ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Specialization Financial Economics

The Environmental Paradox:
Can pollution and green innovation go hand in hand?

Author: P.M. Smeets
Student number: 492734
Thesis supervisor: Dr. J.J.G. Lemmen
Second reader: Dr. R. de Bliek
Finish date: February 2024

PREFACE AND ACKNOWLEDGEMENTS

This dissertation concludes my journey as a Master's student in Economics and Business (with a focus on Financial Economics) at the Erasmus School of Economics. I am grateful to several individuals who have provided their support throughout the process of writing this thesis. My appreciation goes to my thesis supervisor, Jan Lemmen, whose constructive and timely feedback was invaluable. I also express my gratitude to the Erasmus Data Service Centre for their guidance during my data collection phase. Lastly, I want to thank my parents, family, and friends for their support throughout my academic journey and beyond. This thesis signifies the end of an educational, occasionally challenging, but ultimately enjoyable five-year period as a student. I am eagerly anticipating what lies ahead in my future.

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

ABSTRACT

This study examines the relationship between Environmental, Social, and Governance (ESG) scores and green innovation among US public firms (2010-2022), with a focus on the Energy Sector. It specifically compares the green patent outputs of firms with lower ESG ratings to those with higher ratings. The results reveal that firms in the Energy Sector with low scores for Emission and Resource Use - unaffected by reverse causality - are key contributors to green technology. Specifically, firms with low Emission scores produced 4.88 times more patents, and were cited 3.46 times more often than mid/high-scoring peers. Those with low Resource Use scores published 1.97 times more patents, though without a significant difference in citation frequency. Additionally, Energy Sector industries outperformed non-Energy industries on average, producing times 1.95 as many patents and receiving factor 1.67 more citations. These findings highlight a paradox where companies often seen as environmental burdens are actually significant drivers of green innovation.

Keywords: Green Innovation, Patents, Corporate Sustainability, ESG, Energy Sector

JEL Classification: G11, G30, G32, O31, O32, O34

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

Recent decades have seen a sharp increase in natural disasters, and experts widely recognize carbon emissions as a key driver (UN News, 2021). Richard Heede of the Climate Accountability Institute asserts that just twenty corporations, mostly in the fossil fuel industry, are responsible for an astonishing 35% of global, energy-related carbon dioxide and methane emissions (Heede, 2019). These large corporations, due to their massive carbon footprints, stand out as major contributors to the current environmental crisis. The recognition and awareness of the negative environmental impact that corporations can have, has spread through the financial sector. Consequently, many investors are now opting to divest from major polluters, relying heavily on a company's Environmental, Social, and Governance (ESG) ratings to guide their investment decisions (Friede, 2019). Instead, preference is given to investment-opportunities involving firms with higher ESG ratings.

While studies about financial returns from ESG investing are numerous, fewer research has been done on the actual impact on the environment, caused by investing sustainably. Sachs (2021) and Perez (202) argue that green innovation is the only way to achieve the scale and speed of emissions reductions needed to avoid catastrophic climate change. They argue we need to invest in research and development for clean technologies, such as solar, wind, and battery storage, and that we need to create market incentives for these technologies to be developed. In agreement, Bill Gates (2021) asserts that innovation is our only chance to meet the climate challenge. It is therefore surprising that minimal research has been conducted on ESG ratings and investing accordingly in relation to its impact on green innovation. This is especially true considering that the primary objective of sustainable investing is to mitigate the effects of global warming.

Despite this, the intersection of ESG scores and firm innovation has not received sufficient scholarly attention in the past (Dicuonzo, 2022). Indirectly though, research has found a relationship between ESG ratings and innovation. Previous studies have shown that firms with lower ESG ratings, which may indicate higher carbon emissions, are often excluded from major investment portfolios and penalized with higher equity costs (Garzón-Jiménez & Zorio-Grima, 2021; Zhang & Weber, 2022). In contrast, environmentally-focused firms benefit from lower loan rates and favorable capital costs (Sharfman and Fernando, 2008; Ambec and Lanoie, 2008). Financial constraints imposed on low ESG firms appear to hinder their innovation (Savignac, 2008). Considering the established links between ESG-driven divestments, capital costs, and firm innovation, this raises the question: Does ESG-driven investing hinder or help green innovations? Green innovations, linked to greenhouse gas reductions, benefit not only the innovators but also can significantly reduce the carbon footprints of many firms (Galbreath et al., 2016). Given that green innovation is crucial to addressing the global climate crisis, and there are direct and indirect links between ESG and green innovation, this research aims to thoroughly examine the relationship between ESG ratings and green innovations.

Most studies that do examine this relationship have used R&D expenditures as a proxy to measure firm innovation. This research aims to fill this gap by separately examining the impact of several ESG metrics on firm innovation, using the number of green patents published as an indicator of quantity and citations as an indicator of quality of innovation. Wang et al. (2023) and Lin (2023) both research whether incorporating ESG ratings encourages corporate innovation. They investigate whether firms which are included in the ESG rankings of an independent rating agency innovate more than firms which have not received an ESG score. However, their focus is not on the height of firm-level ESG scores and the effect of the ESG performance on corporate green innovation, but rather on the mere inclusion of such a rating. In another recent study, Fafaliou et al. (2022) investigated the propensity of innovative firms to adopt sustainable practices and their influence on firm's ESG reputation. While their study provides useful insights, this thesis takes a different approach by critically evaluating the validity of either high, or low ESG ratings from an innovation perspective. To my knowledge, Cohen, Gurun, and Nguyen (2020) is the only study that researched the relation between ESG performance and patenting activity, with the latter as dependent variable. Their research highlighted the discrepancy in ESG ratings within the Energy Sector, where companies leading in green patent production held the lowest ESG ratings. The Energy Sector is an umbrella term used in Cohen et al. (2020) and this study, for some of the most pollutive industries: metal mining, coal mining, oil and gas extraction, minerals, petroleum and coal products. This research will start with an analysis of a sample including all industries, then focus specifically on the Energy Sector, and last compare the Energy Sector with other industries through answering the following question:

Are ESG performance and environmental reputation negatively related to innovation, as measured by the number of green patent publications and citations?

To do so, particular focus shall be put on the environmental pillar (E of ESG) in relation to the publication of green patents and citations. This study separately examines three sub-metrics of the environmental pillar – the Resource Use score, the Emission score, and the Environmental Innovation score – as well as a weighted E score, comprising these sub-metrics. Moreover, this study will use unique data on green patents and citations received from the US Patent and Trademark Office spanning the years 2010 to 2022. The timeframe was chosen because the divestiture trend started around 2012 with divestiture campaigns from fossil fuel stocks. The United States are particularly interesting since it is among the largest energy producers and consumers worldwide, and the largest developed economy. While the US has experienced steady economic growth over the past few decades, it also has faced severe problems with pollution and the consumption of resources. The following paragraph discusses the main findings.

This study initially finds that there is no significant relationship between low ESG scores and the number green patent publications or quality of patents when evaluating the entire sample including all industries. This contradicts the findings of Cohen et al. (2020) that lower ESG scores would result in a higher frequency of green patent publications. However, when solely evaluating firms in the Energy Sector, firms with low Emission scores and low Resource Use scores (indicating high emission, high resource use companies) actually perform better on innovation measures. Companies within the Energy Sector with low Emission scores produce 4.88 times as many patents that are also cited 3.45 times as often compared to companies with higher scores in the Energy Sector. Companies in the Energy Sector with low Resource Use scores produce 1.97 times as many patents but do not collect significantly more citations, compared to the companies that acquire higher Resource Use ratings. All aforementioned results are highly significant. For firms with lower Environmental Innovation scores and weighted E scores the opposite relationship is found, with a significant negative result for both patent publications and citations. These contractionary results between the different ESG metrics are due to reverse causality in the regressions concerning the Environmental Innovation scores and the weighted E score. Now, when comparing the Energy and non-Energy Sector as a whole, industries in the Energy Sector on average produce 1.95 times the number of patents an average non-Energy industry does, which are cited 1.67 times as often and have a combined total impact (product of patent publications and citations) of 2.80. Remarkably, the Energy Sector, despite having generally lower ESG scores, showed a higher frequency of green patent publications and citations compared to the non-Energy Sector. This challenges the prevalent view on sustainability efforts in pollution-sensitive industries and highlights the potential for significant progress in environmental technology within the Energy Sector.

This thesis represents a unique addition to the academic discourse on the influence of environmental (E) scores, part of the broader Environmental, Social, and Governance (ESG) framework, on the publication and citation of green patents. Several factors point to the originality, which are not found in previous research by Cohen et al. (2020), Hoang et al. (2020), Dicuonzo et al. (2022), Wang et al. (2023) or Lin (2023). Firstly, the dataset employed is different, uniquely merging Refinitiv's ESG data with patent information from The Lens database. The integration was achieved by matching company names, due to the absence of common identifiers, resulting in a novel dataset. Unlike previous papers, this paper employs the Refinitiv ESG database, while other authors, whose research is summarized in Table 1, used the MSCI ESG Ratings and Sustainalytics ESG Ratings databases. Since rating agencies notoriously report different ESG scores, it is interesting to see that the analyses still yield similar results as Cohen et al. (2020), using a different ESG dataset. Further distinctiveness of the dataset comes from the use of the Y02-classification for selecting green patents. Initiated in 2010, this classification identifies technologies that contribute to climate change mitigation or adaptation, offering a more precise filter than traditional classifications. Secondly, this research uses the publication date instead of the application date of patents to ensure that patents are truly unique.

Thirdly, this study employs a combined total impact variable of Patents×Citations, which multiplies the count of patent publications by the number of citations received. Fourthly, this thesis delves into the E score's components with high granularity, dissecting the ESG score to extract the Environmental Innovation score and its constituent themes. This level of segmentation in assessing the relationship between ESG and innovation is a pioneering step in the literature about the interplay between ESG and green innovation. Fifthly, the study addresses potential endogeneity concerns that emerge from this detailed dissection of the environmental pillar. This aspect of the research is particularly original, as it not only identifies the issue but also employs robustness checks using lagged independent variables to mitigate it. Sixth, this is the only research using the negative binomial regression model. Finally, despite the challenges of endogeneity, this research supports the findings of Cohen, Gurun, and Nguyen (2020), who identified a paradox within the Energy Sector where firms with the highest green patent output receive lower ESG ratings on average. By analyzing a more recent dataset, running negative binomial models and employing different data sources, this thesis aligns with and extends their findings, reinforcing the notion that ESG ratings may not fully reflect firms' contributions to green innovation.

In sum, this thesis not only bridges critical gaps in the existing literature but also sheds new light on the complex dynamics between ESG scores and green patent activities, offering valuable insights for academics, investors, and policymakers alike. The findings of this research challenge prevailing narratives and introduce a balanced perspective on this topic by examining the potential innovation capacities of firms and industries which score low on ESG metrics. Practically, these insights urge policymakers and investors to re-evaluate current investment strategies, encouraging a more nuanced approach that recognizes the innovative potential of firms with lower ESG ratings. For example, companies could receive tax breaks or subsidies only if they meet certain benchmarks in reducing emissions and further increase investment in green technologies. This could lead to more inclusive and effective environmental policies and investment criteria that better capture a company's contributions to sustainable innovation. Green innovation, with its long investment cycles and substantial risk of failure as outlined by Holmstrom (1989), necessitates a stable and predictable environment for investment. It is crucial for governments and investors to contribute to predictability and support, encouraging pollution-intensive firms to confidently enter into green innovation initiatives. This approach is essential to mitigate the inherent risks and initiate the transition towards sustainable practices.

The rest of this paper is organized as follows. Chapter 2 presents the theoretical framework, providing definitions and theories related to the topic. Additionally, it reviews existing literature and proposes hypotheses. Chapter 3 discusses the data collection, sample construction and variables, followed by Chapter 4, which details the research methodology. The findings from the empirical analyses and a robustness tests are presented in Chapter 5. Finally, the last chapter concludes the paper, offers suggestions for future research, and evaluates any limitations of this study.

CHAPTER 2 Literature Review

2.1 Innovation

2.1.1 Defining Innovation

Innovation has long been a crucial driver of economic activity. The argument that different aspects of innovation create a unique and superior business combination, goes back to Schumpeter (1942). The economics of innovation has become a recognized field within applied economics that encompasses various economic orientations, such as macroeconomics, microeconomics, industrial organization, and international economics. Its multidisciplinary nature shows the growing importance of innovation as a research area.

For a comprehensive examination of the relationship between ESG performance and green innovation, it is important to clearly define "innovation". According to Baregheh et al. (2009), "Innovation is the multi-stage process whereby organizations transform ideas into new and improved products, service or processes, in order to advance, compete and differentiate themselves successfully in their marketplace" (p.1334). This progress could result in improvements such as efficiency, sustainability, or overall effectiveness. More recently, these innovations might entail new technologies, processes, or business models designed to lessen environmental impact and promote sustainability.

2.1.2 Measuring Innovation

Innovation is a complex concept to measure, as evidenced by the various existing measures and econometric models that attempt to capture innovative activities. One commonly used measure or proxy for innovation is a company's R&D expenditure, often scaled by factors such as sales (Griliches, 1980; Mansfield, 1988; Shefer & Frenkel, 2005; Dicuonzo et al., 2022). A firm's spendings on R&D serve as an estimate for the yearly capital investment that contributes to the accumulation of knowledge (Hall et al., 1986). The underlying rationale is that investments in R&D have an impact on a firm's ability to innovate. However, using R&D expenditure as a direct measure of innovation may result in biased estimates, as not all research and development efforts lead to successful innovations. Rather than measuring the output of a firm's innovation, R&D expenditures reflect the level of commitment a firm has towards R&D. Additionally, different investments in R&D may serve different purposes but are still reported as expenditure in R&D in annual reports.

Another approach to measuring innovation is by using the number of patents as a proxy (Anokhin & Schulze, 2009; Aghion et al., 2009; Lin, 2023; Wang et al., 2023). A patent is a legally approved document that grants exclusive rights to use a product, service, or process for a specific period of time (Griliches, 1998). According to Katila (2000), patents are a valuable measure of innovation as they represent the output of new ideas and inventions. Additionally, patents serve as an early indicator of technological change, highlighting their significance in assessing innovativeness. Using the number of patents has the advantage of capturing the output of R&D, as each patent represents a specific innovation. Moreover, this measure has high comparability, as there is available data for a large number of firms. However, this measure fails to distinguish between incremental and transformative innovation, disregarding the quality of innovation (Griliches, 1986). According to Brandler et al. (2008), patents do not inherently signify innovation but rather invention, which may not always have value. A firm may have a single patent that disrupts the entire market, while another firm may have countless patents that are collectively less valuable.

To overcome the previous limitation associated with the number of patents, patent citations have been introduced as a measure that presents the effect of a patent on successive innovation. By considering the amount of citations a patent receives after its approval, patent citations not only capture the economic importance of a patent but also correlate with its market value (Hall, Jaffe, & Trajtenberg, 2005). Cohen et al. (2020) suggest that the number of citations a patent accrues reflects the underlying innovation's quality. Brandler et al. (2008) propose an approach to measure innovativeness by calculating the average of patent citations and patent originality at the company level.

In conclusion, patents and their citations offer an objective and quantifiable method to gauge innovation. In this study, an additional total output variable shall be used which is the product of patents and citations. When researching the relationship between a company’s ESG score and its level of innovation, these measures can provide valuable insights into the interplay between sustainability and innovation.

2.1.3 Green Innovation

Green innovation, a type of innovation that aligns with sustainability goals, is increasingly important. Standard innovation is typically defined as the consistent enhancement of a company's ability to create new products that satisfy market demands (Garel, 2021). However, the transition to green innovation requires a shift in the conventional understanding of technological innovation, with a focus on eco-innovation or environmental innovation (Demirel & Kesidou, 2019). The inclusion of environmental innovation in a company's strategy is becoming more common and is seen as a factor that can boost corporate competitiveness (Surroca et al., 2010). It promotes long-term value creation for shareholders and improves the company's engagement with relevant stakeholders (Lin et al., 2019).

Besides the benefits for shareholder and stakeholders, green innovation in particular has the potential to significantly reduce emissions across businesses worldwide, making it a critical tool in the fight against global warming. This highlights the need for ongoing and future innovative solutions that focus on reducing environmental emissions and optimizing energy usage (Garel, 2021). By investing in green innovations, firms not only become more sustainable, but they also future-proof their operations in terms of meeting stakeholder demands and preparing for, as well as adapting to, new sustainability policies. Thus, green innovation serves as an effective tool for stimulating the sustainable success of an industry while preserving its environmental advantages (Zhang et al., 2019; Zhang et al., 2020b).

2.2 ESG scores

2.2.1 Defining ESG conceptually

ESG, an acronym developed in 2004, stands for Environmental, Social, and Governance. It refers to how corporations and investors address these aspects in their operations (Gillan et al., 2021). Environmental factors include the company's use of natural resources, pollution, waste management, energy consumption, sustainability initiatives, and related areas. Social factors encompass employment-related issues and broader societal concerns such as human rights, data protection, and community engagement. Governance factors pertain to the system of rules and policies that guide a company's operations, including accurate reporting methods, board member selection, and regulatory compliance (Boffo & Patalano, 2020).

The concept of ESG has experienced significant growth in recent decades. There has been a growing demand for measurable data on how companies use their various forms of capital, both non-financial and financial, to provide their products and services, as well as how their practices impact the environment and society through negative externalities. This demand has led to the establishment of ESG reporting standards (Kotsantonis & Serafeim, 2019), which aim to provide transparency and accountability in these areas.

ESG is a significant non-financial metric since it is used extensively by private and institutional investors to compile their portfolios (Friede, 2019). ESG investing also provides significant value to investors (Engle et al., 2020). Addressing ongoing ESG concerns is a risk reduction policy for long-term investors. Moreover, ignoring these issues may lead to firms becoming less valuable in the future (FD, 2022). Good ESG performance not only benefits investors but also the firms themselves. Studies conducted in 2013 showed a positive association between strong ESG performance and financial growth (Clark et al., 2015). As ESG continues to evolve and gain traction, more investors and managers are recognizing the importance of considering non-financial responsibilities in their investment and business practices, leading to an expansion of its relevance in academic literature.

2.2.2 Measuring ESG

Environmental, Social, and Governance (ESG) metrics are tools for assessing a firms' ethical impact and sustainability practices. The measurement of ESG is multifaceted, incorporating a diverse range of factors across its three pillars. For instance, the Refinitiv ESG score is calculated using a subset of 186 metrics, derived from over 630 ESG measures (Refinitiv, 2022). The provision of underlying category evaluations of E, S and G, besides the ‘overall score’, enables users to select and use the scoring system that best aligns with their needs, obligations, or investment criteria.

Measuring ESG performance is a complex task. Various ESG rating providers assess and score companies’ ESG performance, but their methodologies and transparency can vary (Poh, 2019). Since non-financial reporting is frequently voluntary, companies have more flexibility and can selectively disclose ESG-related information (Paradis & Schiehll, 2021). Moreover, differences in frameworks, data usage, key indicators, metrics, and subjective judgment contribute to variations in ESG ratings. This poses challenges for investors in distinguishing reliable ratings and for researchers in conducting unbiased studies. When conducting academic research, it is important to understand the components of ESG index scores and their relationship to the market. In the data section, I shall further elaborate on the concrete components that make up the ESG score.

2.2.3 Environmental score

This paper places a deliberate emphasis on the environmental aspect of the ESG (Environmental, Social, and Governance) framework. The primary reason for this focus is our study's core topic: sustainable innovation, which we measure using green patents. Since our interest lies in the sustainability features of innovation, it makes sense to equally concentrate on the environmental factors within the ESG spectrum to understand how these elements relate to each other. Furthermore, it's commonly accepted that the environmental component is more impactful than the social or governance aspects in the context of ESG assessments. Recent media coverage has highlighted that the social and governmental aspects are often overshadowed by the environmental pillar (FD, 2022). Investors also seem to prioritize the environmental rating when evaluating potential investments, indicating that this aspect could significantly influence a firm's ability to secure funding and its capacity to innovate. This is consistent with the introductory suggestion that the cost of capital has a considerable influence on the production of green technologies. Additionally, focusing on just the environmental part allows us to delve deeper into its specific sub-metrics and how they correlate with green patents and their citations. We will explore these sub-metrics more thoroughly in Section 3.3.2. In summary, the environmental component is the most pertinent for our investigation into the link between ESG and innovation.

Generally, for most rating agencies, the environmental pillar is made up from several sub scores. These scores often include a resource use score, an emission reduction score and an environmental innovation score. The score for resource use evaluates a company's ability and performance in decreasing the consumption of energy, water, or materials, and in improving eco-efficiency through better supply chain management. The emission reduction score assesses a company's commitment and effectiveness in lowering environmental emissions during its operational and production processes. Lastly, the environmental innovation score signifies a company's potential to lessen environmental costs and burdens for its consumers, thus creating new business opportunities through eco-friendly processes, or products designed with environmental consideration.

2.3 Industry level ESG

One of the primary motivations for this research is to explore whether companies that are excluded from investor portfolios, based on low ESG scores, also exhibit lower levels of green innovation. In the previous sections, E(SG) scores were discussed on the company level. ESG scores can have significant impacts on firms since the ratings are used extensively by private and institutional investors to compile their portfolios (Friede, 2019). Namely, investor divestment might result in increased capital costs and financial limitations for these companies, potentially hindering their innovation efforts. However, it is important to note that companies may be excluded not solely on the basis of their individual ESG scores but also because of their industry's overall reputation for pollution. For example, industries associated with negative environmental impact, such as the oil & gas industry are often omitted from investment portfolios as a whole (Fabozzi & Oliphant, 2008). According to more recent research by Serafeim (2018), financial institutions are increasingly moving away from high carbon investments, replacing them with low carbon alternatives to avoid the risk of stranded assets. The evolving investment landscape has led to a certain stigmatization of the oil industry and other high pollution industries, often branding the industries in their entirety, as having low ESG performance. This brings to light the concept of Industry-level ESG, which posits that industries known for encompassing predominantly polluting companies could be perceived as having a collectively low ESG score. In turn, this collective image could negatively affect the individual firms within these industries as well.

Particularly for the oil, gas, and energy industry, firms are often explicitly excluded from ESG funds' investment universe. Their environmental image is highly scrutinized. Cohen, Gurun, and Nguyen (2020) too find certain industries to be very pollutive, despite their potential to innovate. In their research, these industries are the companies for which the first two digits of their Standard Industrial Classification (SIC) are 10 (Metal, Mining), 12 (Coal Mining), 13 (Oil & Gas Extraction), 14 (Non-metallic Minerals, Except Fuels), 29 (Petroleum & Coal Products), or 49 (Electric, Gas, & Sanitary Services). The Energy Sector is an umbrella term which was used in Cohen et al. (2020) and shall be used in this study to collectively label the industries mentioned above.

In line with Cohen et al. (2020), this study also seeks to understand whether the companies typically divested and excluded from investment portfolios, often belonging to these so-called polluting industries, engage more or less in green innovations. This approach allows for a refined understanding of the relationship between industry-level reputation and individual firm innovation. By examining both the direct impact of ESG scores on a firm's innovation and the broader industry-level

ESG context, this research aims to contribute to a more complete understanding of the interplay between industry reputation and innovation.

2.4 Theories

When examining firm-level innovation in the context of sustainability and ESG scores, the theories below provide a valuable perspective. Some theories may be used to argue in favor of a positive relationship between ESG and innovation while other may be used to argue for a negative relation. Different theories predict different firm behavior which underlines this thesis' relevance.

2.4.1 Resource-Based Theory

Wernerfelt's (1984) concept of firms being collections of resources highlights the core principle of the Resource-Based View (RBV): a firm's competitive advantage is determined by the uniqueness and strategic use of its resources and capabilities. RBV asserts that companies possess distinct resources that differentiate them from their competitors (Becker & Dietz, 2004). These resources extend beyond physical assets to intangible attributes that shape a firm's identity and strategy. The uniqueness of these resources is influenced by historical conditions and firm-specific decisions. Barney (1995) and Dierickx & Cool (1989) emphasize that for a resource to truly provide a competitive advantage, it must possess characteristics such as scarcity, inimitability, relevance and appropriability to the firm.

The power of RBV becomes evident when considering intangible resources, as argued by Grant (1991) and Itami & Roehl (1991). These assets, though not prominently displayed in financial statements, hold significant strategic importance. Intangible resources include the knowledge, expertise, and collective wisdom of employees, a firm's absorptive capacity for new knowledge, cultural nuances, and industry reputation. Due to their nature, competitors find it challenging to replicate intangible resources, making them valuable sources of sustainable competitive advantage.

The resource-based view provides valuable insights into innovation. Innovation capabilities are not solely dependent on external technological acquisitions but are deeply rooted in a firm's internal resources (Barney, 1991). For example, a firm's historical technological achievements, culture of continuous learning, and network of collaborations significantly influence its innovation trajectory. Galende (2006) argues that a firm's potential for industry profitability is not solely based on its ability to adapt external technologies but heavily relies on its internal resources, enabling the creation of novel innovations.

Within the RBV framework, ESG ratings can be viewed as strategic, primarily intangible resources. Examples of such intangible assets that directly impact a firm's ESG score include responsible marketing, diversity and inclusion, Corporate Social Responsibility (CSR) strategies, managerial capabilities, and organizational culture (Galende, 2006; Refinitiv, 2020). ESG scores may indicate a firm's commitment to and culture of sustainability, accumulated environmental knowledge and a network of green partnerships. The aforementioned resources could provide a firm with a competitive advantage. In the resource-based view, an interplay between ESG and innovation arises.

Namely, another source of competitive advantage that high ESG firms possess is their access to financing. As mentioned in Section 2.2.3, particularly high environmental (E from ESG) scores may affect access to financing. As ESG has become an important criterion for investments, highly rated firms can secure funds more easily than their low ESG counterparts. According to the USSIF (2020) report, as of 2020, sustainable investing accounted for more than 33% of total assets under management in the US, amounting to $51.4 trillion. Notably, sustainable investing has experienced significant growth of more than 42% since 2017. Abundant financial resources can be directed towards innovation. Low ESG firms face more expensive financing options (Zhang & Weber, 2022). Savignac (2008) confirms that financing constraints negatively affect innovation likelihood.

Conversely, using the RBV framework, one could also argue that firms with lower ESG might innovate more. Lacking recognized green resources, these firms may adopt more aggressive strategies to catch up or strategically diverge from environmentally questionable practices. Their drive to innovate could be seen as a means to create new resources that may serve as critical assets in an evolving business landscape. For instance, oil & gas companies' resources for their current operations (e.g. fossil fuels) are becoming increasingly scarce which might incentivize them to explore alternative, greener fuel (e.g. hydrogen or biodiesels). Additionally, in the face of climate change and rising environmental consciousness among consumers, there is another risk for oil & gas companies: consumers might shift from fossil-based to zero-emission fuels. These low ESG firms, aware of potential threats to their operations, may invest in innovations to diversify or transform their core operations. For example, within the petroleum industry, part of the Energy Sector in this research, hydrogen investments were recently ramped up to meet ESG goals (S&P Global, 2021). From an RBV perspective, this behavior can be understood as a strategic move to develop new and unique resources, ensuring future competitiveness and relevance.

The RBV, considering both tangible and intangible assets, offers a robust framework to understand why firms exhibit different innovative behaviors based on their ESG scores. Whether leveraging existing green resources or developing new ones in response to depletion of resources, firms' strategic maneuvers in innovation can be effectively understood through the RBV lens. In this research, we will empirically examine the extent to which firms with high versus low ESG scores lead innovative behaviors.

2.4.2 Institutional Theory

Organizations do not operate in isolation. The Institutional Theory posits that firms align their actions with the norms, values, and rules of their environment (Scott, 2008). This environment is shaped by regulatory pressures, industry norms, and societal expectations regarding environmental stewardship.

Over time, the Institutional Theory has evolved to recognize not only formal structures but also shared routines and cultural practices. At its core, the theory suggests that organizations seek legitimacy and often conform to established standards and behaviours set by institutional factors.

Within the context of ESG and innovation, the Institutional Theory emphasizes the role of external pressures in shaping firm behaviour. ESG reflects industry norms and evolves based on societal expectations, regulatory standards, and environmental stewardship practices. Dyck et al. (2019) found that equity investors can exert pressure on corporations to pursue environmental and social strategies, particularly in countries with strong community beliefs and institutional quality, such as developed countries.

Drawing on the Institutional Theory, one can argue that a higher ESG score may indicate a firm's conformity to environmental norms, suggesting a reactive approach to external pressures. If these firms already comply with prevailing norms, their motivation to innovate may be relatively lower compared to those with lower ESG scores. High scoring firms may already meet legislative requirements, resulting in limited urgency to deviate from or extensively innovate beyond existing practices.

The societal shift of importance towards higher ESG ratings implies increased pressure on companies to improve their scores. Firms with environmental concerns have greater likelihood of investor activism (Akey & Appel, 2019) and the development and adoption of stricter environment-related policies increases significantly (Ilhan et al., 2021). These external pressures can act as powerful catalysts for innovation, particularly in technology output, as firms strive to meet and exceed evolving standards. Low ESG rated companies, often those with higher carbon emissions, are more susceptible to regulations and environment related lawsuits (Hsu et al., 2022). Such regulations can stimulate green innovations by breaking firms' inertia and encouraging exploration of new technological fields, as suggested by Van der Linde (1993). By investing in green innovations, firms invest in becoming more sustainable and therefore create operations that are prepared for future sustainability policies.

Conversely, Bartram et al. (2022) find that financially constrained US firms transfer their emissions activities from regulated to unregulated states to cope with environmental and climate-related policies. In line with this, Dai et al. (2021) find that firms with low relocation costs facing high local regulatory pressures relocate their plants and facilities to regions with less stringent environmental policies. This would imply that the green patenting efforts of firms with low ESG scores, as a result of institutional pressures are indistinguishable from those with high ESG scores.

In conclusion, the Institutional Theory provides a complete framework for examining the relationship between ESG scores and innovation. It highlights the various ways in which external pressures, societal expectations, and regulatory norms intersect to drive companies towards or away from innovation. While higher ESG firms may feel less compelled to innovate due to their existing alignment with environmental norms, the pressure to improve ESG ratings can paradoxically stimulate profound innovations, particularly among firms with initially lower ESG scores.

2.4.3 Stakeholder Theory

Freeman (1984) introduced the Stakeholder Theory, which asserts that managers should satisfy the interests of all stakeholders in order to thrive (McWilliams & Siegel, 2001). This challenges Friedman's neoclassical perspective, which argues that businesses should prioritize shareholder satisfaction. Stakeholders include shareholders, suppliers, employees, customers, investors, government, communities, environmental organizations, and the media (Clarkson, 1995). However, companies often prioritize the interests of influential stakeholders, neglecting the core principles of the Stakeholder Theory (Deegan & Unerman, 2006). Jensen (2001) proposes the Enlightened Stakeholder Theory, allowing managers to make necessary trade-offs and maximize long-term value. Post, Preston, and Sachs (2002) argue that a company's success is determined by its relationship with stakeholders, and Freeman (1984) emphasizes the importance of managing stakeholder relationships for long-term success.

Companies with high ESG scores may be more attuned to a wider range of stakeholder needs, particularly those who value environmental sustainability. The act of engaging with stakeholders can guide companies towards adopting sustainable practices in their innovation activities, as noted by the European Commission (2008) and supported by findings from Carrasco & Buendía-Martínez (2016). Lin et al. (2014) also provide evidence that stakeholder pressures promote green innovation. McDougall et al. (2019) state that firms fulfilling their environmental responsibilities are more likely to develop resources for pollution prevention, product stewardship and clean technologies.

Thus, the Stakeholder Theory sheds light on how ESG scores, as a reflection of stakeholder engagement, influence the direction and depth of innovation. The assumption is that the same stakeholders that drove high ESG companies to obtain a high ESG score, will also push for these companies to be more innovative. Stakeholder Theory could explain that firms which want to satisfy the interests of their sustainable investors, also engage in innovation more.

2.5 Hypotheses

Having established a theoretical framework in Section 2.4, we now introduce the hypotheses. The Meta table (Table 1) gives a concise summary of studies closely related to this paper's topic. Research by Dicuonzo (2022) highlights the role of innovation in augmenting ESG practices, implying that innovation and ESG can be mutually reinforcing. Although this does not directly address green patent production, it emphasizes the mutually beneficial relationship between innovation and ESG ratings. Another research conducted by Wang et al. (2023) investigated the impact of including ESG ratings, conducted by an external third party, on corporate green innovation. Their study differs by focusing on whether companies received an ESG rating, regardless of its value. The findings showed a positive influence on innovation from simply having an ESG rating, but did not consider the rating's magnitude. Yet, the mere exploration of this topic suggests a perceived connection between ESG ratings and green innovation. The question remains what direction is, of this relationship.

Several studies imply a positive relation between green innovation and ESG ratings. Recent research by Zhang & Chen (2023) shows that ESG scores can strengthen the relationship between governments and firms, leading to more resources for green innovation. This suggests that higher ESG scores are linked with increased efforts and resources dedicated to green innovation, potentially leading to an increase in green patent production. In this line of thought, firms with low ESG scores innovate less. According to Zhang & Weber (2022), low ESG companies face divestment by socially responsible investors and consequently, tend to face higher capital costs. This means that environmentally beneficial projects, if conducted by high emission companies, face greater financial obstacles (Heinkel et al., 2001). Such financial constraints on firms with low ESG ratings impede corporate innovation (Savignac, 2008).

Yet, some researchers believe that the relation between green innovation and ESG ratings is negative. Firms with unfavorable ESG scores fall victim to environmental lawsuits and regulations more easily (Hsu et al., 2022). According to Van der Linde (1993), such regulations can stimulate green innovations by breaking firms' inertia and encouraging exploration of new technological fields.

In conclusion, while the provided research results hint at a correlation between ESG scores and green innovation, researchers disagree on the direction of the effect. More specific data or research would be required to establish a direct link between high ESG scores and an increase in green patent production as a precise proxy for innovation. Therefore, this thesis shall research the following hypotheses:

Hypothesis 1: *Low E (ESG) firms publish more green patents than high E (ESG) firms.*

Hypothesis 2: *Low E (ESG) firms publish higher quality green patents than high E (ESG) firms, measured by citations.*

The hypotheses above analyze the relationship between the magnitude of ESG scores and innovation without considering the industry of the firm in question. A significant issue with ESG scores is that they are usually relative scores within different industries (Berg et al., 2022). Evaluating ESG scores within the same industry might provide a better context for relative analysis compared to an absolute score. Rating agencies often use different indicators for different industries. For a more accurate and contextual assessment of the E score, it is preferable to compare a firm's environmental performance with another firm within the same industry, rather than between firms from different industries. Gyönyörová et al. (2023) conducted an exploratory factor analysis of the S&P Global 1200 index, containing many companies from the sample in this thesis, and found that the consistency and convergent validity of ESG data varied significantly depending on the industry. Therefore, Kotsantonis & Serafeim (2019) pose that benchmarking companies against peers and establishing how ESG ranking parties define companies' peer groups, is essential in establishing the performance ranking of a firm. This process of ranking companies should be harmonized across industries. Thus, relying solely on primary ESG scores may lead to misleading conclusions about the relationship between ESG and innovation. This is why, for the next two hypotheses, a sub-sample of firms shall be compared with their industry peers. For this analysis, the Energy Sector shall be evaluated, which was discussed more elaborately in Section 2.3. All though the sample for the next two hypotheses differs from that used in Hypotheses 1 and 2, the underlying inquiry remains: Is there a positive or a negative correlation between ESG and firm innovation?

On the one hand, there is a possibility that Energy Sector companies with leading ESG ratings outpace their peers in green patent production. A high ESG score indicates a firm's commitment to sustainability, environmental stewardship, and adept management of stakeholder relationships. According to Wang & Sengupta (2016), from a resource-based view, effectively managing stakeholder relations can create intangible value in labor resources, organizational culture, and sustainable innovation. The efficient usage of these intangible assets can enable the development of a competitive advantage compared to industry peers (Surroca et al., 2010; Wang & Sengupta, 2016). For example, firms that engage in socially responsible practices in human resources tend to attract better applicants, leading to the accumulation of human capital. This accumulation can then translate into a competitive advantage and positively impact financial performance (Surroca et al., 2010). Nowadays, it is particularly important for firms to possess strong reputations in order to attract talented young professionals. Particularly in the Energy Sector, which is widely perceived to be pollutive (Dario & Heede, 2021). ESG controversies can decrease scores and scare off applicants, making them choose for the more sustainable, responsible company in that branch. ESG controversies refer to news events like questionable social conduct or product-related scandals that draw media attention to a company, thereby attracting the notice of investors or potential employees (Aouadi & Marsat, 2018). Besides, high ESG scores often correlate with investor and management's long-term thinking (Fafaliou et al., 2022). Embracing innovative practices can be a way to ensure the company's future profitability in an industry facing potential decline due to environmental concerns.

On the other hand, it could also be that Energy Sector firms with low ESG scores, innovate more. Firstly, it might be the case that low-scoring firms invest their cash flows directly into R&D and innovation projects, whereas high ESG firms might allocate a significant portion of their resources to maintain their ESG standards. Secondly, low ESG firms might have fewer sustainability commitments and partnerships that restrict them, allowing for more flexibility in trying out new strategies and technologies. Lastly, facing criticism and backlash, these firms might see innovation as a tool to better their public image and counter negative perceptions. This could explain a relative increase of patent production by low ESG firms, throughout the sample period.

Hypothesis 3: *Within the Energy Sector, firms with low E (ESG) scores publish more green patents than high E (ESG) firms.*

Hypothesis 4: *Within the Energy Sector, firms with low E (ESG) scores publish higher quality green patents than high E (ESG) firms, measured by citations.*

For Hypotheses 1, 2, 3 and 4, each hypothesis shall be further split up into four sub-hypotheses to account for the various ESG metrics which are evaluated in relation to patent publications or citations.

The objective of this research is to compare firms and sectors with low and higher ESG scores and find the extent to which they produce (high-quality) green innovation. A problem is that ESG data, in general, as well as in my sample, is incomplete. For most firms, various scores are missing for several years, resulting in missing firm-year observations. To avoid this issue in the fifth, sixth and seventh hypotheses, industries that are generally perceived to be low ESG shall be compared with high ESG industries, irrespective of individual firms' ESG scores. This way, this particular part of the thesis can refrain from using actual ESG scores. For pollution-prone or carbon-intensive industries, there is a wide consensus on the ESG of firms operating in these industries. Particularly, their environmental image is scrutinized. As discussed in Section 2.3 Industry level ESG, industries belonging to the Energy Sector, are widely considered low ESG. Therefore, this paper shall compare low ESG industries (the Energy Sector) to high ESG industries. The question remains: on an aggregate level, which one produces more, and higher quality, green patents?

Companies in industries that are perceived to be more sustainable, the non-Energy Sector, may be more willing to fulfill various stakeholder needs. Engaging in green innovation can be a way for firms to meet these diverse demands. Lin et al. (2014) assert that companies often engage in eco-innovations in response to government regulations and consumer demands. McDougall et al. (2019) suggest that companies that fulfill their environmental responsibilities are inclined to create environmentally-friendly technological assets for the responsible management of products and to advance clean and sustainable technologies. In this line of reasoning, high ESG industries innovate more, whereas low ESG industries are less attuned to various stakeholder needs, have a higher focus on short-term financial gain and innovate less. In turn the low ESG Energy Sector might innovate less on average.

However, according to Cainelli et al. (2015), the polluting corporations should be seen as potential contributors to environmental solutions, given their capacity for innovative activities. Companies in these industries, thus the Energy Sector as a whole, tends to have high carbon emissions but also possesses the needed financial means to innovate. Energy firms, due to their size and scale advantage, may adopt a strategy similar to the Stackelberg approach. This strategy entails observing other firms that are more innovative and are taking the lead in introducing a new technology tree, and then capitalize the opportunities once they arise (Chamley & Gale, 1992). Additionally, Kumar (2020) finds that Energy Sector firms have repeatedly shown to be profit maximizing entities with the ultimate goal of being long-lived global energy providers for the next decades. Given the increasingly stringent environmental regulations on resource usage and emissions, it is possible that these pollutive companies are working on potential solutions to continue to exist. Innovation could be a means to adapt to a changing business environment by pivoting to less pollutive operations. To test both views, the following hypotheses shall be tested:

Hypothesis 5: *The low E (ESG) industries of the Energy Sector publish more green patents compared to other industries.*

Moreover, Cohen et al. (2020) find that green patents produced by the energy industries are cited more highly than the average green patent. This leads to the hypothesis:

Hypothesis 6: *The low E (ESG) industries of the Energy Sector publish higher quality green patents compared to other industries, measured by citations.*

Lastly, this research shall evaluate a composite measure that captures both the quantity and influence of these patents which leads to the final hypothesis:

Hypothesis 7: *The low E (ESG) industries of the Energy Sector publish higher combined volume and impact patents compared to other industries, measured by publications and citations.*

Table 1: Highlights of prior research on the relation between ESG ratings and innovation, using patents as a proxy of innovation

Author	**Region and Time Period**	**Method**	**Control variables**	**Results**
Cohen, Gurun, & Nguyen (2020).	United States. 1980 -2020 (for patents) 2008 – 2020 (for financial data)	OLS regressions Poisson models Non-parametric approach to identify the earliest movers within each category of green patenting. Natural language processing techniques to analyze the similarity of patent language between energy and non-energy firms.	R&D Investment Firm Age (establishment age) Total Assets Book Leverage Cash	Energy firms are producing higher quality and more impactful green patents than non-energy firms, and that they roughly have an 80% higher chance of being the earliest "pioneer-patent" in a given green technology class
Hoang, Przychodzeń, Przychodzen, & Segbotangni (2020)	United States. 2007 - 2016	OLS regressions	R&D Investment, Firm age, Book Leverage, Sector Dummy, SD of monthly returns during last 36 months, Total Assets, Earnings Retention	The global financial crisis increased the environmental transparency of firms with green patents but negatively impacted their price to earnings ratio
Dicuonzo, Donofrio, Ranaldo & Dell'Atti (2022).	France, Germany, Italy, Spain, the United Kingdom and the United States. 2013 - 2020	Fixed-effects regression model, Random-effects regression model, Pooled OLS	Log of Total R&D Investment ROA Market Capitalization	Positive relationship between ESG practices and innovation. Companies investing more in R&D and patents have better ESG performance.
Lin (2023).	China. 2009 - 2021	Two-stage difference-in-differences (DID) approach, Instrumental Variable (IV) approach	R&D expenditures/ Total Assets Total Assets ROA Book Leverage Operating Income Tobin's Q Board Independence Institutional Ownership	ESG ratings have a positive impact on patent applications and citations promote firm innovation through three channels: reducing financing constraints, attracting R&D staff and by luring mutual funds
Wang, Ma, Dong & Zhang (2023).	China. 2013 - 2019	Multi-period difference-in-differences (DID)	R&D Investment Firm Age (years listed) Total Assets Book Leverage ROA Revenue Growth PPE/Total Assets Shareholder concentration	Firms covered by the ESG rating agency increase green patent output by 3.9%

CHAPTER 3 Data

3.1 Data collection

This section elaborates on the data sources used, the data preparation and the collection process. The less standard data sources used in this research are explained in more detail. For all data collected, the time frame used is 2010 to 2022, for US publicly listed firms only. To explore the research question, all information on green innovation, ESG and financial variables is needed at firm level.

Patent & Citation Data

The initial step in the data collection process involved obtaining patent data. The patent data is used for Hypothesis 1 to Hypothesis 7. Moreover, Hypotheses 2, 4, 6 and 7 also make use of the patent citations data.

Numerous datasets containing patent data and information on patent characteristics are available, although most of them are not publicly accessible for free. Patent offices such as the European Patent Office (EPO) and the United States Patent and Trademark Office (USPTO), offer comprehensive data on European and American patents respectively. However, these databases primarily provide detailed information on individual patents rather than offering an overview of the number of patents at the patent owner level. For this research, it is necessary to have basic information on individual patents to determine the number of citations received, classify the patent as "green" or not, identify the year of publication, the publisher, and the patent owner(s).

For all patent data, including patent citations, The Lens has been used. The Lens is a patent database that is widely used for academic research and patent analytics. It contains data from the USPTO, EPO and the World Intellectual Property Organization (WIPO). Its main goal is to provide open and secure access to patents and scholarly work from around the world while prioritizing the free dissemination of information (Lens, 2023). The database contains around 149.6 million patent records with 365 million citations from 17.2 million applicants. The Lens database is an important resource for patent analytics and is considered a commendable initiative to democratize patent data (WIPO, 2023). In addition to offering a comprehensive collection of patent records, the platform provides extensive search capabilities with various filters, ensuring researchers can easily access and analyze the specific information they need.

The most important feature of The Lens is that it allows you to search for patents based on Cooperative Patent Classification (CPC) codes. The possibility to specifically extract patent records for a set of patent classifications is essential for this thesis that focusses on green innovation only. This thesis utilizes the CPC scheme, which includes a classification symbol for patents related to new technological developments (Y). Within this scheme, a subset (Y02) specifically captures technologies and applications for mitigating or adapting to climate change (European Patent Office, 2007). The Y02 classification was established through collaboration between the United Nations Environmental

Program, the International Center on Trade and Sustainable Development, and the EPO in 2010 (Angelucci et al., 2018). This classification scheme is preferred as it better distinguishes patents related to green innovations and other commonly used classification schemes, such as the International Patent Classification (IPC) or the United States Patent Classification Systems (USPC) (Cohen et al., 2020). For this thesis, only patents published at the USPTO are considered since the geographical scope is limited to listed US firms.

To determine the number of patent records and related citations, it is necessary to identify a specific date for each patent. Various dates can be used, including the publication date, filing date, earliest priority date or granted date. The most important dates in the patenting process, as determined by the USPTO (2018), are as follows: Filing or priority date. This date is assigned once the application contains all necessary documentation. Within the following 12 months, adjustments and expansion to other countries can be made. Secondly, the Publication date which take place 18 months after the filing date against future patent applications for similar inventions. Lastly, the Grant date is established if the patent office determines that the invention and application meet the requirements. The patent can then be granted, and its effectiveness dates back to the publication date.

The paper of Griliches et al. (1986) argues that the application year, the year in which the patent was first published at a patent office, should be preferred as it represents the actual timing of an innovation. However, counterarguments state that firms cannot immediately diffuse an invention from the application date. It is important to highlight that there exists a temporal lag between the instance of a firm's patent application for an invention and the subsequent utilization or dissemination of the said invention. Consequently, a patent cannot be unambigiously seen as an exact indicator of the timing of an innovation. Another problem in using the patent application date is that it is only impossible to establish the uniqueness of the patent application until after the patent publication, 18 months later. Focusing solely on filing or application dates may introduce biases, as some patents may not progress to publication if filing dates are exclusively used. As a consequence of the problems mentioned above, this research used the patent's publication date. Patent publication dates follow a standardized timing, allowing for easier comparisons and analysis across a wide range of patents and innovations.

The "Cited by Patent Count" variable, also known as forward citation, is used to determine the number of citations a patent has received (Lens, 2023). To decrease the size of the dataset and since the focus of the research is on the significance of the patents published by the firm, the backward citations were deleted.

In the dataset provided by the Lens, the column that shows the owner(s) of a patent record often contains more than one name, followed by a date, in between brackets. Although it is not specified on the Lens's website, customer support of Lens.org provided a clarification of the exact meaning. Patents can change owners/assignees over time and the date indicates the earliest record date. It is noteworthy that patent ownership is frequently exchanged within the subsidiaries of the same corporation. Occasionally patents are acquired by external companies. Thus, these different owners at various dates could either be distinct entities within the same firm or different firms that have acquired the patent. Galende (2006) argues that a firm's potential for industry profitability is not solely based on its ability to adapt external technologies but heavily relies on its internal resources, enabling the creation of novel innovations. Therefore, despite knowledge on further acquisition of patents, this study only focusses on the publishers of patents: the initial owners. This way the focus is on who created the patent organically, meaning who actually made initial technological impact as opposed to firms that acquire them at a later stage.

The data extraction process from The Lens database shall now be discussed. Having, selected the CPC code "Y02", several additional filters were applied to obtain the desired dataset. First, the date range of the "Published Date" was selected from 2010-01-01 to 2022-12-31. Second, several flags can be selected to further filter the sample. The only flag selected is "Has Owner". This way only patents where a company has done the application can be selected. Third, for Document Type "Granted patent" and "Patent application" are selected. Fourth, to only extract US patent records, the jurisdiction is set to "United States". Once the search criteria are selected, the website automatically provides information on the patents that meet the requirements. A sample of 707,870 patent records remains.

The maximum number of patent downloads per extract is 50.000. The 707.870 patent records are downloaded in 17 extracts, covering several full months at a time. This process is carried out for the months January 2010 up to and including December 2022. The data has been extracted using the date criteria to ensure that the files containing the extracted patents are mutually exclusive. This approach captures all patents in an exhaustive manner without duplication from double downloads.

Each dataset can be exported to a CSV file with nearly fifty thousand rows of individual patents and with the columns containing information on each patent such as: Lens ID, abstract, priority date, simple family size, legal status, URL, etc. Due to the heaviness of the file, all irrelevant information was deleted. The remaining data of interest is: owners, applicants, citations and publication date. Lastly, all results of all US firms which published a patent between 2010 and 2022 are merged. Unfortunately, The Lens database only provides company names instead of identifiers. How this is solved is further discussed in the Sample Construction Section 3.2.

ESG Data

The second step in the data collection process involved obtaining ESG data, with particular emphasis on the environmental pillar. The ESG data is used for the analysis in Hypotheses 1, 2, 3 and 4. This research makes use of the Refinitiv ESG database, which was previously known as Thomson Reuters ESG Research Data. It is a leading provider of ESG scores. The database has been used to obtain firm-level ESG data from 2010 to 2022. For each instrument, represented by ISINs, several scores are included. For the Environmental Pillar these are: Resource Use score, Emissions score and

Environmental Innovation score. For the Social Pillar these are: Product Responsibility score, Workforce score, Human Rights score and Community score. For the Governance Pillar these are: CSR Strategy score, Management score and Shareholders score.

The scores from Refinitiv are ranked on a 0 to 100 scale, where a higher score represents superior performance in environmental, social, and governance aspects. The scores are calculated on a percentile basis, meaning a score of 70 suggests a company outperforms 70% of its industry peers. Namely, company scores are benchmarked against The Refnitiv Business Classifications (TRBC – Industry Group) for the social and environmental categories and the separate ESG controversies score. For all governance categories, the percentile rank scores are measured against the country in which the firm is incorporated. This research has laid particular emphasis on the environmental scores. Refinitiv (2020) determines industry peers of a firm using a multiple-level hierarchical classification system with increasing levels of detail at each level. For this research, only the Economic Sector is evaluated which is the broadest level, representing high-level sectors like Financials, Energy, or Healthcare. Companies are assigned to these categories based on their primary revenue-generating activities. This system allows for a granular and precise comparison of firms operating in similar areas of business. Although these specific sectors are used for determining individuals ESG scores, this thesis has only considered the economic sector for splitting the sample in the Energy and non-Energy Sector. In order to achieve this, firms have been grouped in their industries first, based on the first 2 digits of the SIC code.

Company Data

The third step in the data collection process was to obtain company data for the required Standard Industry Classification (SIC) codes and firm age. The SIC codes were used for Hypotheses 3 until 7. The most suited identifier for extracting financial data, CUSIP, was used for all hypotheses. Additionally, the company data included information on firm age.

The datafile of US public companies, containing company names, CUSIP codes, SIC codes and the year of the company's initial public offering (IPO), was extracted from Compustat North America. By matching on CUSIP codes, firm age and SIC codes were added to the sample.

Financial Data

The final step in the data collection process was to acquire the firm-level financial data to construct the required control variables. The financial data is used to control for other factors affecting patent output in the analysis of Hypotheses 1 to 7. Two separate data documents were extracted, for the period of 2010 until 2022, using the firms' CUSIPs. Firstly, several financial ratios were obtained through WRDS (Beta): return on assets, leverage ratio, cash ratio and R&D to sales. Secondly, the firm's total assets were obtained from Compustat Capital IQ, Annual fundamentals for North American firms. An overview of these variables is given in 10 Appendix B

3.2 Sample Construction

As discussed in the previous section there are four main datasets: Patent data file (including company names, without identifiers), ESG scores data file (with company names and identifiers), US listed firms data file (with company names and identifiers), Financial data file (with company names and identifiers). The first step was to merge the Patent data file with the ESG scores data file. Secondly, the CUSIPs of the ESG dataset were used to match with the US listed firms' dataset to acquire SIC codes and firm age. Thirdly, using CUSIPs, (financial) control variables could be obtained from Compustat and matched to compile a panel dataset. Ending up with a list of company names of all firms which have published a Y02 classified patent between 2010 and 2022.

The goal was to acquire a dataset containing the number of green patents and citations received on those patents for each company, for each year. More intuitively, a new dataset could be created that counts the total number of published green patents and the sum of citations received for each company for every year. The dataset on ESG scores includes a unique identifier per firm; however, the patent data lacks such an identifier. Hence, the matching process relied on the firm names, which differ for the patent and the ESG data. A combination of techniques was used to maximize the number of matched firms between the two datasets. The matching process was augmented through manual matching and verification.

The first step involved matching the Patent data file with the ESG scores data file. Initially, due to inconsistencies in company names between The Lens and the ESG datasets, fuzzy matching was implemented along with a string-matching loop to compare the similarity of the two strings. Matches were made if the similarity exceeded a predetermined threshold, which was optimized to ensure the highest possible number of correct name combinations while avoiding incorrect matches. When matching the full names of the ESG and the patent data, one difficulty that arose, was assessing the similarity of the full strings. The addition of words like "Inc", "GmbH", "Ltd" etc. means that even for companies that are really the same, their names in the ESG and the patent dataset respectively can be quite dissimilar from a statistical point of view when looking at string similarity. This is particularly relevant in the patent dataset where a lot of words are added onto the company names. Having dropped capitalizations or the inclusion of "Inc.", "GmbH", "Ltd.", new sample checks were conducted. Another problem that arose was that firm names, starting with abbreviations were often matched on close similarity, despite being completely different firms. Fuzzy matching has its weaknesses in the sense that existing companies with very similar names are matched despite being unrelated conglomerates. Additionally, given the enormous amount of patent records and the immense processing power required, each matching attempt took several days to process. In conclusion, many fuzzy matching variations were attempted but most yielded fewer and/or less accurate matches than the alternative method which I shall now discuss.

Fortunately, this alternative, less complex method did yield satisfactory results. More matches were found and sample-based checks showed more accurate matching. Using the first two words of the string “patent applicants” in the Patent data file and the first two words of company names from the ESG data file, a company ID was created. Matching of both files was done based on year and company ID. Initially, each matching attempt was done based on the first word in both IDs. If there was no match, the second word was considered as well. This process eliminated words that did not provide significant information about the firm’s identity (such as “Nucor Corporation”), typically occurring after the firm’s name. In the ESG dataset, only the first words that appeared once were retained to avoid matching multiple ESG firms to the same patent firms. The fact that words in the ESG dataset appear more than once implies they are unlikely to be the most useful for identifying the company. For example, if NextEra Energy, Inc. (NEE) and NextEra Mining, LP (NMI) were to be checked for resemblance, the word “NextEra would be disregarded and the second word would be evaluated, resulting in no match. However, in the Patent dataset, first words that appeared more than once were kept since different variations of a company’s name may exist in the dataset. For example, there are roughly 60 different firm names in the patent database that start with “Ford”, a great many of which are just different ways to refer to the car manufacturer. Namely, patents are very often transferred to or published by different subsidiaries within the same corporation. By matching these words to the unique words in the ESG dataset, it ensured that meaningless words were filtered out to not be included in the final dataset. Afterwards, the first words in the ESG and patent datasets were matched together.

For the second method, only exact matches were initially considered where the first two words in the ESG and patent datasets are the same. Later, some minor adjustments were made to further improve the accuracy of the sample. For instance, all capital letters were converted to lowercase letters and all dots were replaced with spaces. This process was conducted to address potential variations in company naming styles between the ESG and the Patent data file. Moreover, if the firm name in the Patent data file only consisted of one character or started with “Gen”, “The” or “US” observations were dropped before-hand. These character combinations yielded many incorrect matches. It should be acknowledged that despite efforts to match all firms, it is possible that some firms were not matched successfully. However, the chosen matching process reduced the likelihood of introducing systematic biases into the results. There is no apparent reason why firms with specific ESG scores or numbers of patents would be more or less likely to be matched based on first-word-matching. Therefore, even if some firms are excluded from the analysis due to unsuccessful matching, it should not significantly impact the results as long as a sufficient number of firms are matched to yield a robust dataset. This was the case.

An additional transformation was undertaken to investigate the hypotheses. Hypotheses 1, 3 and 5 examine the relationship between ESG scores and the number of published patents. The original dataset considered a published patent as a unit of measurement represented by a single row. To explore this relationship, the data was transformed into a panel data format, aggregating the count of patents published per firm on a yearly basis. This yields a similar dataset as the dataset used for hypothesis 3 in the research by Hoang et al. (2020). For Hypotheses 2, 4, and 6, which investigate the relationship between patent quality (measured by the number of citations received) and the ESG score of the filing firm, a sum was taken of all citations on all patents published by a company in a given year. For the final hypothesis, Hypothesis 7, the product of the number of published patents and the citation count was taken.

For the aggregated Patent-ESG dataset, the 9-digit CUSIPs were converted to 8-digit CUSIPs in order to match with control variable dataset. All financial data was then merged with the aggregated Patent-ESG data, using company ID and Year. Ultimately, the final sample consisted of 2566 firm-year observations. This sample was used for the first 16 regressions of this study. Note that only firm-year observations which held at least one score of any of the ESG metrics were included in the sample. For the final three main regressions (Model 5, 6 and 7) in this study, a more extensive sample was employed. All valid data points of firms with at least one known ESG score between 2010 and 2022 were included as well. Based on this sample, firm-year observations were aggregated based on the first 2 digits of each firm's SIC code and year. The observations were averaged to determine the industry-means of green patents, citations and financial control variables.

3.3 Defining Variables

3.3.1 Dependent Variables

The dependent variables in this research are green patent publications and citations. The patent variable *GreenPatents* is a count variable and naturally represents the quantity of innovation produced within an interval of one year. *Citations* is a count variable and serves as a measure of the technological impact and potential economic significance of inventions. The variable is named "Cited by Patent Count" in Lens. The variable is often referred to as forward citation. The number of citations a patent attracts typically signifies its technological relevance and commercial value, effectively addressing the issue of heterogeneity of patents' value. There is a proven correlation between the value of a patent and the quantity of its forward citations (Lens, 2023). The patent records in this study include all relevant publications between January 1st 2010 until December 31st 2022. The citations are measured as of September 22nd 2023.

For Hypotheses 1, 2, 3 and 4, the patents and citations are counted on firm-level. The variable "$GreenPatents_{i,t}$" denotes the total amount of patent record publications made by firm i in year t.

The variable "$Citations_{i,t}$" denotes the count of patent citations received for all patent records made by firm i in year t. The patent record publications represent the quantity of innovation whereas the citation count indicates the quality of innovation.

For Hypotheses 5, 6 and 7, the patents and citations are counted on industry level. The variable "$AverageGreenPatents_{j,t}$" denotes the industry's average number of patents published per firm in industry j, in year t. The variable "$AverageCitations_{j,t}$" denotes the industry's average number of citations on green patents received per firm in industry j, in year t. The variable "$AveragePublications * Citations_{j,t}$" denotes the industry's average number of patent publications multiplied by the citations on green patents received per firm in industry j, in year t. A comprehensive overview of all variables employed in this study can be found in Table 2.

3.3.2 Independent Variables

There are several independent variables in this study: different variations of the ESG score for the first until the fourth hypothesis and a dummy variable for the Energy Sector, for the fifth, sixth and the seventh hypothesis. As discussed in Section 3.1, the scores are percentile rank ratings against The Refnitiv Business Classifications (TRBC) industries (Refinitiv, 2022). They are thus a relative performance measure of the firm compared to its industry-peers.

For the methodology of Hypotheses 1, 2, 3 and 4 of this research, firms are categorized into two groups based on their ESG scores: those below the 33rd percentile cutoff (low ESG scorers) and those above it (mid/high ESG scorers). This division is applied across all independent variables. This approach focuses on comparing the impact of ESG performance on green innovation between these two distinct groups. A more elaborate explanation for this set-up shall be given in the Section 4.2 Methodology. In the statistical analysis of Hypotheses 1-4, dummy variables are assigned based on the tercile classification of ESG scores. Specifically, firms in the bottom tercile, which represent the category of primary interest, are assigned a dummy variable value of 1. Conversely, firms in the middle and upper terciles are assigned a value of 0 for these dummy variables. This distinction allows for focused analysis on the group with the lowest ESG scores, which is the main subject of this study.

For the first four Hypotheses, four different independent variables shall be used. A more specific overview of the composition of each ESG score is given in Table 3 in the Appendix A.

Firstly, the weighted E score is discussed. The E score aggregates the metrics from the categories of Emissions score, Resource Use score and Environmental Innovation score, which shall be discussed hereafter, to provide an overall assessment of a company's environmental performance. Since the ESG dataset which was used for this study did not offer a score for the environmental pillar, this research uses a self-constructed variable, based on weights given in Refinitiv. According to Refinitiv (2022), the Environmental pillar is compiled of Emissions score for 35.5%, Resource Use score for 35.5% and 29% Environmental Innovation score. The variable "$dLowE_{i,t}$" is a dummy variable that equals 1 for firm i, in year t if the company belongs to the lowest tertile of the weighted E score. Otherwise, if the firm is above the 33rd percentile threshold, the dummy equals 0.

Secondly, the Emission score measures a company's commitment and effectiveness in reducing environmental emissions in production and operational processes. It reflects how well a company is managing its carbon footprint and waste production, the release of pollutants, and overall efforts in emission control. This score accounts for both the volume of emissions relative to the company's size and the effectiveness of their policies and practices aimed at reducing emissions. The score also considers transparency of reporting. To be extra clear, a low Emission score is indicative of high carbon emissions. The variable "$dLowEmissions_{i,t}$" is a dummy variable that equals 1 for firm i, in year t if the company belongs to the lowest tertile of the Emission score. Otherwise, if the firm is above the 33rd percentile threshold, the dummy equals 0.

Thirdly, the Resource Use score reflects a company's performance and capacity to reduce the use of materials, energy, or water and to find more eco-efficient solutions, such as improving supply chain management. This can include initiatives like recycling, sustainable sourcing, and conservation programs. This score evaluates how efficiently a company exploits its resources and is mainly focused on water and energy consumption. The variable "$dLowResourceUse_{i,t}$" is a dummy variable that equals 1 for firm i, in year t if the company belongs to the lowest tertile of the Resource Use score. Otherwise, if the firm is above the 33rd percentile threshold, the dummy equals 0.

Fourth, the Environmental Innovation score is indicative of a company's capacity to reduce environmental costs and burdens for its customers, creating new market opportunities through new environmental technologies and processes or eco-designed products. Concretely, the Environmental Innovation score is partly determined by "Product innovation", which is a self-reported measure. Firms who report to have developed at least one product line or service that is designed to have positive effects on the environment or which is environmentally labeled and marketed earn points for this metric. The second theme which determines the category score for Innovation is a set of financial metrics: Green revenues, green research and development and green capital expenditures (Capex). The variable "$dLowEInnovation_{i,t}$" is a dummy variable that equals 1 for firm i, in year t if the company belongs to the lowest tertile of the Environmental Innovation score. Otherwise, if the firm is above the 33rd percentile threshold, the dummy equals 0.

To conclude the first four predictors; the first score is the Environmental Pillar score which aggregates the second, third and fourth "sub-score". These are included to measure a more nuanced effect and to see how each sub-score influences green innovation output and quality. All scores are evaluated for hypotheses 1, 2, 3 and 4 with the only difference being the data that is fed into the regression.

For Hypotheses 5, 6 and 7 another ESG measure is taken as the main predictor. The variable "$dEnergySector_j$" is a dummy variable that equals 1 for industry j, if the SIC belongs to the Energy

Sector, thus for companies that are considered to operate in unsustainable industries. Otherwise, the dummy equals 0, for all other industries.

As explained more elaborately in Section 2.3, Cohen et al. (2020) categorize industries for which the first two digits of their Standard Industrial Classification (SIC) are 10 (Metal, Mining), 12 (Coal Mining), 13 (Oil & Gas Extraction), 14 (Non-metallic Minerals, Except Fuels), 29 (Petroleum & Coal Products), or 49 (Electric, Gas, & Sanitary Services) as being part of the Energy Sector. However, this study excludes SIC 49, primarily related to energy utilities and services, from its analysis. Over the entire sample, the number of patents (Hypothesis 5), citations (Hypothesis 6) and the combined patent metric (Hypothesis 7) are aggregated to distil year-averages per industry. Following Cohen et al. (2020), it employs a dummy variable to assess the impact of sectoral affiliation on green patent outputs. This requires averaging all control variables for all industries. Diverging from Cohen et al., this thesis also categorizes firms into 'low ESG' and 'mid/high ESG' groups, with the division line set at the 33rd percentile. This approach facilitates testing of Hypotheses 1 through 4.

3.3.3 Control Variables

Since the outcome variables of this study are influenced by more factors than the ESG predictors, various control variables are included in the analysis. These control variables include firm size, leverage, Return on Assets (ROA), cash, Research and Development (R&D) to sales and firm age. The specific terms for finding these variables in various databases are explained in Appendix A Table 4. Previous literature has considered these control variables as potential contributors to patent production. An explanation of these variables and their expected impact on patent production is provided below.

i. Ln(Assets):

The variable "$Ln(Assets)_{i,t}$" denotes the natural logarithm of the book value of total assets of firm i at the end of year t. It represents the size of the firm. Assets represent the total value of assets reported on the balance sheet and the natural logarithm is taken to control for outliers. Research by Lin et al. (2019) in the automobile industry shows that smaller firms are more prone to exploit resources which they have access to. This led to smaller-sized firms showing higher green innovation investment returns. However, larger companies are more likely to have favorable conditions and resources that increase the chances of successful innovation (Schumpeter, 1942). Additionally, Andersen et al. (2019) find that smaller firms face resource constraints (own fewer assets) and stronger competition. Consequently, they are more hesitant when considering investments in environmental sustainability innovation, as these may inadvertently increase operational costs and potentially compromise their competitive edge. Following this reasoning, a positive relationship between firm size and innovation is expected (+).

ii. Leverage

The variable “$Leverage_{i,t}$” denotes the ratio of book value of total debt as a fraction of the book value of total assets of firm i at the end of year t. This is the most common way of compiling the variable Leverage (Cohen, Gurun & Nguyen, 2020; Wang et al., 2023; Lin, 2023). The variable Leverage indicates what share of a firm’s assets is financed with debt. Intuitively, as the debt ratio increases, interest expenses too increase. Due to reduced cash flow, R&D expenses too decrease, thereby curbing innovation. Furthermore, Leverage serves as an indicator of the firm's financial health and, when considered alongside other factors, signals to financial institutions whether the firm has the ability to assume additional debt. This ability can be referred to as debt flexibility and enables firms to invest, should a good investment opportunity come by. Leverage is a measure of financial constraint (Arugaslan & Miller, 2006). Since financial constraints are found to impede firm innovation (Savignac, 2008), it is likely that there is a negative relationship between Leverage and innovation. (-)

iii. ROA

The variable “$ROA_{i,t}$” denotes Operating income before depreciation as a fraction of total assets of firm i at the end of year t. ROA is a measure of a firm’s profitability and can influence innovation in several ways. For instance, a strong ROA lays the groundwork for long-term strategic planning. Companies with solid profits have the luxury to strategize for the long haul, including channeling resources into inventive initiatives that might take years to bear fruit. Moreover, profitable businesses are typically more appealing to skilled employees, the driving force behind innovation. Such companies can offer appealing remuneration, invest in talent development, and develop a workspace that stimulates creative thought and experimentation. As shown in the Meta table (Table 1), various previous studies on the relation between ESG and innovation use ROA as a control variable. Dicuonzo et al. (2022) too use company profitability, as measured by ROA but did not find significant results. However, Wang et al. (2023) find a coefficient of 0.136 of ROA on green patents, significant at the 1% level. Thus, ROA is assumed to positively impact firm innovation (+).

iv. Cash

The variable “$Cash_{i,t}$” denotes the ratio of cash and short-term investments as a fraction of current liabilities of firm i at the end of year t. The Cash ratio is a measure of liquidity and can too imply a stronger financial cushion. Consequently, firms with higher Cash ratios are often better positioned to take risks associated with innovation. Innovation, especially in the field of eco-friendly technology, is frequently accompanied by considerable risk and unpredictability. A firm possessing robust financial resources is more capable of managing the risks linked to the development of novel environmental technologies and procedures, subsequently resulting in the generation of superior quality patents. Interestingly, Cohen et al. (2020) only find Cash to have a significant influence on Citations, yielding a 0.280 coefficient for an OLS Model and a stronger coefficient of 0.519 when applying a Poisson Model, both significant at the 1% level. Their research did not find a significant relation for innovation output, measured by green patent applications. However, intuitively, a positive relationship is expected between Cash ratio and both measures of innovation (+).

v. R&D

The variable "$R\&D_{i,t}$" denotes the ratio of research & development expenditures as a fraction of sales of firm i at the end of year t. Wang et al. (2023) calculate R&D as the ratio of R&D expenditures to sales revenue as well, whereas Lin (2023) take total assets as the denominator. Albeit R&D is a suboptimal measure of innovative output, it is worth including it as control variable, to isolate the interplay of ESG scores and green patent production. Engaging in research & development should increase the probability of introducing new innovations to the market. Therefore, R&D is expected to have a strong positive impact on patent output and citations. (+)

vi. Age

The variable "$Age_{i,t}$" denotes the year of patent publication, minus the number of years that firm i has been listed on the exchange resulting in year t. Firm age is computed by subtracting the year of the company's initial public offering (IPO) from the year in which the patent is published. The maximum firm age is at 72 because 1950 was the cutoff in the database. In most research papers, firm age is taken as the number of years since the company's establishment (Cohen, 2020). However, in studies involving publicly traded companies, the term "age" refers to the number of years since the firm's stock became available for public trading (Wang et al., 2023). Since the panel data set of this thesis only includes US listed firms, the age measurement is chosen accordingly. The literature presents mixed findings regarding the impact of firm age on the relationship between ESG scores and firm innovation. Some studies suggest that older firms tend to be more innovative (Withers et al., 2011), while others indicate that younger firms exhibit higher levels of innovation. In general, one could expect older firms to be more innovative by having more experienced employees and more established relations with key stakeholders. More specifically, a split could be made between quantity (patents) and quality (patent citations) of innovation. According to Kotha et al. (2010), when older firms venture into new technological niches like green innovation, they tend to produce a greater quantity of innovative output compared to younger firms. This would mean that firm age has a (+) relation to the amount of patent publications. Yet, Coad et al. (2016) discovered that younger firms engage in more radical innovation activities, which offer greater rewards if successful, potentially leading to higher cited patents. For that reason, the relationship between firm age and patent citations is expected to be negative (-).

vii. Year Fixed Effects

The variable "$YearFE_t$" denotes a set of dummy variables representing each year t for 2010 to 2022. These variables are typically used to control for specific time-related effects that could potentially impact the dependent variable, thereby safeguarding the results from being skewed by distinctive characteristics or events unique to a specific year. In this study, the primary rationale behind including Year Fixed Effects is to mitigate a natural bias. Considering the nature of our dependent variables, YearFE play a vital role in controlling for inherent time-related biases. Namely, over time, patents inevitably accumulate more citations. Without accounting for this natural progression, our analysis could incorrectly attribute the increase in citations to ESG scores rather than the mere passage of time. Thus, YearFE helps ensure that our findings accurately reflect the relationship between ESG scores and innovation, independent of the confounding effect of time.

viii. Averages

For the regression analyses in Hypotheses 5, 6 and 7 for each year, for each control variable, industry-averages are taken. The methodology employed here closely follows that of Cohen et al. (2020), with two notable exceptions: the adoption of a negative binomial model and the incorporation of additional control variables. In this approach, companies are cumulated based on their SIC codes, facilitating the calculation of the average number of patent publications and citations per industry per year. Each of these industry-year averages constitutes a single observational data point. The underlying logic of this method is multifaceted. Aggregating companies based on SIC codes to calculate industry-year averages for patent publications and citations allows for a more balanced comparison between the Energy and non-Energy Sector. This method effectively mitigates the skewed distribution of firm-year observations across various industries. By using industry averages, each industry is given equal weight, irrespective of the number of firm-year observations it contains. This approach avoids the overrepresentation of certain industries, ensuring that the analysis reflects sector-wide innovation trends, and not just those of a few dominant industries. For example, the industries with the first two-digit SIC codes 28, 35, 36, 37 and 38 contain 1751 firm-year observations, which is 48,4% of the non-Energy Sector population. An elaborate overview of the sample size, mean patent publications and standard deviation can be found in Table 5 Appendix B. Running a regression in this manner essentially leads to a comparison between the Energy Sector and these specific 5 SIC codes, rather than a comparison between the Energy Sector and the non-Energy Sector. Closer alignment with the research question is achieved by calculating industry averages and subsequently comparing the Energy and the non-Energy Sector. The variable descriptions can be found in Table 2. Additionally, the expected direction of the relation between the averages of the predictors and the outcome variables are expected to be the same on industry-level as on firm-level.

Table 2: Variable definitions

Variable	Definition
Measure of Innovation	
$GreenPatents_{i,t}$	Number of individual green patents published by firm i in year t.
$Citations_{i,t}$	The total number of firm i's citations received on the firm's green patent publications in year t.
$AverageGreenPatents_{j,t}$	The average number of green patents published per firm in industry j, in year t.
$AverageCitations_{j,t}$	The average number of citations on green patents received per firm in industry j, in year t.
$AveragePatents \times Citations_{j,t}$	The average number green patents published, multiplied by the number of citations on green patents received per firm in industry j, in year t.
Measures of ESG	
$dLowE_{i,t}$	Dummy variable, equals 1 for firm i, in year t if the company belongs to the lowest tertile of weighted Environmental score. Otherwise, if the firm is above the 33rd percentile threshold, dummy equals 0.
$dLowEmissions_{i,t}$	Dummy variable, equals 1 for firm i, in year t if the company belongs to the lowest tertile of Emission score. Otherwise, if the firm is above the 33rd percentile threshold, dummy equals 0.
$dLowResourceUse_{i,t}$	Dummy variable, equals 1 for firm i, in year t if the company belongs to the lowest tertile of Resource Use score. Otherwise, if the firm is above the 33rd percentile threshold, dummy equals 0.
$dLowEInnovation_{i,t}$	Dummy variable, equals 1 for firm i, in year t if the company belongs to the lowest tertile of Environmental Innovation score. Otherwise, if the firm is above the 33rd percentile threshold, dummy equals 0.
$dEnergySector_j$	Dummy variable, equals 1 for industry j, if the SIC belongs to the Energy Sector. Otherwise, dummy equals 0.
Measures of Controls	
$Ln(Assets)_{i,t}$	The natural logarithm of the book value of total assets of firm i at the end of year t.
$Leverage_{i,t}$	The ratio of book value of total debt as a fraction of the book value of total assets of firm i at the end of year t.
$ROA_{i,t}$	Operating income before depreciation as a fraction of total assets of firm i at the end of year t.
$Cash_{i,t}$	The ratio of cash and short-term investments as a fraction of current liabilities of firm i at the end of year t.
$R\&D_{i,t}$	The ratio of R&D expenditures as a fraction of sales of firm i at the end of year t.
$Age_{i,t}$	The year of patent publication, year t minus the number of years that firm i has been listed on the exchange.
$YearFE_t$	A set of dummy variables representing each year t for 2010 to 2022.
$AverageLn(Assets)_{j,t}$	The average natural logarithm of the book value of total assets of firms in industry j, at the end of year t.
$AverageLeverage_{j,t}$	The average ratio of book value of total debt as a fraction of the book value of total assets of firms in industry j, at the end of year t.
$AverageROA_{j,t}$	The average operating income before depreciation as a fraction of total assets of firms in industry j, at the end of year t.

$AverageCash_{j,t}$	The average ratio of cash and short-term investments as a fraction of current liabilities of firms in industry j, at the end of year t.
$AverageR\&D_{j,t}$	The average ratio of R&D expenditures as a fraction of sales of firms in industry j, at the end of year t.
$AverageAge_{j,t}$	The average number of years that firms in industry j have been listed on the exchange at the year of their patent publications, year t.

3.4 Descriptive Statistics

3.4.1 Summary Statistics

The following section discusses the descriptive statistics of the variables incorporated in the current study. The characteristics of these variables shall be most elaborately discussed in their state, prior to their consolidation at the industry level. Table 6 presents the descriptive statistics of the variables as they will be regressed in Models 1a-1d (addressing Hypothesis 1) and Models 2a-2d (addressing Hypothesis 2). These are essential in investigating the correlation between ESG scores and firm-level innovation, as measured through patents and patent citations. Further refinement is presented in Table 7, located in Appendix B, where a distinction is drawn to compare firms within the Energy Sector against those outside of it, providing a view of the sample pertinent to Models 3a-3d (for Hypotheses 3) and Models 4a-4d (for Hypotheses 4).

In Table 8 Appendix B, the descriptive statistics are presented for the aggregated dataset. These data points are aggregated per SIC to calculate industry-year averages and serve as the foundation for the regression analyses of Models 5, 6 and 7. Complementing this, Table 9 in Appendix B offers a comparative analysis between the two subpopulations of the Energy and non-Energy Sector, to give insight into the data concerning fifth, sixth and seventh Hypotheses.

Given that the aggregated data in Table 8 and Table 9 are derived from firm-year observations, Table 6 will receive a more thorough discussion in the forthcoming section, providing the foundational data from which subsequent aggregations are built. Thus, if not mentioned otherwise, the statistics discussed in the text of Section 3.4.1 can be assumed to be found in Table 6. The industry condensations shall be more briefly discussed in the final paragraphs of this section.

<u>Patents</u>

The mean number of annual patent publications per firm is at 20.91 (SD = 80.95). Naturally, the minimum amount of patent publications in the sample is 1, since the data initially extracted is a list of patent records. The values span a broad range, with the maximum GreenPatents count reaching 1835, signifying a substantial variation in the volume of patents published by diverse firms. Notably, there is a significant discrepancy between the median (*Mdn* = 4) and the mean (*M* = 20.91), which is indicative of substantial skewness in the data. Indeed, skewness amounts to 11.83, meaning the data is heavily skewed. This suggests that a minority of firms are responsible for the majority of patent production. The kurtosis value is extraordinarily high at 188.18, indicating a "leptokurtic" distribution. This means the distribution has a sharp peak and heavy tails, suggesting substantial outliers with values considerably higher than the majority of the data.

Firms have an average of 154.23 citations, with a significant dispersion (*SD* = 569.77), almost six times as large as the mean. The lower limit for the number of citations is zero, as there exist patent records that are not referenced by any other patents, potentially suggesting minimal innovative significance. The maximum number of citations attributed to a firm amounts to 9322. Citations, like GreenPatents, exhibit a right-skewed and leptokurtic distribution, though less pronounced. This indicates a sharp peak and fat tails in the data, with most firms having low citation counts and a few outliers with exceptionally high counts.

In conclusion, there is substantial variation in the innovative performance of North American firms. This is demonstrated by the high standard deviations for both innovation output measures: patent publications (*SD* = 80.95) and patent citations (*SD* = 569.77). Moreover, given that observations are not normally, but leptokurtic distributed, the methodological framework shall account for this accordingly. This shall be further discussed in Section 4.1 where the decision for the negative binomial model is discussed.

ESG scores

In analyzing the environmental performance of firms, our study reveals noteworthy findings. The weighted E score, which is compiled of three sub-metrics, has an average score of 58.51 (*SD* = 20.11). The score indicates that a company is performing better than 58.51% of its peers and is characterized by a standard deviation of 20.11, indicating a moderate level of variability among firms. The Resource Use score, averaging 55.65, indicates the average firm in the sample surpasses 55.65% of its industry peers in resource efficiency, with scores ranging widely from 0.2 to 99.9. The Emissions score, at an average of 52.00 exhibits relatively high variability (*SD* = 29.11), reflecting the differences in emissions-related performances across firms. Conversely, the Environmental Innovation score with an average of 51.331 and the lowest standard deviation of 24.36 among the sub-scores, indicates moderate variability in innovation activities among firms.

The total sample size for our analysis, including all control and independent variables, encompasses 2566 observations. However, the sample size for the weighted E score is smaller, with 1487 observations, due to the requirement of having all three sub-scores for its computation.

Financial and Operational Metrics

Having inspected the percentiles and distribution of the data, extreme outliers were identified within the financial ratio control variables. To effectively control the potential skewing effect of these outliers, a winsorization was applied at the 2.5^{th} and 97.5^{th} percentiles. Consequently, all ratios underwent this winsorization process to ensure data consistency. Additionally, to further mitigate the impact of outliers present within the firm size variable, the natural logarithm was applied resulting in Ln(Assets). The natural logarithm of total assets has a mean value of 8.67 and a maximum value of 13.22, meaning the average firm in the sample has a book value of total assets amounting to 5.8 billion USD and a maximum value 551.8 billion USD, in absolute terms.

Leverage across firms shows a moderate mean value of 0.57, with a broad range from 0.10 to 1.17, indicating diverse financial structures. Despite the broad range of values that Leverage takes on, the distribution of the data exhibits acceptable levels of skewness and kurtosis for both Ln(Assets) and Leverage. The skewness values are close to zero (-0.15 and 0.23 respectively) and kurtosis values fall within a range that implies normality.

The average ROA is 0.119, indicating that on average, firms generate a return of 11.9% on their assets. The average return on assets is very close the median of 12.9%. However, the variability is notable, stretching from -0.57 to 0.32.

The mean cash ratio (Cash) is 1.08. Despite winsorization, the maximum cash ratio is still high with a range from 0.04 to 13.83, indicating high variability in the liquidity positions of firms. Moreover, high levels of kurtosis are observed in the return on assets (ROA) and cash ratio (Cash), standing at 13.87 and 22.89 respectively. This pronounced kurtosis, especially in the case of Cash, indicates a distribution with tails that are more substantial than those of a normal distribution. The distributions are further characterized by the skewness detected for ROA and Cash, which are -2.56 and 3.84 respectively. Such outcomes are not uncommon among publicly traded firms, which often demonstrate relatively uniform return profiles due to standardized market and industry practices.

The mean R&D to sales ratio is 0.12, showing the degree of investment in research and development in relation to sales. The extreme kurtosis of 156.20 and maximum value of 8.07, despite winsorization, are mainly caused by an over-presence of zero values in the sample. The data shows a strong "zero-inflation" effect as can be seen by the zero values for "Min" "p5" and "p25" in Table 6. Further investigation showed that for 31.0% of firms the variable R&D equals zero and that 43.3% of firms, invest less than 1% of their sales revenue in R&D. The results could be caused by several factors. Firstly, it could simply be that many of the firms in the sample do not really engage in R&D. This could be the case for companies in the industries such as: Construction and Real Estate, Retail and Wholesale Trade, certain Service Industries or Transportation. Nevertheless, these are also present in the sample and have relevance in researching the relation between ESG ratings and innovation. Secondly, R&D expenditures are not always properly reported in the financial statements and can thus be inaccurate in Compustat. Thirdly, given that R&D is a ratio, it could be that firms in the sample have very low sales revenue, thereby inflating the R&D to sales variable. These firms could be companies that have only recently been IPO'd and are still in a growth phase. Since these firms are also representative for a comprehensive image of the US stock market and their capacity to innovate, they are not excluded from the sample.

The mean Age of the companies included in this study is 33.8 years. The sample contains both established and younger companies. The maximum firm age is at 72 because 1950 was the cutoff in the database.

Table 6: Summary statistics of sample containing firm-year observations

Variables	N	Mean	SD	Min	p5	p25	Median	p75	p95	Max	Skewness	Kurtosis
GreenPatents	2566	20.91	80.95	1	1	2	4	12	76	1835	11.83	188.18
Citations	2566	154.23	569.77	0	0	3	16	80	642	9322	8.80	99.52
E	1487	58.51	20.11	10.96	23.15	43.02	61.01	75.16	86.99	97.44	-0.31	2.12
Emissions	2266	52.00	29.11	0.18	6.41	26.40	52.18	77.54	96.32	99.82	-0.03	1.76
ResourceUse	2296	55.65	29.75	0.20	6.16	30.51	57.69	83.25	97.37	99.90	-0.20	1.78
EInnovation	1662	51.33	24.36	0.49	12.38	34.62	50.00	71.62	92.15	99.32	0.09	2.14
Ln(Assets)	2566	8.67	1.77	1.06	5.85	7.50	8.57	9.91	11.63	13.22	-0.15	3.31
Leverage	2566	0.57	0.21	0.10	0.20	0.44	0.56	0.69	0.92	1.17	0.23	3.40
ROA	2566	0.12	0.12	-0.57	-0.06	0.09	0.13	0.17	0.28	0.32	-2.56	13.87
Cash	2566	1.08	1.62	0.04	0.06	0.25	0.52	1.19	4.06	13.83	3.84	22.89
R&D	2566	0.12	0.57	0	0	0	0.02	0.09	0.32	8.07	11.86	156.20
Age	2566	33.81	20.38	1	6	18	28	53	68	72	0.39	1.86

Note: this table shows the descriptive statistics of firm-year observations of the dependent, independent and control variables over the entire sample period of 2010 to 2022. The descriptives of GreenPatents and Citations can be interpreted straightforward as number of green patents and citations. The variables E, Emissions, ResourceUse and EInnovation represent ESG scores between 0 and 100. Firm size Ln(Assets) represents the natural logarhythm of total assets in billions. The leverage ratio (Leverage), return on assets by sales (ROA), cash ratio (Cash), research & development investment relative to sales (R&D), each represent ratios and should be interpreted accordingly. The firm age (Age) is denoted in years.

Population-split of firm-year sample (Models 3 and 4)

In Table 7 of Appendix B, an additional split is made so the difference between non-Energy Sector and Energy Sector can be observed for the firm-level data. This way, the summary statistics of the sample for Models 3a-3d (Hypotheses 3) and Models 4a-4d (Hypotheses 4) are shown.
The average number of green patent publications is 23.07 (*SD* = 34.31) for Energy Sector firms and 20.73 (*SD* = 83.57) for non-Energy companies. However, the maximum for GreenPatents is much higher for the non-Energy Sector (*Max* = 1835) than for the Energy Sector (*Max* = 167). For Citations, the average count is 220.50 (*SD* = 460.63) for Energy Sector firms and only 148.93 (*SD* = 577.63) for non-Energy companies. Again, the maximum count is much higher for the non-Energy

population (*Max* = 9322) compared to the Energy Sector (*Max* = 3898). The higher maxima for the non-Energy population are attributed to a larger number of firms in this population, leading to greater variance and a higher likelihood of outliers.

It is striking that the mean E score and EInnovation score in the Energy Sector (73.43 and 65.60 respectively) are much higher than these ESG metrics in the non-Energy Sector (57.67 and 50.55 respectively). Contrarily, the Resource Use and Emission score of both sectors are of similar magnitudes. The discrepancy between the means of the dependent variables between the two populations does not pose an issue since the non-Energy Sector is disregarded in Models 3 and 4.

For the financial and operational control variables, there are several differences between the two sectors. Firms in the Energy Sector are on average larger *(M* = 1.15) compared to the non-Energy firms ($M_{non\text{-}Energy}$ = 0.29). Besides, relative to non-Energy firms, Energy Sector firms have a lower leverage ratio ($M_{non\text{-}Energy}$ = 0.003 and M_{Energy} = *-0.099* respectively), a lower cash ratio ($M_{non\text{-}Energy}$ = -0.677 and M_{Energy} = *-1.054* respectively), and lower R&D expenditures by sales ($M_{non\text{-}Energy}$ = -0.27 and M_{Energy} = -0.396 respectively). On average, firms in the Energy Sector are 3.16 years older (M_{Energy} = 7.71) than firms in the non-Energy Sector ($M_{non\text{-}Energy}$ = 4.55). The return on assets (ROA) of the populations do not differ much.

Aggregate industry-year sample summary (Models 5, 6 and 7)

Table 8 of Appendix B gives insight into the aggregate data sample where the datapoints are industry-year observations. This way, the summary statistics of the sample for Models 5, 6 and 7 are shown. To acquire this dataset, all firm-year observations were cumulated per industry (based on SIC), per year. After, the average values for GreenPatents, Citations and the financial control variables were calculated per SIC. The result is a sample of 515 non-missing observations. In Table 9 of Appendix B, the data of Table 8 is split into two populations: the Energy Sector and the non-Energy Sector. The average number of green patent publications is 21.10 (*SD* = 18.81) for the Energy Sector and 12.27 (*SD* = 27.17) for the non-Energy Sector. In line with the non-aggregated sample containing firm-year data, the maximum for AverageGreenPatents is almost 5 times higher for the non-Energy Sector (*Max* = 299) compared to the Energy Sector (*Max* = 65.25). For AverageCitations, the count is 173.17 (*SD* = 206.17) for Energy Sector firms and only 104.46 (*SD* = 278.80) for non-Energy companies. Again, the maximum count is much higher for the non-Energy population (*Max* = 3821) compared to the Energy Sector (*Max* = 664). The financial and operational control variables are very similar for both populations.

3.4.2 Correlation Statistics

The correlation Table 10 shows the Pearson Correlations (under the diagonal) and Spearman Correlations (above the diagonal). The comparison of Pearson (pairwise) and Spearman (non-parametric) coefficients yields insights into the relationships among the variables within the sample. The data for which the correlation statistics are displayed in Table 10 is on univariate level. Potential multicollinearity shall be separately discussed in Section 4.4.1 where several necessary assumptions for the regression model are evaluated.

The number of green patent publications (GreenPatents) and the citations received (Citations) showcase a strong relationship. The two dependent variables have a Pearson correlation coefficient of .76 (p<.01) and a Spearman coefficient of .78 (p<.01). This correlation is explained by the fact that firms with a higher count of green patents tend to accrue more citations, simply because they have a higher number of patents to earn citations on. Since these variables serve as dependent variables in separate regression analyses, not as predictors in the same model, their high correlation does not influence the estimation of regression coefficients or compromise the validity of the model's results.

When examining the relationship between Green Patents, and the dummy variables representing low ESG scores (dLowE, dLowResourceUse, dLowEmissions, dLowEInnovation), there is a consistent negative association in both Pearson and Spearman correlations. For example, the negative correlations for GreenPatents with dLowE (Pearson: r = -.14, p<.01; Spearman: r = -.22, p<.01) suggest that firms with lower environmental performance tend to publish fewer green patents. All correlation coefficients of the ESG metrics with GreenPatents are significant at the 1% level except dLowEInnovation which is insignificant. Turning to Citations, similar trends are observed with the ESG scores, where lower scores are correlated with fewer citations. Specifically, dLowE shows a Pearson coefficient of -.13 (p<.01) and Spearman coefficient of -.18 (p<.01), which may imply that firms with weighted E scores are cited less. Albeit the Spearman and Pearson correlations are not far apart, the rank-orders tend to show stronger relations.

The firm size, as measured by Ln(Assets), shows a weak positive correlation of r = .25 (p<.01) for both Green Patents and Citations in the Pearson analyses. The Spearman correlations for firm size (Ln(Assets)) are slightly stronger with r = .39 (p<.01) for GreenPatents and with r = .40 (p<.01) for Citations. The robust Spearman correlations indicate that when the influence of outliers is minimized and the linear assumption is relaxed, the connection between firm size and both patent output and citation counts becomes more pronounced. Although the relationship may not be strictly linear, there is a consistent pattern of larger firms typically publishing more patents and receiving more citations. The Spearman method, with its resilience to outliers and non-linear relationships, captures this trend more accurately. Leverage does not show a significant relationship with Citations in the pairwise correlation analysis but has a weak negative correlation in the Spearman analysis (r = -.04, p<.10), indicating that higher debt levels might marginally impact the firm's citation count. The ROA's positive, strongly significant correlation coefficient of .06 in Pearson and .16 for Spearman shows that more profitable firms are associated with higher citation counts. The relation between ROA and is weaker for GreenPatents, albeit strongly significant still for the Spearman model. Interestingly, the correlation between Citations and R&D is not significant in the Pearson analysis but is positively significant in the Spearman analysis (r = .18, p<.01). This could suggest that while there isn't a linear relationship between R&D spending and citations, there is a monotonic relationship when considering the ranks of the data. Similarly, for GreenPatents, only the Spearman correlation coefficient shows significance (r = .23, p<.01) for its relation with R&D. These results are likely caused by the significant outliers, which were discussed in Section 3.4.1. The Age of the firm shows a consistent positive correlation in both analyses, for both outcome variables, indicating that older firms tend to publish more green patents and also receive more citations.

In conclusion, while both Pearson (pairwise) and Spearman (non-parametric) correlations offer valuable insights, the Spearman method seems to reveal stronger and potentially more reliable relationships within the data. The consistently high significance levels across most correlations confirm the robustness of these findings.

Table 10: Pearson Correlations (under the diagonal) and Spearman Correlations (above the diagonal) of regression variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) GreenPatents	1.000	.779***	-.223***	-.16***	-.172***	-.102***	.394***	.065**	.08***	.022	.227***	.137***
(2) Citations	.762***	1.000	-.176***	-.104***	-.128***	-.071***	.401***	-.037*	.155***	.038*	.18***	.125***
(3) dLowE	-.141***	-.133***	1.000	.654***	.614***	.322***	-.499***	-.111***	-.028	.101***	.013	-.176***
(4) dLowResourceUse	-.119***	-.121***	.654***	1.000	.654***	.614***	.322***	-.499***	-.111***	-.028	.101***	.013
(5) dLowEmissions	-.126***	-.125***	.614***	.520***	1.000	.091***	-.401***	-.071***	-.032	.064**	.025	-.142***
(6) dLowEInnovation	-.039	-.024	.322***	.130***	.085***	1.000	-.188***	-.029	.102***	.087***	.097***	-.03
(7) Ln(Assets)	.254***	.254***	-.486***	-.427***	-.505***	-.168***	1.000	.191	.03	-.145***	-.104***	.358***
(8) Leverage	.057***	-.022	-.123***	-.170***	-.139***	-.063**	.258***	1.000	-.106***	-.486***	-.298***	.089***
(9) ROA	.030	.060***	-.046*	-.075***	-.169***	.057**	.326***	.043**	1.000	.145***	.101***	.096***
(10) Cash	-.019	.012	.091***	.096***	.109***	.120***	-.276***	-.466***	-.338***	1.000	.559***	-.27***
(11) R&D	-.004	-.004	.029	.008	.104***	.054**	-.194***	-.125***	-.467***	.411***	1.000	-.175***
(12) Age	.130***	.109***	-.167***	-.242***	-.238***	-.053**	.407***	.113***	.205***	-.263***	-.156***	1.000

Note: The pairwise correlations (Pearson) are shown in the lower triangle and the non-parametric correlations (Spearman) in the upper triangle. Significance levels are denoted by asterisks: * p<.10; ** p<.05; *** p<.01.

CHAPTER 4 Methodology

4.1 Negative Binomial Model

For all hypotheses, the Negative Binomial Model (NBM) is used. The selection of the NBM for regression analysis is grounded in both the nature of the data and the specific requirements of the hypotheses under investigation. Several models were considered before selecting the NBM.

Firstly, the potential applicability of a pooled ordinary least squares (OLS) regression was considered. This model is typically suited for analyzing datasets where the dependent variable is continuous and follows a logarithmic distribution. However, the primary dependent variables in this study are count data, specifically the number of green patents published and their associated citation counts. Count data inherently follow a discrete distribution, making an OLS regression less suitable.

Secondly, the Poisson regression model appeared to be a viable alternative, given its suitability for count data, which theoretically follow a Poisson distribution. However, a critical assumption of the Poisson model is the equivalence of the mean and variance. In our dataset, this assumption does not hold, as evident from the summary statistics: the mean of patent publications is 2.91, with a considerably higher standard deviation of 8.95. Similarly, for patent citation counts, the mean stands at 154.23 with a standard deviation of 569.77, indicating over-dispersion in the data.

Lastly, the NBM was considered. This method addresses the issue of over-dispersion by allowing for a variance that exceeds the mean. This is a common situation in count data where the occurrence of an event can vary significantly across observations. This tolerance for over-dispersion makes it particularly well-suited for our data, where the standard deviation significantly exceeds the mean for both patent publications and citation counts. Given the extreme kurtosis and high skewness as shown in the summary statistics of Table 6, a method such as the NBM is thus required in this research. Moreover, the NBM is adaptable to scenarios where the dependent variable is a count variable, such as the average counts of patents and patent citations in Hypotheses 5, 6 and 7. These hypotheses involve a binary independent variable representing industry classification, specifically comparing the Energy Sector against non-Energy Sector.

Further validation of the NBM’s appropriateness was obtained through the examination of deviance residuals, which showed satisfactory results. Therefore, the negative binomial regression model was selected as the most appropriate for analyzing the relationship between ESG scores and firm-level innovation in the context of this study.

4.2 Grouping

In the methodology of this research, an approach is adopted which involves the comparison of groups based on ESG scores. Specifically, for each regression analysis, firms were categorized into two distinct groups: those with lower ESG scores, falling below the 33rd percentile cutoff, and those with higher ESG scores, above this threshold. This split was consistently applied across various ESG metrics, including the weighted Environmental score, Emissions score, Resource Use score, and Environmental Innovation score. This split yielded the variables dLowE, dLowEmissions, dLowResourceUse and dLowEInnovation. Note throughout the entire thesis that a negative relationship between a firms' ESG performance and innovation is equivalent to a positive relationship between the variables low ESG and innovation.

The splitting of the data into two groups is justified in two ways. Firstly, the nature of ESG scores as percentile ranks inherently enables this division. Ranking the firms against their industry peers allows for a split between lower-performing and higher-performing firms in terms of ESG criteria. Besides, the skewness values ranging between -0.31 and 0.09 suggest that the ESG scores are fairly symmetrically distributed, with no significant skewness to the left or right. Kurtosis values between 1.76 and 2.14 indicate a distribution that is not excessively peaked or flat. These statistics imply that the ESG scores do not exhibit extreme values or asymmetry, which allows for group-splits.

The desirability of segmenting firms into low and mid/high ESG groups is justified by several strategic reasons. Firstly, this grouping strategy improves the clarity and interpretability of the results. By comparing the lowest ESG scorers with their higher-scoring counterparts, the research can more effectively isolate and examine the impact of being burdened with a low ESG rating on green innovation. This contrasting approach aligns well with investment practices in the real world, where investors often categorize firms into "sustainable" or "non-sustainable" based on certain ESG score thresholds. By mirroring this real-world categorization, the research design gains practical relevance and applicability.

In conclusion, this method makes the results more straightforward for policy implications and business strategy formulation. The comparative analysis between these two distinct groups can shed light on whether there is a threshold effect in ESG ratings impacting innovation, which is a critical aspect for both investors and policymakers.

4.3 Regressions Equations

4.3.1 Firm-level Regressions

Multiple regressions have been conducted to test the effect of various ESG scores on green innovation output and quality. As mentioned earlier, four measures of ESG have been used for the independent variable: the weighted E score, the Emission score, the Resource Use score and the Environmental

Innovation score. The independent variable of green innovation will take on two forms: the count of patents published and a count of citations. To isolate the effect of ESG score on firm innovation, six control variables were added to each regression: firm size (Ln(Assets)), leverage ratio (Leverage), return on assets by sales (ROA), cash ratio (Cash), research and development to sales (R&D) and firm age (Age). Additionally, Year Fixed Effects were included.

Regressions 1a, 1b, 1c and 1d were conducted to test Hypothesis 1: *Low E (ESG) firms publish more green patents than high E (ESG) firms*.

Testing Hypothesis 1 with weighted E score as a measure of E (ESG):
1a: $GreenPatents_{i,t} = \alpha_0 + \beta_1 dLowE_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

Testing Hypothesis 1 with Emission score as a measure of E (ESG):
1b: $GreenPatents_{i,t} = \alpha_0 + \beta_1 dLowEmission_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

Testing Hypothesis 1 with Resource Use score as a measure of E (ESG):
1c: $GreenPatents_{i,t} = \alpha_0 + \beta_1 dLowResourceUse_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

Testing Hypothesis 1 with Environmental Innovation score as a measure of E (ESG):
1d: $GreenPatents_{i,t} = \alpha_0 + \beta_1 dLowEInnovation_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

The coefficients of interest are the β_1 coefficients before the dummy variables for the various ESG scores, which represent the effect of low ESG scores on the number of patent publications. The significance of the coefficient of the dummy variable was assessed using a z-test (using the normal distribution), where the null hypothesis (H_0) is that $\beta_1 \leq 0$ indicating no positive effect of low ESG scores on patent filings. Meaning that low E firms do not publish more or fewer green patents than their higher E counterparts. Either they publish the same or fewer green patents. The alternative hypothesis (H_1) asserts that: $H_1: \beta_1 > 0$. This alternative hypothesis posits that low E firms publish more green patents than high E firms. If the coefficient β_1 is statistically significantly greater than 0, the null hypothesis is rejected in favor of the alternative hypothesis.

The second set of hypotheses examines the relationship between the various ESG scores and the total number of firm i's citations received on the firm's green patent publications in year t. These hypotheses propose that patents published by low ESG firms receive more citations than those published by mid/high ESG firms.

Regressions 2a, 2b, 2c and 2d were used to test Hypothesis 2: *Low E (ESG) firms publish higher quality green patents than high E (ESG) firms, measured by citations.*

Testing Hypothesis 2 with weighted E score as a measure of E (ESG):
2a: $Citations_{i,t} = \alpha_0 + \beta_1 dLowEscore_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

Testing Hypothesis 2 with Emission score as a measure of E (ESG):
2b: $Citations_{i,t} = \alpha_0 + \beta_1 dLowEmission_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

Testing Hypothesis 2 with Resource Use score as a measure of E (ESG):
2c: $Citations_{i,t} = \alpha_0 + \beta_1 dLowResourceUse_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

Testing Hypothesis 2 with Environmental Innovation score as a measure of E (ESG):
2d: $Citations_{i,t} = \alpha_0 + \beta_1 dLowEInnovation_{i,t} + \beta_2 Ln(Assets)_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 ROA_{i,t} + \beta_5 Cash_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 Age_{i,t} + YearFE_t + \varepsilon_{i,t}$

The coefficients of interest are the β_1 coefficients before the dummy variables for the various ESG scores, which represents the effect of low ESG scores on the count of citations. The significance of the coefficient of the dummy variable was tested using a z-test (using the normal distribution), where the null hypothesis (H_0) is that $\beta_1 \leq 0$ indicating no positive effect of low ESG scores on citations. In other words, low E firms do not receive more citations on their green patents than their higher E counterparts. The alternative hypothesis (H_1) would assert that: $H_1: \beta_1 > 0$. This alternative hypothesis posits that low E firms receive more citations on their green patents than high E firms. If the coefficient β_1 is statistically significantly greater than 0, the null hypothesis is rejected in favor of the alternative hypothesis.

All regressions, concerning Hypotheses 3 and 4 were tested using the same models, with the only difference being the data fed into the models for acquiring results. In fact, the regression formulae of Hypotheses 1a-d and Hypothesis 2a-d are exactly the same as Hypotheses 3a-d and Hypotheses 4a-d, respectively. Namely, the first and second hypotheses focus on the entire sample of firms while Hypotheses 3 and 4 focus on firms operating in the Energy Sector only. Hypotheses 3 and 4 are:

Hypothesis 3: *Within the Energy Sector, companies with low E (ESG) scores publish more green patents than high E (ESG) firms.*

Hypothesis 4: *Within the Energy Sector, companies with low E (ESG) scores publish higher quality green patents than high E (ESG) firms, measured by citations.*

4.3.2 Industry-level Regressions

Furthermore, an industry-level analysis was conducted to examine whether the quantity and quality of green patent production in the Energy Sector differs from the non-Energy Sector. In contrast to the third and fourth hypotheses which only considered the Energy Sector, Hypotheses 5, 6 and 7 consider the total sample of both sectors again. The Energy Sector is now used as a proxy for having low ESG scores. Again, parts of the sample were compared. To compare the Energy Sector with other industries, a dummy variable is operated which equals 1 for industry j, if the first two digits of its SIC are 10 (Metal, Mining), 12 (Coal Mining), 13 (Oil & Gas Extraction), 14 (Non-metallic Minerals, Except Fuels) or 29 (Petroleum & Coal Products). Otherwise the industry does not belong to the Energy Sector and the dummy equals 0.

The independent variable of green innovation took on three forms: the average count of patents published in an industry, the industry-average count of citations on green patents and a product variable of these two. The last variable is a metric which multiplies the average patent count and the average citations count in a given industry in a given year. Observations are industry-year averages. To isolate the effect of the Energy Sector on innovation, six control variables were added to each regression: firm size (AverageLn(Assets)), leverage ratio (AverageLeverage), return on assets (AverageROA), cash ratio (AverageCash), research and development to sales (AverageR&D) and firm age (AverageAge). For all control variables, industry-averages are taken. Additionally, Year Fixed Effects were added.

Regression 5 is deployed to test Hypothesis 5*: The low E (ESG) industries of the Energy Sector publish more patents compared to other industries.*

Testing Hypothesis 5 with a dummy variable for Energy Sector as a measure for E (ESG):

$$5: AverageGreenPatents_{j,t} = \alpha_0 + \beta_1 dEnergySector_j + \beta_2 AverageLn(Assets)_{j,t} + \beta_3 AverageLeverage_{j,t} + \beta_4 AverageROA_{j,t} + \beta_5 AverageCash_{j,t} + \beta_6 AverageR\&D_{j,t} + \beta_7 AverageAge_{j,t} + \beta_8 YearFE_t + \varepsilon_{i,t}$$

Regression 6 is deployed to test Hypothesis 6*: The low E (ESG) industries of the Energy Sector publish higher quality patents compared to other industries, measured by citations.*

Testing Hypothesis 6 with a dummy variable for Energy Sector as a measure for E (ESG):

6: $AverageCitations_{j,t} = \alpha_0 + \beta_1 dEnergySector_j + \beta_2 AverageLn(Assets)_{j,t} + \beta_3 AverageLeverage_{j,t} + \beta_4 AverageROA_{j,t} + \beta_5 AverageCash_{j,t} + \beta_6 AverageR\&D_{j,t} + \beta_7 AverageAge_{j,t} + \beta_8 YearFE_t + \varepsilon_{i,t}$

Hypothesis 7*: The low E (ESG) industries of the Energy Sector publish higher combined volume and impact patents compared to other industries, measured by publications and citations.*

Testing Hypothesis 7 with a dummy variable for Energy Sector as a measure for E (ESG):

7: $AverageProduct_{j,t} = \alpha_0 + \beta_1 dEnergySector_j + \beta_2 AverageLn(Assets)_{j,t} + \beta_3 AverageLeverage_{j,t} + \beta_4 AverageROA_{j,t} + \beta_5 AverageCash_{j,t} + \beta_6 AverageR\&D_{j,t} + \beta_7 AverageAge_{j,t} + \beta_8 YearFE_t + \varepsilon_{i,t}$

For Hypotheses 5, 6 and 7, the coefficient of interest is β_1 of the dummy variable for the Energy Sector. The significance of the coefficient of the dummy variable was tested using a z-test (using the normal distribution), where the null hypothesis (H_0) is that $\beta_1 \leq 0$ indicating no positive effect of being part of the Energy Sector on patent publications or citations. In other words, low ESG (Energy Sector) industries do not publish more or higher quality green patents than their higher ESG (non-Energy Sector) counterparts. If the null hypothesis is accepted, the Energy Sector publishes the same amount or even fewer green patents than industries that do not belong to the Energy Sector. The alternative hypothesis (H_1) would assert that: $H_1: \beta_1 > 0$ .This alternative hypothesis posits that on average, the Energy Sector publishes more green patents (Hypothesis 5) that are more highly cited (Hypothesis 6) than other, more sustainable industries. If the coefficient β_1 is statistically significantly greater than 0, the null hypothesis is rejected in favor of the alternative hypothesis.

4.4 Assumption Tests

Section 4.4 discusses several preliminary tests which were conducted to evaluate potential violations of assumptions for the regression model.

4.4.1 Multicollinearity

Multicollinearity is a statistical phenomenon where two or more explanatory variables in a regression model are strongly correlated. Comprehension of the correlation between the explanatory variables is important as multicollinearity can affect the reliability of the regression coefficients. Therefore, the independent variables should not exhibit excessive interdependence. As per the guidelines laid out by Senaviratna and Cooray (2019), if the correlation coefficient between two variables exceeds 0.8, multicollinearity becomes a significant concern. As shown in Table 10, none of the explanatory variables exceeds this threshold, suggesting that multicollinearity is unlikely. The variables that come closest to this threshold are the dependent variables: Citations and GreenPatents. Albeit unlikely, the high correlation coefficients within the ESG scores and with Ln(Assets) could pose a serial correlation issue. It is possible to have multicollinearity even with lower correlations if there is a combination of multiple variables that together create redundancy. For a more extensive analysis, variance inflation factors (VIFs) are often calculated to measure how much the variance of the estimated regression coefficients increases due to collinearity. Therefore, all regressions were checked by inspecting values of VIF. VIF provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity. A VIF of 5 or 10 is often used as a threshold for significant multicollinearity concerns, according to Craney and Surles (2002). No issues with multicollinearity were found as all VIF values were below the lower bound of 5. Additionally, the separate regressions for both dependent variables with each ESG score reduce the potential for multicollinearity even further.

4.4.2 Heteroskedasticity and Non-Linearity

To assess whether heteroskedasticity could be an issue, the residual plots of Figures 1 until 7 in Appendix C were visually evaluated. All residual plots depicted in Appendix C illustrate the relationship between linear predictors (X-axis) and deviance residuals (Y-axis). The residual plots have been composed for all regression models. Some of the residual plots exhibit variations in the spread of residuals against linear predictions, suggesting the presence of heteroskedasticity in the data. Specifically, regressions considering the entire sample, shown in Figure 1 (Patents for total sample) and Figure 2 (Citations for total sample) show a fan shape, widening as the value of the linear prediction increases, which is indicative of heteroskedasticity. In Figure 3 (Patents in the Energy Sector) and Figure 4 (Citations in the Energy Sector), the residuals are more evenly distributed and do not indicate heteroskedasticity. The plot in Figure 5 (Patents between sectors) which concerns the comparison of the Energy Sector with the non-Energy Sector, also shows a fan shape, implying potential heteroskedasticity. However, Figure 6 (Citations between sectors) displays a cloud-like spread without a clear pattern, suggesting no concern for heteroskedasticity. Finally, Figure 7 (Patents×Citations between sectors) presents a more pronounced fan shape, especially at higher values of linear prediction, strongly suggesting the presence of heteroskedasticity. This pattern of unequal variance across the range of predictors necessitates adjustment. To address this, we have employed robust, heteroskedasticity-consistent standard errors to mitigate any potential impact on the regression model.

While the plots reveal slight heteroskedasticity, they also confirm the absence of non-linear relationships; the residuals show no clear curve, suggesting that the relationships between the independent variables and the dependent variable are linear. Additionally, the residual plots do not reveal any extremely influential cases, indicating that our findings are not affected by outliers. These assessments are based on a multivariate level, ensuring a comprehensive evaluation of the model's assumptions and the integrity of the results.

4.4.3 Autocorrelation

<u>Autocorrelation concerns</u>

Autocorrelation refers to the degree of correlation of a variable at a point in time and the same variable at a successive time interval. If this correlation is high, past values of the dependent variable are predictive of its future values, which is problematic for model-specification. The Wooldridge test was used to test for autocorrelation in the panel data. The test has been conducted on the residuals of all regression models to detect the presence of first-order autocorrelation. The null hypothesis (H_0) of this test assumes no first-order autocorrelation. The Wooldridge tests yielded mixed results.

Note that each regression-model is aligned with the corresponding hypothesis. Specifically, the regression executed in Model 1a is designed to test Hypothesis 1a. The results for Model 1a indicate a highly significant presence of autocorrelation with an F-statistic valued at 18.78 (p<.01), leading to the rejection of H_0. For Model 3a, which focuses on patents within the Energy Sector, the Wooldridge test also suggests autocorrelation, significant at 5% (F = 8.50, p<.05). Again, the results of Model 5 show significant autocorrelation (F = 5.68, p<.05). Albeit serial correlation in Models 3a and Model 5 is not as strong as in Model 1a, the results necessitate further investigation or model adjustments to account for this autocorrelation. In contrast, for Model 2a, the test does not indicate the presence of autocorrelation (F = 2.03, p = 0.16). Similarly, for Model 4a related to citations in the Energy Sector, the test results do not provide enough evidence to reject the null hypothesis of no autocorrelation at conventional significance levels (F = 3.55, p=0.109), albeit the p-value indicates that caution should be taken in interpreting the results. Hypothesis 6 yields an F-statistic of 2.893 at the margin of significance (p = 0.095), suggesting potential autocorrelation that may need to be addressed. Lastly, the analysis for Hypothesis 7 reveals an F-statistic of 9.50 (p = 0.003), significant at the 1% level, indicating that previous values of the dependent variable have a strong effect on the current values. Thus, particularly for the regressions with patent publications (GreenPatents) as the dependent variable (marginal) autocorrelation is detected. The presence of autocorrelation in several models indicates the need for careful model specification and potentially the use of autocorrelation-consistent standard errors or model adjustments to mitigate this issue. In the Robustness Section 5.6.3, additional analyses have been conducted to account for these issues.

Decision for non-lagged dependent variables

Despite the findings of the Wooldridge tests, non-lagged dependent variables were analyzed in the main regressions to obtain the results of Section 5.1 until Section 5.5. In conducting the main regression analysis of this study, the decision to forego correction for autocorrelation using lagged dependent variables, namely lags of GreenPatents, Citations and Patents×Citations, was based on several carefully considered factors.

Firstly, the dataset presents an inherent limitation in that many firms do not have consecutive yearly observations; therefore, generating lagged values is not always feasible. The necessity for a firm to have consistent annual outputs or citations to create a lag sequence resulted in a substantial number of missing observations. On average 32% of observations were lost for each of the regression Models when lagging either GreenPatents, Citations, or Patents×Citations.

Secondly, introducing lagged variables into the analysis could lead to a selection bias, systematically excluding those entities that do not publish patents or receive citations regularly. This would favor larger firms which are more active in patent production, thereby distorting the representativeness of the sample.

Thirdly, the regression models already include Year Fixed Effects to control for time-invariant heterogeneity within the sample. These fixed effects serve to mitigate potential autocorrelation by absorbing shocks common to all firms within a given year, thereby preserving the integrity of the individual variations that are of primary interest to this study.

Fourthly, the study's primary objective is to examine the cross-sectional differences in firms' innovative outcomes as they relate to ESG scores, not the time-series progression of their innovations. While the assumption of no autocorrelation has been compromised, this does not necessarily invalidate the study's comparative approach.

Despite these considerations, it is wise to conduct robustness checks to confirm the stability of the findings. While lagged dependent variables are not used in the main analysis due to the aforementioned reasons, including potential observation loss and selection bias, they are considered in the robustness checks for the remaining variables. This allows for a thorough examination of the impact of autocorrelation without compromising the sample size and diversity. The outcomes of these robustness checks will be discussed in detail in Section 5.6.2, providing additional validation for the study's results.

CHAPTER 5 Results

Chapter 5 concerns the empirical findings gathered from executing all the previously mentioned regression analyses. For all regressions the NMB was used. However, interpretation of coefficients in the NBM is not straightforward. Simply put, if the coefficient of a predictor variable value is "x", this number must be exponentiated (calculating e^x) to interpret the result practically. This outcome of calculating e^x can be interpreted as the multiplicative effect of a one-unit increase of the independent variable on the outcome variable. If the exponentiated coefficient is $B < 0$ (after calculating e^x), the multiplicative effect is negative. If the exponentiated coefficient is $B > 0$, the predictor has a positive effect on the outcome variable. Concretely, if the coefficient of dLowEscore is 0.6, the count of green patents published by firms within the lowest tercile of the weighted Environmental score is expected be a factor of 1.82 ($e^{0.6}$) higher than firms in the middle and higher terciles, holding all other variables constant.

For all hypotheses, the respective models are indicated by the same number and letter in between parentheses. For example, Hypothesis 3a is tested for in the regression analysis of Model 3a.

Additionally, note throughout the entire thesis that a negative relationship between a firms' ESG performance and innovation is equivalent to a positive relationship between the variable low ESG and innovation. When specifically discussing coefficients, the relationship between the variables is addressed (with low ESG) but when discussing the implications of this, the reverse is stated (regarding ESG performance).

5.1 ESG scores on Patent Publications

The regression analysis as presented in Table 11 investigates the relationship between green patent publications and various ESG scores, measured by the weighted Environmental score (Model 1a), the Emission score (Model 1b), the Resource Use score (Model 1c) and the Environmental Innovation score (Model 1d). The combination of the number 1 to 4 followed by the letter a, b, c or d refers to a regression model, belonging to the respective hypothesis. The analysis spans the period from 2010 to 2022. By use of the NBM, the first hypothesis seeks to ascertain if firms with lower ESG scores demonstrate a higher frequency of green patent publications as indicated by Hypothesis 1: *Low E (ESG) firms publish more green patents than high E (ESG) firms.*

In Model 1a, the coefficient of dLowE ($\leq 33^{rd}$ percentile) is -0.11 (*SD* = 0.12). This negative coefficient suggests that firms with a low weighted Environmental score do not publish more green patents than firms with higher weighted Environmental scores, although this result is not statistically significant (*p*>.10). Similar results are found for Model 1b, 1c and 1d, with negative coefficients of -0.089, -0.059 and -0.061 respectively. Again, these findings are not statistically significant (*p*>.10). All Models indicate a negative association of the ESG score with the number of green patents published. Considering these results, Hypothesis 1 is not supported as there is no statistically significant evidence that firms with lower ESG scores publish more green patents. The negative coefficients for the ESG scores, although not significant, counter the expected positive relationship stated in Hypothesis 1.

In all four models, the control variable firm size (represented by Ln(Assets)) has a positive coefficient of at least 0.45 on the green patent count, significant at 1%. This result implies that with a one-unit increase in Ln(Assets), the expected green patent count is approximately 56.8% higher (ceteris paribus). Leverage also has a positive effect on the dependent variable with a coefficient of 0.78 yet, only significant for the weighted Environmental score, dLowE (Model 1a). The cash ratio (Cash) has a positive significant effect of 0.12 (Model 1c) and 0.15 (Model 1d) but is insignificant for Model 1a and 1b. In Model 1a, with the weighted E score as dummy variable, R&D has a huge positive coefficient of 5.77, significant at the 1% level. The results imply that larger firms and those investing more heavily in R&D are more likely to publish green patents, aligning with existing literature on innovation and firm characteristics. The R-squared values, which provide insight into the model's explanatory power, indicate that the independent variables explain a moderate proportion of the variability in green patent filings. However, given the lack of statistical significance in the coefficients of interest, the R-squared values could also explain that additional factors not included in the model may play a significant role in influencing the publication of green patents.

In summary, while the regression analysis does not provide evidence to support Hypothesis 1, the data provides evidence that firm size (represented by Ln(Assets)) and R&D investment and to a lesser extent Leverage, are meaningful predictors of green patent citations (Citations).

Table 11: Negative binomial regression results for the relationship between E (ESG) scores and the quantity of patent publications

Variables	1a	1b	1c	1d
		GreenPatents		
dLowE	-.111			
	(.119)			
dLowEmmissions		-.0892		
		(.0989)		
dLowResourceUse			-.0591	
			(.0957)	
dLowEInnovation				-.0611
				(.125)
Ln(Assets)	.496***	.474***	.536***	.446***
	(0.0394)	(0.0313)	(0.0310)	(0.0408)
Leverage	0.797***	0.224	0.299	0.288
	(0.258)	(0.230)	(0.211)	(0.286)
ROA	0.437	-0.571	-0.222	-0.339
	(0.633)	(0.427)	(0.409)	(0.654)
Cash	-0.0682	0.0611	0.122***	0.146**
	(0.0655)	(0.0388)	(0.0375)	(0.0658)
R&D	5.774***	0.0587	0.184*	1.254
	(0.801)	(0.0749)	(0.0985)	(1.344)
Age	0.00384*	0.00141	0.00134	0.00416*
	(0.00208)	(0.00194)	(0.00195)	(0.00227)
Constant	-2.810***	-1.910***	-2.665***	-1.858***
	(0.485)	(0.370)	(0.350)	(0.503)
YearFE	controlled	controlled	controlled	controlled
N	1,487	2,266	2,296	1,662
Chi^2	445.7	390.8	463.5	296.9
p	< .001	< .001	< .001	< .001
R^2_{pseudo}	0.0720	0.0621	0.0681	0.0564

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent publications over the period 2010 - 2022. Dummy variables for low E score (Model 1a), low Emissions score (Model 1b), low Resource Use score (Model 1c), and low Environmental Innovation score (Model 1d) represent firms within the lowest tercile of each ESG category compared to firms in the middle and higher terciles, the reference group (>33 pct.). Control variables include the log of total assets (Ln(Assets)), leverage ratio (Leverage), return on assets (ROA), cash ratio (Cash), research & development investment relative to sales (R&D), and firm age (Age). All variables are elaborately defined in Table 2. Coefficients indicate the expected change in the log count of green patents published for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the green patents count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

5.2 ESG scores on Patent Citations

Table 12 provides the regression results for the relationship between citations received, on a firm's green patent records (Citations) and various ESG scores, measured by the weighted Environmental score (Model 2a), the Emission score (Model 2b), the Resource Use score (Model 2c) and the Environmental Innovation score (Model 2d). The analysis spans the period from 2010 to 2022. By exploiting the NBM, the study seeks to ascertain if firms with lower ESG scores demonstrate a higher frequency of green patent citations as indicated by Hypothesis 2: *Low E (ESG) firms publish higher quality green patents than high E (ESG) firms, measured by citations*.

In Models 2a, 2b, 2c and 2d the coefficients for the dummy low E score, low Emissions score and low Resource Use score are -0.21, -0.14, -0.19 and -0.01 respectively. None of the coefficients is statistically significant. According to these models, firms with lower ESG scores do not demonstrate a significant difference in the number of green patent citations compared to firms with higher ESG scores. Thus, Hypothesis 2 is not supported by any of the models and therefore rejected.

For the control variables, the Ln(Assets) shows a positive and highly significant impact across all models with coefficients ranging from 0.48 to 0.53 (p<.01), indicating that larger firms tend to receive more citations. The implication of this finding is that, with a one-unit increase in Ln(assets), the expected count of citations is at least 61.05% higher. The R&D investment relative to sales is significant in Models 2a and 2c with coefficients of 6.39 (p<.01) and 0.20 (p<.05) respectively, suggesting a positive relationship between R&D intensity and citation counts. Return on assets (ROA) has a positive effect on patent citation count in Model 2a but is insignificant for the other ESG scores. Cash presents a significant, but inconsistent pattern across the models. For the cash ratio in Model 2a, where firms are split based on the weighted E score, the coefficient of Cash is -0.13. Yet, for Models 2b, 2c and 2d the coefficients are 0.14, 0.18 and 0.11, potentially indicating that more liquid firms have the resources to invest in higher quality or more impactful green innovations. The R-squared values are relatively low, indicating that the models explain a moderate proportion of the variance in patent citations. The Chi-squared statistics show overall model significance, confirming that the predictors collectively relate to green patent citation counts.

In summary, Hypotheses 2 is rejected based on all models. While the ESG scores under investigation do not exhibit a significant influence on the citation counts, the data does provide evidence that firm size (Ln(Assets)) and R&D investment are meaningful predictors of green patent citations.

Table 12: Negative binomial regression results for the relationship between E (ESG) scores and the quality of patent publications, measured by citation count

Variables	2a	2b	2c	2d
	Citations (patent quality)			
dLowE	-0.213			
	(0.138)			
dLowEmmissions		-0.140		
		(0.129)		
dLowResourceUse			-0.190	
			(0.114)	
dLowEInnovation				-0.007
				(0.148)
Ln(Assets)	0.506***	0.482***	0.531***	0.478***
	(0.044)	(0.035)	(0.036)	(0.044)
Leverage	-0.053	-0.209	-0.157	-0.375
	(0.302)	(0.290)	(0.271)	(0.321)
ROA	1.298**	-0.001	0.156	-0.303
	(0.654)	(0.438)	(0.469)	(0.748)
Cash	-0.133*	0.141**	0.184***	0.111**
	(0.069)	(0.062)	(0.057)	(0.050)
R&D	6.389***	0.120	0.204**	1.596
	(0.937)	(0.101)	(0.096)	(1.296)
Age	0.003	0.001	0.000	0.003
	(0.002)	(0.002)	(0.002)	(0.003)
Constant	0.675	1.236***	0.721*	1.510***
	(0.542)	(0.434)	(0.396)	(0.540)
YearFE				
N	1,487	2,266	2,296	1,662
Chi^2	397.3	595.2	615.1	385.9
p	< .001	< .001	< .001	< .001
R^2_{pseudo}	0.0511	0.0469	0.0479	0.0423

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent citations over the period 2010 - 2022. Dummy variables for low E score (Model 2a), low Emissions score (Model 2b), low Resource Use score (Model 2c), and low Environmental Innovation score (Model 2d) represent firms within the lowest tercile of each ESG category compared to firms in the middle and higher terciles, the reference group (>33 pct.). Control variables include the log of total assets (Ln(Assets)), leverage ratio (Leverage), return on assets (ROA), cash ratio (Cash), research & development investment relative to sales (R&D), and firm age (Age). All variables are elaborately defined in Table 2. Coefficients indicate the expected change in the log count of citations for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the citations count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

5.3 ESG scores on Patent Publications in the Energy Sector

The regression results presented in Table 13 offer insights into the relationship between ESG scores and green patent publications within the Energy Sector over the period from 2010 to 2022. The various ESG metrics utilized are the weighted Environmental score (Model 3a), the Emission score (Model 3b), the Resource Use score (Model 3c) and the Environmental Innovation score (Model 3d). Through usage of the NBM, the research analyzes if firms in the Energy Sector demonstrate a higher frequency of green patent publications, as indicated by Hypothesis 3: *Within the Energy Sector, companies with low E (ESG) scores publish more green patents.*

Model 3a shows a negative coefficient for dLowE (B = -0.89, p<.05), suggesting that firms in the lowest tertile of the weighted Environemtal score category file fewer green patents compared to their counterparts with higher scores. However, in Model 3b, the low Emissions score has a positive coefficient of 1.59 (p<.01), highly significant at the 1% level, indicating that firms with high emissions actually file more green patents. This result also means that within the Energy Sector, the mid and highly ranked firms on Emission score, the companies that pollute less, tend to produce fewer green patents relatively. The dummy low Resource Use score in Model 3c also has a positive and significant coefficient of 0.68 at the 1% level, pointing to a similar trend of more publications for patents by firms with lower scores in this category. On the other hand, the dLowEInnovation in Model 3d shows a negative coefficient of -0.45 significant at the 5% level. Therefore, concluding on the acceptance of Hypothesis 3 is not straightforward. Due to the contradicting findings, the third Hypothesis is neither rejected nor accepted.

Firm size, as measured by the natural logarithm of assets, demonstrates a consistently positive and significant effect across all models (coefficients ranging from 0.63 to 0.91, p<.01), implying that larger firms tend to produce more green patents. The impact of Leverage varies, with no significant effect in Model 3a, a negative relationship in Models 3b and 3c (B = -1.60, p<.05 and B = -1.74, p<.10), and a positive relationship in Model 3d (B = 2.65, p<.01), suggesting the influence of financial structure on innovation is complex and context-dependent. Profitability, indicated by ROA, is only significant in Model 3b (B = 2.77, p<.10), where higher profitability correlates with increased green patent publications. Conversely, the cash ratio's effect of the variable Cash is mixed, showing a negative correlation in Models 3a and 3b (coefficients -0.70 and -0.10, p<.01 and p<.10) but turning positive in Model 3d (coefficient 0.22, p<.01), reflecting different liquidity dynamics across ESG scores. R&D intensity robustly predicts green patent activity (coefficients range from 61.39 to 84.67, p<.01), which affirms the role of research investment in innovation. Age, while generally not significant, shows a negative effect in Model 3b and Model3c (B = -0.014, p<.10 and B = -0.017, p<.05), hinting at younger firms' propensity for higher green patent activity within specific ESG categories. The models' pseudo R-squared values, though moderate (R^2_{pseudo} = 0.072 to R^2_{pseudo} = 0.14), along with significant chi-squared statistics across all models, affirm the models' overall predictive validity.

In summary, while the models provide evidence supporting some of the hypotheses, they also highlight the complexity of the relationship between ESG scores and innovation in the Energy Sector. Each ESG metric has a different impact on green patent publications and the control variables indicate that firm size (Ln(Assets)), Leverage, and R&D intensity are important factors influencing this relationship.

Table 13: Negative binomial regression results for the relationship between E (ESG) scores and the quantity of patent publications in the Energy Sector

	3a	**3b**	**3c**	**3d**
Variables		**GreenPatents**		
dLowE	-0.894**			
	(0.356)			
dLowEmmissions		1.586***		
		(0.278)		
dLowResourceUse			0.678***	
			(0.261)	
dLowEInnovation				-0.453**
				(0.180)
Ln(Assets)	0.644***	0.630***	0.625***	0.905***
	(0.095)	(0.067)	(0.093)	(0.049)
Leverage	1.271	-1.598**	-1.738*	2.651***
	(1.142)	(0.814)	(0.927)	(0.960)
ROA	3.283	2.772*	-0.334	0.843
	(2.241)	(1.681)	(1.733)	(2.082)
Cash	-0.698***	-0.095*	-0.118	0.222***
	(0.251)	(0.053)	(0.176)	(0.035)
R&D	61.387***	84.673***	73.750***	66.120***
	(16.733)	(23.931)	(25.865)	(11.844)
Age	-0.009	-0.014*	-0.017**	-0.003
	(0.008)	(0.007)	(0.008)	(0.005)
Constant	-4.445***	-3.761***	-2.523**	-8.611***
	(1.495)	(0.879)	(1.136)	(0.761)
YearFE	controlled	controlled	controlled	controlled
N	79	175	170	86
Chi^2	765.9	252.6	155.1	773.4
p	< .001	< .001	< .001	< .001
R^2_{pseudo}	0.134	0.0845	0.0724	0.140

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent publications over the period 2010 – 2022. Dummy variables for low E score (Model 3a), low Emissions score (Model 3b), low Resource Use score (Model 3c), and low Environmental Innovation score (Model 3d) represent firms within the lowest tercile of each ESG category compared to firms in the middle and higher terciles, the reference group (>33 pct.). Control variables include the log of total assets (Ln(Assets)), leverage ratio (Leverage), return on assets (ROA), cash ratio (Cash), research & development investment relative to sales (R&D), and firm age (Age). All variables are elaborately defined in Table 2. Coefficients indicate the expected change in the log count of green patents published for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the green patents count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

5.4 ESG scores on Patent Citations in the Energy Sector

The regression results of Table 14 explore the relationship between E (ESG) scores and the quality of patent publications within the Energy Sector, as measured by citation count. The analysis spans the period from 2010 to 2022. The table presents four models, each assessing a different aspect of ESG scores: the weighted Environmental score (4a), Emissions score (4b), Resource Use score (4c) and Environmental Innovation score (4d). *Within the Energy Sector, companies with low E (ESG) scores publish higher quality green patents, measured by citations.*

In Model 4a and 4d, the coefficients for the dummies dLowE score and dLowEInnovation are -1.50 ($p<.05$) and -1.43 ($p<.01$) respectively. As a result, it follows that the companies ranked below the 33rd percentile for these ESG metrics only receive around 23% of the citations that firms in higher terciles accrue. Put differently, firms in the lowest group of E scores and Environmental Innovation scores receive five times fewer citations compared to firms with higher E scores. In contrast, Model 4b shows a significant positive effect of 3.46 for the Low Emissions score (B = 1.24, $p<.01$). Hence, it can be deduced that the amount of citations received by firms with Low Emissions scores is expected to be 346% higher than that of the mid/high ranked groups. Model 4c reveals no significant effect for the coefficient of dLowResourceUse. Thus, drawing a definitive conclusion regarding the acceptance of Hypothesis 4 presents a challenge due to the contradictory findings of Models 4a and 4d on the one hand, and Model 4b on the other hand. As such, the third hypothesis cannot be unequivocally rejected or accepted.

Firm size (denoted by Ln(Assets)) is slightly negative (B = -0.04, $p>.10$) for the weighted E score but does not hold any significance. However, for Models 4b, 4c and 4d, the variable Ln(Assets) is significant at the 1% level and has a positive coefficient of 0.54, 0.43 and 0.64 respectively. Note that the sample only includes US-listed firms and does not necessarily provide evidence for firm size being an important contributor to innovation in non-listed firms. The leverage ratio (Leverage) is insignificantly negative for all models except Model 4d. In the regression of the Environmental Innovation score and Citations, Leverage has a positive coefficient of 3.23 ($p<.01$). This suggests that firms in the lowest tercile of Environmental Innovation scores with a higher leverage ratio, are associated with an increase in green patent citations. In Model 4b, which is designed to measure patent quality through citation count, return on assets (ROA) exerts a significant positive effect evidenced by a coefficient of 4.34 ($p<.05$). This denotes that pollution-prone firms (with low Emissions scores), with higher profitability tend to produce higher quality patents. The cash ratio (Cash), interestingly, exerts a substantial negative effect in Model 4a indicated by a coefficient of -1.76 ($p<.01$). This implies that firms with higher liquidity may file fewer green patents. The relative research and development (R&D) investment to sales exhibits varied influence across the models. Again, all models show a positive relationship between the variables R&D and Citations. In Model 4b, the coefficient stands at a highly significant 87.22 ($p<.01$), demonstrating an extremely strong positive correlation between R&D intensity and patent quality. Model 4c and 4d also have positive significant coefficients of 64.43 ($p<.05$) and 35.77 ($p<.10$). This supports the argument that R&D investments play a critical role in strengthening the impact and recognition of green innovations. The control variable Age does not exhibit significant influence across any of the models, suggesting that the age of a firm does not significantly impact the citations green patents of low ESG firms receive in the Energy Sector. The pseudo R-squared values of Table 14, ranging from 0.058 to 0.066, suggest a moderate fit for the models. The Chi-squared values: 199.4, 278.5, 257.9, and 190.4 (all with $p<.01$) indicate the models' overall significance.

In summary, a low weighted E score and Environmental Innovation score significantly reduce the count of patent citations. In contrast, the Emissions score has the opposite effect. Additionally, the data suggests that firm size (Ln(Assets)) and R&D investment are significant predictors of green patent citations, having a positive effect for the sub-scores (Models 4b, 4c and 4d). Lastly, for the weighted E score the Cash ratio has a significantly negative impact.

Table 14: Negative binomial regression results for the relationship between E (ESG) scores and the quality of patent publications, measured by citation count in the Energy Sector

	4a	4b	4c	4d
Variables		**Citations (patent quality)**		
dLowE	-1.497**			
	(0.640)			
dLowEmmissions		1.237***		
		(0.368)		
dLowResourceUse			0.211	
			(0.333)	
dLowEInnovation				-1.427***
				(0.344)
Ln(Assets)	-0.041	0.537***	0.434***	0.641***
	(0.209)	(0.096)	(0.116)	(0.116)
Leverage	-2.140	-1.096	-0.988	3.229***
	(1.839)	(0.997)	(1.039)	(1.171)
ROA	7.271	4.338**	-1.043	-0.047
	(4.723)	(2.110)	(2.130)	(3.401)
Cash	-1.761***	0.026	-0.268*	0.385***
	(0.568)	(0.062)	(0.151)	(0.065)
R&D	38.685	87.224***	64.432**	35.767*
	(33.809)	(29.847)	(28.474)	(18.658)
Age	-0.003	-0.010	-0.006	0.004
	(0.015)	(0.008)	(0.009)	(0.006)
Constant	8.306***	-0.406	1.963	-2.208
	(3.123)	(1.274)	(1.347)	(1.783)
YearFE	controlled	controlled	controlled	controlled
N	79	175	170	86
Chi^2	199.4	278.5	257.9	190.4
p	< .001	< .001	< .001	< .001
R^2_{pseudo}	0.0656	0.0623	0.0580	0.0652

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent publications over the period 2010 – 2022. Dummy variables for low E score (Model 4a), low Emissions score (Model 4b), low Resource Use score (Model 4c), and low Environmental Innovation score (Model 4d) represent firms within the lowest tercile of each ESG category compared to firms in the middle and higher terciles, the reference group (>33 pct.). Control variables include the log of total assets (Ln(Assets)), leverage ratio (Leverage), return on assets (ROA), cash ratio (Cash), research & development investment relative to sales (R&D), and firm age (Age). All variables are elaborately defined in Table 2. Coefficients indicate the expected change in the log count of citations for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the citations count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

5.5 Industry ESG on Patent Publications and Citations

The regressions presented in Table 15 employ the Energy Sector dummy variable, dEnergySector, as a proxy for low ESG scores. This differs from the methodology of the previous four hypotheses, which used actual third-party ESG ratings. The dEnergySector variable is aligned with the Industry level ESG concept introduced in Section 2.3, serving as an implicit indicator of an industry's low ESG score. Models 5, 6, and 7 explore innovation potential within the Energy Sector, often criticized for its environmental impact.

Through usage of the NBM, this study seeks to ascertain if industries in the Energy Sector demonstrated a higher frequency of green patent publications as indicated by Hypothesis 5: *The low E (ESG) industries of the Energy Sector publish more green patents compared to other industries.* In Model 5, the coefficient for the dummy variable termed dEnergySector stands at 0.67 (p<.01), thereby revealing a robust and positive correlation. This suggests that, on average, entities within the Energy Sector produced 95.42% more green patents compared to industries in the non-Energy Sector. The sample comprises 515 observations, reflecting industry-year averages. Pertaining to the control variables, the natural logarithm of assets (Ln(Assets)) has a coefficient of 0.16 (p <.01). This indicates that industries comprising larger firms had a propensity to publish a greater number of green patents on average. AverageLeverage and AverageROA, and AverageCash are all negative, albeit insignificant. The average R&D intensity (AverageR&D) is positively significant (B = 1.62, p<.05), reinforcing the proposition that industries which invested more heavily in R&D were more likely to generate green patents. The firm age (AverageAge) coefficient stands at a mere 0.03 (p<.01) but is highly significant. The pseudo R-squared value of 0.05 indicates that the model accounts for approximately 9% of the variability in the dependent variable. The Chi-squared statistic of 163.0 (p<.01) is highly significant, which suggests a strong overall fit of the model.

Hypothesis 6 asserts that the Energy Sector tends to produce patents of superior quality in comparison to their more reputable counterparts in the non-Energy Sector. Hypothesis 6: *The low E (ESG) industries of the Energy Sector publish higher quality green patents compared to other industries, measured by citations.* Evidence supporting this proposition can be found in Model 6, where the coefficient for dEnergySector takes on 0.51 (p<.01).This suggests that industries within the Energy Sector, on average, accumulated a 66.53% higher count of citations for their patents, which is indicative of superior patent quality. This increase in citation counts lends credence to Hypothesis 6. Similarly to the findings in Model 5, Model 6 finds that the size of firms within an industry exerts a significant positive influence on the average number of citations received per green patent per firm. This is shown by an Average Ln(Assets) coefficient of 0.18 (p<.01). Conversely, Leverage is found to negatively impact the propensity of the Energy industries to receive citations on green patents but does not yield significant results (B = -0.776, p>.10). Given the coefficient of 1.529 (p<.10), the industry-average R&D expenditures (by sales) exerts a positive influence on patent citations. Average firm age had a small, yet strongly significant positive effect on citations (*B* = 0.03, *p*<.10). Interestingly, Cohen et al. also find that, industries that are older and have higher R&D investments seem to have higher green innovation production on average.

Hypothesis 7 states that the Energy Sector tends to produce a larger quantity of patents of superior quality in comparison to the non-Energy Sector. Hypothesis 7: *The low E (ESG) industries of the Energy Sector publish higher combined volume and impact patents compared to other industries, measured by publications and citations.* Evidence supporting this proposition can be found in Model 7. In Model 7, the coefficient for the dummy variable termed dEnergySector stands at 1.03 (*p*<.01). thereby revealing a robust and positive correlation. This suggests that, on average, entities within the Energy Sector produced more green patents that are more highly cited compared to firms in the other sector. The sample comprises 515 observations, reflecting industry-year averages. The coefficients of the control variables are quite similar to Model 5 and 6. However, in line with Cohen et al. (2020), the AverageCash is negatively related to innovation (*B* = -0.26, *p* <.01), which was insignificant for Models 5 and 6.

In summation, these findings suggest that industries within the Energy Sector, typically associated with lower E (ESG), are more productive in the publication of green patents. Furthermore, these patents tend to be of superior quality, as denoted by an increased accumulation of citations. This challenges the conventional view that pollution-prone industries always correlate with lower sustainability efforts and highlights the nuanced dynamics within the industries of the Energy Sector in their pursuit of green innovation.

Table 15: Negative binomial regression results for the relationship between the Energy Sector and quantity (Model 5), quality (Model 6) and total impact (Model 7) of green patent production

Variables	5 AverageGreenPatents	6 AverageCitations	7 AveragePatents×Citations
dEnergySector	0.671***	0.510***	1.028***
	(0.161)	(0.179)	(0.352)
AverageLn(Assets)	0.159***	0.182***	0.082*
	(0.058)	(0.063)	(0.128)
AverageLeverage	-0.501	-0.776	-1.808*
	(0.406)	(0.513)	(0.998)
AverageROA	-0.445	-0.620	-1.601
	(0.576)	(0.767)	(1.933)
AverageCash	-0.091	-0.026	-0.257***
	(0.050)	(0.057)	(0.092)
AverageR&D	1.621**	1.529*	5.150*
	(0.723)	(0.874)	(2.771)
AverageAge	0.031***	0.026***	0.093***
	(0.005)	(0.006)	(0.012)
Constant	0.212	3.456***	6.219***
	(0.533)	(0.591)	(1.269)
YearFE	Controlled	Controlled	Controlled
N	515	515	515
Chi^2	163.0	322.6	330.3
p	<.001	<.001	<.001
R^2_{pseudo}	0.0495	0.0520	0.0388

Note: This table presents the results of the negative binomial regressions, examining the dependent variables average count of green patent publications 5, count of citations on those patents (6) and product of the publications and citations (7) in a given industry in a given year, over the period 2010 - 2022. The independent variable dEnergySector, is a dummy variable which equals 1 if the first two digits of Standard Industrial Classification (SIC) are 10 (Metal, Mining), 12 (Coal Mining), 13 (Oil & Gas Extraction), 14 (Nonmetallic Minerals, Except Fuels) or 29 (Petroleum & Coal Products). Control variables include industry-year averages of the log of total assets (AverageLn(Assets)), leverage ratio (AverageLeverage), return on assets (AverageROA), cash ratio (AverageCash), research & development investment relative to sales (AverageR&D) and firm age (AverageAge). The robust standard errors are in parentheses. Significance levels are denoted by asterisks (* *p*<.10; ** *p*<.05; *** *p*<.01).

5.6 Robustness Checks

5.6.1 Endogeneity

In this research, one of the ways in which endogeneity could arise is through reverse causality. Reverse causality is present when the expected outcome or dependent variable influences the predictor, creating a scenario where the direction of cause and effect becomes contrary to the hypothesized relationship. Occurrences of reverse causality challenge the conventional assumption of unidirectional causality from the independent to the dependent variable and require the use statistical techniques to identify causal relationships accurately. Albeit there is no discussion of this specific issue in previous literature, there is one ESG metric which has the potential to cause reverse causality in the relationship between the magnitude of the ESG rating and green patent publications or citations.

Namely, part of the Environmental Innovation score (independent variable) is derived from metrics that could be influenced by a firm's green patent activities (dependent variable). In turn, the weighted E score, which is determined by the Environmental Innovation score for 29%, could also be affected by the dependent variable GreenPatents. Regarding the dependent variables, all patents in this study’s sample classify as Y02, a classification which specifically captures technologies and applications for mitigating or adapting to climate change (European Patent Office, 2007). Regarding the independent variables (explained Section 3.3.2), the environmental pillar (E of ESG) consists of three categories: Emission, Environmental Innovation and Resource use. Each of these categories is comprised from several Themes (Appendix A, Table 3). The Environmental Innovation score in the Refinitiv ESG dataset is made up from two themes. It consists of several P&L items (Green revenues, green R&D expenditures and Capex) and a self-reported data point about environmental products or services. The latter resembles the patents (Y02) characteristics.

Specifically, the fact that a company publishes a green patent, meaning an increase in the dependent count variable GreenPatents, might also cause the firm to report the presence of a product line or service that is designed to have positive effects on the environment or which is environmentally labeled and marketed. In turn, by reporting ‘yes’ to the question about environmental products, Refinitiv shall award points for the Environmental Innovation category. Therefore, the regressions with the explanatory variable GreenPatents could be prone to endogeneity since a patent publication could indirectly lead to a positive answer to this ESG questionnaire. Models 2 and 4 which centered on the outcome variable Citations are slightly less susceptible to reverse causality since receiving citations on green patents (dependent variable Citations) would not immediately give rise to answering ‘yes’. Conveniently, four different ESG metrics are considered in this research, which allows for a segmented analysis of the independent variable. Emissions and Resource Use are not directly influenced by patent publications. These two ESG metrics are distinctly unaffected by green patent production which adds robustness to the findings of Models b and c. Also, Models 5, 6 and 7 are unaffected by this. The aforementioned endogeneity concerns might explain that in several regression analyses, the weighted E score (Models a) and Environmental Innovation score (Models d) show a positive relationship with green patent publications and citations, whereas the ESG ratings without these endogeneity concerns; Resource Use score (Models b) and Emission score (Models c) show that a low rating is associated with more green innovation.

Thus, reverse causality seems more likely in Models 1a, 1d, 3a and 3d (dependent variable = GreenPatents), and to a lesser extent in Models 2a, 2d, 4a and 4d (dependent variable = Citations). In Summary, the endogeneity issue related to the Environmental Innovation score is unlikely a concern for the robustness in the Models with Emission score (Models b) and Resource Use score (Models c) as independent variables. Its occurrence is more likely in the EInnovation score (Models d), and consequently in the weighted E score (Models a) which is a weighted average of the three sub-scores. Besides, even though a patent publication could result in answering yes to Refinitiv's transparency question about environmental products or services, this does not necessarily yield a one-on-one relationship between GreenPatents and EInnovation. However, to be sure, this concern shall be addressed through lagging the independent variables in Section 5.6.2.

5.6.2 Lagged Independent Variables

In addressing the potential endogeneity concerns discussed in Section 5.6.1, specifically reverse causality, all nineteen models were regressed with one-year lags on all independent variables, including control variables. This method was emloyed to determine whether ESG score classifications (low or mid/high) influenced the number of green patents or citations, instead of the other way around. Applying lags ensured that the measurement of patents and citations follows after the assessment of ESG scores, thus reducing the likelihood of reverse causality. Key findings are summarized in Appendix B, Tables 12 and 13, focusing on ESG variables: Lag(dLowE), Lag(dLowEmissions), Lag(dLowResourceUse), Lag(dLowEInnovation), and Lag(dEnergySector). In evaluating the key findings, results that are in line with the direction of the coefficient found in the lagged regression and attain a similar level of significance are deemed robust for reverse causality. Similarity of significance is concluded either if both effects have a significance level of at least 5% ($p<0.05$) or if both effects are insignificant ($p \geq 0.05$).

For Model 1, all lagged results were consistent with original findings. In Model 2a, the previously insignificant coefficient of dLowE shifted to a significant coefficient of -0.42 ($p<.01$). Models 2b, 2c, and 2d showed no notable changes.

The lagged regressions of Model 3 presented mixed outcomes. In lagged Models 3a, 3b and 3c, the coefficients were nearly similar compared to the main regression and all significance level remained the same. The coefficient of dLowE changed from -0.89 ($p<.05$) to -1.03 ($p<.05$) for Lag(dLowE). The coefficient of dLowEmissions (Model 3b) is 1.59 ($p<.01$) compared to 1.62 ($p<.01$) for its lag and the coefficient for dLowResourceUse (Model 3c) is 0.68 ($p<.01$) compared to 0.89

(p<.01) for its lag. In contrast, dLowEInnovation (Model 3d) was -0.45 at 5% and became insignificant with a coefficient of 0.002 for its Lag(dLowEInnovation).

For the Energy Sector population of Model 4a, the weighted E score dropped from dLowE -1.50 (p<.01) to Lag(dLowE) -0.48 and lost its significance. In Model 4b, the coefficient of Lag(dLowEmissions) remained positive at 0.99 (p<.05) compared to the slightly higher dLowEmissions coefficient of 1.24 (p<.01). Both lagged and non-lagged results of Model 4c were insignificant and therefore results are robust. In Model 4d, the dLowEInnovation coefficient is -1.43 (p<.01) compared to -0.60 (p<.05) for the variable Lag(dLowEInnovation). For Models 5, 6, and 7, comparing industry averages, the introduction of Lag(dEnergySector) led to similar results, suggesting a consistent direction of effect.

Aside from Model 2a, Models 1a-d and Models 2b-d did not produce coefficients that attained the minimum threshold for statistical significance, set at the 5% level. Given the lack of significance in both regression models, this lends further credibility to the robust rejection of Hypotheses 1 and 2. Conversely, Models 3a, 3b and 3c which concern the relationship between ESG and GreenPatents in the Energy Sector sample do not give rise to reverse causality concerns. The altered significance levels and coefficients regarding the variable Citations in the Energy Sector Models 4a do suggest risk of reverse causality. However, significance levels for the Emission score (Model 4b), Resource Use score (Model 4c) and Environmental Innovation score (Model 4d) remained stable. Also, the minimal changes in Models 5, 6, and 7 imply absence of reverse causality. Overall, the consistency of results across most models supports the robustness of the main findings against reverse causality concerns.

5.6.3 Lagged Dependent Variables

Based, on the findings of the Wooldridge tests in Section 4.4.3, it appears that particularly Model 1,3,5 and 7 could suffer from autocorrelation. Therefore, through the robustness checks of this section, the study explores the potential issue of autocorrelation by including lagged dependent variables as independent variables across different models. In the context of this research, autocorrelation suggests that the number of patent publications for firm i or industry j, in year $t = 0$, is influenced by the number of patent publications for the same company (Models 1-4) or industry (Models 5-7) in the previous year $t = -1$. Autocorrelation, where error terms in a regression model are not independent across observations, can lead to inefficient estimates and may impact the inference drawn from the model. By incorporating lagged outcome variables, we can control for the possibility that past values of the dependent variable influence its current value. Thus, providing a more rigorous test of the model's stability over time. Table 14 in Appendix C presents the results for the first four models including lagged dependent variables as predictors. These lagged dependent variables represent the number of patents and citations in the year prior to the actual dependent variable in the regression. In Table 15 in Appendix C, the results for lagged regression Models 5, 6 and 7 are shown. Key findings are deemed robust for autocorrelation if they align with the lagged regression's coefficient direction and achieve a similar statistical significance. The latter means that either both effects reach least 5% significance, or that both results do not attain a p-value below .05.

In Model 1, lagged results aligned with the original findings, except for Model 1a, where Lag(dLowE) exhibited a significant coefficient of -0.22 ($p<.05$), contrasting with its non-significant effect in the main regression. Similarly, in Model 2a, the previously insignificant coefficient of dLowE shifted to a significant coefficient of -0.42 ($p<.01$). Models 2b, 2c, and 2d showed no notable changes. Thus, applying lags yielded significant negative coefficients for the weighted E score on both patents and citations. Yet, in line with the non-lagged regressions, the one-year lags of the Emission, Resource Use and Environmental Innovation regressions (Models 1b-d and Models 2b-d) did not exceed 5% significance.

The following paragraph concerns the robustness of findings within the Energy Sector (Models 3 and 4). In Models 3a-d, which entail the effect of ESG on GreenPatents, the differences in coefficients and significance negligible between the main and the lagged regressions. Therefore, it is unlikely that autocorrelation is present in these models. The subsequent Models 4a-d examine the impact of various ESG ratings, including a one-year lag of Citations in the robustness test models, on the dependent variable Citations. In Models 4b and 4d, the main results were robust to autocorrelation, as the coefficients retained the same direction and remained highly significant. Therefore, these main findings are consistent when controlling for autocorrelation. In Model 4a, which explores the relationship between the weighted E score (dLowE) and citations, the negative effect of dLowE ($B = -1.50$, $p<.05$) disappeared when Lag(Citations) was added to the regression, making the coefficient insignificant ($B = -0.76$, $p>.10$). Conversely, in the relationship between the dLowResourceUse score and Citations (Model 4c), the previously insignificant coefficient ($B = 0.21$, $p>.10$) became stronger and significant ($B = 0.75$, $p<.05$) when accounting for autocorrelation. Thus, the effect, initially absent, emerged upon controlling for autocorrelation by including the variable Lag(Citations).

Turning to Models 5, 6, and 7, which compare industry averages, the lagged regressions yielded slightly weaker coefficients for the variable dEnergySector but maintained their significance and direction. The coefficients shift from 0.67 ($p<.01$), 0.51 ($p<.01$), and 1.03 ($p<.01$), in the main regressions to 0.39 ($p<.01$), 0.25 ($p<.05$), and 0.79 ($p<.01$), in the lagged regressions. These similar results in the lagged regressions for the industry averages indicate a level of consistency that suggests autocorrelation is not a predominant issue affecting the validity of the findings.

In summary, the lagged dependent variable analysis undertaken in this section confirms the robustness of the main regression findings for all ESG metrics on GreenPatents in the Energy Sector (Model 3a-d), for the Emission and Environmental Innovation on Citations (Model 4b and 4d), and for the Energy Sector dummy on the three innovation metrics in Model 5-7. While the presence of autocorrelation identified by the Wooldridge tests in Section 4.4.3 for Models 1, 3, 5, and 7 initially prompted the undertaking of this robustness check, the subsequent analysis only indicates that the autocorrelation significantly alters the study's key findings for Model 1a, out of these aforementioned Models. Moreover, the regression regarding the effect of the weighted E score on Citations (Models 2a and 4a) show significant autocorrelation risk. Overall, while some shifts in coefficient values and significance levels are observed, the overarching patterns and implications of the key findings remain intact. This reinforces confidence in these original regressions and suggests that autocorrelation does not unduly compromise the study's conclusions.

5.6.4 Alternative Cutoff

In the initial analysis of Hypotheses 1a through 4d, comparisons were made between two distinct groups. Each of the 16 models involved a population comprising a low score group and a mid/high score group based on an ESG metric, with the dividing cutoff set at the 33rd percentile. For the next robustness check, this cutoff is adjusted to the 10th percentile to examine whether an alternative group split yields consistent or divergent results. In the original models, a dummy variable indicated membership in the first tercile (below the 33rd percentile); now, for the robustness test, this dummy signifies inclusion in the lower 10th percentile. Regression models employing the 10% threshold for ESG metrics are denoted with an 'r' suffix. For instance, Model 1a examines the relationship between GreenPatents and the dummy variable dLowE at a 33rd percentile split, while Model 1ar investigates the same relationship but at a 10th percentile split. The regression results are given in Table 20 (Models 1ar-1dr), Table 21 (Models 2ar-2dr), Table 22 (Models 3ar-3dr) and Table 23 (Models 4ar-4dr).). In evaluating the key findings, results that are in line with the direction of the coefficient found in the main regression and attain a similar level of significance are deemed robust. Similarity of significance is concluded either if both effects have a significance level of at least 5% (p<0.05) or if both effects are insignificant ($p \geq 0.05$).

The results are largely consistent across most models. For the total sample analyses of ESG on green patent output (Models 1ar, 1br, and 1cr), the coefficients of the ESG metrics remain insignificant, mirroring the findings of Models 1a, 1b, 1c, and 1d. However, a notable exception is observed in the 10% threshold for Model 1dr, where the coefficient for the independent variable of interest, dLowEInnovation, is -0.571 (p<.01), indicating high significance. The only divergence among the control variable coefficients is in the 10% Model 1br, where Cash becomes marginally significant (B =0.07, p<.10). Similar patterns are observed regarding the effect of ESG metrics on the variable Citations (Models 2). The 10% cutoff regressions of the weighted E score (2ar), Emission score (2br), and Resource Use score (2cr) remain insignificant, consistent with the main regressions (Models 2a-c). Yet, the total sample analysis of the Environmental Innovation score with 10% cutoff (Model 2dr) presents a coefficient of -0.66 (p<.01), contrasting with the 33rd percentile cutoff main result (Model 2d). Other control variables maintain near-identical coefficients, magnitudes, and significance levels.

Similarly, only a very low degree of variation in ESG metrics is found in Models 3 and 4. The 10% cutoff Models 3ar, 3br, 3cr, and 3dr yield coefficients of -1.57 (p<.01), 1.21 (p<.01), 0.82 (p<.05), and -0.60 (p<.10), respectively. This barely contrasts with the 33% cutoff Models 3a, 3b, 3c, and 3d, which have coefficients of -0.89 (p<.05), 1.59 (p<.01), 0.68 (p<.01), and -0.45 (p<.05), respectively. However, since a minimum significance level of 5% is required, the findings of Model 3d are not robust. Control variables in these models remain largely consistent as well.

In Models 4ar, 4br, and 4dr, all results retain their significance as in Models 4a, 4b, and 4d. The most pronounced difference is observed in Model 4ar, where the coefficient (B = -4.22, p<.01) is nearly triple that of the 33rd percentile Model 4a (B = -1.50, p<.05).

In conclusion, robustness tests largely affirm the initial findings, with the exception of Model 1d and 2d, suggesting that the relationship between ESG ratings and green innovation is indeed robust to different threshold settings. The most significant changes and results are observed in Models 1dr and 2dr, where the dEInnovation becomes highly significant which carried no significance in the main regression Models 1d and 2d.

CHAPTER 6 Conclusion and Discussion

The motivation for this thesis arose from the observation that investors divesting from high-polluting companies, based on ESG ratings, may unintentionally hinder those companies' ability to innovate. These low ESG firms often face financial challenges due to exclusion from major investment portfolios and consequently higher equity costs. However, their innovative potential in developing solutions to combat climate change is substantial. The current ESG-driven investment strategies may inadvertently hinder these critical green innovations. Therefore, the research question of this thesis was: *Are ESG performance and environmental reputation negatively related to firm innovation as measured by the number of green patents and citations?* This thesis has critically examined the relationship between ESG scores and patenting activity, specifically in green patent publications and citations, with a focus on the Energy Sector.

In Table 24, an overview is presented of all hypotheses, showcasing the results from the main regression analyses. Additionally, the multiplicative effect of the dummy variables is displayed which enables immediate interpretation of the coefficients. The table also indicates whether the results remain robust against reverse causality, autocorrelation, and an alternative cutoff. Only when the main regression and the aforementioned three adaptations consistently yield effects in the same direction, under a significance level of 5%, the hypothesis is accepted. This approach effectively mitigates the risk of accepting false positives.

Hypotheses 1 and 2 posited that lower ESG scores would result in a higher frequency of green patent publications and citation count, respectively. None of the main regressions in question yielded significant results. For the alternative operationalizations where the insignificant results turned significant and the main results did not appear robust, all results yielded negative effects. Therefore, these models did not support the notion that throughout the total sample, firms with low ESG scores produced more (GreenPatents) or higher quality (Citations) patents and as a result, Hypotheses 1 and 2 are rejected.

Hypothesis 3 centered on the Energy Sector, proposing that within this sector specifically, companies with lower ESG scores would publish more green patents. The findings were mixed, with some ESG metrics showing positive and others negative relationships. Therefore, the Hypothesis 3 was partially accepted. For Models 3b and 3c, representing firms scoring low on Resource Use and Emissions, the hypothesis was accepted; none of the robustness tests confounded the results. Concretely, firms in the lowest tercile of dLowEmission published five times as many green patents compared to those in the mid/high group. This finding also implies that within the Energy Sector, firms ranked mid/high on the Emission score (those that pollute less) tend to produce fewer green patents. For companies scoring low on Resource Use, patent publications were twice as high. Contrarily, the results for the weighted E score (Hypothesis 3a) and Environmental Innovation score

(Hypothesis 3d) suggest that low ratings in these categories did *not* lead to a higher output of green patents as was hypothesized. This result holds robustly for Hypothesis 3a, while Hypothesis 3d lacks robustness due to inconsistencies in the one-year lag of the independent variable and the alternative 10% threshold.

Hypothesis 4 focused on patent quality within the Energy Sector, hypothesizing that lower ESG scores correlate with higher citation counts. Similar to Hypothesis 3, the results were heterogenous across different ESG group splits leading to partially accepting this hypothesis. Firms within the Energy Sector with low Emission scores (Hypothesis 4b) received substantially (factor 3.46) more citations than their higher ranked peers on the green patents which they published. The acceptance of Hypothesis 4b indicates that low Emission scores, typically associated with high carbon emissions, had a positive effect on citation counts. Contrary to the hypothesized positive effect, fewer citations were found for the group with lowest Environmental Innovation scores (Hypothesis 4d) and the lowest weighted E scores (Hypothesis 4a). Their multiplicative factors are 0.24 and 0.22 respectively, suggesting that low scoring firms received less than a fourth of the citations that the mid/high scoring group did, ceteris paribus. However, the result of for the weighted E score (Model 4a) is not robust and may be biased due to autocorrelation and reverse causality.

Hypotheses 5, 6 and 7 explored the number of green patent publications and citations in the Energy Sector in comparison to the industries belonging to the non-Energy Sector. The Energy Sector produced more green patents and generated higher citation counts, indicating both quantity and quality in green innovation. On average, Energy Sector industries produced 1.95 times the number of patents in a year compared to average number of patents in non-Energy sector industries and were cited 1.67 times as often and had a combined total impact (Patents×Citations) of 2.80 the magnitude. As a result, Hypotheses 5, 6 and 7 were accepted and their acceptance was further supported by subsequent robustness checks. As explained before, the Energy Sector consists of the following Compustat classified industries: Metal Mining, Coal mining, Oil & Gas Extraction, Nonmetallic Minerals and Petroleum & Coal Products. Thus apparently, firms in industries traditionally associated with low ESG scores are outperforming less pollutive industries in the field of green innovation. This points to the potential of the Energy Sector to drive significant advancements in environmental technology, despite their low ESG image.

The influence of the financial and operational control variables in the various ESG-related operationalizations, on the variables GreenPatents and Citations are now discussed. Across both the full sample and specifically within the Energy Sector, firm size (Ln(Assets)) emerges as a critical determinant of patent output and quality. A consistently positive relationship is observed across nearly all operationalizations, with the exception of the weighted E score's impact on Citations within the Energy Sector. When comparing the Energy Sector to the non-Energy Sector, the average firm size of an industry maintains a positive influence on innovation, albeit to a slightly lesser extent. Another financial variable which demonstrates a broadly positive impact on green innovation outcomes across the study is R&D. This effect is particularly pronounced for the weighted E score within the full sample. Focusing exclusively on the Energy Sector sample, the positive contribution of R&D expenditures to innovation is evident across all ESG metrics, affecting both patent production and citation count. These results are in line with Andersen et al. (2019), who found that control variables R&D expenses and firm size (Ln(Assets)) were a huge contributor to green patent production. The variable of Leverage shows a positive effect on the variables GreenPatents and Citations in several models, especially those involving the weighted E score and Environmental Innovation score. However, overall the effects related to Leverage, as well as profitability (ROA), are mixed in direction and significance. Contrary to expectations, both Cash holdings and firm Age play a minor role in influencing innovation outcomes within this study. This nuanced understanding of how financial and operational variables interact with ESG considerations offers valuable insights into the drivers of green innovation.

Table 24: Summary of hypotheses testing

						Robust for			
H	**Dependent Var.**	**Independent Var.**	**Effect**	**Coef.**	**Sign.**	**Reverse Causality**	**Auto-cor.**	**10% Cutoff**	**Conclusion**
1a	GreenPatents	dLowE	0.89	-0.11	Insignifct.	Yes (0)	No (-)	Yes (0)	Reject
1b	GreenPatents	dLowEmissions	0.91	-0.09	Insignifct.	Yes (0)	Yes (0)	Yes (0)	Reject
1c	GreenPatents	dLowResourceUse	0.94	-0.06	Insignifct.	Yes (0)	Yes (0)	Yes (0)	Reject
1d	GreenPatents	dLowEInnovation	0.94	-0.06	Insignifct.	Yes (0)	Yes (0)	No (-)	Reject
2a	Citations	dLowE	0.81	-0.21	Insignifct.	No (-)	No (-)	Yes (0)	Reject
2b	Citations	dLowEmissions	0.87	-0.14	Insignifct.	Yes (0)	Yes (0)	Yes (0)	Reject
2c	Citations	dLowResourceUse	0.83	-0.19	Insignifct.	Yes (0)	Yes (0)	Yes (0)	Reject
2d	Citations	dLowEInnovation	0.99	-0.01	Insignifct.	Yes (0)	Yes (0)	No (-)	Reject
3a	GreenPatents	dLowE	0.41	-0.89	$p<0.05$	Yes (-)	Yes (-)	Yes (-)	Reject
3b	GreenPatents	dLowEmissions	4.88	1.59	$p<0.01$	Yes (+)	Yes (+)	Yes (+)	Accept
3c	GreenPatents	dLowResourceUse	1.97	0.68	$p<0.01$	Yes (+)	Yes (+)	Yes (+)	Accept
3d	GreenPatents	dLowEInnovation	0.64	-0.45	$p<0.05$	No (0)	Yes (-)	No (0)	Reject
4a	Citations	dLowE	0.22	-1.50	$p<0.05$	No (0)	No (0)	Yes (-)	Reject
4b	Citations	dLowEmissions	3.46	1.24	$p<0.01$	Yes (+)	Yes (+)	Yes (+)	Accept
4c	Citations	dLowResourceUse	1.23	0.21	Insignifct.	Yes (0)	No (+)	Yes (0)	Reject
4d	Citations	dLowEInnovation	0.24	-1.43	$p<0.01$	Yes (-)	Yes (-)	Yes (-)	Reject
5	Avg.Greenpatents	dEnergySector	1.95	0.67	$p<0.01$	Yes (+)	Yes (+)	n.a.	Accept
6	Avg.Citations	dEnergySector	1.67	0.51	$p<0.01$	Yes (+)	Yes (+)	n.a.	Accept
7	Avg.Patents×Cit.	dEnergySector	2.80	1.03	$p<0.01$	Yes (+)	Yes (+)	n.a.	Accept

Note: This table summarizes the main conclusions for each hypothesis, indicated by the first column titled H, which sums all hypotheses. In the fifth column, the regression coefficients for the main regressions of the independent variables (third column) are given. The numbers in the column titled “Effect” represent the multiplicative effect of a one-unit increase of the independent variable on the outcome variable. In the seventh, eighth and ninth column, the yes/no answer is stated to the question whether the main regression results are robust for reverse causality, autocorrelation and an alternative group split, respectively. Besides, between brackets, a ‘+’ is shown if the effect found was significantly positive ($p<.05$), a ‘-‘ is shown if it was significantly negative ($p<.05$), and ‘0’ if the effect was insignificant ($p\geq 0.05$). If and only if all coefficients are positive and at least significant at the 5% level, the hypothesis was accepted.

There are two main limitations in this study. This study's first limitation stems from potential bias introduced by reverse causality. Part of the Environmental Innovation score (independent variable) is derived from metrics that could be influenced by a firm's green patent activities (dependent variable). In turn, the weighted E score, which is determined by the Environmental Innovation score for 29%, could also be (reversely) affected by the dependent variable GreenPatents. Consequently, identification issues arise in the relationship between the Environmental Innovation score and E score on the one hand, and green patent publications and citations on the other hand. As detailed in Table 24, the negative effects between scoring low on the Environmental Innovation score and patent output (Model 3d) and between the weighted E score and patent quality (Model 4a) were not robust for reverse causality. Yet, the negative effects between the low weighted E score and patent output (Model 3a) and between the Environmental Innovation score and patent quality (Model 4d) did remain robust for reverse causality, when lagging the independent variables. Despite the robust results for lagging the independent variables, the identification issues mentioned earlier might have caused the outcomes of Models 3a and 4d to deviate from the hypothesized direction. Theoretically, the observed negative effects of scoring low on the Environmental Innovation and weighted E score not illogical: publishing more patents that are more highly cited yields by definition a higher Environmental Innovation score and consequently a higher weighted E score. Namely, the positive relationship between ESG and innovation was triggered by overlap in operationalization with the dependent variables' patents and citations. Advantageously to the validity of this study, four different ESG metrics are considered in this thesis. Two of these ESG metrics are distinctly unaffected by green patent production which adds robustness to the findings of for the Emission and Resource Use scores (Models b and c). The Environmental Innovation category was particularly prone for issues with reverse causality and identification, which might explain that in several regression analyses, Environmental Innovation and the weighted E score show a positive relationship with green patent publications and citations, whereas the ESG ratings without these endogeneity concerns (Resource Use score and Emission score) show that a low rating is associated with more green innovation as hypothesized.

The second limitation of this study is the potential presence of a sample selection bias, which is particularly evident when examining the characteristics of the firm-year sample. A disparity is observed between the mean values of the weighted E score and the Environmental Innovation score for the Energy and non-Energy Sector. In the Energy Sector, the mean weighted E score was higher compared to the Environmental Innovation Score, while in the non-Energy Sector, these ratings were on average much lower. This discrepancy may be attributed to the initial selection criteria of the sample, which only included companies with at least one patent publication, thereby excluding firms or firm-years without patents from the analysis. This selection criterion inherently biased the sample towards companies with patent publications, which, in the Energy Sector, apparently resulted in a sample of firms with higher-than-average Environmental Innovation and weighted E scores. The alternative approach of including companies into the sample without any patent publications would have introduced other challenges, such as zero inflation, corrupt data points caused by zero imputation for firms with undetected patents, issues with fuzzy matching due to discrepancies in naming variations of subsidiaries and introducing potential biases caused by patent data availability. Faced with choosing between two imperfect options, the decision was made to select a sample with known data, but containing only firms with one or more patents published to be able to compare the number of patents and citations.

Nevertheless, this limitation did not substantially undermine the key findings. For Hypotheses 1 and 2, the influence of the sample selection bias—leading to inflated Environmental Innovation scores for the Energy sector—is minimal given the Energy sector's relatively small representation in the total sample, comprising less than 10% of firms. For Hypotheses 3 and 4, which exclusively concern the Energy Sector sample, the discrepancy in ESG scores between the two sectors is not relevant, as these hypotheses were investigated within the Energy Sector only. Lastly, firm-year observations with missing ESG scores were included during the analyses of Hypotheses 5, 6, and 7, which makes it unclear if this bias was also present comparing the innovation between the Energy and non-Energy sectors. Nevertheless, the lack of significant effects of the weighted E scores on the number of green patents and citations in the total sample (Hypotheses 1 and 2), reduces the likelihood that lower E or Environmental Innovation scores in the non-Energy sector influenced the overall levels of green innovation. While acknowledging the presence of a sample selection bias, it can be concluded that it is unlikely to have invalidated the overall results of the study.

Keeping in mind these limitations, it is now possible to assert with reasonable certainty that the answer to the research question is *yes*. ESG performance and environmental reputation are *not* necessarily positively related to firm innovation as measured by the number of green patents and citations. Within the Energy Sector, the ESG ratings that are devoid of an endogeneity problem, specifically the Resource Use score and Emission score relate negatively to the number of patents and citations. Besides, in a cross-sector comparison, the Energy Sector which is perceived as being low ESG, demonstrates a propensity for higher green patent publications and citations. Therefore, it can be concluded that ESG performance is often negatively related to green innovation, challenging conventional views on sustainability efforts in pollution-prone industries. In line with the resource-based view, companies that have high emissions and resource use might realize that in order to survive in the long-run, in the face of more stringent environmental regulations or because natural resources dry up, they have no option but to innovate and diversify their operations. Besides, given their size, firms in the Energy Sector more often possess the resources that increase the chances of successful innovation.

Building on this thesis' insights into green innovation and ESG performance, several avenues for future research emerge. Firstly, further research could explore whether companies with specific ESG scores primarily develop or actively purchase patents. This could be done by considering externally acquired patents, which is possible given that the Lens database provides all owners of each patent over time. This extension would offer a broader view of firms' strategies in acquiring patents versus developing green technologies in-house. Besides the interplay between firm ESG ratings and acquisition or creation of green technologies, a company's profitability could be considered as well. By combining innovation and profitability, one could test whether a firm's profitability potential is primarily based on its ability to adapt external technologies based on acquired patents or heavily relies on its internal resources, enabling the creation of novel innovations. Secondly, in extension, one could research the impact of patents produced in particular industries on other industries through the citation frequency of patents across different industries. For example, patents published in the Oil & Gas Extraction industry might be cited most in Organic Textile Manufacturing, reflecting the cross-industry application of innovations developed within the Oil & Gas Extraction industry. Thirdly, this research only focuses on US-listed firms, presenting findings primarily relevant to large North American companies. Future studies could extend this analysis to smaller, privately-held firms in Europe and other regions, examining whether the observed relationships between ESG performance and green innovation hold across different business sizes and geographical areas. This would be particularly interesting given the growing trend of sustainability ratings for non-listed firms in Europe (European Commission, 2021). More specifically, future studies could explore the relationship of sustainability scores and green innovation for Small and Medium-sized Enterprises (SMEs). This way, one could assess how emerging sustainability ratings influence the financing opportunities and innovation contributions of SMEs. Albeit, the results of this thesis cannot be generalizable across all firms, the findings are relevant for US-listed firms at least. In the next paragraph, the research' implications are discussed and a final conclusion is given.

Investors often prioritize firms with strong ESG performance, frequently employing negative screening based on ESG ratings due to its efficiency. However, this study suggests that such an exclusionary approach is outdated, particularly in industries crucial for environmental innovation. The contrasting methodological approach of low and mid/high groups aligns well with investment practices in the real world, where investors often categorize firms into "sustainable" or "non-sustainable" based on certain ESG score thresholds. By mirroring this real-world categorization, the research design gains practical relevance and applicability. Notably, energy producers, often with lower ESG scores, are found to generate significant green innovation, indicating a need to reassess investment strategies that solely focus on ESG ratings. Zagos and Brad (2020) show that patent metrics are suitable to enhance ESG factors and thus can be used for equity selection in financial products. This research underlines the importance of innovation, especially in "low ESG" firms in the Energy Sector, in driving sustainable solutions. Although high E (ESG) scores might indicate reduced emissions for a single firm, green innovations hold the potential to benefit entire industries in reducing their carbon footprints. Such technological advancements have the potential to significantly reduce emissions on a global scale and are of paramount importance in the fight against global warming. This research advocates for a more inclusive investment approach that emphasizes innovation metrics more than ESG ratings. Moreover, policy makers and financial backers should ensure more predictive operating environments so that barriers to the long-term and high-risk investments into green innovation can be lowered. In uncovering the 'Environmental Paradox', this thesis reveals a compelling truth: Sometimes, the path to green innovation paradoxically intertwines with the very industries often criticized for pollution. This cues us to recognize the potential of the traditionally low ESG industries in solving global environmental challenges. Such insights should prompt policymakers and investors to revise their strategies to support sustainable investment without stifling essential innovation. In essence, this study is not an endorsement of pollution-prone firms as environmental saviors, but rather a call for awareness of their innovative capabilities.

REFERENCES

Abrams, D., Akcigit, U. & Popadak, J. (2013). Patent value and citations: Creative destruction or strategic disruption? *National Bureau of Economic Research Working Paper* No. 19647.

Aghion, P. & Howitt, P. (2009). The economics of growth, *Massachusetts Institute of Technology*. The MIT Press.

Akey, P. & Appel, I. (2019). Environmental externalities of activism, *Social Science Research Network*. https://doi.org/10.2139/ssrn.3508808

Andersén, J., Jansson, C. & Ljungkvist, T. (2019). Can environmentally oriented CEOs and environmentally friendly suppliers boost the growth of small firms? *Business Strategy and the Environment* No 29(2), 325–334. Wiley. https://doi.org/10.1002/bse.2366

Angelucci, S., Hurtado-Albir, F. J. & Volpe, A. (2018). Supporting global initiatives on climate change: The EPO's "Y02-Y04S" tagging scheme, *World Patent Information*, No 54, 85-S92.

Anokhin, S. & Schulze, W.S. (2009). Entrepreneurship, innovation, and corruption, *Journal of Business Venturing*, 24(5), 465-476.

Ambec, S. & Lanoie, P. (2008). Does it pay to be green? A systematic overview, *The Academy of Management Perspectives*, 22(4), 45-62

Aouadi, A. & Marsat, S. (2018). Do ESG Controversies Matter for Firm Value? Evidence from International Data, *Journal of Business Ethics*, 151(4), 1027–1047. https://doi.org/10.1007/s10551-016-3213-8

Baregheh, A., Rowley, J. & Sambrook, S. (2009), "Towards a multidisciplinary definition of innovation", *Management Decision*, 47(8), 1323-1339. https://doi.org/10.1108/00251740910984578

Barney, J. (1991). Firm resources and sustained competitive advantage, *Journal of Management*, 17(1), 99-120.

Barney, J. B. (1995). Looking inside for competitive advantage, *Academy of Management Perspectives*, 9(4), 49-61.

Bartram, S. M., Hou, K. & Kim, S. (2022). Real effects of climate policy: Financial constraints and spillovers, *Journal of Financial Economics*, 143(2), 668-696.

Becker, W. & Dietz, J. (2004). R&D cooperation and innovation activities of firms—evidence for the German manufacturing industry, *Research Policy*, 33(2), 209-223.

Berg, F., Kölbel, J. F. & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings, *Review of Finance*, 26(6), 1315-1344. https://doi.org/10.1093/rof/rfac033

Boffo, R. & R. Patalano (2020), "ESG Investing: Practices, Progress and Challenges", OECD Paris, www.oecd.org/finance/ESG-Investing-Practices-Progress-and-Challenges.pdf

Brander, J. A., Egan, E. J. & Hellmann, T. F. (2008). Government sponsored venture capital in Canada: effects on value creation, competition and innovation, *NBER conference on International Differences in Entrepreneurship* No. 4.

Cainelli, G., De Marchi, V. & Grandinetti, R. (2015). Does the development of environmental innovation require different resources? Evidence from Spanish manufacturing firms, *Journal of Cleaner Production*, *94*, 211-220.

Carrasco, I. & Buendía-Martínez, I. (2016). From the business strategy result to a source of economic development: the dual role of CSR, *Journal of Small Business Strategy*, 26(1), 69-86.

Clarkson, M. E. (1995). A stakeholder framework for analyzing and evaluating corporate social performance, *Academy of Management Review*, 20(1), 92-117.

Cohen, L., Gurun, U. G. & Nguyen, Q. H. (2020). The ESG-Innovation Disconnect: Evidence from Green Patenting, *National Bureau of Economic Research Working Paper* No. 27990. https://doi.org/10.3386/w27990.

Craney, T. A. & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values, *Quality engineering*, 14(3), 391-403.

Dai, L. & Ngo, P. (2021). Political uncertainty and accounting conservatism, *European Accounting Review*, 30(2), 277-307.

Dario, K. & Heede, R. (2021). White Knights, or Horsemen of the Apocalypse? Prospects for Big Oil to align emissions with a 1.5°C pathway, *Energy Research & Social Science*.

Deegan, C. & Unerman, J. (European edition). (2006). *Financial Accounting Theory, Berkshire: McGraw-Hill Education.*

Demirel, P. & Kesidou, E. (2019). Sustainability-oriented capabilities for eco-innovation: Meeting the regulatory, technology, and market demands, *Business Strategy and the Environment*, 28(5), 847–857. https://doi.org/10.1002/bse.2286

Dicuonzo, G., Donofrio, F., Ranaldo, S. & Dell'Atti, V. (2022). The effect of innovation on environmental, social and governance (ESG) practices, *Meditari Accountancy Research*, 30(4), 1191-1209. https://doi.org/10.1108/medar-12-2020-1120

Dierickx, I. & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage, *Management Science*, 35(12), 1504-1511.

Downing, P. B. & White, L. J. (1986). Innovation in pollution control. *Journal of environmental economics and management*, *13*(1), 18-29.

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

European Commission (2008), *European Competitiveness Report*, European Commission Enterprise and Industry, available at: www.eurosfaire.prd.fr/7pc/doc/1251892016_competitivite_ue_2008.pdf

European Commission, (2021). *Sustainable Finance Disclosure Regulation (SFDR),* available at: https://finance.ec.europa.eu/sustainable-finance/disclosures/sustainability-related-disclosure-financial-services-sector_en

Fafaliou, I., Giaka, M., Konstantios, D. & Polemis, M. (2022). Firms' ESG reputational risk and market longevity: A firm-level analysis for the United States, *Journal of Business Research*, No. 149, 161-177.

Freeman, E. R. (1984). Strategic Management: A Stakeholder Approach, *Cambridge University Press*.

Friede, G. (2019). Why don't we see more action? A metasynthesis of the investor impediments to integrate environmental, social, and governance factors, *Business Strategy and the Environment*, 28(6), 1260-1282.

Galbreath, J., Charles, D. & Oczkowski, E. (2016). The Drivers of Climate Change Innovations: Evidence from the Australian Wine Industry, *Journal of Business Ethics*, No 135, 217-231. https://doi.org/10.1007/S10551-014-2461-8.

Gale, D. & Chamley, C. (1992). *Information Revelation and Strategic Delay in a Model of Investment* No. 10.

Galende, J. (2006). Main determinants of product and process R&D cooperation: an empirical analysis for Spanish firms, *Technovation*, 26(1), 77-88.

Garel, A. & Petit-Romec, A. (2021). Investor rewards to environmental responsibility: Evidence from the COVID-19 crisis, *Journal of Corporate Finance*, No. 68, 101948. https://doi.org/10.1016/j.jcorpfin.2021.101948

Garzón-Jiménez, R. & Zorio-Grima, A. (2021). Effects of Carbon Emissions, Environmental Disclosures and CSR Assurance on Cost of Equity in Emerging Markets, *Journal of Sustainability*. https://doi.org/10.3390/SU13020696.

Gillan, S. L., Koch, A. & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance* ,No. 66, 101889.

Grant, R. M. (1991). The resource-based theory of competitive advantage: implications for strategy formulation, *California Management Review*, 33(3), 114-135.

Griliches, Z. (1980). Patents and r&d at the firm level: A first report, *Economic letters*, 5(4), 377–381.

Griliches, Z. (1998). Patent statistics as economic indicators: a survey, *Journal of Economic Literature*, *28*(4), 1523-1543.

Griliches, Z., Pakes, A. & Hall, B. H. (1986). The value of patents as indicators of inventive activity, *NBER Working paper*, (w2083).

Gyönyörová, L., Stachoň, M. & Stašek, D. (2023). ESG ratings: relevant information or misleading clue? Evidence from the S&P Global 1200, *Journal of Sustainable Finance & Investment*, 13(2), 1075-1109.

Hall, B., Griliches, Z. & Hausman, J. (1986). Patents and R and D: Is there a lag? *International Economic Review*, *27*(2), 265-283.

Hall, B.H. & Rosenberg, N. (2010), “Chapter 1 – Introduction to the Handbook”, in B.H. Hall, and N. Rosenberg, (eds.), *Handbook of the Economics of Innovation*, No. 1, North-Holland, UK.

Harrison, J. S., Phillips, R. A. & Freeman, R. E. (2020). On the 2019 Business Roundtable “Statement on the Purpose of a Corporation.”, *Journal of Management*, 46(7), 1223–1237. https://doi.org/10.1177/0149206319892669

Heede, R. (2019). Carbon Majors: Accounting for carbon and methane emissions 1854-2010 Methods & Results Report, *Climate Accountability Institute*. ISBN 978-3-659-57841-0, 148.

Hoang, T., Przychodzen, W., Przychodzen, J. & Segbotangni, E.A. (2020). Does it pay to be green? A disaggregated analysis of U.S. firms with green patents, *Business Stratetgy and the Environment*; 29, 1331–1361. https://doi.org/10.1002/bse.2437

Hsu, P. H., Li, K. & Tsou, C. Y. (2022). The pollution premium. *Journal of Finance*, forthcoming.

Ilhan, E., Sautner, Z. & Vilkov, G. (2021). Carbon tail risk, *The Review of Financial Studies*, 34(3), 1540-1571

Itami, H. & Roehl, T. W. (1991). Mobilizing invisible assets. *Harvard University Press*.

Jensen, M. (2001). Value maximisation, stakeholder theory, and the corporate objective function, *European financial management*, 7(3), 297-317.

Jongsma, M. (2022). Zijn wapens, olie en gas nu toch een verantwoordelijke belegging? *Het Financiële Dagblad*. https://fd.nl/financiele-markten/1437011/zijn-wapens-olie-en-gas-nu-toch-een-verantwoorde-belegging-mce2caS7RjZX

Katila, R. (2000). Using patent data to measure innovation performance, *International Journal of Business Performance Management*, 180-193.

Khan, M., Serafeim, G. & Yoon, A. (2016). Corporate sustainability: First evidence on materiality, *The accounting review*, *91*(6), 1697-1724.

Kotha, R., Zheng, Y. & George, G. (2010). Entry into New Niches: The Effects of Firm Age and the Expansion of Technological Capabilities on Innovative Output and Impact, *ERN: Innovation*. https://doi.org/10.1002/SMJ.915.

Kumar, K. (2020). A History of the Solar Cell, in Patents, *Intellectual Property Magazine*.

Lens. (2023). Patent Cited By Patents Tabs. https://support.lens.org/knowledge-base/patent-cited-by-patents-tabs/

Lin, Z. (2023). Sustainability as a Catalyst for Innovation: Evidence from ESG Rating Event. http://dx.doi.org/10.2139/ssrn.4586511

Lin, W., Cheah, J., Azali, M., Ho, J. & Yip, N. (2019). Does firm size matter? Evidence on the impact of the green innovation strategy on corporate financial performance in the automotive sector, *Journal of Cleaner Production.* https://doi.org/10.1016/J.JCLEPRO.2019.04.214.

Lin, H., Zeng, S.X., Ma, H.Y., Qi, G.Y. & Tam, V.W. (2014). Can political capital drive corporate green innovation? Lessons from China, *Journal of Cleaner Production*, 64, 63–72. https://doi.org/10.1016/j.jclepro.2013.07.046

Magat, W. (1979). The Effects of Environmental Regulation on Innovation, *Law and Contemporary Problems*, 43(1), 4-25.

Mansfield, E. (1988). Industrial R&D in Japan and the United States: A comparative study, *The American Economic Review*, 78(2), 223-228.

McDougall, N., Wagner, B. & MacBryde, J. (2019). An empirical explanation of the natural-resource-based view of the firm, *Production Planning & Control*, 30(16), 1366–1382. https://doi.org/10.1080/09537287.2019.1620361

McWilliams, A. & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective, *Academy of management review*, 26(1), 117-127.

Milliman, S. R. & Prince, R. (1989). Firm Incentives to Promote Technological Change in Pollution Control, *Journal of Environmental Economics and Management*, 247-265.

Paradis, G. & Schiehll, E. (2021). ESG Outcasts: Study of the ESG Performance of Sin Stocks. *Sustainability*, *13*(17), 9556.

Poh, J. (2019). Conflicting ESG Ratings Are Confusing Sustainable Investors. *Bloomberg*. https://www.bloomberg.com/news/articles/2019-12-11/conflicting-esg-ratings-are- confusing-sustainable-investors

Post, J. E., Preston, L. E. & Sauter-Sachs, S. (2002). Redefining the corporation: Stakeholder management and organizational wealth, *Stanford University Press*.

Refinitiv (2022): Environmental, Social and Governance scores from Refinitiv - May 2022. Refinitiv, an LSEG business. Available online at https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf , checked on 10/10/2022.

Reitzig, M. (2004). Improving patent valuations for management purposes: Validating new indicators by analyzing application rationales, *Research Policy*, 33(67), 939–957.

Savignac, F. (2008). "Impact of financial constraints on innovation: What can be learned from a direct measure?", *Economic Innovation and New Technology*, 17(6), 553–569.

Schumpeter, J. A. (1942). Socialism, capitalism and democracy, *Harper and Brothers*.

Scott, W. R. (2008). Institutions and Organizations: Ideas, Interests, and Identities, *Sage Publications*.

Senaviratna, N. A. M. R. & Cooray, T. M. J. A. (2019). Diagnosing multicollinearity of logistic regression model, *Asian Journal of Probability and Statistics*, *5*(2), 1-9.

Sharfman, M. P. & Fernando, C. S. (2008). Environmental risk management and the cost of capital, *Strategic management journal*, 29(6), 569-592.

Shefer, D. & Frenkel, A. (2005). R&D, firm size and innovation: an empirical analysis, *Technovation*, 25(1), 25-32.

Surroca, J., Tribó, J. A. & Waddock, S. (2010). Corporate responsibility and financial performance: The role of intangible resources, *Strategic management journal*, *31*(5), 463-490.

Swann, P.G.M. (2009), The Economics of Innovation, *Edward Elgar Publishing, Oxfordshire, England*.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance, *Strategic Management Journal*, 28(13), 1319-1350.

Trajtenberg, M. (1990). Economic analysis of product innovation: The case of CT scanners, No. 160, *Harvard University Press*.

UN News (2021) Climate and weather related disasters surge five-fold over 50 years, but early warnings save lives, WMO. https://news.un.org/en/story/2021/09/1098662

USSIF Foundation (2020), Report on US Sustainable and Impact Investing Trends.

Van der Linde, C. (1993). The Microeconomic Implication of Environmental Regulation: A Preliminary Framework, *Environmental Policies and Industrial Competitiveness*. Paris. OECD, 69-77.

Wang, J., Ma, M., Dong, T. & Zhang, Z. (2023). Do ESG ratings promote corporate green innovation? A quasi-natural experiment based on SynTao Green Finance's ESG ratings, *International Review of Financial Analysis*, No. 87, 102623.

Wernerfelt, B. (1984). A resource-based view of the firm, *Strategic Management Journal*, 5(2), 171-180.

Withers, M., Drnevich, P. & Marino, L. (2011). Doing More with Less: The Disordinal Implications of Firm Age for Leveraging Capabilities for Innovation Activity, *Journal of Small Business Management*, No. 49, 515-536. https://doi.org/10.1111/j.1540-627X.2011.00334.x.

Zhang, C. & Chen, D. (2023). Do environmental, social, and governance scores improve green innovation? Empirical evidence from Chinese-listed companies, *PloS one*, 18(5), e0279220.

Zhang, D., Rong, Z. & Ji, Q. (2019). Green innovation and firm performance: evidence from listed companies in China, *Resources, Conservation and Recycling*, No. 144, 48-55.

Zhang, Y., Sun, J., Yang, Z. & Wang, Y. (2020b). Critical success factors of green innovation: technology, organization and environment readiness, *Journal of Cleaner Production*, No. 264, 121701, doi:10.1016/j.jclepro.2020.121701

Zhang, Y. & Weber, O. (2022). Investors' Moral and Financial Concerns—Ethical and Financial Divestment in the Fossil Fuel Industry, *Sustainability*, 14(4), 1952.

APPENDIX A

Table 3: ESG themes per category with respective data points evaluated as proxies of ESG

Pillar	Category	Theme	Data Points	Weighting Method
Environmental	Emission	Emissions	TR.AnalyticCO2	Quant industry median
		Waste	TR.AnalyticTotalWaste	Quant industry median
		Biodiversity	None	n.a.
		Environmental management systems	None	n.a.
	Innovation	Product Innovation	TR.AnalyticEnvProducts[1]	Transparancy weights
		Green revenues, R&D, Capex	TR.AnalyticEnvRD	Quant industry median
	Resource-Use	Water	TR.AnalyticWaterUse	Quant industry median
		Energy	TR.AnalyticEnergyUse	Quant industry median
		Sustainable Packaging	None	n.a.
		Environmental supply chain	None	n.a.

Table 4: Overview of control variables

Financial Ratios	Variable Name	Variable Name from source	Category	Formula	Source
Research and Development/Sales	R&D	rd_sale	Other	R&D expenses as a fraction of Sales	Compustat, financial ratios
Cash Ratio	Cash	cash_ratio	Liquidity	Cash and Short-term Investments as a fraction of Current Liabilities	Compustat, financial ratios
Return on Assets	ROA	roa	Profitability	Operating Income Before Depreciation as a fraction of average Total Assets based on most recent two periods	Compustat, financial ratios
Total Debt/Total Assets	Leverage	debt_assets	Solvency	Total Debt as a fraction of Total Assets	Compustat, financial ratios
Firm Age	Age	age	n.a.	n.a.	Comustat company data
Firm Size	Ln(Assets)	at	Other	Total Assets	Compustat, annual fundamentals

[1]Code: TR.EnvProducts, Title: Environmental Products, Description: Does the company report on at least one product line or service that is designed to have positive effects on the environment or which is environmentally labeled and marketed? - in focus are the products and services that have positive environmental effects, or marketed as which solve environment problems. https://developers.lseg.com/en/article-catalog/article/Analyze_EU_Taxonomy_climate_change

APPENDIX B

Table 5: Summary statistics of sample by 2-digit SIC

SIC2	N	Mean	SD
1	2	0	0
15	13	15.92	19.76
16	18	29.72	34.54
17	2	6.50	9.19
20	71	25.56	36.34
21	16	21.56	38.65
22	19	28.00	57.22
23	13	36.00	57.55
24	13	6.54	9.14
25	20	75.05	86.05
26	44	69.48	140.83
27	22	25.55	35.01
28	384	75.07	213.99
30	31	75.90	115.54
31	13	7.08	11.27
32	12	47.67	31.23
33	88	24.17	43.31
34	62	9.19	18.41
35	373	129.01	304.36
36	485	358.24	1015.23
37	232	473.78	1434.63
38	277	84.87	192.89
39	20	29.20	55.22
40	10	18.20	21.09
41	7	88.00	98.03
42	30	66.40	99.18
44	8	72.50	92.61
45	30	218.73	321.06
46	1	0	.
48	74	60.16	106.96
49	146	121.31	308.01
50	41	46.63	111.95
51	24	21.33	21.74
52	1	0	.
53	12	95.67	171.08
54	3	4.00	4.58
55	6	29.50	26.77
56	35	49.71	141.35
57	23	35.35	50.62
58	19	26.11	65.10
59	6	5.67	9.18
60	114	55.45	196.86
61	25	28.24	45.34

62	71	50.20	101.88
63	34	21.18	34.69
64	25	114.88	240.18
65	25	110.88	212.82
67	177	52.84	147.74
70	6	11.68	14.75
72	1	1.00	.
73	312	90.60	330.97
75	9	34.44	81.10
78	4	3.50	3.70
79	28	38.50	69.84
80	14	14.14	16.63
82	21	16.38	22.59
83	6	46.33	33.39
87	23	13.35	21.66
99	14	1117.79	1146.36

Table 7: Summary statistics of dataset containing firm-year observations by sector (dEnergySector)

Variables, **dEnergy: 0**	N	Mean	SD	Min	Max
GreenPatents	2376	20.73	83.57	1	1835
Citations	2376	148.93	577.36	0	9322
E	1408	57.67	20.01	10.96	97.44
Emissions	2091	51.81	28.98	0.18	99.82
ResourceUse	2126	55.58	29.44	0.30	99.9
EInnovation	1576	50.55	24.22	0.49	99.32
Ln(Assets)	2376	8.61	1.73	1.06	13.22
Leverage	2376	0.57	0.21	0.10	1.17
ROA	2376	0.12	0.12	-0.57	.32
Cash	2376	1.10	1.64	0.04	13.83
R&D	2376	0.13	0.59	0	8.07
Age	2376	33.58	20.28	1	72

Variables, **dEnergy: 1**	N	Mean	SD	Min	Max
GreenPatents	190	23.07	34.31	1	167
Citations	190	220.50	460.63	0	3898
E	79	73.43	15.71	26.02	91.88
Emissions	175	54.29	30.60	1.47	99.69
ResourceUse	170	56.44	33.47	0.20	99.78
EInnovation	86	65.60	22.46	21.88	90.74
Ln(Assets)	190	9.46	2.03	4.31	12.80
Leverage	190	0.47	0.15	0.10	1.16
ROA	190	0.12	0.09	-0.17	0.32
Cash	190	0.73	1.32	0.04	13.83
R&D	190	0.01	0.01	0	0.03
Age	190	36.73	21.38	2	71

Table 8: Summary statistics of dataset containing industry-year observations

Variables	N	Mean	SD	Min	Max
AverageGreenPatents	607	13.05	26.64	1	299
AverageCitations	607	110.57	273.69	0	3821
AverageLn(Assets)	553	8.75	1.44	1.40	12.54
AverageLeverage	559	0.59	0.16	0.12	1.17
AverageROA	559	0.12	0.09	-0.57	.32
AverageCash	521	0.86	1.10	0.04	9.11
AverageR&D	559	0.05	0.17	0	2.21
AverageAge	607	32.25	14.21	1	68

Table 9: Summary statistics of dataset containing industry-year observations by sector (dEnergySector)

Variables, **dEnergy: 0**	N	Mean	SD	Min	Max
AverageGreenPatents	553	12.27	27.17	1	299
AverageCitations	553	104.46	278.80	0	3821
AverageLn(Assets)	508	8.73	1.40	1.40	12.54
AverageLeverage	514	.61	0.16	0.12	1.17
AverageROA	514	.12	0.09	-0.57	0.32
AverageCash	476	.86	1.11	0.04	9.11
AverageR&D	514	.05	0.17	0	2.21
AverageAge	553	32.46	14.31	1	68

Variables, **dEnergy: 1**	N	Mean	SD	Min	Max
AverageGreenPatents	54	21.10	18.81	1	65.25
AverageCitations	54	173.17	206.17	0	664
AverageLn(Assets)	45	8.99	1.79	4.59	11.73
AverageLeverage	45	0.48	0.12	0.15	0.87
AverageROA	45	0.12	0.07	-0.04	0.32
AverageCash	45	0.87	1.05	0.11	6.14
AverageR&D	45	0.002	0.002	0	0.01
AverageAge	54	30.06	13.08	10	66.67

APPENDIX C

Table 16: Lagged regression results for the relationship between GreenPatents (Models 1 and 3) or Citations (Models 2 and 4) and the one-year lag of the independent variables. The model includes lagged independent variables to correct/control for reverse causality

	Lagged independent variables			
Dependent variables	**(Models a) Lag(dLowE)**	**(Models b) Lag(dLowEmissions)**	**(Models c) Lag(dLowResourceUse)**	**(Models d) Lag(dLowEInnovation)**
Models 1 (Total sample)				
GreenPatents	-0.223*	-0.0131	-0.0975	-0.112
	(0.116)	(0.104)	(0.0951)	(0.1000)
N	1,032	1,508	1,529	1,146
Chi^2	468.8	328.6	434.8	408.3
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.0725	0.0613	0.0672	0.0685
Models 2 (Total sample)				
Citations	-0.416***	-0.064	-0.188*	-0.104
	(0.138)	(0.129)	(0.112)	(0.113)
N	1,032	1,508	1,529	1,146
Chi^2	613.2	457.7	687.5	563.2
p	< .001	< .001	< .001	< .001
R^2_{pseudo}	0.0534	0.0480	0.0506	0.0517
Models 3 (Energy Sector)				
GreenPatents	-1.032**	1.619***	0.891***	0.002
	(0.471)	(0.292)	(0.323)	(0.191)
N	58	122	116	63
Chi^2	387.7	174.4	122.1	603.4
p	< .001	< .001	< .001	< .001
R^2_{pseudo}	0.129	0.0736	0.0717	0.138
Models 4 (Energy Sector)				
Citations	-0.481	0.996**	0.228	-0.600**
	(0.704)	(0.434)	(0.373)	(0.306)
N	58	122	116	63
Chi^2	318.1	208.7	266.4	345.9
p	< .001	< .001	< .001	< .001
R^2_{pseudo}	0.0846	0.0613	0.0623	0.0781

Note: This table presents the results of the negative binomial regressions, examining the association between the one-year lags of the firms' ESG scores and the count of green patent publications (Models 1 and 3) or the citation count (Models 3 and 4). Dummy variables for the one-year lags of Low E score (Model 1a), Low Emissions score (Model 1b), Low Resource Use score (Model 1c), and Low Environmental Innovation score (Model 1d) represent firms within the lowest tercile of each ESG category compared to firms in the middle and higher terciles, the reference group (>33 pct.). Control variables include the log of total assets (Ln(Assets)), leverage ratio (Leverage), return on assets (ROA), cash ratio (Cash), research & development investment relative to sales (R&D), firm age (Age) and Year Fixed effects. All independent variables are lagged to adjust for reverse causality. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

Table 17: Lagged regression results for the for the relationship between the one-year lag of Energy Sector the quantity (Model 5), quality (Model 6) and product of quantity and quality (Model 7) of green patent production. The model includes lagged independent variables to correct/control for reverse casuality.

	(Model 5) AverageGreenPatents	**(Model 6) AverageCitations**	**(Model 7) AveragePatents×Citations**
Lagged IV			
Lag(dEnergySector)	0.668***	0.444**	0.838***
	(0.167)	(0.172)	(0.331)
N	440	440	440
Chi^2	128.1	330.4	248.9
p	<.001	<.001	<.001
R^2_{pseudo}	0.0450	0.0562	0.0364

Note: This table presents the results of the negative binomial regressions, examining the dependent variables average count of green patent publications (5), count of citations on those patents (6) and product of the publications and citations (7) in a given industry in a given year, over the period 2010 - 2022. The independent variable Lag(dEnergySector), is the one-year lag of the dummy variable which equals 1 if the first two digits of Standard Industrial Classification (SIC) are 10 (Metal, Mining), 12 (Coal Mining), 13 (Oil & Gas Extraction), 14 (Nonmetallic Minerals, Except Fuels) or 29 (Petroleum & Coal Products). Control variables include one-year lags of industry-year averages of log of total assets (AverageLn(Assets)), leverage ratio (AverageLeverage), return on assets (AverageROA), cash ratio (AverageCash), research & development investment relative to sales (AverageR&D), firm age (AverageAge). The robust standard errors are in parentheses. Significance levels are denoted by asterisks (* $p<.10$; ** $p<.05$; *** $p<.01$).

Table 18: Lagged regression results for the relationship between green patents or citations and ESG scores, including the one-year lag of GreenPatents (Models 1 and 3) or the one-year lag of Citations (Models 2 and 4) and the control variables. The models include lagged dependent variables as predictors to correct/control for autocorrelation

	(Models a) dLowE	**(Models b) dLowEmissions**	**(Models c) dLowResourceUse**	**(Models d) dLowEInnovation**
Models 1 (Total sample)				
GreenPatents	-0.185**	-0.116*	-0.0703	-0.128*
	(0.0785)	(0.0651)	(0.0626)	(0.0655)
N	1,096	1,610	1,629	1,194
Chi^2	520.2	484.6	620.7	455.7
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.129	0.132	0.134	0.128
Models 2 (Total sample)				
Citations	-0.364***	-0.168*	-0.143	-0.146
	(0.107)	(0.095)	(0.093)	(0.095)
N	1,096	1,610	1,629	1,194
Chi^2	630.9	712.6	794.9	591.2
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.0759	0.0757	0.0753	0.0752
Models 3 (Energy Sector)				
GreenPatents	-0.994***	0.678***	0.354**	-0.360**
	(0.283)	(0.218)	(0.204)	(0.166)
N	64	130	125	69
Chi^2	554.1	379.2	259.5	868.4
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.182	0.159	0.150	0.182
Models 4 (Energy Sector)				
Citations	-0.763	1.402***	0.751**	-1.133***
	(0.637)	(0.303)	(0.329)	(0.312)
N	64	130	125	69
Chi^2	433.2	343.7	274.4	386.1
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.0937	0.0871	0.0849	0.104

Note: This table presents the results of the negative binomial regressions, examining the association between the firms' ESG scores, including the one-year lags of the count of green patent publications (Models 1 and 3) or the one-year lags of the citation count (Models 3 and 4), on the dependent variables GreenPatents and Citations. Dummy variables for the Low E score (Model 1a), Low Emissions score (Model 1b), Low Resource Use score (Model 1c), and Low Environmental Innovation score (Model 1d) represent firms within the lowest tercile of each ESG category compared to firms in the middle and higher terciles, the reference group (>33 pct.). Control variables include the lagged dependent variable (number of patents or citations) to adjust for autocorrelation, the log of total assets (Ln(Assets)), leverage ratio (Leverage), return on assets (ROA), cash ratio (Cash), research & development investment relative to sales (R&D), firm age (Age) and Year Fixed effects. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

Table 19: Lagged regression results for the for the relationship between the quantity (Model 5), quality (Model 6) or product of quantity and quality (Model 7) of green patent production and the dummy Energy Sector, including the one year lag of AverageGreenPatents (Model 5), AverageCitations (Model 6) or AveragePatents×Citations (Model 7) as independent variable, respectively. The model includes lagged dependent variables as predictors to correct/control for autocorrelation.

	Dependent Variables		
	(Model 5) (AverageGreenPatents)	**(Model 6) (AverageCitations)**	**(Model 7) (AveragePatents×Citations)**
dEnergySector	0.387***	0.250**	0.790***
	(0.130)	(0.149)	(0.253)
N	438	438	438
Chi^2	252.7	376.6	242.6
p	<.001	<.001	<.001
R^2_{pseudo}	0.131	0.0885	0.0559

Note: This table presents the results of the negative binomial regressions, examining the one-year lags of the dependent variables average count of green patent publications (5), count of citations on those patents (6) and product of the publications and citations (7) in a given industry in a given year. Lagged dependent variables were included as predictors to account for autocorrelation. The independent variable dEnergySector, is a dummy variable which equals 1 if the first two digits of Standard Industrial Classification (SIC) are 10 (Metal, Mining), 12 (Coal Mining), 13 (Oil & Gas Extraction), 14 (Nonmetallic Minerals, Except Fuels) or 29 (Petroleum & Coal Products). Control variables include industry-year averages of log of total assets (AverageLn(Assets)), leverage ratio (AverageLeverage), return on assets (AverageROA), cash ratio (AverageCash), research & development investment relative to sales (AverageR&D), firm age (AverageAge). The robust standard errors are in parentheses. Significance levels are denoted by asterisks (* *p*<.10; ** *p*<.05; *** *p*<.01).

Table 20.: Negative binomial regression results for the relationship between E (ESG) scores and the quantity of patent publications with low-group cutoff at first decile

	1ar	**1br**	**1cr**	**1dr**
Variables		**GreenPatents**		
dLowE	0.0256			
	(0.295)			
dLowEmmissions		0.0702		
		(0.140)		
dLowResourceUse			-0.0716	
			(0.126)	
dLowEInnovation				-0.571***
				(0.169)
Ln(Assets)	0.514***	0.489***	0.541***	0.439***
	(0.0325)	(0.0304)	(0.0330)	(0.0403)
Leverage	0.794***	0.253	0.303	0.358
	(0.269)	(0.232)	(0.211)	(0.281)
ROA	0.390	-0.545	-0.215	-0.308
	(0.657)	(0.422)	(0.409)	(0.620)
Cash	-0.0712	0.0668*	0.122***	0.146**
	(0.0621)	(0.0395)	(0.0383)	(0.0631)
R&D	5.858***	0.0585	0.184*	1.146
	(0.808)	(0.0753)	(0.0980)	(1.261)
Age	0.00384*	0.00152	0.00137	0.00418*
	(0.00209)	(0.00195)	(0.00190)	(0.00230)
Constant	-3.024***	-2.114***	-2.719***	-1.770***
	(0.415)	(0.356)	(0.370)	(0.504)
YearFE	controlled	controlled	controlled	controlled
N	1,487	2,266	2,296	1,662
Chi^2	461.3	393.2	474.3	315.1
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.0719	0.0620	0.0681	0.0580

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent publications over the period 2010 - 2022. Dummy variables for low E score (Model 1ar), low Emissions score (Model 1br), low Resource Use score (Model 1cr), and low Environmental Innovation score (Model 1dr) represent firms within the lowest decile of each ESG category compared to firms in the higher nine deciles, the reference group (>10 pct.). Control variables include Log Total Assets, Leverage ratio, return on assets (ROA), Cash ratio, R&D investment relative to sales, and Firm Age. Coefficients indicate the expected change in the log count of green patents published for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the green patents count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

Table 21.: Negative binomial regression results for the relationship between E (ESG) scores and the quality of patent publications, measured by citation count, with low-group cutoff at first decile

	2ar	**2br**	**2cr**	**2dr**
Variables		**Citations (patent quality)**		
dLowE	-0.076			
	(0.344)			
dLowEmmissions		-0.157		
		(0.154)		
dLowResourceUse			-0.169	
			(0.137)	
dLowEInnovation				-0.655***
				(0.194)
Ln(Assets)	0.536***	0.496***	0.547***	0.467***
	(0.038)	(0.035)	(0.038)	(0.044)
Leverage	-0.085	-0.248	-0.162	-0.298
	(0.312)	(0.290)	(0.272)	(0.321)
ROA	1.244*	-0.044	0.181	-0.230
	(0.674)	(0.440)	(0.470)	(0.695)
Cash	-0.142**	0.138**	0.185***	0.109**
	(0.067)	(0.059)	(0.057)	(0.048)
R&D	6.483***	0.119	0.218**	1.542
	(0.946)	(0.106)	(0.099)	(1.162)
Age	0.003	0.001	0.001	0.004
	(0.002)	(0.002)	(0.002)	(0.003)
Constant	0.346	1.104***	0.525	1.655***
	(0.470)	(0.406)	(0.411)	(0.549)
YearFE	controlled	controlled	controlled	controlled
N	1,487	2,266	2,296	1,662
Chi^2	492.5	590.9	662.7	480.3
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.0509	0.0469	0.0477	0.0432

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent citations over the period 2010 - 2022. Dummy variables for low E score (Model 1a), low Emissions score (Model 1b), low Resource Use score (Model 1c), and low Environmental Innovation score (Model 1d) represent firms within the lowest decile of each ESG category compared to firms in the higher nine deciles, the reference group (>10 pct.). Control variables include Log Total Assets, Leverage ratio, return on assets (ROA), Cash ratio, R&D investment relative to sales, and Firm Age. Coefficients indicate the expected change in the log count of citations for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the citations count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* *p*<.10; ** *p*<.05; *** *p*<.01).

Table 22: Negative binomial regression results for the relationship between E (ESG) scores and the quantity of patent publications in the Energy Sector, with low-group cutoff at first decile

	3ar	**3br**	**3cr**	**3dr**
Variables		**GreenPatents**		
dLowE	-1.568***			
	(0.557)			
dLowEmmissions		1.214***		
		(0.270)		
dLowResourceUse			0.820**	
			(0.438)	
dLowEInnovation				-0.598*
				(0.338)
Ln(Assets)	0.578***	0.466***	0.591***	0.896***
	(0.097)	(0.067)	(0.084)	(0.056)
Leverage	1.010	-2.559***	-1.924**	2.576***
	(1.003)	(0.849)	(0.851)	(0.946)
ROA	0.451	0.342	-0.919	0.637
	(2.512)	(1.643)	(1.681)	(2.258)
Cash	-0.825***	-0.128*	-0.039	0.179***
	(0.275)	(0.070)	(0.156)	(0.037)
R&D	69.145***	82.707***	69.423***	72.150***
	(11.980)	(26.579)	(26.372)	(12.621)
Age	0.002	-0.013*	-0.017**	-0.005
	(0.005)	(0.007)	(0.009)	(0.005)
Constant	-4.162***	-0.713	-1.963**	-8.411***
	(1.512)	(0.769)	(0.794)	(0.831)
YearFE	controlled	controlled	controlled	controlled
N	79	175	170	86
Chi^2	512.2	176.4	200.1	771.5
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.133	0.0784	0.0716	0.139

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent publications over the period 2010 – 2022. Dummy variables for low E score (Model 3ar), low Emissions score (Model 3br), low Resource Use score (Model 3cr), and low Environmental Innovation score (Model 3dr) represent firms within the lowest decile of each ESG category compared to firms in the higher nine deciles, the reference group (>10 pct.). Control variables include Log Total Assets, Leverage ratio, return on assets (ROA), Cash ratio, R&D investment relative to sales, and Firm Age. Coefficients indicate the expected change in the log count of green patents published for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the green patents count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

Table 23.: Negative binomial regression results for the relationship between E (ESG) scores and the quality of patent publications, measured by citation count in the Energy Sector, with low-group cutoff at first decile

Variables	**4ar**	**4br**	**4cr**	**4dr**
	Citations (patent quality)			
dLowE	-4.233***			
	(0.377)			
dLowEmmissions		1.049***		
		(0.336)		
dLowResourceUse			0.286	
			(0.516)	
dLowEInnovation				-1.138**
				(0.544)
Ln(Assets)	-0.246*	0.405***	0.425***	0.651***
	(0.140)	(0.083)	(0.094)	(0.140)
Leverage	-1.586	-1.522	-0.984	3.028**
	(1.333)	(1.042)	(0.994)	(1.383)
ROA	-1.727	2.097	-1.043	-0.743
	(2.958)	(2.140)	(2.106)	(3.919)
Cash	-2.417***	-0.007	-0.228	0.261***
	(0.356)	(0.061)	(0.146)	(0.065)
R&D	54.172***	86.886***	60.558**	56.290**
	(18.784)	(33.152)	(27.126)	(22.131)
Age	0.019***	-0.011	-0.006	0.001
	(0.007)	(0.009)	(0.009)	(0.008)
Constant	9.944***	1.961**	2.015*	-2.567
	(2.122)	(0.934)	(1.057)	(2.014)
YearFE	controlled	controlled	controlled	controlled
N	79	175	170	86
Chi^2	576.0	290.1	308.8	184.5
p	<.001	<.001	<.001	<.001
R^2_{pseudo}	0.0845	0.0616	0.0580	0.0572

Note: This table presents the results of the negative binomial regressions, examining the association between firms' ESG scores and the count of green patent publications over the period 2010 – 2022. Dummy variables for low E score (Model 1a), low Emissions score (Model 1b), low Resource Use score (Model 1c), and low Environmental Innovation score (Model 1d) represent firms within the lowest decile of each ESG category compared to firms in the higher nine deciles, the reference group (>10 pct.). Control variables include Log Total Assets, Leverage ratio, return on assets (ROA), Cash ratio, R&D investment relative to sales, and Firm Age. Coefficients indicate the expected change in the log count of citations for a one-unit change in the predictor variable. Exponentiating the coefficients will give the multiplicative effect on the citations count. Robust standard errors are in parentheses. Significance levels are denoted by asterisks (* p<.10; ** p<.05; *** p<.01).

APPENDIX D

All residual plots depicted on this page illustrate the relationship between linear predictors (X-axis) and deviance residuals (Y-axis).

Figure 1: E-score on GreenPatents for the entire sample (Model 1)

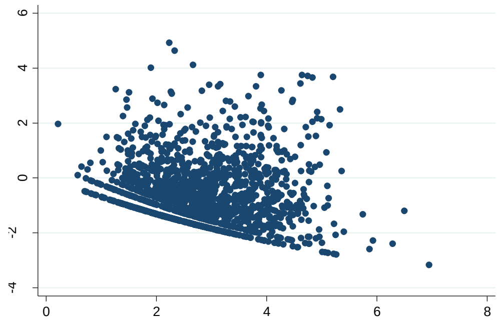


Figure 2: E-score on Citations for the entire sample (Model 2)

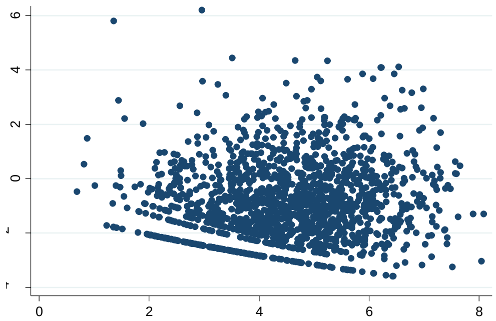


Figure 3: E-Score on GreenPatents within the Energy Sector sample (Model 3)

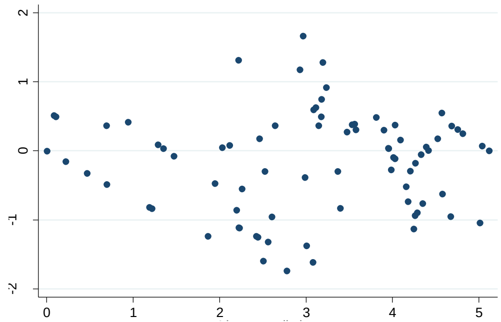


Figure 4: E-Score on Citations within the Energy Sector sample (Model 4)

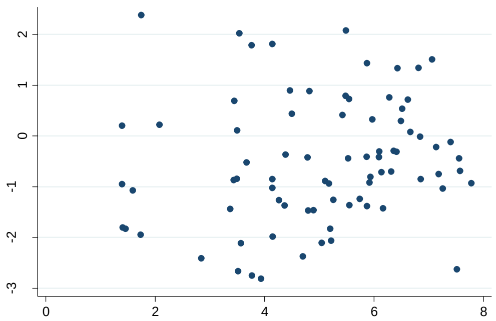


Figure 5: Energy Sector on AveragePatents (Model 5)

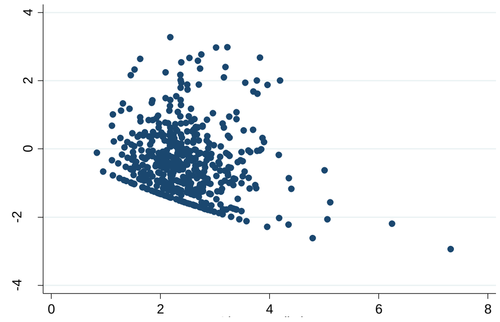


Figure 6: Energy Sector on AverageCitations (Model 6)

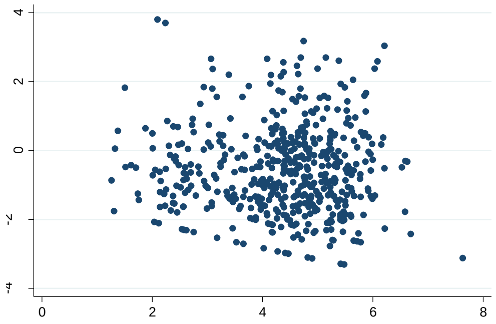
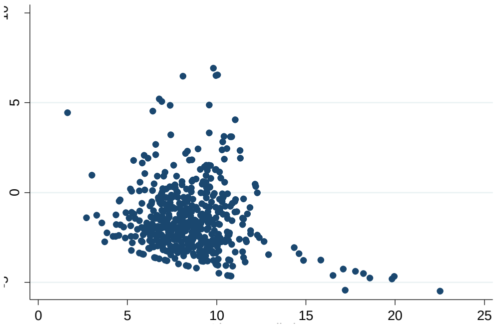


Figure 7: Energy Sector on AveragePatents×Citations (Model 7)



Note: All residual plots depicted on this page illustrate the relationship between linear predictors (X-axis) and deviance residuals (Y-axis). In Figure 1,2,3 and 4, the residual regressions of Models 1a, 2a, 3a and 4a are given. The ‘a-Models’ strongly corresponded with the residual plots of the other scores.