most cases for

regressions that are applied to time series data. Since this study mostly relies on performing cross-sectional regressions and computing time series averages that are quite likely to have correlated error terms over time, the use of estimators such as the Newey-West estimator is required to demonstrate robust statistical significance.⁶³

5.1. Univariate portfolio analysis

In order to perform our first portfolio analysis – the univariate portfolio analysis – we form quintile portfolios for every month between July 2005⁶⁴ and March 2019 by sorting corporate bonds according to their downside risk values. Quintile 1 contains bonds with the lowest VaR values, while quintile 5 contains bonds with the highest VaR values. The portfolios are weighted using size as weights. Table 2 shows the average VaR for each quintile as well as average next-month excess returns⁶⁵. This allows to illustrate the relationship that exists between downside risk and future returns. As the average VaR grows from 0.82% for quintile 1 to 4.77% for quintile 5, we observe a similar phenomenon for next-month average bond excess returns which grow from 0.11% to 0.37%. The difference between high VaR portfolios and low VaR portfolios is both economically and statistically significant with an average next-month excess return difference of 0.26% and t-statistic of 5.10.

In addition to providing the next-month average excess return for every quintile, we also use control variables to understand if this relationship can be explained by other factors that have previously been developed in the literature. We regress next-month portfolio excess returns on the well-known stock market factors.

⁶³ This specific subject is addressed in Petersen (2009).

⁶⁴ Even though we study TRACE data from July 2002 to March 2019, developing our measure for value-at-risk required at least 36 months of previous monthly returns, which means that our values for VaR can only be obtained from July 2005 onwards.

⁶⁵ To compute excess returns, we subtract the risk-free rate from corporate bond returns. We use the risk-free measure available on Kenneth French's website <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>

Table 2

Panel A - Corporate Bond market univariate portfolios										
Quintiles	Average	Average	Stock Factors	Bond Factors	All Factors	Average Portfolio Characteristics				
	VaR	return	alpha	alpha	alpha	Bbond	ILLIQ	Rating	Maturity	Size
Low VaR	0.82	0.11	0.10	0.09	0.09	0.34	0.15	6.90	3.21	370.48
		<i>3.06</i>	<i>3.06</i>	<i>3.22</i>	<i>2.27</i>					
2	1.50	0.18	0.16	0.14	0.14	0.54	0.30	7.29	4.74	414.49
		<i>3.45</i>	<i>2.79</i>	<i>3.12</i>	<i>3.23</i>					
3	2.21	0.23	0.21	0.19	0.19	0.74	0.47	7.92	7.36	390.86
		<i>3.67</i>	<i>3.13</i>	<i>3.52</i>	<i>3.37</i>					
4	3.11	0.28	0.26	0.22	0.23	0.96	0.63	8.10	11.54	397.51
		<i>3.97</i>	<i>3.43</i>	<i>3.99</i>	<i>3.56</i>					
High Var	4.77	0.37	0.34	0.32	0.32	1.27	0.78	8.99	15.32	392.07
		<i>4.91</i>	<i>4.28</i>	<i>4.73</i>	<i>4.43</i>					
High-Low	3.95***	0.26***	0.25***	0.23***	0.23***					
	<i>22.08</i>	<i>5.10</i>	<i>4.64</i>	<i>4.43</i>	<i>4.40</i>					

Panel B - Climate Bond univariate portfolios										
Quintiles	Average	Average	Stock Factors	Bond Factors	All Factors	Average Portfolio Characteristics				
	VaR	return	alpha	alpha	alpha	Bbond	ILLIQ	Rating	Maturity	Size
Low VaR	0.90	0.15	0.13	0.12	0.12	0.32	0.18	8.40	5.32	367.44
		<i>2.85</i>	<i>3.10</i>	<i>2.59</i>	<i>2.85</i>					
2	1.60	0.21	0.18	0.15	0.15	0.55	0.41	8.78	6.26	404.60
		<i>3.68</i>	<i>3.22</i>	<i>2.52</i>	<i>2.19</i>					
3	2.20	0.24	0.22	0.21	0.21	0.72	0.48	8.75	8.43	506.50
		<i>3.78</i>	<i>3.46</i>	<i>2.74</i>	<i>2.79</i>					
4	2.93	0.28	0.25	0.21	0.21	0.92	0.61	8.54	14.89	482.19
		<i>2.80</i>	<i>2.65</i>	<i>2.61</i>	<i>2.52</i>					
High Var	4.41	0.34	0.31	0.25	0.28	1.26	0.81	8.38	21.00	429.19
		<i>3.66</i>	<i>3.04</i>	<i>2.29</i>	<i>2.49</i>					
High-Low	3.52***	0.19**	0.18**	0.13	0.16*					
	<i>15.69</i>	<i>2.53</i>	<i>2.23</i>	<i>1.52</i>	<i>1.86</i>					

These are the excess-stock market return (MKTstock), the size factor (SMB), the book-to-market factor (HML), the profitability factor (RMW), the investment factor (CMA), the momentum factor (MOM) and the liquidity risk factor (LIQ), which have been developed and computed in Fama and French (1993), Carhart (1997) and Pastor and Stambaugh (2003)⁶⁶. We once again obtain economically and statistically significant stock alphas, which grow from an average of 0.10% for the lowest VaR quintile to 0.34% for the highest VaR quintile, with an

⁶⁶ Much like the risk-free rate, the excess-stock market return (MKTstock), the size factor (SMB), the book-to-market factor (HML), the profitability factor (RMW), the investment factor (CMA), the momentum factor (MOM) and the liquidity risk factor (LIQ) can be found online, either on Kenneth French's website <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/> or Lubos Pastor's website <https://faculty.chicagobooth.edu/lubos-pastor/research>

economically and statistically significant difference of 0.25% in terms of return difference and a t-statistic of 4.64.

We also regress the next-month portfolios excess returns on well-known bond market factors: the bond market excess return (MKTbond), the term factor (TERM) and the default factor (DEF). We obtain the default factor (DEF) by computing the difference between the monthly returns of a market portfolio of long-term corporate bonds and monthly long-term government bond returns, and the term factor (TERM) by computing the difference between monthly long-term government bond returns and the risk-free rate. Much like for our stock market factors, we obtain statistically significant alphas that grow as the downside risk grows. Finally, we also perform a regression of the next-month portfolio excess returns on all our factors. Our results are economically and statistically significant when regressing on our bond factors and on all our factors combined, with a difference in returns of 0.23% for both type of regressions, and t-statistics of 4.43 and 4.40 respectively.

Finally, we also compute the averages for the market beta, illiquidity, rating, maturity and size for each VaR quintile and observe that some relations might also exist between next-month excess bond returns and some of these variables. Average bond market beta grows from an average of 0.34 for VaR 1 to 1.27 for VaR 5, illiquidity goes from an average of 0.15 to 0.78, and time-to-maturity grows from an average of 3.21 years to 15.32 years with VaR. The existence of such relation is less clear regarding size, as growth of average size for VaR quintiles is not constant, even though size does grow from 370.48 million in average for VaR 1 to 392.07 million for VaR 5. Credit rating diminishes⁶⁷ as value-at-risk grows for the market, with credit ratings going from 6.90 (which approximately corresponds to an A-/A3 rating) for the lowest VaR quintile to an 8.99 (which approximately corresponds to a BBB/Baa rating) for highest Var quintile.

⁶⁷ Our credit ratings are ranked from 1 to 27, 1 being the best possible rating (AAA or Aaa) and 27 being the worst. Therefore, as our value for credit rating grows, it illustrates a worsening of the actual credit risk of the bond.

Comparing our results for the market sample with the results for the climate bond sample, we do observe some differences in terms of statistical significance, as none of the differences between Var 1 and Var 5 in terms of next-month excess return or alphas have a statistical significance at the 1% level. More specifically, the bond factor alpha is strictly not statistically significant, even though the difference in average alphas for cross-sections with all factors alphas is statistically significant at the 10% level. This lower statistical significance could be attributed to the fact that there are much less bond observations in our climate bond sample, which could lead to less significant results. Economic significance in returns and alphas between average VaR 1 quintile portfolios and VaR 5 quintile portfolios is lower for climate bonds with 0.19% in average excess return difference and 0.16% when controlling for the ten stock and bond market factors. This is not surprising given the fact that climate bonds in our sample have both lower average returns and lower downside risk than the market sample. Finally, average portfolio characteristics for our climate bond sample behave in a similar fashion than for the market sample with regards to bond market beta, illiquidity, maturity and size, while the relationship between VaR, excess return and credit rating seems to disappear as credit rating for VaR 1 and for Var 5 are similar (8.40 and 8.38 respectively). This could in part be explained by the lack of variation in credit ratings within our climate bond sample.

Following Bai, Bali and Wen (2019), these apparent interactions between downside risk and other variables leads us to test whether the relationship between downside risk and returns holds when controlling for market beta, credit rating, maturity, liquidity and size using both a bivariate portfolio analysis and Fama-Macbeth (1973) bond-level regressions.

5.2. Bivariate portfolio analysis

Table 3 exposes the results of our bivariate portfolio analysis. Much like for the previous univariate analysis, we sort our bonds monthly according to specific characteristics. However, the process here is a bit different, as we first sort bonds according to either credit rating (Panel

A), maturity (Panel B), size (Panel C) or liquidity (Panel D) into five different quintiles. Once this initial process has been performed, we further sort each quintile into five quintiles based on downside risk. We therefore have 25 distinct portfolios per month, and each portfolio has very similar characteristics while downside risk maintains some dispersion. We compute value-weighted average returns for each portfolio. Much like the univariate sort, portfolios are weighted using size as weights. We then compute average returns for every VaR quintile category (portfolios with the highest VaR quintile being Var quintile 5 and the lowest being Var quintile 1). We therefore have five portfolios per month corresponding to each VaR category. We perform time series regressions for each of these monthly portfolios using the 10 stock and bond factors previously computed for our univariate portfolio analysis. Much like it was the case for our univariate sorts, we refer to the Newey-West adjusted t-statistic for statistical significance.

As we can see from our results, the relationship between downside risk and next month excess returns holds for every characteristic, both for our market sample and our climate bond sample after controlling for the ten stock and bond market factors. Looking at the market sample, difference between VaR 5 portfolios and VaR 1 portfolios are both economically and statistically significant, with differences of 0.19% when controlling for credit rating, 0.15% when controlling for maturity, 0.18% when controlling for size and 0.15% when controlling for liquidity, with each result having a statistical significance at the 1% level. Results for the climate bond sample only vary slightly, with lower economic significance for differences in excess returns when controlling for credit rating (0.13%) and maturity (0.07%) and overall lower statistical significance, as the difference in excess returns when controlling for size is the only one that has a statistical significance at the 1% level.

5.3. Bond-level Fama-MacBeth regressions

Once we have tested the relationship between downside risk and corporate bond returns through portfolio analyses, we perform bond-level analysis through the application of Fama-Macbeth (1973) regressions. The Fama-MacBeth procedure consists in running a cross-sectional regression every month in order to obtain monthly coefficients, and then perform a time series average to obtain average coefficients for the entire time period considered. Once again, in this procedure, the use of Newey-West adjusted t-statistics is essential to establish statistical significance. Following Bai, Bali and Wen (2019) we perform ten distinct Fama-French regressions, regressing each factor individually and with control variables and then regressing all four factors with and without control variables. Our control variables are bond exposure to both the term factor (β^{TERM}) and the default factor (β^{DEF}), time-to-maturity, size and lagged excess return. Table 4 shows the results for the bond-level Fama-MacBeth regressions.

In regression (1), the average slope coefficients of monthly regressions of excess bond returns on VaR without control variables is 0.078 with a t-statistic of 4.90. The economic significance of this effect resembles that of the univariate portfolio analysis in which we found a spread of 3.95 for our VaR values between low Var and high Var portfolios. Multiplying this spread by the slope coefficient, we obtain a monthly downside risk premium of 30 basis points. Slope coefficients for regressions (3), (5) and (7), which correspond to the regressions of excess bond returns on credit rating, illiquidity and market beta individually and without control variables, are also positive and statistically significant. This also corresponds to our results for the univariate portfolio analysis, in which we can observe a positive difference between low Var quintiles and high Var quintiles in terms of average credit rating, illiquidity and bond market beta.

Performing these same regressions but controlling for beta term, beta default, maturity, size and lagged excess returns does not change our results. Controlling for these characteristics

maintains significantly positive and significant results for all our risk variables and does not affect their positive relation with next-month bond returns. Surprisingly, however, controlling for these factors even enhances our results for regressions (2), (4) and (8).

Our results diverge slightly for Bai, Bali and Wen (2019) when all our variables are regressed together. In regression (9), though our VaR measure is still positive and statistically significant, our liquidity measure is neither economically nor statistically significant, but our credit rating measure is. However, when controlling for our different bond characteristics in regression (10), economical and statistical significance for credit rating disappears and illiquidity becomes positive and statistically significant. Much like Bai, Bali and Wen (2019), statistical significance for bond market beta disappears for both regression (9) and (10), and the return reversal control variable is very significant statistically and has a negative relation with next-month excess return. Finally, we observe that many of our control variables are statistically significant when regressing variables independently. More specifically, bond exposure to the default factor is always both positive and statistically significant at the 1% level. This is also the case for bond exposure to the term factor for regressions (4) and (6).

Regarding our climate bond sample (Panel B), our bond-level Fama-Macbeth regression analysis is limited by the lack of monthly data available for our climate bond sample⁶⁸, and we are only able to perform individual regressions without control variables. This is sufficient however to detect positive and statistically significant slope coefficients for regression of one-month ahead excess return on VaR (1), illiquidity (3) and bond market beta (4) and that no such results are obtained for the regression of one-month ahead excess returns on credit rating (3). These results correspond to results obtained in our univariate portfolio analysis for the climate bond sample, where we established a positive statistically significant difference in terms of

⁶⁸ Our number of monthly observations for Climate Bonds have a median value of 61, while it can be considered as a rule of thumb that 10 to 20 observations are needed per estimated parameter (Harrell,2015). This would mean that in order to perform any of the regressions with control variables, a strict minimum of 60 observations are required.

VaR between VaR 5 portfolios and VaR 1 portfolios, but also observed such positive differences for average quintile liquidity and average bond market beta, while this relation did not exist for our average credit ratings. Much like it was the case for the univariate portfolio analysis, this result could well be due to the lack of variation in terms of credit ratings in our climate bond sample. Furthermore, by multiplying our slope coefficient of 0.072 for regression (1) by the difference in downside risk in our univariate portfolio analysis of 3.52%, we obtain a monthly down-side risk premium for climate bonds of 27,5 basis points, slightly under our results for the market sample.

Table 4

Panel A - Bond-level Fama-MacBeth cross-sectional regressions (Market)

	Intercept	5% VaR	Rating	ILLIQ	BBond	BDEF	BTERM	Maturity	Size	REV	Adj. R2
(1)	0.446 <i>60.516</i>	0.078 <i>4.899</i>									0.025
(2)	0.363 <i>2.408</i>	0.080 <i>6.592</i>				0.322 <i>4.621</i>	0.079 <i>0.524</i>	0.016 <i>0.647</i>	-0.017 <i>-1.630</i>	-0.185 <i>-28.493</i>	0.158
(3)	0.484 <i>61.602</i>		0.034 <i>3.535</i>								0.022
(4)	-0.026 <i>-0.161</i>		0.071 <i>5.488</i>			0.594 <i>4.657</i>	0.558 <i>2.733</i>	0.087 <i>3.057</i>	-0.037 <i>-3.143</i>	-0.191 <i>-28.118</i>	0.165
(5)	0.484 <i>81.291</i>			0.026 <i>3.747</i>							0.018
(6)	-0.239 <i>-1.282</i>			0.021 <i>3.868</i>		1.032 <i>5.438</i>	0.542 <i>2.259</i>	0.082 <i>2.604</i>	0.017 <i>1.131</i>	-0.164 <i>-30.171</i>	0.176
(7)	0.472 <i>92.110</i>				0.018 <i>5.808</i>						0.035
(8)	0.399 <i>2.794</i>				0.068 <i>4.782</i>	0.289 <i>2.776</i>	0.064 <i>0.309</i>	0.067 <i>2.356</i>	-0.033 <i>-2.900</i>	-0.188 <i>-28.801</i>	0.158
(9)	0.424 <i>56.756</i>	0.057 <i>6.299</i>	0.045 <i>2.747</i>	0.009 <i>1.471</i>	0.029 <i>1.303</i>						0.112
(10)	0.252 <i>1.678</i>	0.062 <i>7.908</i>	0.010 <i>1.842</i>	0.056 <i>3.830</i>	0.025 <i>0.958</i>	0.223 <i>1.467</i>	0.221 <i>1.423</i>	0.000 <i>0.004</i>	0.000 <i>0.025</i>	-0.174 <i>-46.298</i>	0.214

Panel B - Bond-level Fama-MacBeth cross-sectional regressions (Climate Bonds)

	Intercept	5% VaR	Rating	ILLIQ	BBond	BDEF	BTERM	Maturity	Size	REV	Adj. R2
(1)	0.448 <i>56.660</i>	0.072 <i>2.404</i>									0.083
(2)	0.478 <i>40.930</i>		0.039 <i>1.329</i>								0.020
(3)	0.484 <i>71.040</i>			0.026 <i>2.386</i>							0.046
(4)	0.464 <i>21.020</i>				0.026 <i>2.495</i>						0.094

6. Common risk factors in the cross-section of climate bond returns

6.1. Determining the bond-level risk factors

In order to construct our bond factors, we follow Fama and French (2015) and rely on independent sorts. To construct the downside risk factor for corporate bonds, for each month from July 2004 to December 2016, we form bivariate portfolios by independently sorting bonds into five quintiles based on their credit rating and five quintiles based on their downside risk (measured by 5% VaR). The downside risk factor, DRF, is the value-weighted average return difference between the highest-VaR portfolio and the lowest-VaR portfolio across the rating portfolios. The credit risk factor, CRF VaR, is the value-weighted average return difference between the lowest-rating (i.e., highest credit risk) portfolio and the highest-rating (i.e., lowest credit risk) portfolio across the VaR portfolios. The liquidity risk and the return reversal factors are constructed similarly using independent sorts. The liquidity risk factor, LRF, is the value-weighted average return difference between the highest-illiquidity and the lowest- illiquidity portfolios across the rating portfolios. The return reversal factor, REV, is the value-weighted average return difference between the short-term loser and the short- term winner portfolios (losers-minus-winners) across the rating portfolios. The above independent sorts used to construct LRF and REV produce two additional credit risk factors, CRF ILLIQ and CRF REV. The credit risk factor CRF is defined as the average of CRF VaR , CRF ILLIQ , and CRF REV. As table 5 reports, over our sample period the corporate bond market risk premium (MKT^{Bond}) is 0.58% per month with a t-statistic of 3.28 and the value-weighted downside risk factor (DRF) has an economically and statistically significant risk premium of 0.22% with a t-statistic of 4.11.

Table 5

Panel A - Bond factors (summary statistics)		
	Mean	t-stat
MKTbond	0.58	3.28
Downside risk factor (DRF)	0.22	4.11
Liquidity risk factor (LRF)	0.18	3.50
Credit risk factor (CRF)	0.44	11.93
Return reversal factor (REV)	0.14	2.44

Panel B - Bond Factors (ten-factor alpha)		
	Alpha	t-stat
Downside risk factor (DRF)	0.19	3.40
Liquidity risk factor (LRF)	0.14	2.43
Credit risk factor (CRF)	0.10	1.73
Return reversal factor (REV)	0.43	10.99

All other risk factors are also economically and statistically significant : our liquidity risk factor (LRF) has a risk premium of 0.18% with a t-statistic of 3.60, our credit risk factor (CRF) has a risk premium of 0.44% with a t-statistic of 11.93 and our return reversal factor (REV) has a risk premium of 0.14 with a t-stat of 2.44.

We then explore whether stock and bond factors used to test our results in our univariate and bivariate portfolio analyses participate in explaining our bond-specific factors. We use the exact same factors to compute factors alphas in table 5. These consist of seven stock factors that are the excess-stock market return (MKTstock), the size factor (SMB), the book-to-market factor (HML), the profitability factor (RMW), the investment factor (CMA), the momentum factor (MOM) and the liquidity risk factor (LIQ) and three bond factors that the bond market excess return (MKTbond), the term factor (TERM) and the default factor (DEF).

Our factor alphas are also positive and statistically significant. Our ten-factor alpha for the downside risk factor is of 0.19% with a t-statistic of 3.40, our alpha for the liquidity risk factor is of 0.14% with a t-statistic of 2.43, our alpha for the credit risk factor is of 0.10% with a lower statistical significance of 1.73 and our return reversal factor is of 0.43% with a t-statistic of 10.99. Overall, these results demonstrate that that the ten stock and bond factors previously

developed by the literature do not explain the new bond-specific factors, which therefore capture a consequent proportion of corporate bond return variation.

6.2. Size/maturity-sorted and industry-sorted portfolios

Continuing to follow Bai, Bali and Wen (2019), we apply the authors' most performant model – the four-factor model composed of the excess bond market return (MKT^{Bond}), the downside risk factor (DRF), the credit risk factor (CRF) and the bond specific liquidity factor (LRF) – to both size/maturity-sorted portfolios and industry-sorted portfolio.

Though our results differ from the authors', the four-factor model, when applied to our dataset, results in an average adjusted R-square of 0.49 for size/maturity-sorted portfolios and 0.50 for industry-sorted portfolios, and outperforms other models proposed by Bai, Bali and Wen (2019). However, results for the four-factor model do differ in our case, since we obtain a positive and statistically significant average alpha of 0.15 with a t-statistic of 3.99 for size/maturity-sorted portfolios of the market sample. In addition, model performance, when using adjusted R-square as a measure, seems to increase as maturity increases. This leads the four-factor model to obtain an adjusted R-square of 0.91 for the portfolio containing the largest and most mature bonds (size quintile 5 and maturity quintile 5). This pattern is also recognizable in our size/maturity-sorted portfolio constructed from the climate bond sample. Looking at industry sorted portfolios, we notice some industries have much higher adjusted R-squared than others. As an example, the "Electric" and "Transportation" industries having adjusted R-squared of 0.76 and 0.74 respectively, while industries such as "Water" and "Savings and Loans" have adjusted R-squared of 0.26. Similar differences can be observed in our climate bond sample. Results for our size/maturity-sorted portfolios and industry-sorted portfolios are available in table A1 and A2 of the Appendix.

7. Conclusion

The literature on the common risk factors of bond returns is still quite scarce, and Bai, Bali and Wen (2019) are the first to identify bond factors conceptualized from the specific profile of bond investors. Their study could well be the first of a series of studies on the cross-section of US corporate bonds that refer to rich bond-specific databases such as TRACE and FISD in order to obtain more precise information and results on the market.

First and foremost, this paper participated in this literature by exploring the authors' methodology in depth in order to examine if results of this previous research were replicable. Overall, we obtain very positive results on the relationship between downside risk and next-month excess-bond returns, and the Bai, Bali and Wen's four-factor model performs very well on size/maturity- and industry-sorted portfolios.

Furthermore, the specific methodology developed by the authors gives us a favorable framework to study our sample of climate bonds and allows us to understand if these financial products behave differently than their more "traditional" equivalents. We notice some dissimilarities, but overall climate bonds seem only to represent a small, specific sample of the overall US corporate bond market. We consider this study only as a first step towards understanding how climate-aligned firms behave. As urgency regarding climate adaptation and mitigation grows, so will transparency on the environmental impact of business activities and the firms that participate in these activities. This will undoubtedly start to impact how these firms are viewed by investors in the market. In this sense, this paper established preliminary work for future research in this field.

Lastly, as financial economists are starting to question the dense literature on the "zoo of factors" that have been discovered to explain the cross-section of stock returns and the methodologies that are used to determine such factors, there seems to be a need for more advanced and robust statistical tools in financial economy, the likes of which could well be provided by machine learning statistical approaches and learning methods.

Appendix

A.1 Size/Maturity-sorted portfolios

Panel A - Explanatory power of factor models for size/maturity sorted portfolios (Market)

Maturity Quintile	Maturity Quintile	Alpha	t-stat	Adjusted R-squared
1	1	0.22	5.97	0.09
1	2	0.12	4.22	0.21
1	3	0.08	2.86	0.25
1	4	0.06	2.39	0.26
1	5	0.08	2.35	0.16
2	1	0.24	6.87	0.28
2	2	0.18	5.28	0.31
2	3	0.12	4.04	0.43
2	4	0.10	2.52	0.44
2	5	0.10	2.23	0.34
3	1	0.27	6.59	0.30
3	2	0.20	5.00	0.48
3	3	0.15	4.13	0.59
3	4	0.12	2.65	0.57
3	5	0.12	1.73	0.50
4	1	0.32	8.19	0.43
4	2	0.21	6.46	0.60
4	3	0.15	3.34	0.69
4	4	0.10	2.44	0.73
4	5	0.06	1.16	0.70
5	1	0.32	8.32	0.51
5	2	0.20	5.82	0.69
5	3	0.10	3.58	0.83
5	4	0.03	1.15	0.90
5	5	0.02	0.52	0.91
	Average	0.15	3.99	0.49

Panel B - Explanatory power of factor models for size/maturity sorted portfolios (Climate Bonds)

Size quintile	Size quintile	alpha	tstat	adjR
1	1	0.06	0.96	0.17
2	1	0.13	3.61	0.32
3	1	0.10	2.16	0.33
1	2	0.05	0.92	0.38
2	2	0.22	2.60	0.47
3	2	0.04	0.54	0.62
1	3	0.02	0.33	0.50
2	3	0.09	1.51	0.63
3	3	0.03	0.59	0.72
	Average	0.08	1.47	0.46

A.2 Industry-sorted portfolios

Industry Code	Description	Alpha	t-stat	Adjusted R-squared
10	Manufacturing	0.09	1.58	0.66
11	Media/Communications	0.07	1.43	0.64
12	Oil and Gas	0.05	0.88	0.56
13	Railroad	0.07	1.53	0.64
14	Retail	0.10	2.04	0.68
15	Service/Leisure	0.08	1.30	0.63
16	Transportation	0.08	1.52	0.74
17	Mining/Refining	0.01	0.04	0.45
20	Banking	0.08	1.18	0.50
21	Credit/Financing	0.10	1.60	0.43
22	Financial Services	0.06	0.74	0.42
23	Insurance	0.07	1.60	0.69
24	Real Estate	0.07	1.40	0.59
25	Savings And Loan	0.05	0.50	0.26
26	Leasing	0.18	2.56	0.29
30	Electric	0.08	1.63	0.76
31	Gas	0.06	1.05	0.57
32	Telephone	0.09	1.12	0.58
33	Water	0.08	1.15	0.26
40	Foreign Agencies	-0.13	-0.98	0.24
42	Supranationals	0.12	1.87	0.27
44	U.S. Agencies	0.05	0.27	0.17
60	Miscellaneous	0.09	0.97	0.33
99	Unassigned	-0.03	-0.35	0.59
	Average	0.07	1.11	0.50

Industry Code	Description	Alpha	t-stat	Adjusted R-squared
10	Manufacturing	-0.01	-0.23	0.54
13	Railroad	0.07	1.44	0.65
16	Transportation	0.12	2.07	0.71
30	Electric	0.21	1.76	0.34
26	Leasing	0.07	1.15	0.29
33	Water	0.12	1.39	0.25
	Average	0.10	1.26	0.46

References

- Anton, W.R.Q., Deltas, G. and Khanna, M., 2004. Incentives for environmental self-regulation and implications for environmental performance. *Journal of environmental economics and management*, 48(1), pp.632-654.
- Bachelet, M. J., Becchetti, L., & Manfredonia, S. (2019). The green bonds premium puzzle: The role of issuer characteristics and third-party verification. *Sustainability*, 11(4), 1098.
- Bai, J., Bali, T.G. and Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3), pp.619-642.
- Bao, J., Pan, J. and Wang, J., 2011. The illiquidity of corporate bonds. *The Journal of Finance*, 66(3), pp.911-946.
- Barney, J., 1991. Firm resources and sustained competitive advantage. *Journal of management*, 17(1), pp.99-120.
- Bhattacharyya, A. and Cummings, L., 2015. Measuring corporate environmental performance–stakeholder engagement evaluation. *Business Strategy and the Environment*, 24(5), pp.309-325.
- Bessembinder, H., Kahle, K.M., Maxwell, W.F. and Xu, D., 2008. Measuring abnormal bond performance. *The Review of Financial Studies*, 22(10), pp.4219-4258.
- Buchner, B., Herve-Mignucci, M., Trabacchi, C., Wilkinson, J., Stadelmann, M., Boyd, R., Mazza, F., Falconer, A. and Micale, V., 2013. Global landscape of climate finance 2015. *Climate Policy Initiative*.
- Busch, T. and Friede, G., 2018. The robustness of the corporate social and financial performance relation: A second-order meta-analysis. *Corporate Social Responsibility and Environmental Management*, 25(4), pp.583-608.
- Chen, L., Lesmond, D.A. and Wei, J., 2007. Corporate yield spreads and bond liquidity. *The Journal of Finance*, 62(1), pp.119-149.
- Climate Bonds Initiative, 2018. Bonds and Climate Change: The State of the Market.
- Council of the European Union, 2019. “Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the establishment of a framework to facilitate sustainable investment - Approval of the final compromise text”
- Delmas, M.A., Etzion, D. and Nairn-Birch, N., 2013. Triangulating environmental performance: What do corporate social responsibility ratings really capture?. *Academy of Management Perspectives*, 27(3), pp.255-267.
- Dick-Nielsen, J., 2009. Liquidity biases in TRACE. *The Journal of Fixed Income*, 19(2), pp.43-55.
- Dick-Nielsen, J., 2014. How to clean enhanced TRACE data. *Available at SSRN 2337908*.
- Dick-Nielsen, J., Feldhütter, P. and Lando, D., 2012. Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics*, 103(3), pp.471-492.
- Donaldson, T. and Preston, L.E., 1995. The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of management Review*, 20(1), pp.65-91.
- Escrig-Olmedo, E., Muñoz-Torres, M.J., Fernández-Izquierdo, M.Á. and Rivera-Lirio, J.M., 2017. Measuring corporate environmental performance: A methodology for sustainable development. *Business Strategy and the Environment*, 26(2), pp.142-162.
- European Commission (2018) *Action Plan: Financing Sustainable Growth*.
- EU Technical Expert Group on Sustainable Finance, 2020. “Taxonomy: Final report of the Technical Expert Group on Sustainable Finance.”

- Fama, E.F. and MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3), pp.607-636.
- Fama, E.F. and French, K.R., 2015. A five-factor asset pricing model. *Journal of financial economics*, 116(1), pp.1-22.
- Flammer, C., 2018. Corporate green bonds.
- Friede, G., Busch, T. and Bassen, A., 2015. ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), pp.210-233.
- Friedman, M., 1970. September 13. The social responsibility of business is to increase its profits. *New York Times Magazine*, 33, p.126.
- Ge, W. and Liu, M., 2015. Corporate social responsibility and the cost of corporate bonds. *Journal of Accounting and Public Policy*, 34(6), pp.597-624.
- GSIA, 2018. Global Sustainable Investment Review, GSIA Report, April.
- Hachenberg, B. and Schiereck, D., 2018. Are green bonds priced differently from conventional bonds?. *Journal of Asset Management*, 19(6), pp.371-383.
- Hartmann, J. and Vachon, S., 2018. Linking environmental management to environmental performance: The interactive role of industry context. *Business Strategy and the Environment*, 27(3), pp.359-374.
- Jostova, G., Nikolova, S., Philipov, A. and Stahel, C.W., 2013. Momentum in corporate bond returns. *The Review of Financial Studies*, 26(7), pp.1649-1693.
- Kapraun, J. and Scheins, C., 2019. (In)-Credibly Green: Which Bonds Trade at a Green Bond Premium?. Available at SSRN 3347337.
- Kotsantonis, S. and Serafeim, G., 2019. Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, 31(2), pp.50-58.
- Lin, H., Wang, J. and Wu, C., 2011. Liquidity risk and expected corporate bond returns. *Journal of Financial Economics*, 99(3), pp.628-650.
- Linklaters, 2019. Sustainable Finance. The rise of green loans and sustainability linked lending.
- Markowitz, H., 1959. *Portfolio selection: Efficient diversification of investments* (Vol. 16). New York: John Wiley.
- Misani, N. and Pogutz, S., 2015. Unraveling the effects of environmental outcomes and processes on financial performance: A non-linear approach. *Ecological economics*, 109, pp.150-160.
- Newey, W.K. and West, K.D., 1986. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix.
- Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1), pp.435-480.
- Polbennikov, S., Desclée, A., Dynkin, L. and Maitra, A., 2016. ESG ratings and performance of corporate bonds. *The Journal of Fixed Income*, 26(1), pp.21-41.
- Porter, M.E. and Kramer, M.R., 2006. The link between competitive advantage and corporate social responsibility. *Harvard business review*, 84(12), pp.78-92.
- Post, C., Rahman, N. and McQuillen, C., 2015. From board composition to corporate environmental performance through sustainability-themed alliances. *Journal of Business Ethics*, 130(2), pp.423-435.

- Ren, S., He, D., Zhang, T. and Chen, X., 2019. Symbolic reactions or substantive pro-environmental behaviour? An empirical study of corporate environmental performance under the government's environmental subsidy scheme. *Business Strategy and the Environment*, 28(6), pp.1148-1165.
- Sarig, O. and Warga, A., 1989. Bond price data and bond market liquidity. *Journal of Financial and Quantitative Analysis*, 24(3), pp.367-378.
- Schiereck, D., Friede, G. and Bassen, A., 2019. Financial Performances of Green Securities. In *The Rise of Green Finance in Europe* (pp. 95-117). Palgrave Macmillan, Cham.
- SIFMA, 2019. Capital Markets Fact Book 2019.
- Trumpp, C., Endrikat, J., Zopf, C. and Guenther, E., 2015. Definition, conceptualization, and measurement of corporate environmental performance: A critical examination of a multidimensional construct. *Journal of Business Ethics*, 126(2), pp.185-204.
- Trumpp, C. and Guenther, T., 2017. Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. *Business Strategy and the Environment*, 26(1), pp.49-68.
- Xie, S. and Hayase, K., 2007. Corporate environmental performance evaluation: a measurement model and a new concept. *Business Strategy and the Environment*, 16(2), pp.148-168.
- Xue, B., Zhang, Z. and Li, P., 2020. Corporate environmental performance, environmental management and firm risk. *Business Strategy and the Environment*, 29(3), pp.1074-1096.
- Zerbib, O.D., 2019. The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of Banking & Finance*, 98, pp.39-60.

Chapter 6

An Introduction to the Pricing of Climate Bonds via Machine Learning

\

Abstract

In this study, we attempt to determine if corporate bonds issued by climate-aligned firms are priced differently by corporate bond markets. Relying on trade-based data from the TRACE database to obtain the most indicative monthly prices on corporate bonds, we apply machine learning methods and attempt to predict future corporate bond returns by using a set of predictors taken or adapted from the bond pricing and stock pricing literature. We add a dummy variable that identifies climate bonds to see if our machine learning algorithms use this information to predict future bond returns. We find that the climate dummy variable is almost never used to price corporate bonds throughout hundreds of different model fits, indicating that to date, the fact that an issuer is climate-aligned is not taken under account by corporate bond markets to price corporate bonds.

Keywords: Machine Learning; Green Bonds; Sustainable Finance; Return Prediction; Cross-Section of Returns; Ridge Regression; Lasso; Elastic Net; Random Forest

1. Introduction

Data science and machine learning are changing the entire financial landscape, and machine learning methods and algorithms are increasingly viewed by academics in finance as the future of empirical financial research. In the first sentence of his very well received book on the subject, Marcos Lopez de Prado, one of the leading scholars on machine learning (ML) in finance, states that “today’s machine learning (ML) algorithms have conquered the major strategy games, and are routinely used to execute tasks once only possible by a limited group of experts. Over the next few years, ML algorithms will transform finance beyond anything we know today.” Amongst the different comments on Lopez de Prado’s work is that of Campbell Harvey, former President of the American Finance Association and leading scholar in the asset pricing literature, which reads: “The first wave of quantitative innovation in finance was led by Markowitz optimization. Machine learning is the second wave, and it will touch every aspect of finance” further stating that Lopez de Prado’s work on the subject is “essential for readers who want to be ahead of the technology rather than being replaced by it.” (Lopez de Prado, 2018).

As we will see in the following section, machine learning methods are slowly making their way into the stock pricing literature. To date, however, these methods have not been used in the context of the corporate bond pricing literature, which is a more recent and less developed. This is mostly due to the fact that historically, obtaining reliable corporate bond prices has been much more complex⁶⁹. In addition, data on corporate bond prices is generally either matrix-based or quote-based instead of purely transaction-based, and therefore less reliable. Fortunately, the Trade Reporting and Compliance Engine (TRACE) database opened in 2002,

⁶⁹ Data on corporate bond prices has almost entirely disappeared between 1997 and 2004 as the Lehman Brothers Fixed Income Database closed in December 1997 and the Mergent Fixed Income Securities Database (FISD) opened in 2004.

providing academics with purely trade-based data on the US corporate bond market and leading to the first studies on corporate bond asset pricing. In the context of this study, we rely mainly on Bai, Bali and Wen (2019) to construct our bond-specific predictors. In their study, the authors use TRACE data to construct their monthly prices and put forward a model based on four predictors that best explains the cross-section of US corporate bond returns. We include these predictors and include or adapt all the control variables that were used in their study to test the robustness of their results in our set of predictors.

Finally, we rely on Gu, Kelly and Xiu (2020) to construct our methodology and include some predictors adapted from the stock literature in our models. In this study, the authors provide a comparison of the main machine learning methods that are available for academics that wish to better understand and predict future stock returns. We adapt this methodology as well as some stock-based predictors to our study of corporate bond pricing.

We run our different machine learning methods on a sample of 1,279,861 bond-month return observations for 40,657 US corporate bonds over the 165-month period going from June 2005 to February 2019. We first run our models using a set of 23 predictors without including the climate dummy variable, in order to understand the impact of adding the climate predictor to our models in a second step. All our predictors are determined on a monthly basis, and we separate our dataset in a training sample of 60 months, a validation sample of 20 months, and 85 months of monthly test samples. We apply a cross-validation technique that respects the temporal structure of our data.

When the climate variable is excluded from our predictors, only one machine learning method – the Principal Component Regression (PCR) method - succeeds in obtaining a positive average out-of-sample R-squared over the 85 monthly-test samples. Throughout the different

machine learning methods that can select the best performing predictors⁷⁰, the most impactful predictor is a bond's time to maturity, followed by its beta with the profitability factor. Along with the bond's beta with the stock market risk factor, three of the four factors developed in Bai, Bali and Wen (2019) are in the six most impactful predictors for our models. However, these predictors do not suffice for our models to constantly obtain positive out-of-sample R-squared throughout our 85 test-month samples.

Overall, adding a climate variable worsens our results, slightly improving our worst performing traditional linear methods and worsening our best performing machine learning models. With the climate dummy variable included, none of our machine learning methods succeed in obtaining a positive average out-of-sample R-squared over the 85 monthly-test samples. Our analysis of variable importance throughout the machine learning models provide us with interesting insight: the climate bond variable is barely used at all by any of our machine learning methods throughout hundreds of different model fits. The results of this study lead us to conclude that the fact that an issuer is climate-aligned is not taken under account by corporate bond markets to price corporate bonds.

2. Literature Review

In this work, we follow the definition of “machine learning” provided by Gu, Kelly and Xiu (2020). In their study on machine learning in asset pricing, the authors define the term as “(a) a diverse collection of high-dimensional models for statistical prediction, combined with (b) so-called “regularization” methods for model selection and mitigation of overfit, and (c) efficient algorithms for searching among a vast number of potential model specifications”. The high-dimensional aspect of ML methods provides more flexibility compared to more

⁷⁰ These are the machine learning technique that allow for variable selection and dimension reduction techniques such as Elastic Net, Principal Component Regression and Partial Least Square.

traditional econometric methods and allows for better approximation of risk premiums. This however comes at a price of higher risk of overfit, which, as we will see in this section, represents a central subject today in the field of asset pricing. To minimize the risks of overfit, ML techniques include tools specifically designed to enhance out-of-sample performance. The final challenge that can be resolved with ML methods consists in finding the optimal model permutations when dealing with a high number of variables, a challenge that was also central to the asset pricing literature which had relied on logic to put forward the potential predictors it though were most likely to impact the cross-section of stock returns.

Within the financial literature, machine learning methods have mostly emerged in the context of the asset pricing research. A lasso regression⁷¹ method was applied to predict international stock returns and understand the role of the United States in these global market returns (Rapach, Jack, Strauss and Zhou, 2013). Regression trees were used to predict credit card risk (Butaru, Florentin, Chen, Clark, Das, Lo and Siddique, 2016) and model consumer credit-risk (Khandani, Amir, Kim, Lo, 2010). Neural networks have been used to price and hedge derivatives (Hutchinson, James, Lo and Poggio, 1994), model mortgage prepayment and delinquencies (Sirignano, Justin, Sadhwani and Giesecke, 2016), and to perform portfolio selection (Heaton, Polson and Witte, 2017).

Focusing more precisely on the cross-section of expected returns, several technical reports for the University of Chicago and Duke University have applied machine learning techniques to improve their regression models. These include using the bootstrap procedure (Harvey and Liu, 2019) as well as dimension reduction methods for factor pricing models (Giglio and Xiu, 2017; Kelly, Pruitt and Su, 2017). Tree-based models have also been used to sort portfolios in order to establish a relationship between past and future stock returns (Moritz and

⁷¹ Lasso (least absolute shrinkage and selection operator) regression is a machine learning method that consists in performing both variable selection and regularization to optimize the accuracy and interpretability of the produced statistical model.

Zimmermann, 2016). Other reports from the University of Michigan (Kozak, Nagel and Santosh, 2017) and the University of Wisconsin-Madison (Freyberger, Neuhierl and Weber, 2017) have worked on applying selection methods and shrinkage in the context of the study of the cross-section of expected returns.

Looking at the machine learning literature, many authors have realized that machine learning methods could be applied to the consequential challenge of using times series data to predict stock market returns, and that their models could be used to develop profitable strategies regardless of the efficient market hypothesis (Henrique, Sobreiro, Kimura, 2019). The efficient market hypothesis was developed by Malkiel and Fama (1970) but was revised more than 20 years later by Fama (1991) when the author established that financial markets followed random directions and were therefore unpredictable. More recently, the efficiency of financial markets has been challenged many times, and authors both from machine learning research (Atsalakis and Valavanis, 2009) and finance research (Malkiel, 2003) have addressed this issue by summarizing different works in their field. Given the non-stationary, non-linearity, noisy and chaotic nature of price time series in financial markets, and the fact that these can be influenced by a variety of economic, financial, political or even psychological factors, research on the subject has generated much interest in the machine learning community (Henrique, Sobreiro, Kimura, 2019). Prediction models developed by the machine learning academic community are applied to financial distress estimation (Chen and Du, 2009) and stock trading prediction (Chang, Liu, Lin, Fan and Ng, 2009; de Oliveira, Nobre, and Zarate, 2013), with a series of studies successfully identifying profitable opportunities (Doeksen, Abraham, Thomas, and Paprzycki, 2005; Huang, Yang and Chuang, 2008; Patel, Shah, Thakkar and Kotecha, 2015).

To date, there are few studies that focus on applying machine learning methods to bond markets. Bianchi, Buchner and Tamoni (2019) apply such methods to estimate bond risk premia following the work on equity risk premia by Gu, Kelly and Xiu (2019) on Treasury

Bond data. Other studies focus on more specific technical subjects, such as using machine learning to identify clients that are most likely to be interested in a given bond (Wright, Capriotti, Lee, 2018), or to improve the quality of corporate bond yield data (Kirczenow, Hashemi, Fathi and Davison, 2018). Given the positive results of machine learning on market price prediction both for equity and treasury bond markets, it seems clear that machine learning methods will increasingly be used in financial research.

3. Data

3.1. Required databases for corporate bond and climate bond pricing

For this paper, we use three distinct sources of data. We obtain transaction data for the US corporate bond market through the TRACE database, which we use to compute corporate bond returns. We use the FISD database to obtain information on bond characteristics. Finally, we obtain a list of climate bond issuers from the Climate Bond Initiative, which we use to identify climate bonds through the FISD database.

Since June 2002, all corporate bond transactions in the secondary market have been made available with the TRACE system through the Trade Reporting Compliance Engine⁷². Before this new dataset was made available, most studies that focused on the corporate bond market only used daily quotes and matrix prices for corporate bonds, which could bias results. In their paper on the subject, Sarig and Warga (1989) explore the fact that there can be liquidity-driven noise errors in daily prices for corporate bonds since daily prices are given even on days when bonds have not been traded for multiple days. When this is the case, brokers set matrix prices based on similar bonds issued by issuers with similar characteristics, which creates bias. More

⁷² The Trade Reporting and Compliance Engine (TRACE) was developed by the National Association of Securities Dealers (NASD) and is used for over-the-counter (OTC) transactions for fixed-income securities. This trade-based data is made available to academics through the TRACE database.

recently, Dick-Nielsen (2009) shows that this bias still exists with prices from Datastream⁷³. This gives TRACE data a considerable edge when focusing on daily prices that can be used for microstructure research such as event studies, as well as an edge for weekly and monthly prices.

Two distinct versions of TRACE exist: the standard TRACE data that censors trading volumes that are greater than \$5 million for investment grade bonds and greater than \$1 million for speculative grade bonds and that usually has a three-month lag for the availability of data, and the enhanced TRACE data, that has information on all transaction volumes but has an 18 month lag for the availability of data. In this paper, we refer to enhanced TRACE data, which spans from June 2002 to March 31st 2019.

Considering bond characteristic, FISD is considered as the most comprehensive database focusing on bonds and contains detailed information on the characteristics of more than 140,000 US bonds, either corporate bonds, supranational bonds, U.S. Agency Bonds and U.S. Treasury Bonds. The FISD database contains essential information on bond issuers as well as specific bond issues such as bond issue date, maturity, size, coupon, type, and any information that can be used to identify and categorize US bonds.

Finally, we use a list of climate bond issuers provided directly by the Climate Bond Initiative (CBI) research team. This list was used by CBI in their study on the climate bond market (Climate Bond Initiative, 2018). To identify climate bonds, CBI identified issuers that originated at least 75% of their revenues from green business lines in either clean energy, low-carbon buildings and transport, water and waste management and sustainable land use. Climate bonds were included if they were issued after the 1st of January 2005 and before the end of Q2 2018. Therefore, we use the list of issuers provided by the Climate Bond Initiative to identify all climate bond transactions available on the Enhanced TRACE database since June 2002 to

⁷³ As it is explained in Jostova, Nikolova, Philipov, Stahel (2013): “Extensive discussions with the DataStream support team about the source of their data confirmed that most U.S. corporate bond prices are dealer quotes reported by market-makers. These data are further augmented with trading prices for traded bonds. Like Lehman, DataStream provides no indication of whether a price is based on a quote or a trade.”

the 31st of March 2019. Our database for traditional bonds uses all available Enhanced TRACE data on the same period.

3.2. From raw intraday transaction data to monthly corporate bond returns

As I continue to follow Bai, Bali and Wen (2019), I remove bonds that are not traded in in the US public market, bonds that are structured notes, mortgage or asset-backed, agency backed or equity-linked, convertible bonds, bonds with floating coupon rates, bonds that trade below \$5 or above \$1000 as well as bonds that have less than one year to maturity. Regarding transaction data, much like Bai, Bali and Wen, I remove bond transactions that are labeled as “when issued”, “locked-in”, that have special sales conditions and that have more than a two-day settlement. Furthermore, I also remove transactions that have a volume inferior to \$10,000. Finally, I follow the filtering methodology developed in Dick-Nielsen (2014) to remove cancellations and corrections made on the TRACE Enhanced database, as well as inter-dealer transactions that are reported twice in the trace database. Much like the author, I apply the two distinct approaches for pre- and post-2012 Enhanced TRACE data. After applying these different filters, my dataset is composed of 91,294,517 intraday transactions spanning from June 2002 to March 2019.

Once we obtain this intra-day information on bond transactions, we compute a daily price by referring to the “trade-weighted price, all trades” approach of Bessembinder et al. (2008), which consists in performing a value-weighted average of all intraday transaction prices. This methodology puts more emphasis on trades from institutional investors that benefit from lower execution costs, therefore providing a better accuracy for daily prices. Once we apply this methodology, our daily price dataset consists of 13,770,971 daily prices⁷⁴.

74 In Bessembinder et al (2009), four distinct strategies are used to compute daily prices from TRACE data: using the last price reported in TRACE for the day (“last price, all trades”), construct daily prices by weighing each trade by size (“trade-weighted price, all trades”), eliminating all trades below \$100,000 dollars and using the last available price of the day (“last price, trade $\geq 100k$ ”), and eliminating all trades below \$100,000 and weighing each trade by size (“trade-weighted price, trade

In order to obtain monthly prices from daily prices, I develop a methodology similar to that of Bai, Bali and Wen (2019). I first identify the bond price for the last trading day for each bond-month. If the last trading day is one of the last five trading days within month t , then this daily price is used as a monthly price. If that is not the case, I identify the bond price for the first trading day of month $t+1$. If this trading day is one of the first five trading days of month $t+1$, it is used as a price for month t . This allows for a more complete dataset of bond monthly prices in order to be able to compute more monthly returns using data from months t and $t+1$, where returns are either computed using (1) end of month $t-1$ to end of month t daily prices (2) start of month t to end of month t daily prices (3) start of month t to start of month $t+1$. Using the methodology, we compute 1,761,543 monthly prices for our dataset.

The monthly corporate bond return at time t is computed as:

$$r_{i,t} = \frac{(P_{i,t} + AI_{i,t} + Coupon_{i,t}) - (P_{i,t-1} + AI_{i,t-1})}{P_{i,t-1} + AI_{i,t-1}}$$

where $P_{i,t}$ is the transaction price, $AI_{i,t}$ is accrued interest, and $Coupon_{i,t}$ is the coupon payment, if there is any, of bond i for month t . We obtain the necessary information to compute corporate bond returns from the FISD database.⁷⁵ In order to compute our returns, month prices need to be available for adjacent months. Our final dataset is composed of 1,530,745 corporate bond returns.

4. Variables

In order to select and compute our variables, we follow Gu, Kelly and Xiu (2018). In their study, the authors apply a series of machine learning methods to a sample of 30,000 individual

$\geq 100k$). Even though we do apply a “trade-weighted price, all trades” in this paper, we do exclude trades below \$10,000 from our sample following Bai, Bali and Wen (2019).

⁷⁵ This includes the coupon rate, the dated date marking the start of the interest period, as well as the interest frequency.

stocks from 1957 to 2016 in order to predict stock market returns. Out of the 900 possible factors that are used to run these models and predict stock market returns, the authors find that the most successful types of factors are - in order from most informative to least informative - price trends factors, liquidity factors and volatility factors. Price trends factors include either stock or industry momentum and short-term reversals. Liquidity predictors include market value, dollar volume and bid-ask spreads. Finally, predictors linked to volatility include return volatility, idiosyncratic volatility, market beta or beta squared.

Going into more detail concerning the predictive power of each predictor, the authors provide a list illustrating the importance of each characteristic for each machine learning model they apply to their sample, which in turn provides a precise ranking of the impact of each individual variable in predicting stock market return. Furthermore, citations of the papers that first developed each predictor are made available by the authors.

There are some important differences between the approach of Gu, Kelly and Xiu (2020) and the chosen approach for this study. First and foremost, this study focuses on predicting corporate bond market returns, and therefore predictors need to be adapted to the corporate bond market. Much like Gu, Kelly and Xiu (2020) refer to the asset pricing literature that focuses on stock markets, I refer here to corporate bond literature to develop the predictors in this study. Predictors are either directly taken from the corporate bond pricing literature or adapted to corporate bonds from the stock pricing literature. Another important difference resides in the fact that the author is limited in terms of computing power and cannot reasonably run complex machine learning models on such a large number of predictors. Regardless, given the insight provided by Gu, Kelly and Xiu (2020), we select the most important factors in terms of price trends, liquidity and volatility in their sample of factors and adapt them to the corporate bond market when possible. A second important difference between our approach and that of Gu, Kelly and Xiu (2020) is that in this study, we develop factors that are available on a

monthly basis to develop models that focus on predicting next-month bond returns. This means that the factors used in Gu, Kelly and Xiu (2020) have to be determined monthly. This information is also provided by the authors.

Supplementing these selected variables are some other factors that are either taken directly from Bai, Bali and Wen (2019), or adapted from the study. Through their research, Bai, Bali and Wen (2019) observed that the literature that focus on the cross-section of corporate bond returns relied on stock market factors, based on the rationale that the stock and bond markets are integrated, since both bonds and stocks value similar underlying assets. Their work goes against this more traditional approach by focusing on risk factors that are specific to the bond market. In addition to these bond specific factors, the authors use a series of control variables in their studies to demonstrate the explanatory power of their factors. Amongst these are the most renowned risk factors in the literature. These are the excess-stock market return (MKTstock), the size factor (SMB), the book-to-market factor (HML), the profitability factor (RMW), the investment factor (CMA), the momentum factor (MOM) and the liquidity risk factor (LIQ), which have been developed and computed in Fama and French (1993), Carhart (1997) and Pastor and Stambaugh (2003)⁷⁶. These factors were developed at portfolio levels. Following Gu, Kelly and Xiu (2018), I apply machine learning methods in this study that can be applied at bond-level and do not need to use control variables at a portfolio level. Much like it is performed in Bai, Bali and Wen (2019), I create bond-level factors by computing betas for each of these portfolio-level factors. In the following subsections we explore four categories of factors in accordance with the results of Gu, Kelly and Xiu (2018): price-trend factors, volatility factors, liquidity factors, and a subsection addressing other types factors we have

⁷⁶ Much like the risk-free rate, the excess-stock market return (MKTstock), the size factor (SMB), the book-to-market factor (HML), the profitability factor (RMW), the investment factor (CMA), the momentum factor (MOM) and the liquidity risk factor (LIQ) can be found online, either on Kenneth French's website <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/> or Lubos Pastor's website <https://faculty.chicagobooth.edu/lubos-pastor/research>

chosen to take under consideration. The last subsection explores the summary statistics of our dataset. In Table 1, we summarize our selected factors, provide the reference of the original paper and a description of how we compute the predictor with our data. Information on whether the factor is monthly or yearly is also provided.

4.1. Price-trend factors

Price trends taken from the stock asset pricing literature that focus on monthly returns are relatively straightforward to adapt to the corporate bond market, since the processes that need to be performed are similar and only require monthly corporate bond returns. Most of the price-trend factors Gu, Kelly and Xiu (2020) are momentum factors. However, some factors rely on accounting data that is only available on a yearly basis and cannot be used in this study. Some examples of these include the sales to price ratio factor (sp), the number of earnings increases factor (nincr) and the earnings to price factor (ep). Momentum factors we include in our study are the 1-month (mom1m), 6-month (mom6m), 12-month (mom12m) and 36-month (mom36m) momentum factors, each corresponding to the sum of returns in the period previous to month- t , with 1-month momentum corresponding to monthly returns in months $t-1$. The original 12-month momentum factor was discovered by Jegadeesh (1990), and the 1-month, 6-month and 36-month momentum factors were developed in Jegadeesh and Titman (1993). Other momentum factors include the industry momentum factor, which was developed by Moskowitz and Grinblatt (1999) and consists in computing 6-month momentums for industry portfolios, and the momentum change factor, which consists in computing the difference between 6-month momentum and momentum between month $t - 12$ and momentum between month $t - 6$ (Gettleman and Marks, 2006).

The last price trend factor that we choose to adapt from Gu, Kelly and Xiu (2020) is the maximum daily return factor, and consists, like its name indicates, in finding the maximum daily return for every month. In order to compute this predictor, we use our dataset of daily

prices computed from TRACE and determine daily returns using these daily prices. The original max daily return was developed by Bali, Cakici & Whitelaw (2011).

Finally, I include the Value-at-risk (VaR) factor in our sample of factors. VaR measures the proportion by which the price of an asset could decline over a certain period of time as a result of variations in market rates or prices. To develop our measure for VaR we follow Bai, Bali and Wen (2020) a create a proxy by selecting the second lowest monthly return in a 36-month period prior to month t. Our measure is then multiplied by -1 in order to have positive values for VaR.

4.2. Liquidity factors

The liquidity measure that we use in this study is that developed by Bao, Pan and Wang (2011), which is also used in the study on the cross-section of corporate bond returns by Bai, Bali and Wen (2019).

The authors define P_t as a bond's clean price (meaning the price of a bond without accounting for accrued interests and coupon payment) at time t and p_t as the log price, and assume that p_t consists of the following components :

$$p_t = f_t + u_t$$

Where f_t represents the fundamental value of the log price in the absence of friction, and u_t is generated from the impact of illiquidity. Therefore, in the author's framework, u_t characterizes illiquidity on the market.

Finally, the authors define their measure for illiquidity, which is aimed at extracting this component in the observed clean log price, as follows:

$$\gamma = - Cov(\Delta p_t, \Delta p_{t+1})$$

$$\Delta p_t = p_t - p_{t-1}$$

With . Bao et al. (2011) develop this model under the assumption that follows a random walk, so that γ only depends on u_t and therefore increases when u_t increases.

4.3. Volatility factors

Our selected volatility factors include the bond market beta developed by Bai, Bali and Wen (2019) and the tradition stock market beta from Fama-French (1993). Following the authors, we first determine bond market excess return by computing value-weighted average returns of all corporate bonds in our sample and subtract the one-month treasury bill rate from these values, and then perform a time-series regression of every bond's excess returns on the bond market excess return using a 36-months rolling-window. In order to compute the stock market beta, we export the stock market excess return ($Mkt-Rf$) from Kenneth French's website and perform the same operation with each corporate bond's excess monthly returns.

4.4. Other factors

Other factors are included in this study that are not considered price trends, liquidity or volatility factors. This includes the credit quality factor from Bai, Bali and Wen (2019), which consists in computing the monthly average of S&P, Moody's and Fitch ratings for every bond. In addition, I also include the two bond-level control variables that are used by the authors: bond maturity, which corresponds to the number of years left at month t before bond reaches maturity, and bond size, which corresponds to the amount outstanding of the bond.

Finally, I include several factors that consist in computing the betas of several renown bond-level factors that are used in the literature and by Bai, Bali and Wen (2019) as control variables. These are the betas of the aforementioned stock market return (MKTstock), size factor (SMB), book-to-market factor (HML), profitability factor (RMW), investment factor (CMA), momentum factor (MOM) and liquidity risk factor (LIQ) developed and computed in Fama and French (1993), Carhart (1997) and Pastor and Stambaugh (2003). We compute betas

by performing time-series regressions of individual bond excess returns on each risk factor using a 36-month rolling-window.

5. Methodology

Throughout this paper, we follow the methodology developed in Gu, Kelly and Xiu (2020). In their work, the authors compare the performance of different ML methods in predicting stock returns. This includes (1) linear models, including ordinary least squares (OLS); (2) generalized linear models that apply penalization methods such as elastic net, LASSO⁷⁷ and ridge regressions; (3) methods that apply dimension reduction techniques such as principal component regressions and partial least squares; (4) regressions trees and (5) neural networks.

For each type of model, Gu, Kelly and Xiu (2020) provide a statistical model describing how the method is adapted to risk premium predictions, an objective function to estimates the parameters of the model and the computational algorithms corresponding to the model. The authors' objective for each model is to minimize the mean square predictions error (MSE). In the context of this study, the authors describe an asset's excess return as an additive prediction error model:

$$r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1}$$

where

$$E_t(r_{i,t+1}) = g^*(z_{i,t})$$

And stocks are indexed as $i = 1, \dots, N_t$ and months by $t = 1, \dots, T$. With this approach, the authors attempt to isolate $E_t(r_{i,t+1})$ as a function of variables that maximizes the out-of-sample explanatory power for $r_{i,t+1}$. $z_{i,t}$ represents the P-dimensional vector of variables and $g^*(\cdot)$ is a flexible function of these variables.

⁷⁷ LASSO stands for least absolute shrinkage and selection operator

Table 1

From Gu, Kelly and Xiu (2020)

Acronym	Predictor	Reference	Type of predictor	Computation
mom1m	1-month momentum	Jegadeesh & Titman (1993)	Price-trend	Return in month t-1
mom12m	12-month momentum	Jegadeesh (1981)	Price-trend	Cumulative returns between month t-12 and t-1
chmom	Change in 6-month momentum	Gettleman & Marks (2006)	Price-trend	Cumulative returns between month t-6 and t-1 minus by Cumulative returns between month t-12 and t-6
maxret	Maximum daily return	Bali, Cakici & Whitelaw (2011)	Price-trend	Maximum daily return of month t-1
indmom	Industry momentum	Moskowitz & Grinblatt (1999)	Price-trend	Cumulative returns of industry portfolio between t-6 and t-1
dolvol	Dollar trading volume	Chordia, Subrahmanyam & Anshuman (2001)	Liquidity	The natural logarithm of the dollar volume of trading in the security in month t-2
mom6m	6-month momentum	Jegadeesh & Titman (1993)	Price-trend	Cumulative returns between month t-6 and t-1
mom36m	36-month momentum	Jegadeesh & Titman (1993)	Price-trend	Cumulative returns between month t-36 and t-1
ill	Amihud liquidity measure	Amihud (2002)	Liquidity	Sum of absolute daily returns divided by daily dollar volume over each month

From Bai, Bali and Wen (2019)

- Main predictors

VaR	5% Value at risk	Bai, Bali and Wen (2019)	Price trend	Second lowest monthly return observation over the past 36 months
meangrade	Credit Quality	Bai, Bali and Wen (adapted from)	Other	Average monthly rating between S&P Ratings, Moody's Ratings and Fitch Ratings
Illiq	Bond Illiquidity	Bai, Bali and Wen (2019)	Liquidity	Bao liquidity measure
β bond	Bond Market Beta	Bai, Bali and Wen (2019)	Volatility	Time-series regressions of individual bond excess returns on the bond market excess returns using a 36-month rolling window

- Control Variables

timetomat	Time to maturity of the bond	Bai, Bali and Wen (2019)	Other	Number of years left until bond reaches maturity
Size	Bond size	Bai, Bali and Wen (2019)	Other	Amount outstanding of the bond at month t - 1
β DEF	Default Beta	Fama and French (1993) (adapted from)	Volatility/Other	Time-series regressions of individual bond excess returns on the Fama-French default factor using a 36-month rolling window
β TERM	Term Beta	Fama and French (1993) (adapted from)	Volatility/Other	Time-series regressions of individual bond excess returns on the Fama-French term factor using a 36-month rolling window
β Mkt-Rf	Stock Market Risk Beta	Fama and French (1993) (adapted from)	Volatility	Time-series regressions of individual bond excess returns on the Fama-French stock market risk factor using a 36-month rolling window
β SMB	Size Beta	Fama and French (1993) (adapted from)	Volatility/Other	Time-series regressions of individual bond excess returns on the Fama-French size factor using a 36-month rolling window
β HML	Value Beta	Fama and French (1993) (adapted from)	Volatility/Other	Time-series regressions of individual bond excess returns on the Fama-French book-to-market factor using a 36-month rolling window
β MOM	Momentum Beta	Carhart (1997) (adapted from)	Volatility/Price Trend	Time-series regressions of individual bond excess returns on the Carhart momentum factor using a 36-month rolling window
β LIQ	Liquidity Beta	Pastor and Stambaugh (2003) (adapted from)	Volatility/Liquidity	Time-series regressions of individual bond excess returns on the Pastor and Stambaugh liquidity measure using a 36-month rolling window

This approach does have important limitations. $g^*(\cdot)$ does not depend on i or t . Contrarily to standard asset pricing approaches that reestimate a time-series model for each asset or a cross-section model for each time-period, this model maintains the same form through time and across assets, providing more stability with regards to the risk premium estimates of assets. Furthermore, $g^*(\cdot)$ depends only on z through $z_{i,t}$, meaning the model's predictions do not use any information from historical data prior to t or any other asset than the i th.

As we continue to follow the methodology of Gu, Kelly and Xiu (2020) every model description in this study can be found in the authors' work, with the following exceptions: our linear models rely on a standard least squares objective function without its robust extension and we perform our generalized linear model without performing a group lasso⁷⁸.

3.1. Sample splitting and tuning via cross-validation

The application of ML methods requires performing specific preliminary steps. In order to be able to perform regularization - the central tool applied in the context of ML methods to minimize overfitting - a choice regarding hyperparameters needs to be made. Hyperparameters control the complexity of models, and for each ML method applied to data, a choice of hyperparameter⁷⁹ will lead to the best possible result from the model in out-of-sample data. In order to determine the best value for a model's hyperparameter, the data sample is divided in three different time periods maintaining similar temporal ordering. The first sample, the "training" sample, is used to estimate the model that will be subject the hyperparameters. The second sample, the "validation" sample, is used to determine the optimal value for the hyperparameter. In order to do that, forecasted datapoints are determined in the validation sample using the model applied to the training sample, and the optimal hyperparameter is found

⁷⁸ In Gu, Kelly and Xiu (2020), the linear model description is performed in section 1.2, the description of the statistical model for the elastic net is performed in section 1.3, PCR and PLS methods are described in section 1.4 and the generalized linear model is explained in section 1.5.

⁷⁹ Main examples of hyperparameters include penalization parameters for elastic net, the number of trees in boosting or the number of random trees in a forest

iteratively based on forecasts errors in the validation set. It is important to note here that the model used in the training data needs to be re-estimated every time a new hyperparameter is tested. The idea of the validation sub-sample is to stimulate a form of out-of-sample test of the model and find the hyperparameter values that correspond to the optimal level of complexity of the model that produces the best results in the validation sample. Once the optimal model, hyperparameter and model complexity have been determined, the “testing” subsample, which is truly out-of-sample because it has not been used for either estimation or tuning, is used to evaluate the out-of-sample predictive performance of a method.

There are a few differences between this paper and Gu, Kelly and Xiu (2020) that we can benefit from, and which impact our approach to sample splitting and tuning via validation. First and foremost, in the context of this paper, we are more limited in terms of computational power, and therefore limit the ML methods we apply to the less computationally intensive. The methods we apply throughout this paper are linear-based : they are the standard ordinary least square (OLS) method, using either the four factor of Bai, Bali and Wen (2020) or the our entire set of factors, the Elastic Net method, the Principal Component Regression (PCR), the Partial Least Square (PLS) method and the Generalized Linear Model (GLM) method, which we define in the following sub-sections.

Furthermore, the period we study for our sample is smaller than that of Gu, Kelly and Xiu (2020). The authors study stock prices from 1956 to 2016 when our sample spans from 2005 to 2019. This provides with another opportunity which consists in fitting out models monthly instead of yearly. This is even more interesting because our variables are computed monthly and therefore provide new monthly information that can be used by our different models.

The last central difference between our study and Gu, Kelly and Xiu (2020) consists in the fact that we apply a cross-validation technique that maintains the temporal ordering of our data. Given the fact that some of our variables require 36 months of previous monthly data to be

computed, and that data from TRACE is first made available in June 2002, our models run on data from June 2005 to March 2019. Our approach then consists in the following steps:

- We use a 5 years / 60 months period for our training sample.
- We require 20 further months for our validation sample.
- We test our models on the month following the last month of the validation sample

Each month in our validation sample is used to fit a model using a different hyperparameter. This gives our ML methods 20 different samples to try to fit the best possible model. Every time a month in our validation sample is used to fit a model using a new hyperparameter, we then integrate this month in our training sample while maintaining a training sample of 60 months. We then use the following month as our validation sample. This method allows us to perform cross-validation while maintaining the temporal ordering of the data. Since our dataset starts on June 2005 and we maintain a 60-month training sample and a 20 months validation sample, our first test month is on March 2012. We therefore have 7 years (84 months) of different test results for our different models. We explore these results in the following section.

6. Results

6.1. Models without Climate dummy variables

We test each of our machine learning methods for every month from March 2012 to February 2019. Our results vary between months and ML methods and are summarized in Table 2. The full extent of the different obtained out-of-sample R-squared are available in Appendix 1. Monthly out-of-sample R-squared are represented in Figure 1.

Table 2

Month	OLS4 - Roos	OLS - Roos	Elastic Net - Roos	PCR - Roos	PLS - Roos	Glm - Roos
Min	-0.517	-0.782	-0.187	-0.110	-0.130	-0.189
Median	-0.009	-0.065	-0.001	-0.001	-0.004	-0.065
Mean	-0.017	-0.085	-0.003	0.003	-0.003	-0.085
Max	0.290	0.419	0.164	0.087	0.093	0.419
Standard Deviation	0.097	0.134	0.064	0.032	0.065	0.134

Our results demonstrate that our models are unable to constantly obtain positive out-of-sample R-squared using the developed predictors. Though all models obtain both negative mean and median values over the sample period, some perform better than others. Both traditional linear models that used all variables – the Ordinary Least Square approach and generalized linear model approach - have the worst performance, with a mean out-of-sample R-square of -8.5% and a median out-of-sample R-squared of -6.5%. These models also have the highest standard deviation for their performance with standard deviation of 0.134 each.

The OLS4 model, which represents the Ordinary Least Square approach using only the four factors of Bai, Bali and Wen (2019), has a much better performance, with a mean out-of-sample R-squared of -1.7% and a median out-of-sample R-squared of -0.9%. Though Bai, Bali and Wen (2019) do argue that their four-factor model outperform all other models to explain the cross-section of corporate returns, it seems this is still not sufficient to create a model that performs positively in predicting next-month bond returns.

Finally, as it could be expected, ML methods that use shrinkage methods – such as the elastic net – and dimension reduction techniques – such as the principal component regression (PCR) or partial least square (PLS) method – perform better than standard linear models. The elastic net method has a mean out-of-sample R-Square of -0.3% and median out-of-sample R-squared of -0.1%, while the PLS approach has a similar mean out-of-sample R-Square of -0.3% and a slightly lower median out-of-sample R-squared of -0.4%. The PCR method is the only

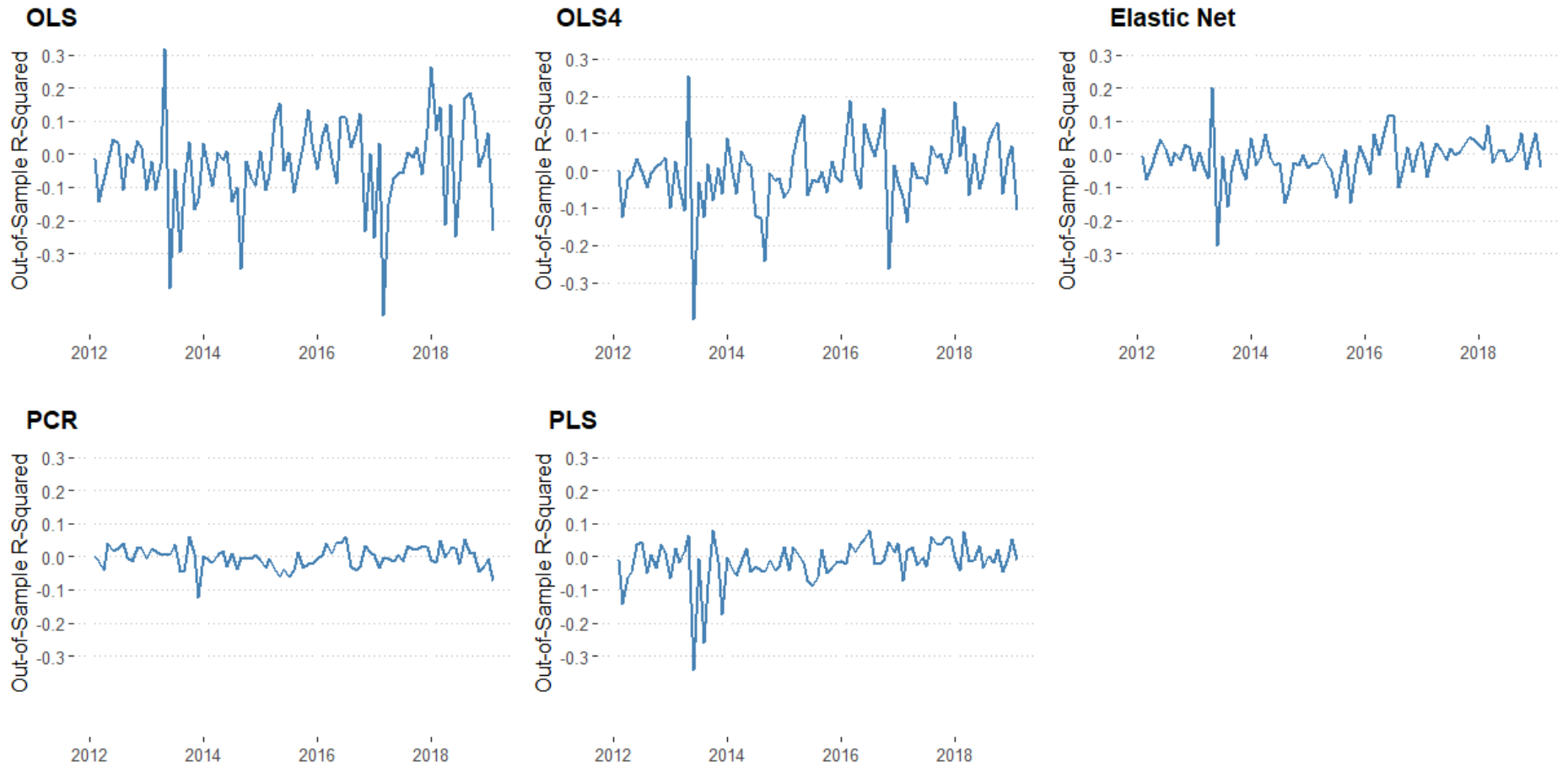
method to have positive mean out-of-sample R-squared of 0.3%, though median out-of-sample R-squared values are slightly in negative territory with -0.1%. The PCR method is also the one with the lowest standard deviation (0.032) compared to the other two methods (0.064 for elastic net and 0.065 for PLS). The fact that the PCR approach outperforms other approach could be explained by the fact that many of our price-trend variables are correlated and dimension reduction technique are best for dealing with correlated variables.

The volatility of monthly out-of-sample R-squared for each method can be visualized in Figure 1. Through these figures, we understand how each test-month showcases different performances for each type of method, and that some test months have quite high yet isolated out-of-sample R-squared. As an interesting example, we can observe that for the test month of May 2013, the standard OLS regression using every variable results in an out-of-sample R-squared of 30%, but then results in a very low out-of-sample R-squared of -40.6% for the next test-month of June 2013. From these representations, we can also notice how standard deviation differs between our models, with Elastic Net, PCR and PLS methods demonstrating much less variation and having out-of-sample R-squared performances varying at a much closer distance from 0.

6.2. Variable importance for models without climate dummy variable

While the traditional linear models that use none of the machine learning methods such as cross-validation, shrinkage or dimension reduction use all of the variables they are provided, Elastic Net, PCR and PLS machine learning methods use specific sets of predictors to obtain the best performing models. As previously mentioned, the Elastic Net method will in most cases use only a subset of provided predictors to create its best models and use the predictors they find to be the most impactful. Machine learning methods that use dimension reduction techniques perform a different type of operation and

Figure 1



learning methods that use dimension reduction techniques perform a different type of operation and construct new components that are composed of different predictors and use these constructed components as new variables for their models. Much like it is the case for the elastic net approach, both PCR and PLS methods use the most impactful predictors to construct these components. In table 2, we show the importance of each predictor in the construction of each methods' final model for every test month.

Table 3

Variables	Elastic Net	PCR	PLS	Mean
timetomat	0.362	1.000	1.000	0.787
betaRMW	0.550	0.376	0.893	0.606
meangrade	0.473	0.412	0.805	0.563
betaMktRF	0.154	0.418	0.974	0.515
VaR	0.212	0.343	0.751	0.435
MKTbond	1.000	0.000	0.304	0.435
betaMom	0.100	0.349	0.809	0.419
betaSMB	0.056	0.289	0.752	0.366
betaterm	0.032	0.325	0.609	0.322
betaHML	0.063	0.230	0.610	0.301
betadef	0.067	0.237	0.548	0.284
mom12m	0.332	0.142	0.314	0.263
maxday	0.317	0.106	0.361	0.261
realindmom	0.381	0.078	0.242	0.234
mom36m	0.113	0.146	0.367	0.209
ILLIQ	0.108	0.121	0.338	0.189
mom1m	0.395	0.029	0.114	0.180
betaCMA	0.073	0.109	0.299	0.160
size	0.040	0.194	0.171	0.135
momt6t1	0.018	0.125	0.220	0.121
betaAgg.Liq.	0.155	0.061	0.123	0.113
chmom	0.160	0.000	0.118	0.093
sd_turn	0.000	0.244	0.000	0.081

These results in terms of variable importance are insightful, both in the context of the work of Bai, Bali and Wen (2019) and of Gu, Kelly and Xiu (2020). In Bai, Bali and Wen (2019), the authors establish that the best model available to date to explain the cross-section of corporate bond returns is composed of 4 variables that they develop in their paper : the Value-at-risk variable (VaR) the credit rating variable (meangrade), and the illiquidity measure of Bao, Pan and Wang

(2011) (ILLIQ) as well as the bond market risk (MKTbond). Table 2 confirms that throughout the hundreds of different model fits performed through these different machine learning methods, although three of these predictors are in the top six most important predictors overall, a bond's time to maturity (timetomat) and a bond's beta with the profitability factor (betaRMW) are the most impactful predictors to determine next-month corporate bond returns. Interestingly, a bond's beta with the stock market risk factor (betaMktRF) is more important than its beta with the corporate bond market risk return (MKTBond).

Regarding Gu, Kelly and Xiu (2020), to date our linear machine learning models are unable to determine future corporate bond returns continually. Furthermore, it seems like predictors based on price trends have little role to play in determining future corporate bond returns compared to stock returns, as the first price trend predictor – twelve-month momentum (mom12m) – is only the twelfth most significant predictor for our different models. According to this study, and unlike Gu, Kellu and Xiu's (2020) on the stock market, the most impactful predictors in our sample are not the traditional price-trend, volatility and liquidity-based predictors but rather bond specific predictors such as time-to-maturity, credit rating, and betas with the literature's most important stock and bond-level risk factors.

It is also interesting to notice that for our elastic net models, the leading predictor is the bond market risk factor when our PCR models do not even use this predictor in constructing their models, which, in turn, seem slightly more efficient.

3.2. Models without Climate dummy variables

Once we have obtained our results, we run our models once again, this time integrating the climate dummy variable. This can help us understand whether or not the fact that a bond is issued by climate-aligned firms has an impact on its pricing and can be used by the machine to better

approximate how the corporate bond's future price. Descriptive statistics of our new dataset are available in Table 4.

Table 4

Month	OLS4 - Roos	OLS - Roos	Elastic Net - Roos	PCR - Roos	PLS - Roos	Glm - Roos
Min	-0.397	-0.486	-0.277	-0.157	-0.345	-0.489
Median	-0.002	-0.022	-0.010	-0.003	-0.013	-0.023
Mean	-0.004	-0.030	-0.011	-0.004	-0.016	-0.030
Max	0.256	0.318	0.200	0.061	0.081	0.318
Standard Deviation	0.097	0.134	0.064	0.035	0.065	0.134

As we can see, adding a climate dummy variable changes some of our results. Interestingly, mean and median out-of-sample R-squared for our classic linear models are higher, when they are reduced for our Elastic Net, PCR and PLS methods. However, standard deviation for each method does not change. Overall, since our Elastic Net, PCR and PLS methods were providing us with the best results for our different predictors, adding a climate bond dummy does reduce our best mean out-of-sample R-squared from 0.3% from our PCR method to -0.4% from both our OLS4 approach and our PCR approach with climate dummy variable.

6.3. Models without Climate dummy variables

We perform a new variable importance analysis on the machine learning methods and models that include a climate dummy variable, and find that the climate variable is the least used variable for all three machine learning methods that are elastic net, PCR and PLS. This means that even though hundreds of models are fitted through the 165 month period, and taking under account that our cross-validation methodology fits hundreds of models to fit its best fit for every month, the climate dummy variable contains strictly no information that can be used by a machine learning algorithm to determine a bond's pricing. This allows us to conclude that climate bonds are not

priced differently from traditional bonds. Variable importance for our new model permutations are available in table 4.

Table 4

Variables	Elastic Net	PCR	PLS	Mean
timetomat	0.37	1.00	1.00	0.79
betaRMW	0.56	0.45	0.89	0.63
meangrade	0.48	0.44	0.80	0.57
betaMktRF	0.17	0.50	0.96	0.54
MKTbond	1.00	0.03	0.36	0.46
VaR	0.23	0.40	0.76	0.46
betaMom	0.11	0.41	0.82	0.45
betaSMB	0.07	0.36	0.77	0.40
betaterm	0.04	0.37	0.64	0.35
betaHML	0.08	0.29	0.64	0.34
betadef	0.08	0.29	0.58	0.32
maxday	0.33	0.14	0.42	0.30
mom12m	0.34	0.17	0.37	0.30
realindmom	0.39	0.12	0.30	0.27
mom36m	0.13	0.18	0.42	0.24
size	0.11	0.35	0.23	0.23
ILLIQ	0.12	0.16	0.39	0.22
mom1m	0.40	0.05	0.19	0.21
betaCMA	0.09	0.16	0.36	0.20
betaAgg.Liq.	0.17	0.09	0.21	0.15
momt6t1	0.03	0.15	0.27	0.15
sd_turn	0.00	0.30	0.10	0.13
chmom	0.17	0.02	0.20	0.13
climate	0.00	0.00	0.00	0.00

7. Conclusion

As the use of Machine Learning methods is progressively growing in popularity amongst financial academics, we dedicate this study to the exploration of these methods when applied to the pricing of corporate bonds, and use a climate dummy variable to determine whether the fact that a corporate bond's issuer is climate-aligned has an impact on the pricing of this corporate bond. We first run our different models without the climate dummy variable and find that time to

maturity predictor that is most used by our ML algorithms. We also find insightful results relative to the bond-specific risk factors developed in Bai, Bali and Wen (2019), as three of these factors are amongst the most used, while the authors' liquidity risk factor is amongst the less used predictors. Both stock market betas and corporate bond market betas are also used by the ML algorithms in a recurring manner. To date, however, only one of our ML method successfully obtains positive average out-of-sample R-squared over the entire test period.

Adding a climate dummy variable worsens our results and brings all our average out-of-sample R-squared into negative territory over the test period. However, adding the climate dummy variable does provide us with interesting insights: the climate dummy variable is by far the least used variable by our ML algorithm and is barely used over hundreds of models fits. This indicates that the fact that a corporate bond is issued by a climate-aligned firm has no influence on its pricing.

This study is only an introduction to the possibilities that are provided by machine learning methods to understand the pricing of corporate bonds by market participants and to understand what differences may exist between climate-aligned firms and their traditional equivalents, as it has some important limitations. Firstly, the author's limited computational power prohibits him, to date, from applying more complex non-linear machine learning methods to this data, such as Random Forest or Neural Networks methods, which usually provide better results. These limitations in terms of computational power also restricts the author in integrating more predictors, specifically volatility, liquidity and macro-economic predictors that have been taken under account in Gu, Kelly and Xiu (2020). Finally, as aforementioned, the corporate bond pricing literature is still much less developed than the stock pricing literature, which means that there might still be many predictors that have not yet been developed by academics in this sector that could successfully predict future corporate bond returns.

References

- Atsalakis, G.S. and Valavanis, K.P., 2009. Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert Systems with applications*, 36(3), pp.5932-5941.
- Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1), pp.31-56.
- Bai, J., Bali, T.G. and Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3), pp.619-642.
- Bali, T.G., Cakici, N. and Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2), pp.427-446.
- Bao, J., Pan, J. and Wang, J., 2011. The illiquidity of corporate bonds. *The Journal of Finance*, 66(3), pp.911-946.
- Bessembinder, H., Kahle, K.M., Maxwell, W.F. and Xu, D., 2008. Measuring abnormal bond performance. *The Review of Financial Studies*, 22(10), pp.4219-4258.
- Bianchi, D., Büchner, M. and Tamoni, A., 2019. What Matters When? Time-Varying Sparsity in Expected Returns. *Time-Varying Sparsity in Expected Returns (August 17, 2019)*.
- Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A.W. and Siddique, A., 2016. Risk and risk management in the credit card industry. *Journal of Banking & Finance*, 72, pp.218-239.
- Carhart, M.M., 1997. On persistence in mutual fund performance. *The Journal of finance*, 52(1), pp.57-82.
- Chang, P.C., Liu, C.H., Lin, J.L., Fan, C.Y. and Celeste, S.P., Ng.(2009). *A neural network with a case based dynamic window for stock trading prediction. Expert Systems with Applications*, 36(3), pp.6889-6898.
- Chen, W.S. and Du, Y.K., 2009. Using neural networks and data mining techniques for the financial distress prediction model. *Expert systems with applications*, 36(2), pp.4075-4086.
- Chordia, T., Subrahmanyam, A. and Anshuman, V.R., 2001. Trading activity and expected stock returns. *Journal of financial Economics*, 59(1), pp.3-32.
- De Prado, M.L., 2018. *Advances in financial machine learning*. John Wiley & Sons.
- Dick-Nielsen, J., 2009. Liquidity biases in TRACE. *The Journal of Fixed Income*, 19(2), pp.43-55.
- Dick-Nielsen, J., 2014. How to clean enhanced TRACE data. *Available at SSRN 2337908*.
- Doeksen, B., Abraham, A., Thomas, J. and Paprzycki, M., 2005, April. Real stock trading using soft computing models. In *International Conference on Information Technology: Coding and Computing (ITCC'05)-Volume II* (Vol. 2, pp. 162-167). IEEE.
- Fama, E.F., 1991. Efficient capital markets: II. *The journal of finance*, 46(5), pp.1575-1617.
- Filkova, M., Boule, B., Frandon-Martinez, C., Giorgi, A., Giuliani, D., Meng, A. and Rado, G., 2018. Bonds and Climate Change. The State of the Market 2018. *Climate Bonds Initiative*.
- Freyberger, J., Neuhierl, A. and Weber, M., 2017. Nonparametric Dissection of the Cross-Section of Expected Stock Returns. *NBER working paper*, 23227.
- Gittleman, E. and Marks, J.M., 2006. Acceleration strategies. *SSRN Electronic Journal*.
- Giglio, S. and Xiu, D., 2017. *Inference on risk premia in the presence of omitted factors* (No. w23527). National Bureau of Economic Research.
- Gu, S., Kelly, B. and Xiu, D., 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), pp.2223-2273.

- Harvey, C.R. and Liu, Y., 2019. Lucky factors. *Available at SSRN 2528780*.
- Heaton, J.B., Polson, N.G. and Witte, J.H., 2017. Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1), pp.3-12.
- Henrique, B.M., Sobreiro, V.A. and Kimura, H., 2019. Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124, pp.226-251.
- Huang, C.J., Yang, D.X. and Chuang, Y.T., 2008. Application of wrapper approach and composite classifier to the stock trend prediction. *Expert Systems with Applications*, 34(4), pp.2870-2878.
- Hutchinson, J.M., Lo, A.W. and Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, 49(3), pp.851-889.
- Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3), pp.881-898.
- Jegadeesh, N. and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1), pp.65-91.
- Kelly, B.T., Pruitt, S. and Su, Y., 2017. Instrumented principal component analysis. *Available at SSRN 2983919*.
- Kirczenow, G., Hashemi, M., Fathi, A. and Davison, M., 2018. Machine Learning for Yield Curve Feature Extraction: Application to Illiquid Corporate Bonds. *arXiv preprint arXiv:1812.01102*.
- Khandani, A.E., Kim, A.J. and Lo, A.W., 2010. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), pp.2767-2787.
- Kozak, S., Nagel, S. and Santosh, S., 2017. *Shrinking the cross section* (No. w24070). National Bureau of Economic Research.
- Malkiel, B.G., 2003. The efficient market hypothesis and its critics. *Journal of economic perspectives*, 17(1), pp.59-82.
- Malkiel, B.G. and Fama, E.F., 1970. Efficient capital markets. *A review of theory and empirical work Journal of Finance* 25, 383, 417.
- Moritz, B. and Zimmermann, T., 2016. Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Available at SSRN 2740751*.
- Moskowitz, T.J. and Grinblatt, M., 1999. Do industries explain momentum?. *The Journal of finance*, 54(4), pp.1249-1290.
- de Oliveira, F.A., Nobre, C.N. and Zarate, L.E., 2013. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, 40(18), pp.7596-7606.
- Pástor, L. and Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3), pp.642-685.
- Patel, J., Shah, S., Thakkar, P. and Kotecha, K., 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, 42(1), pp.259-268.
- Rapach, D.E., Strauss, J.K. and Zhou, G., 2013. International stock return predictability: what is the role of the United States?. *The Journal of Finance*, 68(4), pp.1633-1662.
- Sarig, O. and Warga, A., 1989. Bond price data and bond market liquidity. *Journal of Financial and Quantitative Analysis*, pp.367-378.
- Sirignano, J., Sadhwani, A. and Giesecke, K., 2016. Deep learning for mortgage risk. *arXiv preprint arXiv:1607.02470*.
- Wright, D., Capriotti, L. and Lee, J., 2018. Machine learning and corporate bond trading. *Algorithmic Finance*, 7(3-4), pp.105-110.

Appendix 1 - Model Monthly Performances

Month	OLS4 - Roos	OLS - Roos	Elastic Net - Roos	PCR - Roos	PLS - Roos	Glm - Roos
2012-02-01	0.000	-0.016	-0.007	0.000	-0.011	-0.016
2012-03-01	-0.124	-0.141	-0.075	-0.015	-0.145	-0.141
2012-04-01	-0.025	-0.078	-0.036	-0.042	-0.065	-0.078
2012-05-01	-0.013	-0.025	0.000	0.040	-0.046	-0.025
2012-06-01	0.030	0.045	0.046	0.018	0.037	0.045
2012-07-01	-0.001	0.032	0.012	0.026	0.045	0.032
2012-08-01	-0.044	-0.108	-0.032	0.040	-0.051	-0.108
2012-09-01	-0.007	0.001	0.006	-0.002	0.005	0.001
2012-10-01	0.010	-0.024	-0.016	-0.014	-0.034	-0.024
2012-11-01	0.022	0.039	0.029	0.028	0.036	0.039
2012-12-01	0.036	0.021	0.022	0.023	0.012	0.021
2013-01-01	-0.101	-0.107	-0.048	-0.005	-0.066	-0.107
2013-02-01	0.026	-0.023	0.004	0.026	0.027	-0.023
2013-03-01	-0.046	-0.106	-0.033	0.018	-0.017	-0.106
2013-04-01	-0.107	-0.031	-0.072	0.004	0.018	-0.031
2013-05-01	0.256	0.318	0.200	0.006	0.062	0.318
2013-06-01	-0.397	-0.406	-0.277	0.005	-0.343	-0.406
2013-07-01	-0.029	-0.044	-0.006	0.038	-0.007	-0.044
2013-08-01	-0.123	-0.296	-0.158	-0.044	-0.260	-0.296
2013-09-01	0.020	-0.087	-0.047	-0.047	-0.083	-0.087
2013-10-01	-0.078	0.038	0.015	0.062	0.078	0.038
2013-11-01	0.008	-0.167	-0.044	0.010	-0.014	-0.167
2013-12-01	-0.062	-0.136	-0.078	-0.123	-0.174	-0.136
2014-01-01	0.089	0.033	0.049	0.002	-0.003	0.033
2014-02-01	-0.002	-0.044	-0.032	-0.011	-0.039	-0.044
2014-03-01	-0.061	-0.097	-0.010	-0.017	-0.058	-0.097
2014-04-01	0.052	0.005	0.061	0.004	-0.013	0.005
2014-05-01	0.021	-0.016	-0.012	0.019	0.026	-0.016
2014-06-01	0.014	0.009	-0.034	-0.031	-0.046	0.009
2014-07-01	-0.121	-0.143	-0.024	0.011	-0.030	-0.143
2014-08-01	-0.129	-0.100	-0.148	-0.039	-0.042	-0.100
2014-09-01	-0.241	-0.346	-0.113	-0.003	-0.045	-0.346
2014-10-01	-0.007	-0.023	-0.025	-0.008	-0.008	-0.023
2014-11-01	-0.027	-0.071	-0.034	-0.006	-0.041	-0.071
2014-12-01	-0.020	-0.098	-0.001	0.007	-0.029	-0.098
2015-01-01	-0.073	0.011	-0.042	-0.009	0.027	0.011
2015-02-01	-0.046	-0.109	-0.026	-0.032	-0.041	-0.109
2015-03-01	0.041	-0.061	-0.031	-0.007	0.030	-0.061
2015-04-01	0.108	0.106	0.002	-0.037	0.006	0.106
2015-05-01	0.149	0.154	-0.034	-0.062	-0.020	0.154
2015-06-01	-0.064	-0.047	-0.050	-0.037	-0.072	-0.047
2015-07-01	-0.023	0.007	-0.131	-0.060	-0.088	0.007
2015-08-01	-0.032	-0.115	-0.036	-0.041	-0.059	-0.115
2015-09-01	-0.002	-0.054	0.015	0.013	0.020	-0.054
2015-10-01	-0.060	0.025	-0.146	-0.032	-0.051	0.025
2015-11-01	0.024	0.135	-0.030	-0.021	-0.033	0.135
2015-12-01	-0.017	0.033	0.026	-0.020	-0.018	0.033
2016-01-01	-0.031	-0.045	-0.012	-0.006	-0.014	-0.045
2016-02-01	0.096	0.058	-0.061	0.005	-0.022	0.058
2016-03-01	0.188	0.090	0.059	0.040	0.040	0.090
2016-04-01	0.001	-0.011	-0.002	0.011	0.013	-0.011
2016-05-01	-0.048	-0.088	0.070	0.045	0.041	-0.088
2016-06-01	0.126	0.111	0.114	0.039	0.054	0.111
2016-07-01	0.076	0.109	0.120	0.061	0.081	0.109
2016-08-01	0.039	0.020	-0.100	-0.029	-0.023	0.020
2016-09-01	0.103	0.069	-0.038	-0.041	-0.022	0.069
2016-10-01	0.166	0.124	0.022	-0.031	-0.011	0.124
2016-11-01	-0.264	-0.233	-0.055	0.031	0.045	-0.233
2016-12-01	0.014	0.001	0.014	0.015	0.015	0.001

2017-01-01	-0.030	-0.252	0.035	0.007	0.042	-0.252
2017-02-01	-0.073	0.034	-0.067	-0.035	-0.074	0.034
2017-03-01	-0.139	-0.489	-0.016	-0.002	0.016	-0.489
2017-04-01	0.021	-0.150	0.034	-0.008	0.030	-0.150
2017-05-01	-0.021	-0.072	0.014	-0.015	-0.028	-0.072
2017-06-01	-0.017	-0.058	-0.018	0.006	-0.001	-0.058
2017-07-01	-0.038	-0.057	0.016	-0.013	-0.028	-0.057
2017-08-01	0.068	0.005	-0.001	0.032	0.058	0.005
2017-09-01	0.036	-0.010	0.009	0.023	0.038	-0.010
2017-10-01	0.044	0.020	0.028	0.022	0.037	0.020
2017-11-01	-0.006	-0.060	0.054	0.034	0.056	-0.060
2017-12-01	0.045	0.076	0.040	0.030	0.054	0.076
2018-01-01	0.185	0.262	0.030	-0.010	-0.001	0.262
2018-02-01	0.040	0.071	0.015	-0.020	-0.040	0.071
2018-03-01	0.117	0.142	0.086	0.048	0.076	0.142
2018-04-01	-0.065	-0.213	-0.024	0.000	-0.014	-0.213
2018-05-01	0.046	0.152	0.011	0.024	-0.009	0.152
2018-06-01	-0.048	-0.250	0.013	0.024	0.034	-0.250
2018-07-01	-0.011	-0.086	-0.024	-0.021	-0.035	-0.086
2018-08-01	0.079	0.170	-0.013	0.054	0.002	0.170
2018-09-01	0.112	0.185	0.008	0.009	-0.017	0.185
2018-10-01	0.129	0.135	0.063	0.015	0.020	0.135
2018-11-01	-0.060	-0.039	-0.045	-0.044	-0.047	-0.039
2018-12-01	0.036	0.006	0.019	-0.032	-0.005	0.006
2019-01-01	0.068	0.064	0.064	-0.004	0.051	0.064
2019-02-01	-0.108	-0.232	-0.040	-0.073	-0.011	-0.232

Appendix 2 - Monthly Monthly Performances (With Climate Dummy Variable)

Month	OLS4 - Roos	OLS - Roos	Elastic Net - Roos	PCR - Roos	PLS - Roos	Glm - Roos
2012-02-01	0.000	-0.016	-0.007	0.001	-0.013	-0.016
2012-03-01	-0.124	-0.139	-0.075	-0.016	-0.144	-0.141
2012-04-01	-0.025	-0.078	-0.036	-0.037	-0.065	-0.078
2012-05-01	-0.013	-0.024	0.000	0.038	-0.046	-0.025
2012-06-01	0.030	0.044	0.046	0.016	0.039	0.045
2012-07-01	-0.001	0.032	0.012	0.025	-0.008	0.032
2012-08-01	-0.044	-0.109	-0.032	0.037	-0.056	-0.108
2012-09-01	-0.007	0.002	0.006	-0.003	0.007	0.001
2012-10-01	0.010	-0.023	-0.016	-0.013	-0.034	-0.024
2012-11-01	0.022	0.040	0.029	0.028	0.037	0.039
2012-12-01	0.036	0.022	0.022	0.024	0.011	0.021
2013-01-01	-0.101	-0.107	-0.048	-0.007	-0.064	-0.107
2013-02-01	0.026	-0.022	0.004	0.026	0.026	-0.023
2013-03-01	-0.046	-0.106	-0.033	0.016	-0.018	-0.106
2013-04-01	-0.107	-0.031	-0.072	0.006	0.018	-0.031
2013-05-01	0.256	0.318	0.200	0.008	0.062	0.318
2013-06-01	-0.397	-0.407	-0.277	0.004	-0.345	-0.406
2013-07-01	-0.029	-0.043	-0.006	0.038	-0.008	-0.044
2013-08-01	-0.123	-0.299	-0.158	-0.045	-0.261	-0.296
2013-09-01	0.020	-0.086	-0.047	-0.048	-0.084	-0.087
2013-10-01	-0.078	0.038	0.015	0.061	0.078	0.038
2013-11-01	0.008	-0.167	-0.044	-0.003	-0.014	-0.167
2013-12-01	-0.062	-0.138	-0.078	-0.157	-0.175	-0.136
2014-01-01	0.089	0.034	0.049	-0.003	-0.004	0.033
2014-02-01	-0.002	-0.044	-0.032	-0.050	-0.039	-0.044
2014-03-01	-0.061	-0.097	-0.010	-0.064	-0.059	-0.097
2014-04-01	0.052	0.004	0.061	-0.020	-0.014	0.005
2014-05-01	0.021	-0.015	-0.012	0.022	0.026	-0.016
2014-06-01	0.014	0.009	-0.034	-0.067	-0.045	0.009
2014-07-01	-0.121	-0.143	-0.024	-0.025	-0.031	-0.143
2014-08-01	-0.129	-0.098	-0.148	-0.032	-0.040	-0.100
2014-09-01	-0.241	-0.346	-0.113	-0.021	-0.046	-0.346
2014-10-01	-0.007	-0.025	-0.025	-0.009	-0.008	-0.023

2014-11-01	-0.027	-0.072	-0.034	-0.007	-0.041	-0.071
2014-12-01	-0.020	-0.100	-0.001	0.005	-0.030	-0.098
2015-01-01	-0.073	0.012	-0.042	-0.007	0.028	0.011
2015-02-01	-0.046	-0.110	-0.026	-0.032	-0.041	-0.109
2015-03-01	0.041	-0.061	-0.031	-0.007	0.031	-0.061
2015-04-01	0.108	0.107	0.002	-0.036	0.007	0.106
2015-05-01	0.149	0.154	-0.034	-0.060	-0.018	0.154
2015-06-01	-0.064	-0.048	-0.050	-0.035	-0.071	-0.047
2015-07-01	-0.023	0.007	-0.131	-0.059	-0.087	0.007
2015-08-01	-0.032	-0.118	-0.036	-0.033	-0.059	-0.115
2015-09-01	-0.002	-0.052	0.015	0.009	0.019	-0.054
2015-10-01	-0.060	0.023	-0.146	-0.027	-0.050	0.025
2015-11-01	0.024	0.135	-0.030	-0.020	-0.033	0.135
2015-12-01	-0.017	0.032	0.026	-0.020	-0.018	0.033
2016-01-01	-0.031	-0.044	-0.012	-0.007	-0.015	-0.045
2016-02-01	0.096	0.059	-0.061	0.005	-0.022	0.058
2016-03-01	0.188	0.094	0.059	0.042	0.040	0.090
2016-04-01	0.001	-0.010	-0.002	0.009	0.012	-0.011
2016-05-01	-0.048	-0.088	0.070	0.045	0.041	-0.088
2016-06-01	0.126	0.111	0.114	0.039	0.054	0.111
2016-07-01	0.076	0.110	0.121	0.061	0.081	0.109
2016-08-01	0.039	0.022	-0.099	-0.029	-0.023	0.020
2016-09-01	0.103	0.071	-0.038	-0.041	-0.055	0.069
2016-10-01	0.166	0.124	0.022	-0.032	-0.021	0.124
2016-11-01	-0.264	-0.232	-0.054	0.032	0.047	-0.233
2016-12-01	0.014	0.001	0.013	0.015	0.014	0.001
2017-01-01	-0.030	-0.253	0.036	0.007	0.044	-0.252
2017-02-01	-0.073	0.034	-0.069	-0.036	-0.078	0.034
2017-03-01	-0.139	-0.486	-0.013	-0.003	0.019	-0.489
2017-04-01	0.021	-0.153	0.033	-0.008	0.031	-0.150
2017-05-01	-0.021	-0.072	0.014	-0.016	-0.028	-0.072
2017-06-01	-0.017	-0.058	-0.017	0.007	0.000	-0.058
2017-07-01	-0.038	-0.058	0.016	-0.014	-0.029	-0.057
2017-08-01	0.068	0.006	-0.001	0.032	0.059	0.005
2017-09-01	0.036	-0.010	0.011	0.022	0.039	-0.010
2017-10-01	0.044	0.020	0.027	0.021	0.036	0.020
2017-11-01	-0.006	-0.060	0.054	0.033	0.057	-0.060
2017-12-01	0.045	0.079	0.040	0.030	0.055	0.076
2018-01-01	0.185	0.261	0.030	-0.009	-0.002	0.262
2018-02-01	0.040	0.071	0.014	-0.019	-0.040	0.071
2018-03-01	0.117	0.143	0.086	0.045	0.076	0.142
2018-04-01	-0.065	-0.214	-0.025	0.001	-0.014	-0.213
2018-05-01	0.046	0.152	0.009	0.023	-0.010	0.152
2018-06-01	-0.048	-0.249	0.015	0.025	0.038	-0.250
2018-07-01	-0.011	-0.087	-0.022	-0.009	-0.035	-0.086
2018-08-01	0.079	0.170	-0.011	0.047	0.003	0.170
2018-09-01	0.112	0.185	0.008	0.034	-0.018	0.185
2018-10-01	0.129	0.135	0.063	0.027	0.020	0.135
2018-11-01	-0.060	-0.040	-0.046	-0.029	-0.047	-0.039
2018-12-01	0.036	0.004	0.029	-0.037	-0.005	0.006
2019-01-01	0.068	0.066	0.065	-0.006	0.050	0.064
2019-02-01	-0.108	-0.234	-0.040	-0.082	-0.009	-0.232

General Conclusion

Our PhD dissertation brings with it a series of different concluding remarks. Our work focusing on defining sustainable finance brings us to the conclusion that sustainable finance has no common framework and definition to date, and that the necessary regulation that could lead to true understanding and transparency on the environmental and social impacts of large corporations and their investors has only just begun. The European Union has taken a lead on an issue that will incrementally become a necessity for the proper functioning of the global economy: how can the massive investment potential of capital markets and corporations help mitigate and adapt to climate change to prevent a worldwide environmental and social crisis. But the EU's framework and taxonomy brings with it many important underlying questions. Will asset managers and corporations' submit to this new approach or will they prefer sustaining the EU's sanctions? Will other nations join the EU in creating such as framework? What will be the choices made by the US and China, amongst others? Will the EU's framework be sufficient? In essence, this initial work's conclusion represents an introduction to a new set of issues that will need to be studied thoroughly in the years to come.

Following this first set of new challenges, looking at the financial relationship between corporations and capital markets also provides interesting insights that give rise to all-the-more interesting questions. We confirm through our own study what has been discovered in the literature: stock markets are not financing large corporations, but large corporations are financing stock markets. We find similar results for climate-aligned firms. Is the fact that stock markets are taking cash out of large corporations normal? Should it be expected when trillions of dollars are needed to invest in climate mitigation and climate adaptation projects? If there are important challenges that prevent corporations from using this cash to invest in climate-mitigation, what are

they? How is the cash acquired by stockholders used? Is this cash used to invest in the economy? Is it invested sustainably? How are large cash-rich corporations spending this extra capital? Are they channeling it in climate-aligned economic activities and projects? Are they investing properly in innovation, at a time when innovation might be most precious to face global climate risk? We also find that the corporate bond is a positive source of financing for large firms, but how is this source of finance used by firms? Why did the study of stocks attract more attention than the study of corporate bonds when corporate bonds market are larger and participate much more in financing large firms? Much like our first article, our work on capital markets represents only a series of answers in a set questions that increasingly need to be addressed.

Finally, focusing on developing a dataset for the corporate bond market, with a specific focus on green and climate bonds allows for better understanding of data availability for each market. While green bonds have drawn a lot of attention from financial market practitioners in recent years, studying the precise data for this market makes us understand that robust studies of these products are limited given the small size of this market, both in terms of issues and transaction volumes. The climate-bond market, which is represents a much larger set of different corporate bonds that have been issued for decades, presents a more interesting opportunity to study the potential differences between climate-aligned products and their traditional equivalents. Studying climate-bond data does also make us realize that the current methodology to determine climate-aligned firms is still limited, as only economic activities that are naturally beneficial to climate mitigation are considered, while all other activities in which concrete change need to occur are considered given the lack of available data.

The last and central part of our PhD dissertation focuses on asset pricing and was developed based on the idea that sustainability will inevitably enter the asset pricing literature, which has had

so much impact on the functioning of modern financial markets and economies. One of our first conclusion results from our choice to focus on corporate bonds: this market has barely been explored by academics in asset pricing compared to stock markets. Through the review of the general literature in asset pricing and factor models, we also conclude that sustainability, to date, has not at all entered this literature, with only a few factors being ESG related. Linking this work to our initial discoveries of our first article, we understand that there is a low probability that ESG factors, especially environmental factors, are not currently affecting the decision making process of capital market actors, but that their absence is due to a lack of transparent data. Looking at traditional asset pricing methodologies, and by trying to develop our own competences leads us to another the conclusion concerning the fact that these methodologies have a series of limitations that might make them partially obsolete for the study of modern financial markets.

Going deeper into our understanding and application of machine learning methods also provided its part of interesting insights on the complexities and requirements of these methods. Our main conclusion on this topic – apart from the fact that machine learning methods are extremely complex to implement alone – is that the general machine learning environment has a lot to offer to research in financial economics. These methods have been conceptualized for big data analysis, the scales of which traditional asset pricing studies have never truly reached. With the correct machines set up, the traditional approach of studying monthly portfolio prices could well soon be replaced by studying the daily prices of individual assets. Machine learning methods allow access to powerful tools to prevent overfit and false discoveries at a time when the validity of the statistical robustness of traditional financial models and research is being questioned. The machine learning environment, which continues to develop itself in time, allows for wider and

deeper approaches to data, proposing vast arrays of different models and model specifications, but also for deeper, more powerful algorithms to explore the complexities of financial markets.

The overall conclusion of this work, much like it must be the case for most PhD dissertations, is that much more has yet to be done on these different topics. Today, the fundamental purpose of financial economics is being questioned through the growing need for more sustainable financial markets and economies. Conjointly, the underlying techniques and statistical methodologies that have been used for decades by both practitioners and academics that compose these markets and economies are also being challenged. In a way, most of the previous work that has been performed to construct our modern financial and economic paradigms need to be re-explored through both a sustainability and a machine learning lens. This PhD dissertation represents the author's attempt to start exploring these new paradigms.

Bibliography

- Acharya, V.; Y. Amihud; and S. T. Bharath. "Liquidity Risk of Corporate Bond Returns: Conditional Approach." *Journal of Financial Economics*, 110 (2013), 358–386.
- Aghion, P., Bechtold, S., Cassar, L. and Herz, H., 2014. The causal effects of competition on innovation: Experimental evidence (No. w19987). National Bureau of Economic Research.
- Ahmad, E., 2014. Public finance underpinnings for infrastructure financing in developing countries. Paper for the G, 24.
- Alexander, G. J., A. K. Edwards, and M. G. Ferri. 2000. The determinants of the trading volume of high yield corporate bonds. *Journal of Financial Markets* 3:177–204.
- Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1), pp.31-56.
- Ang, A.; R. J. Hodrick; Y. Xing; and X. Zhang. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance*, 61 (2006), 259–299.
- Anton, W.R.Q., Deltas, G. and Khanna, M., 2004. Incentives for environmental self-regulation and implications for environmental performance. *Journal of environmental economics and management*, 48(1), pp.632-654.
- Atsalakis, G.S. and Valavanis, K.P., 2009. Surveying stock market forecasting techniques—Part II: Soft computing methods. *Expert Systems with applications*, 36(3), pp.5932-5941.
- Bachelet, M. J., Becchetti, L., & Manfredonia, S. (2019). The green bonds premium puzzle: The role of issuer characteristics and third-party verification. *Sustainability*, 11(4), 1098.
- Bai, J., Bali, T.G. and Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3), pp.619-642.
- Baldwin, C.Y., 1991. How capital budgeting deters innovation—and what to do about it. *Research-Technology Management*, 34(6), pp.39-45.
- Bali, T.G., Cakici, N. and Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2), pp.427-446.
- Bao, J., and K. Hou. "Comovement of Corporate Bonds and Equities." Working Paper, Ohio State University (2013).
- Bao, J., Pan, J. and Wang, J., 2011. The illiquidity of corporate bonds. *The Journal of Finance*, 66(3), pp.911-946.
- Barberis, N.; A. Shleifer; and R. Vishny. "A Model of Investor Sentiment." *Journal of Financial Economics*, 49 (1998), 307–343.
- Barney, J., 1991. Firm resources and sustained competitive advantage. *Journal of management*, 17(1), pp.99-120.
- BCG, 2018. "Global Asset Management, 2018. The Digital Metamorphosis."
- Berg, F., Koelbel, J.F. and Rigobon, R., 2020. Aggregate confusion: the divergence of ESG ratings. *Available at SSRN 3438533*.
- Bernard, V., and J. Thomas. "Evidence That Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings." *Journal of Accounting and Economics*, 13 (1990), 305–340.
- Bernard, V., and J. Thomas. "Post-Earnings-Announcement Drift: Delayed Price Response or Characteristic Premium?" *Journal of Accounting Research*, 27 (1989), 1–36.
- Bessembinder, H., and W. Maxwell. 2008. Transparency and the corporate bond market. *Journal of Economic Perspectives* 22:217–34.

- Bessembinder, H., Kahle, K.M., Maxwell, W.F. and Xu, D., 2008. Measuring abnormal bond performance. *The Review of Financial Studies*, 22(10), pp.4219-4258.
- Bhattacharyya, A. and Cummings, L., 2015. Measuring corporate environmental performance—stakeholder engagement evaluation. *Business Strategy and the Environment*, 24(5), pp.309-325.
- Bianchi, D., Büchner, M. and Tamoni, A., 2019. What Matters When? Time-Varying Sparsity in Expected Returns. *Time-Varying Sparsity in Expected Returns* (August 17, 2019).
- Blume, M. E., D. B. Keim, and S. A. Patel. 1991. Returns and volatility of low-grade bonds 1977-1989. *Journal of Finance* 46:49–74.
- Brammer, S., Brooks, C. and Pavelin, S., 2006. Corporate social performance and stock returns: UK evidence from disaggregate measures. *Financial management*, 35(3), pp.97-116.
- Brooks, C. and Oikonomou, I., 2018. The effects of environmental, social and governance disclosures and performance on firm value: A review of the literature in accounting and finance. *The British Accounting Review*, 50(1), pp.1-15.
- Buchner, B., Herve-Mignucci, M., Trabacchi, C., Wilkinson, J., Stadelmann, M., Boyd, R., Mazza, F., Falconer, A. and Micale, V., 2013. Global landscape of climate finance 2015. Climate Policy Initiative.
- Busch, T. and Friede, G., 2018. The robustness of the corporate social and financial performance relation: A second-order meta-analysis. *Corporate Social Responsibility and Environmental Management*, 25(4), pp.583-608.
- Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A.W. and Siddique, A., 2016. Risk and risk management in the credit card industry. *Journal of Banking & Finance*, 72, pp.218-239.
- Cai, N., Helwege, J., Warga, A., 2007. Underwriting in the corporate bond market. *Review of Financial Studies* 20, 2021–2046.
- Campbell, J.Y. and G.B. Taksler, 2003, "Equity Volatility and Corporate Bond Yields," *Journal of Finance* 58, 2321–2349.
- Campbell, J.Y., 1996. Understanding risk and return. *Journal of Political economy*, 104(2), pp.298-345.
- Carhart, M.M., 1997. On persistence in mutual fund performance. *The Journal of finance*, 52(1), pp.57-82.
- CEA, 2016. Benefits of competition and indicators of market power.
- Chang, P.C., Liu, C.H., Lin, J.L., Fan, C.Y. and Celeste, S.P., Ng.(2009). A neural network with a case based dynamic window for stock trading prediction. *Expert Systems with Applications*, 36(3), pp.6889-6898.
- Chen, L., D. A. Lesmond, and J. Wei. 2007. Corporate yield spreads and bond liquidity. *Journal of Finance* 62:119–49.
- Chen, W.S. and Du, Y.K., 2009. Using neural networks and data mining techniques for the financial distress prediction model. *Expert systems with applications*, 36(2), pp.4075-4086.
- Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A., Tong, Q., 2017. Are capital market anomalies common to equity and corporate bond markets? *Journal of Financial and Quantitative Analysis* 52, 1301–1342.
- Chordia, T., Subrahmanyam, A. and Anshuman, V.R., 2001. Trading activity and expected stock returns. *Journal of financial Economics*, 59(1), pp.3-32.
- Christensen, C.M., Kaufman, S.P. and Shih, W.C., 2008. Innovation killers: how financial tools destroy your capacity to do new things. *Harvard business review*, 86(1), pp.98-105.
- Christensen, D., Serafeim, G. and Sikochi, A., 2019. Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings.
- Clarkson, Max E. "A stakeholder framework for analyzing and evaluating corporate social performance." *Academy of management review* 20.1 (1995): 92-117.

- Climate Bonds Initiative, 2018. Bonds and Climate Change: The State of the Market.
- Cooper, M.; H. Gulen; and M. Schill. "Asset Growth and the Cross-Section of Stock Returns." *Journal of Finance*, 63 (2008), 1609–1651.
- Council of the European Union, 2019. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the establishment of a framework to facilitate sustainable investment - Approval of the final compromise text
- Cremers, M., V. Nair, and C. Wei. 2007. Governance Mechanisms and Bond Prices. *Review of Financial Studies* 20:1359–88.
- Daniel, K.; D. Hirshleifer; and A. Subrahmanyam. "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance*, 53 (1998), 1839–1885.
- Dass, N. and M. Massa, 2014, "The Variety of Maturities Offered by Firms and Institutional Investment in Corporate Bonds," *Review of Financial Studies* 27, 2219–2266.
- De Jong, F., Driessen, J., 2007. Liquidity risk premia in corporate bond and equity markets. *Management Science* 53, 1439–1451.
- de Oliveira, F.A., Nobre, C.N. and Zarate, L.E., 2013. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, 40(18), pp.7596-7606.
- De Prado, M.L., 2018. *Advances in financial machine learning*. John Wiley & Sons.
- DeCosta, D., F. Leng, and G. Noronha, 2013, "Minimum Maturity Rules: The Cost of Selling Bonds Before Their Time," *Financial Analysts Journal* 69, 45–56.
- Delmas, M.A., Etzion, D. and Nairn-Birch, N., 2013. Triangulating environmental performance: What do corporate social responsibility ratings really capture?. *Academy of Management Perspectives*, 27(3), pp.255-267.
- Dichev, I. D., and J. D. Piotroski. 2001. The long-run stock returns following bond rating changes. *Journal of Finance* 56:55–84.
- Dick-Nielsen, J., 2009. Liquidity biases in TRACE. *The Journal of Fixed Income*, 19(2), pp.43-55.
- Dick-Nielsen, J., 2014. How to clean enhanced TRACE data. Available at SSRN 2337908.
- Dick-Nielsen, J., Feldhütter, P. and Lando, D., 2012. Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics*, 103(3), pp.471-492.
- Diebecker, J., Rose, C. and Sommer, F., 2019. Spoiled for Choice: Does the Selection of Sustainability Datasets Matter?. Available at SSRN 3359508.
- Doeksen, B., Abraham, A., Thomas, J. and Paprzycki, M., 2005, April. Real stock trading using soft computing models. In *International Conference on Information Technology: Coding and Computing (ITCC'05)-Volume II (Vol. 2, pp. 162-167)*. IEEE.
- Donaldson, T. and Preston, L.E., 1995. The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of management Review*, 20(1), pp.65-91.
- Doyle, T., 2018. Ratings that Don't Rate: The Subjective World of ESG Ratings Agencies. Report, American Council for Capital Formation.
- EBA, 2019. "EBA Report on Undue Short-term pressure from the financial sector on corporations".
- Eberhart, A. C., and A. Siddique. 2002. The long-term performance of corporate bonds (and stocks) following seasoned equity offerings. *Review of Financial Studies* 15:1385–406.^[1]_[SEP]
- Edwards, A., L. E. Harris, and M. S. Piwowar. 2007. Corporate bond market transaction costs and transparency. *Journal of Finance* 62:1421–51.^[1]_[SEP]

EIOPA, 2019. “Potential undue short-term pressure from financial markets on corporates: Investigation on European insurance and occupational pension sectors.”

Ellul, A., C. Jotikasthira, and C.T. Lundblad, 2011, “Regulatory Pressure and Fire Sales in the Corporate Bond Market,” *Journal of Financial Economics* 101, 596-620.^[1]_{SEP}

Escrig-Olmedo, E., Muñoz-Torres, M.J., Fernández-Izquierdo, M.Á. and Rivera-Lirio, J.M., 2017. Measuring corporate environmental performance: A methodology for sustainable development. *Business Strategy and the Environment*, 26(2), pp.142-162.

ESMA, 2019. “Final Report. Guidelines on Disclosure Requirements Applicable to Credit Ratings.”

ESMA, 2019b. “Report. Undue short-term pressure on corporations.”

EU Technical Expert Group on Sustainable Finance, 2020. “Taxonomy: Final report of the Technical Expert Group on Sustainable Finance.”

European Commission, 2018. Action Plan: Financing Sustainable Growth.

European Commission 2020d. “Study on due diligence requirements through the supply chain. Final Report.”

European Commission, 2018. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN CENTRAL BANK, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. Action Plan: Financing Sustainable Growth

European Commission, 2019a. “United in delivering the Energy Union and Climate Action - Setting the foundations for a successful clean energy transition”

European Commission, 2019b. “Communication on the European Green Deal”

European Commission, 2019c. “TEG Report. Proposal for an EU Green Bond Standard”

European Commission, 2020a. “Commission Communication on the Sustainable Europe Investment Plan”

European Commission, 2020b. “Consultation Document. Consultation on the Renewed Sustainable Finance Strategy.”

European Commission, 2020c. “Taxonomy: Final report of the Technical Expert Group on Sustainable Finance”. March 2020.

European Commission, 2020d. “Usability Guide, TEG Proposal for a EU Green Bond Standard”.

Eurosif, 2018. “European SRI Study 2018.”

Fairfield, P. M.; J. S. Whisenant; and T. L. Yohn. “Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing.” *Accounting Review*, 78 (2003), 353–371.

Fama, E. F., and K. R. French. “Business Conditions and Expected Returns on Stocks and Bonds.” *Journal of Financial Economics*, 25 (1989), 23–49.

Fama, E. F., and K. R. French. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics*, 33 (1993), 23–49.

Fama, E. F., and K. R. French. “Dissecting Anomalies.” *Journal of Finance*, 63 (2008), 1653–1678.

Fama, E., French, K., 2012. Size, value, and momentum in international stock returns. *Journal of Financial Economics* 105, 457-472.

Fama, E.F. and French, K.R., 2015. A five-factor asset pricing model. *Journal of financial economics*, 116(1), pp.1-22.

- Fama, E.F. and MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3), pp.607-636.
- Fama, E.F., 1991. Efficient capital markets: II. *The journal of finance*, 46(5), pp.1575-1617.
- Fama, Eugene F. 1996. Multifactor portfolio efficiency and multifactor asset pricing. *Journal of Financial and Quantitative Analysis* 31, 441-465.
- Fama, Eugene F. 1998. Determining the number of priced state variables in the ICAPM. *Journal of Financial and Quantitative Analysis* 33, 217-231.
- Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.
- Fama, Eugene F., and Kenneth R. French, 2015. A five-factor asset pricing model, *Journal of Financial Economics* 116, 1-22.
- Fama, Eugene F., and Kenneth R. French, 2016a. Dissecting anomalies with a five-factor model, *Review of Financial Studies*, 29,69-103.
- Fama, Eugene F., and Kenneth R. French, 2016b. International tests of a five-factor model, manuscript, Tuck School of Business, Dartmouth College.
- Fichtner, J., Heemskerk, E.M. and Garcia-Bernardo, J., 2017. Hidden power of the Big Three? Passive index funds, re-concentration of corporate ownership, and new financial risk. *Business and Politics*, 19(2), pp.298-326.
- Filkova, M., Bouille, B., Frandon-Martinez, C., Giorgi, A., Giuliani, D., Meng, A. and Rado, G., 2018. Bonds and Climate Change. *The State of the Market 2018. Climate Bonds Initiative*.
- Flammer, C., 2018. Corporate green bonds.
- Freyberger, J., Neuhierl, A. and Weber, M., 2017. Nonparametric Dissection of the Cross-Section of Expected Stock Returns. NBER working paper, 23227.
- Friede, G., Busch, T. and Bassen, A., 2015. ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), pp.210-233.
- Friedman, M., 1970. September 13. The social responsibility of business is to increase its profits. *New York Times Magazine*, 33, p.126.
- Ge, W. and Liu, M., 2015. Corporate social responsibility and the cost of corporate bonds. *Journal of Accounting and Public Policy*, 34(6), pp.597-624.
- Gebhardt, W. R., S. Hvidkjaer, and B. Swaminathan. 2005a. The cross-section of expected corporate bond returns: Betas or characteristics? *Journal of Financial Economics* 75:85–114.
- Gettleman, E. and Marks, J.M., 2006. Acceleration strategies. *SSRN Electronic Journal*.
- Giglio, S. and Xiu, D., 2017. Inference on risk premia in the presence of omitted factors (No. w23527). National Bureau of Economic Research.
- Goldstein, Michael A., Edith S. Hotchkiss, and Erik R. Sirri, 2007, Transparency and liquidity: A controlled experiment on corporate bonds, *Review of Financial Studies* 20, 235-273.
- Gómez, J.P., Priestley, R. and Zapatero, F., 2016. Labor income, relative wealth concerns, and the cross section of stock returns. *Journal of Financial and Quantitative Analysis*, 51(4), pp.1111-1133.
- Griffin, P., 2017. CDP carbon majors report 2017. Carbon Majors Database.
- GSIA, 2013. "Global Sustainable Investment Review 2012."

GSIA, 2018. “Global Sustainable Investment Review 2018.”

Gu, S., Kelly, B. and Xiu, D., 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), pp.2223-2273.

Gutiérrez, G. and Philippon, T., 2016. Investment-less growth: An empirical investigation (No. w22897). National Bureau of Economic Research.

Gutiérrez, G. and Philippon, T., 2017. Declining Competition and Investment in the US (No. w23583). National Bureau of Economic Research.

Hachenberg, B. and Schiereck, D., 2018. Are green bonds priced differently from conventional bonds?. *Journal of Asset Management*, 19(6), pp.371-383.

Hand, J. R. M., R. W. Holthausen, and R. W. Leftwich. 1992. The effect of bond rating agency announcements on bond and stock prices. *Journal of Finance* 47:733–52.

Harford, J., 1999. Corporate cash reserves and acquisitions. *The Journal of Finance*, 54(6), pp.1969-1997.

Harford, J., Mansi, S.A. and Maxwell, W.F., 2008. Corporate governance and firm cash holdings in the US. *Journal of financial economics*, 87(3), pp.535-555.

Harris, L. E., and M. S. Piwowar. 2006. Secondary trading costs in the municipal bond market. *Journal of Finance* 61:1361–97.

Hartmann, J. and Vachon, S., 2018. Linking environmental management to environmental performance: The interactive role of industry context. *Business Strategy and the Environment*, 27(3), pp.359-374.

Harvey, C.R. and Liu, Y., 2019. Lucky factors. Available at SSRN 2528780.

Harvey, C.R., Liu, Y. and Zhu, H., 2016. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), pp.5-68.

Heaton, J.B., Polson, N.G. and Witte, J.H., 2017. Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1), pp.3-12.

Henrique, B.M., Sobreiro, V.A. and Kimura, H., 2019. Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124, pp.226-251.

Hirshleifer, D., and S. H. Teoh. “Limited Attention, Information Disclosure, and Financial Reporting.” *Journal of Accounting and Economics*, 36 (2003), 337–386. ^[11] _[SEP]

Hirshleifer, D.; S. S. Lim; and S. H. Teoh. “Limited Investor Attention and Stock Market Misreactions to Accounting Information.” *Review of Asset Pricing Studies*, 1 (2011), 35–73. ^[11] _[SEP]

HLEG, 2018. Financing a Sustainable European Economy, Final Report 2018 by the High-Level Expert Group on Sustainable Finance.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: an investment approach. *Review of Financial Studies* 28, 650-705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2016, A comparison of new factor models, Dice Center Working Paper 2015-05, November.

Huang, C.J., Yang, D.X. and Chuang, Y.T., 2008. Application of wrapper approach and composite classifier to the stock trend prediction. *Expert Systems with Applications*, 34(4), pp.2870-2878.

Hutchinson, J.M., Lo, A.W. and Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, 49(3), pp.851-889.

Hyun Jin Lee & Insook Cho (2016) Corporate Governance and Corporate Bond Liquidity, *Global Economic Review*,45:2, 189-205

IEA, 2016. Renewable Energy Medium-Term Market Report, 2016: Market Analysis and Forecasts to 2021.

IEA, 2017. Perspectives for the Energy Transition: Investment Needs for a Low-Carbon Energy System. International Energy Agency. International Renewable Energy Agency, Paris, Bonn.

Ikenberry, D., J. Lakonishok, and T. Vermaelen. 1995. Market underreaction to open market share repurchases. *Journal of Financial Economics* 39: 181–208^[1]_{SEP}.

IPCC, 2018: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)].

Jagannathan, R. and Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. *The Journal of finance*, 51(1), pp.3-53.

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

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3), pp.881-898.

Jegadeesh, N., and S. Titman. “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” *Journal of Finance*, 48 (1993), 65–91.

Jegadeesh, N., and S. Titman. “Short-Horizon Return Reversals and the Bid–Ask Spread.” *Journal of Financial Intermediation*, 4 (1995), 116–132.

Jostova, G., Nikolova, S., Philipov, A. and Stahel, C.W., 2013. Momentum in corporate bond returns. *The Review of Financial Studies*, 26(7), pp.1649-1693.

Kapraun, J. and Scheins, C., 2019. (In)-Credibly Green: Which Bonds Trade at a Green Bond Premium?. Available at SSRN 3347337.

Kelly, B.T., Pruitt, S. and Su, Y., 2017. Instrumented principal component analysis. Available at SSRN 2983919.

Khandani, A.E., Kim, A.J. and Lo, A.W., 2010. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), pp.2767-2787.

Khang, Kenneth, and Tao-Hsien Dolly King, 2004, Return reversals in the bond market: Evidence and causes, *Journal of Banking and Finance* 28, 569-593.

Kirczenow, G., Hashemi, M., Fathi, A. and Davison, M., 2018. Machine Learning for Yield Curve Feature Extraction: Application to Illiquid Corporate Bonds. arXiv preprint arXiv:1812.01102.

Kotsantonis, S. and Serafeim, G., 2019. Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, 31(2), pp.50-58.

Kozak, S., Nagel, S. and Santosh, S., 2017. Shrinking the cross section (No. w24070). National Bureau of Economic Research.

Lazonick, W. and O'sullivan, M., 2000. Maximizing shareholder value: a new ideology for corporate governance. *Economy and society*, 29(1), pp.13-35.

Lazonick, W., 1990. Organizational capabilities in American industry: The rise and decline of managerial capitalism. *Business and Economic History*, pp.35-54.

- Lazonick, W., 2015. Stock buybacks: From retain-and-reinvest to downsize-and-distribute. Center for Effective Public Management at Brookings, 17.
- Lehmann, B. “Fads, Martingales, and Market Efficiency.” *Quarterly Journal of Economics*, 105 (1990), 1–28.
- Lev, B., and D. Nissim. “The Persistence of the Accruals Anomaly.” *Contemporary Accounting Research*, 23 (2006), 193–226. ^[1]_{SEP}
- Lin, H., Wang, J. and Wu, C., 2011. Liquidity risk and expected corporate bond returns. *Journal of Financial Economics*, 99(3), pp.628-650.
- Linklaters, 2019. Sustainable Finance. The rise of green loans and sustainability linked lending.
- Liu M, Magnan M (2014) Conditional conservatism and underpricing in US corporate bond market. *Appl Financ Econ* 24(20):1323–1334
- Liu, S., Shi, J., Wang, J., Wu, C., 2007. How much of the corporate bond spread is due to personal taxes? *Journal of Financial Economics* 85, 599–636. ^[1]_{SEP}
- Loughran, Tim, and Jay R. Ritter. 1995. The new issues puzzle. *Journal of Finance* 50: 23– 51. ^[1]_{SEP}
- Luo, H.A. and Balvers, R.J., 2017. Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis*, 52(1), pp.365-399.
- Malkiel, B.G. and Fama, E.F., 1970. Efficient capital markets. A review of theory and empirical work *Journal of Finance* 25, 383, 417.
- Malkiel, B.G., 2003. The efficient market hypothesis and its critics. *Journal of economic perspectives*, 17(1), pp.59-82.
- Margolis, J.D., Elfenbein, H.A. and Walsh, J.P., 2007. Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance. *Ann Arbor*, 1001, pp.48109-1234.
- Markowitz, H., 1959. *Portfolio selection: Efficient diversification of investments* (Vol. 16). New York: John Wiley.
- Markowitz, H.M., 1952. (1952). *Portfolio selection*. *Journal of Finance*, 7(1), pp.77-91.
- Misani, N. and Pogutz, S., 2015. Unraveling the effects of environmental outcomes and processes on financial performance: A non-linear approach. *Ecological economics*, 109, pp.150-160.
- Moritz, B. and Zimmermann, T., 2016. Tree-based conditional portfolio sorts: The relation between past and future stock returns. Available at SSRN 2740751.
- Moskowitz, T.J. and Grinblatt, M., 1999. Do industries explain momentum?. *The Journal of finance*, 54(4), pp.1249-1290.
- Nagel, S. “Evaporating Liquidity.” *Review of Financial Studies*, 25 (2012), 2005–2039. ^[1]_{SEP}
- Newey, W.K. and West, K.D., 1986. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.
- OECD (2011), *OECD Guidelines for Multinational Enterprises*, OECD Publishing. <http://dx.doi.org/10.1787/9789264115415-en>
- Official Journal of the European Union, 2014. DIRECTIVE 2014/95/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 October 2014 amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups
- Official Journal of the European Union, 2019a. REGULATION (EU) 2019/2088 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 November 2019 on sustainability-related disclosures in the financial services sector

Official Journal of the European Union, 2019b. REGULATION (EU) 2019/2089 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 November 2019 amending Regulation (EU) 2016/1011 as regards EU Climate Transition Benchmarks, EU Paris-aligned Benchmarks and sustainability-related disclosures for benchmarks

Orlitzky, M., Schmidt, F.L. and Rynes, S.L., 2003. Corporate social and financial performance: A meta-analysis. *Organization studies*, 24(3), pp.403-441.

Pástor, E. and Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3), pp.642-685.

Patel, J., Shah, S., Thakkar, P. and Kotecha, K., 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, 42(1), pp.259-268.

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1), pp.435-480.

Polbennikov, S., Desclée, A., Dynkin, L. and Maitra, A., 2016. ESG ratings and performance of corporate bonds. *The Journal of Fixed Income*, 26(1), pp.21-41.

Porter, M.E. and Kramer, M.R., 2006. The link between competitive advantage and corporate social responsibility. *Harvard business review*, 84(12), pp.78-92.

Post, C., Rahman, N. and McQuillen, C., 2015. From board composition to corporate environmental performance through sustainability-themed alliances. *Journal of Business Ethics*, 130(2), pp.423-435.

Rapach, D.E., Strauss, J.K. and Zhou, G., 2013. International stock return predictability: what is the role of the United States?. *The Journal of Finance*, 68(4), pp.1633-1662.

Ren, S., He, D., Zhang, T. and Chen, X., 2019. Symbolic reactions or substantive pro-environmental behaviour? An empirical study of corporate environmental performance under the government's environmental subsidy scheme. *Business Strategy and the Environment*, 28(6), pp.1148-1165.

Sarig, O. and Warga, A., 1989. Bond price data and bond market liquidity. *Journal of Financial and Quantitative Analysis*, 24(3), pp.367-378.

Schiereck, D., Friede, G. and Bassen, A., 2019. Financial Performances of Green Securities. In *The Rise of Green Finance in Europe* (pp. 95-117). Palgrave Macmillan, Cham.

Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3), pp.425-442.

SIFMA, 2019. *Capital Markets Fact Book 2019*.

Sirignano, J., Sadhwani, A. and Giesecke, K., 2016. Deep learning for mortgage risk. *arXiv preprint arXiv:1607.02470*.

Sloan, R. G. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" *Accounting Review*, 71 (1996), 289–315.^[1]

Stambaugh, R. F.; J. Yu; and Y. Yuan. "Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle." *Journal of Finance*, 70 (2015), 1903–1948.^[1]

Trumpp, C. and Guenther, T., 2017. Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. *Business Strategy and the Environment*, 26(1), pp.49-68.

UNEP, 2019. "Emissions Gap Report 2019."

Trumpp, C. and Guenther, T., 2017. Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. *Business Strategy and the Environment*, 26(1), pp.49-68.

- Trumpp, C., Endrikat, J., Zopf, C. and Guenther, E., 2015. Definition, conceptualization, and measurement of corporate environmental performance: A critical examination of a multidimensional construct. *Journal of Business Ethics*, 126(2), pp.185-204.
- UNFCCC, V., 2015. Adoption of the Paris agreement. I: Proposal by the President (Draft Decision), United Nations Office, Geneva (Switzerland), (s 32).
- United Nations, 2011. UN Guiding Principles on Business and Human Rights.
- United Nations, 2015. Transforming our world: The 2030 agenda for sustainable development. General Assembly 70 session.
- US SIF, 2018. "Report on US Sustainable, Responsible and Impact Investing Trends."
- Wood, D.J., 2010. Measuring corporate social performance: A review. *International journal of management reviews*, 12(1), pp.50-84.
- World Economic Forum, 2020. "Global Risks Report 2020."
- Wright, D., Capriotti, L. and Lee, J., 2018. Machine learning and corporate bond trading. *Algorithmic Finance*, 7(3-4), pp.105-110.
- Xie, S. and Hayase, K., 2007. Corporate environmental performance evaluation: a measurement model and a new concept. *Business Strategy and the Environment*, 16(2), pp.148-168.
- Xue, B., Zhang, Z. and Li, P., 2020. Corporate environmental performance, environmental management and firm risk. *Business Strategy and the Environment*, 29(3), pp.1074-1096.
- Yang, R., 2019. What Do We Learn From Ratings About Corporate Social Responsibility (CSR)?. Columbia Business School Research Paper, (18-37).
- Zerbib, O.D., 2019. The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of Banking & Finance*, 98, pp.39-60.

Table of contents (detailed)

<u>CHAPTER 1 – SUMMARY NOTE</u>	8
1. GENERAL INTRODUCTION	8
1.1. THE CLIMATE CHANGE URGENCY AND ASSET PRICING	8
1.2. IDENTIFYING CLIMATE-ALIGNED FIRMS	11
1.3. ASSET PRICING APPLIED TO CLIMATE BONDS	15
2. DEFINING SUSTAINABLE FINANCE	21
2.1. A LACK OF COMMON DEFINITION AMONGST PRACTITIONERS	21
2.2. A LACK OF COMMON DEFINITION AMONGST ACADEMICS	24
2.3. THE IMPORTANT ROLE OF REGULATORS AND THE EUROPEAN COMMISSION	26
2.4. THE FOCUS ON LARGE FIRMS AND ASSET MANAGERS	28
3. CAPITAL MARKET FINANCING AND BOND DATA FOR TRADITIONAL AND CLIMATE FIRMS	31
3.1. DIVIDENDS, STOCK REPURCHASES AND CORPORATE BOND ISSUANCES FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	32
3.2. CORPORATE BOND DATA FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	36
4. MODERN EMPIRICAL ASSET PRICING AND CLIMATE-ALIGNED FIRMS	38
4.1. FROM PREDICTING LEVELS OF RETURNS TO PREDICTING CROSS-SECTIONAL DIFFERENCES IN RETURNS	39
4.2. WHAT ABOUT CORPORATE BONDS?	41
4.3. POSSIBLE LIMITATIONS OF TRADITIONAL ASSET PRICING METHODOLOGIES	43
4.3.1. CONTROL VARIABLES	43
4.3.2. DATA MINING	44
4.3.3. OUT-OF-SAMPLE DATA	45
4.4. ADVANTAGES OF MACHINE LEARNING AND THE “ZOO OF FACTORS”	46
4.4.1. <i>Machine learning methods for prediction in asset pricing</i>	46
4.4.2. <i>New factors for corporate bond pricing</i>	48
5. LITERATURE REVIEW	49
5.1. LITERATURE ON THE SRI MARKET	50
5.2. THE ACADEMIC LITERATURE ON ESG/CSP AND CEP	50
5.3. CORPORATE FINANCING VIA CAPITAL MARKETS	53
5.4. DATA PROCESSING METHODOLOGY TO STUDY TRADITIONAL AND CLIMATE BOND RETURNS	57
5.5. THE NASCENT CLIMATE-ALIGNED BOND LITERATURE	60
5.6. ASSET PRICING AND TRADITIONAL AND CLIMATE BONDS	62
5.7. MACHINE LEARNING IN ASSET PRICING APPLIED IN THE CONTEXT OF SUSTAINABLE FINANCE	63
6. DATA	65
6.1. DATA TREATMENT METHODS FOR CAPITAL MARKET FINANCING FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	66
6.1.1. <i>Cumulative Cash Flows for the period 2009-2019</i>	67
6.1.2. <i>Average Cash Flows for US 1971-2019 and Euro 1999 - 1919</i>	68
6.1.3. <i>Corporate bonds issuances and reductions 1995 - 2018</i>	68
6.1.4. <i>US and Euro climate-aligned firms</i>	71
6.2. THE CORPORATE BOND DATA TREATMENT METHODOLOGY	72
6.3. DATA USED FOR TRADITIONAL AND CLIMATE BOND PRICING	78
6.4. OTHER TYPES OF DATA ON SUSTAINABLE FINANCE	81
7. METHODOLOGY	82
7.1. STRUCTURED ANALYSIS OF THE LITERATURE ON SUSTAINABLE FINANCE	82
7.1.1. <i>Analysis of the literature on the SRI Market</i>	82
7.1.2. <i>Analysis of the academic literature on ESG</i>	83
7.1.3. <i>Analysis of the academic literature on CEP</i>	83
7.2. THE CROSS-SECTION OF TRADITIONAL AND CLIMATE BOND RETURNS	86
7.2.1. <i>Univariate portfolio analysis</i>	86

7.2.2.	<i>Bivariate portfolio analysis</i>	89
7.2.3.	<i>Fama-Macbeth regressions</i>	89
7.2.4.	<i>Determining bond-level risk factors</i>	90
7.3.	MACHINE LEARNING AND ASSET PRICING FOR TRADITIONAL AND CLIMATE-ALIGNED FIRMS	91
7.3.1.	<i>Sample splitting and tuning via cross-validation</i>	93
8.	RESULTS	95
8.1.	RESULTS FROM THE LITERATURE ON SRI, ESG AND CEP	95
8.2.	RESULTS ON THE CAPITAL MARKET FINANCING OF TRADITIONAL AND CLIMATE-ALIGNED FIRMS	96
8.2.1.	<i>Negative stock market financing, positive long-term debt financing</i>	97
8.2.2.	<i>Corporate bonds as a major source of financing for traditional and climate-aligned firms</i>	99
8.3.	A BETTER UNDERSTANDING OF CLIMATE BOND AND GREEN BOND DATA AVAILABILITY	100
8.4.	INSIGHTFUL RESULTS FROM THE CROSS-SECTION OF TRADITIONAL AND CLIMATE BOND RETURNS	101
8.4.1.	<i>Univariate portfolio results</i>	101
8.4.2.	<i>Bivariate portfolio results</i>	104
8.4.3.	<i>Bond-level Fama-MacBeth regressions</i>	104
8.4.4.	<i>Testing the factors on Size/maturity-sorted and industry-sorted portfolios for both traditional and climate bonds</i>	108
8.5.	THE IRRELEVANCE OF THE CLIMATE FACTOR FOR MACHINE LEARNING METHODS	109
8.5.1.	<i>Models without Climate dummy variables</i>	109
8.5.2.	<i>Variable importance for models without climate dummy variable</i>	112
8.5.3.	<i>Models with a Climate dummy variable</i>	114
8.6.	CONCLUDING REMARKS ON RESULTS	115
9.	FUTURE RESEARCH	117
10.	SUMMARY CONCLUSION	121
 <u>CHAPTER 2 - THE EU TAXONOMY: A NEW SUSTAINABLE FINANCE PARADIGM?</u>		122
1.	INTRODUCTION	125
2.	THE SUSTAINABLE INVESTMENT INDUSTRY AND THE ESG APPROACH	130
2.1.	NEGATIVE/EXCLUSIONARY SCREENING	131
2.2.	ESG INTEGRATION	132
2.3.	CORPORATE ENGAGEMENT AND SHAREHOLDER ACTION	132
2.4.	NORMS-BASED SCREENING	133
2.5.	POSITIVE/BEST IN CLASS SCREENING	133
2.6.	SUSTAINABILITY THEMED INVESTING	133
2.7.	IMPACT AND COMMUNITY INVESTING	134
2.8.	GENERAL REMARKS	134
3.	CORPORATE ENVIRONMENTAL PERFORMANCE	138
3.1.	ESG AND FINANCIAL PERFORMANCE	138
3.2.	ENVIRONMENTAL PERFORMANCE AND FINANCIAL PERFORMANCE	143
3.3.	ENVIRONMENTAL AND SOCIAL DATA	145
3.4.	THE ABSENCE OF ESG FACTORS IN ASSET PRICING	149
4.	THE EU'S FRAMEWORK	151
5.	THE EU TAXONOMY	155
5.1.	ACTIVITY-SPECIFIC TECHNICAL CRITERIA	158
5.2.	ACTIVITY-SPECIFIC DNSH CRITERIA	158
5.3.	MEETING MINIMAL SOCIAL SAFEGUARDS	159
5.4.	GENERAL REMARKS	159
6.	THE TAXONOMY REGULATION	160
6.1.	FINANCIAL MARKET PARTICIPANTS	161
6.2.	COMPANIES	162

6.3.	PRESSURE ON FINANCIAL-MARKET PARTICIPANTS	164
7.	IMPLEMENTATION OF SUSTAINABILITY MEASURES IN THE REST OF THE MARKET	165
7.1.	THE BENCHMARK REGULATION	166
7.2.	EU LABELS	167
7.3.	EU INITIATIVES TO MOBILIZE PRIVATE INVESTMENTS	167
7.4.	CREDIT RATING AGENCIES	168
7.5.	INSURANCES	168
7.7.	PENSION PROVIDERS	169
7.8.	CORPORATE GOVERNANCE	169
7.9.	GENERAL REMARKS	171
8.	CONCLUSION	171

CHAPTER 3 - THE ROLE OF CAPITAL MARKETS IN FINANCING CLIMATE CHANGE MITIGATION 178

1.	INTRODUCTION	180
2.	LITERATURE REVIEW	184
3.	DATA	190
3.1.	CUMULATIVE CASH FLOWS FOR THE PERIOD 2009-2019	192
3.2.	AVERAGE CASH FLOWS FOR US 1971-2019 AND EURO 1999 - 1919	192
3.3.	CORPORATE BONDS ISSUANCES AND REDUCTIONS 1995 - 2018	193
4.	CUMULATIVE CASH FLOWS OF US AND EURO FIRMS	196
4.1.	THE S&P500 AND EURO FIRMS IN THE S&P 350 EUROPE	196
4.2.	US AND EURO CLIMATE-ALIGNED FIRMS	199
5.	AVERAGE HISTORICAL CASH FLOWS OF US AND EURO FIRMS	201
5.1.	SHARE ISSUANCES, REPURCHASES AND DIVIDENDS	201
5.2.	LONG-TERM DEBT AND OVERALL EXTERNAL FINANCING	203
6.	CORPORATE BONDS AND LONG-TERM DEBT FOR US FIRMS	207
7.	CONCLUSION	208

CHAPTER 4 - A DATA PROCESSING METHODOLOGY FOR GREEN BONDS AND CLIMATE BONDS 214

1.	INTRODUCTION	217
2.	LITERATURE REVIEW	220
2.1.	THE NASCENT CLIMATE-ALIGNED BOND LITERATURE	220
2.2.	RECENT STUDIES ON THE CORPORATE BONDS MARKET	222
2.3.	THE CORPORATE BOND DATA CHALLENGE	225
3.	DATA AND METHODOLOGY	227
3.1.	DATABASES OF THE CORPORATE BOND MARKET	227
3.2.	DATA PROCESSING METHODOLOGY	230
3.2.1.	<i>Obtaining a sample of US Green Corporate Bonds</i>	230
3.2.2.	<i>Obtaining a sample of US Climate Bonds</i>	231
3.2.3.	<i>Obtaining Enhanced TRACE data</i>	231
3.2.4.	<i>Developing the Enhanced TRACE data filter</i>	232
3.2.5.	<i>Verifying and applying the TRACE filter</i>	234
3.2.6.	<i>Transforming TRACE clean data into daily bond prices</i>	235
3.2.7.	<i>Transforming TRACE clean data into monthly bond prices</i>	236
3.2.8.	<i>Extracting Bloomberg and Datastream Data</i>	237
3.2.9.	<i>Merging databases</i>	238
3.2.10.	<i>Computing monthly returns using month-end prices</i>	239
4.	RESULTING DATASETS AND POSSIBLE APPLICATIONS	240
4.1.	THE TRACE DATASET WITHOUT REPORTING ERRORS	240

4.2.	THE TRACE DATASET WITHOUT REPORTING ERRORS AND WITHOUT AGENCY TRANSACTIONS	240
4.3.	THE TRACE DATASET WITHOUT REPORTING ERRORS AND WITHOUT AGENCY TRANSACTIONS	241
4.4.	THE TRACE DAILY PRICES DATASET	242
4.5.	THE MONTHLY PRICE AND MONTHLY RETURN DATASET	242
5.	CONCLUSION	243
 <u>CHAPTER 5 - A STUDY OF THE CROSS-SECTION OF CLIMATE BOND RETURNS</u>		248
1.	INTRODUCTION	250
2.	LITERATURE REVIEW	256
3.	DATA	261
3.1.	REQUIRED DATABASES FOR CORPORATE BOND AND CLIMATE BOND PRICING	261
3.2.	FROM RAW INTRADAY TRANSACTION DATA TO MONTHLY CORPORATE BOND RETURNS	263
4.	VARIABLE SELECTION	265
4.1.	DOWNSIDE RISK	265
4.2.	CREDIT QUALITY	266
4.3.	BOND LIQUIDITY	266
4.4.	SUMMARY STATISTICS	268
5.	DOWNSIDE RISK AND NEXT-MONTH EXCESS RETURNS	269
5.1.	UNIVARIATE PORTFOLIO ANALYSIS	270
5.2.	BIVARIATE PORTFOLIO ANALYSIS	273
5.3.	BOND-LEVEL FAMA-MACBETH REGRESSIONS	274
6.	COMMON RISK FACTORS IN THE CROSS-SECTION OF CLIMATE BOND RETURNS	278
6.1.	DETERMINING THE BOND-LEVEL RISK FACTORS	278
6.2.	SIZE/MATURITY-SORTED AND INDUSTRY-SORTED PORTFOLIOS	280
7.	CONCLUSION	280
 <u>CHAPTER 6 - AN INTRODUCTION TO THE PRICING OF CLIMATE BONDS VIA MACHINE LEARNING</u>		286
1.	INTRODUCTION	289
2.	LITERATURE REVIEW	291
3.	DATA	294
3.1.	REQUIRED DATABASES FOR CORPORATE BOND AND CLIMATE BOND PRICING	294
3.2.	FROM RAW INTRADAY TRANSACTION DATA TO MONTHLY CORPORATE BOND RETURNS	296
4.	VARIABLES	297
4.1.	PRICE-TREND FACTORS	300
4.2.	LIQUIDITY FACTORS	301
4.3.	VOLATILITY FACTORS	302
4.4.	OTHER FACTORS	302
5.	METHODOLOGY	303
6.	RESULTS	307
6.1.	MODELS WITHOUT CLIMATE DUMMY VARIABLES	307
6.2.	VARIABLE IMPORTANCE FOR MODELS WITHOUT CLIMATE DUMMY VARIABLE	309
6.3.	MODELS WITHOUT CLIMATE DUMMY VARIABLES	313
7.	CONCLUSION	314
	GENERAL CONCLUSION	321
	BIBLIOGRAPHY	325