Politecnico di Torino

Corso di Laurea Magistrale
A.a. 2024/2025
Sessione di Laurea Marzo/Aprile 2025

Climate and Sustainability Indexes as Drivers of Investment in Green Energy

An Empirical Analysis of CCPI, ESG, and Their Impact on Venture Capital and Country-Level Investment Trends

Relatrice:
Prof. Elisa Ughetto

Candidata:
Martina Agnese

Alla mia famiglia,

per l’amore incondizionato e il sostegno costante.

Senza di voi, non sarebbe stato lo stesso.

Index

1. Abstract

The transition towards a sustainable economy is increasingly shaped by financial incentives and policy-driven indexes such as the Climate Change Performance Index (CCPI) and the Environmental, Social, and Governance (ESG) Index. This thesis investigates the statistical relationship between these indexes and country-level investment dynamics, with a particular focus on Venture Capital investments in green energy companies.

A range of statistical and econometric approaches will be considered to assess how sustainability indicators may influence investment patterns. The analysis will also explore the role of additional economic variables, including market conditions, policy frameworks, and financial incentives, to determine their potential interaction with CCPI and ESG scores. Furthermore, the study will investigate whether these indexes are associated with differences in the geographical distribution of green energy companies and the clustering of VC-backed firms.

By integrating multiple perspectives, this research aims to provide insights into the potential drivers of sustainable investment and the role of policy and financial markets in supporting the green energy transition.

2. European Cleantech Sector: overview and preliminary analysis

A Cleantech is a company that develops, manufactures, or implements technologies and processes aimed at environmental sustainability. These technologies and processes focus on reducing environmental impact through activities such as:

1. Pollution mitigation and remediation: technologies that address air, water, and soil pollution through cleaning or restoration efforts.
2. Resource conservation and efficiency: technologies and practices that promote the sustainable use of natural resources like water, raw materials, and energy.
3. Renewable energy production: companies involved in the development and deployment of clean energy sources like solar, wind, geothermal, or biomass power.
4. Sustainable waste management: technologies and processes that promote responsible waste disposal and recycling, minimizing environmental harm.

The Cleantech Companies analyzed have been sourced from the European Investment Fund database.

The companies have been identified according to three main criteria, namely:

1. Companies located in Europe.
2. Companies which had recorded accounting data for at least one business year.

3. Companies with an available extended business description.

Furthermore, they have been classified among seven different technological categories reflecting the pillars of the European Green Deal and EU Taxonomy. The categories are:

1. Environmental management
    1.1. Air/water/soil pollution abatement/remediation.
    1.2. Waste management.
2. Resources preservation
    2.1. Water conservation/availability.
    2.2. Sustainable agri-food technologies.
    2.3. Sustainable raw materials.
3. Industrial energy management
    3.1. Sustainable energy production.
    3.2. Sustainable fuels.
    3.3. Energy-efficient industrial technologies.
4. Capture, storage, sequestration, or disposal of GHG.
5. Sustainable modes of transport.
6. Sustainable buildings.
7. Other categories

In this analysis we will focus on the third category: Industrial energy management. To be more precise, the first category, *Sustainable energy production*, includes clean energy generation technologies such as wind, solar thermal, photovoltaic, geothermal, marine, hydroelectric, new nuclear technologies, fuel cells, and co-generation technologies.

Sustainable fuels include fuels from renewable sources that minimize the environmental impact, like fuels deriving from renewable biomass or from waste.

The latter category, *Energy-efficient industrial technologies*, includes battery storage, capacitor, thermal storage, technologies related to superconductors, pressurized fluid, mechanical & pumped, recyclable products, and reduction of materials in manufacturing.

2.1 Distribution of the companies by ecosystem segment and technological category

This classification has been performed taking into consideration the supply chain structure of the Cleantech ecosystem. The companies have been divided into two broad categories, namely:

1. "Cleantech innovators": these companies create the clean technology as their core business and are at the center of the supply chain.
2. "Cleantech ecosystem": these companies adopt clean technologies, sell services based on clean technologies or provide inputs for the development if clean technologies. They are further divided into:
    2.1. Experimenters: companies involved in performing experimental tasks that can lead to discoveries and advances in the science of Cleantech supply chain.
    2.2. Manufacturers: companies involved in the Cleantech supply chain, dealing with ancillary services concerning actual innovation
    2.3. Distributors: companies that only distribute or are involved in the commercial provision of certain Cleantech products or technologies
    2.4. Integrators: companies involved in the Cleantech supply chain, dealing with accessory services concerning actual innovation
    2.5. Operators: companies involved in the Cleantech supply chain that deals with the construction, implementation, and maintenance of facilities where clean technology is used.

We analyzed the ecosystem segment distribution of each company, taking into consideration that one company can belong to different segments, as these data are based on the details provided by the companies in their business description. From the Table 2.1.1 and Figure 2.1.1 we can see that the companies belonging to the ecosystem segment are way more than the ones in the innovators segment. This is likely because the ecosystem sector is much broader than the innovators one, as it comprises several sub-categories.

Table 2.1.1: Classification of Cleantech companies into ecosystem segment

Segmentation	# companies	%
Innovators	1 911	17.0%
Ecosystem	9 310	83.0%
Experimenters	29	0.3%
Manufacturers	2 253	20.1%
Distributors	1 403	12.5%
Integrators	2 747	24.5%
Operators	2 875	25.6%
Total	11 221	100%

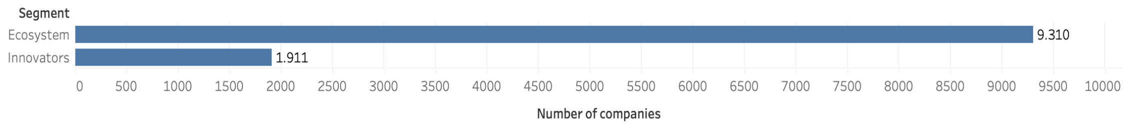

Fig. 2.1.1: companies' distribution according to the ecosystem segment

This assumption concerning the size of the segment is confirmed by the fact that, if we differentiate the company according to their segmentation while expanding the ecosystem segment into its sub-categories, the firms' distribution is much more uniform, as showed in Figure 2.1.2.

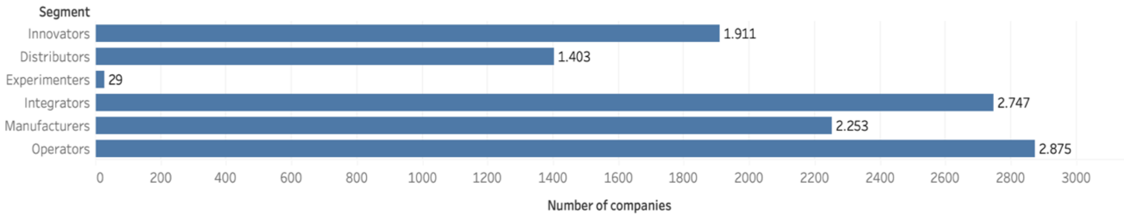

Figure 2.1.2: companies' distribution according to the expanded ecosystem segment

Then we derived the distribution of the Cleantech companies operating in the Energy sector among the technological categories that we identified before, namely *Sustainable energy producers* (3.1), *Sustainable fuel producers* (3.2), and *Energy-efficient industrial technologies.*

The results, reported in Table 2.1.2, highlights that most of the companies categorized as Industrial energy management belongs either to the Sustainable energy production (3.1) category or to the Energy-efficient industrial technologies (3.3) one.

Table 2.1.2: Classification of Cleantech companies into technological category

Technological category	# companies	%
Sustainable energy production (3.1)	6 421	57.2%
Sustainable fuels (3.2)	1 547	13.8%
Energy-efficient industrial technologies (3.3)	4 497	40.1%
Total	11 221	

We can see that the Sustainable energy production and the Energy-efficient industrial technologies categories are the most represented, with respectively the 57.2% and the 40.1%.

Then we derived the distribution of Cleantech companies by technological category considering only those companies classified as innovators. The results are plotted in Table 2.1.3.

Table 2.1.3: Classification of Cleantech companies into technological category considering only those classified as innovators.

Technological category	# innovators	%
Sustainable energy production (3.1)	1 419	74.3%
Sustainable fuels (3.2)	158	8.3%
Energy-efficient industrial technologies (3.3)	451	23.6%
#Innovators in the sample	1 911	

We can see that Sustainable energy production is the category with the largest number of companies overall and with the largest number of companies belonging to the Innovators sector. The results of this cross analysis are reported in figure 2.1.3.

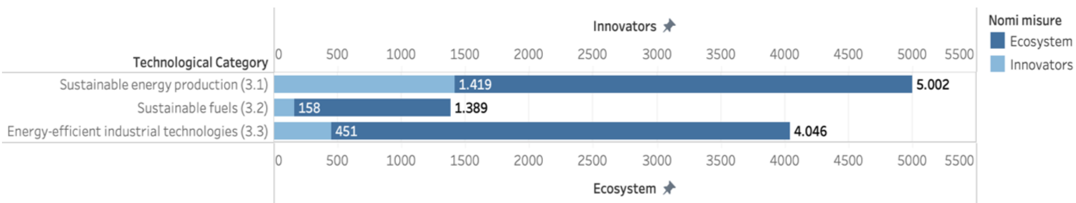

Figure 2.1.3: Classification of companies into segment by technological category

Considering the firms classified also in other technological categories, plotted in Table 2.1.4, we can see that most firms are operating in the waste-management sector (50.6%), followed by Air/water/soil pollution companies (21.4%), and Sustainable buildings ones (15.4%).

Table 2.1.4: Classification of Cleantech companies into technological categories different from Industrial energy management companies

Technological category	#Companies	%
Air/water/soil pollution (1.1)	874	21.4%
Waste management (1.2)	2 068	50.6%
Water conservation / availability (2.1)	217	5.3%
Agri-food (2.2)	63	1.5%
Sustainable raw materials (2.3)	88	2.2%
Capture, storage, sequestration, or disposal of GHG (4)	14	0.3%
Sustainable modes of transport (5)	132	3.2%
Sustainable buildings (6)	627	15.4%
Other categories (7)	0	0.0%
Total	4 083	100%

2.2 Classification of Cleantech companies by geography

Classifying the companies by geography is important as it helps in identifying clusters and understand differences and possible advantages of the various geographic regions.

The first analysis performed concerns the number of Cleantech companies in each European country. The results are showed in Table 2.2.1 and in Figure 2.2.1.

As shown in the table, we can identify 3 main areas of concentrations of Cleantech companies, namely: Germany (19.71%), Italy (16,95%), and France (13,56%), that alone count for more than half of the total number of companies.

This can be due to several factors, which will be analyzed later. Among them we can probably identify some economic factors, political, and institutional differences, such as the ease to obtain funds.

The same observations can be made for the distribution of the cleantech innovators and cleantech ecosystem. Their distribution is showed in Figure 2.2.2 and Figure 2.2.3.

Table 2.2.1: distribution of Cleantech companies by country

	Cleantech companies		Cleantech innovators		Cleantech ecosystem	
Country	#Companies	%	#Companies	%	#Companies	%
Albania	1	0.0%	1	0.1%	0	0.0%
Austria	299	2.7%	54	2.8%	245	2.6%
Belgium	298	2.7%	54	2.8%	244	2.6%
Bosnia and Herzegovina	2	0.0%	0	0.0%	2	0.0%

Bulgaria	151	1.4%	21	1.1%	130	1.4%
Croatia	58	0.5%	14	0.7%	44	0.5%
Cyprus	2	0.0%	1	0.1%	1	0.0%
Czech Republic	292	2.6%	50	2.6%	242	2.6%
Denmark	165	1.5%	38	2.0%	127	1.4%
Estonia	33	0.3%	9	0.5%	24	0.3%
Finland	248	2.2%	45	2.4%	203	2.2%
France	1522	13.6%	209	10.9%	1313	14.1%
Germany	2212	19.7%	355	18.6%	1857	20.0%
Greece	126	1.1%	29	1.5%	97	1.0%
Hungary	185	1.7%	20	1.1%	165	1.8%
Iceland	14	0.1%	1	0.1%	13	0.1%
Ireland	12	0.1%	0	0.0%	12	0.1%
Italy	1902	17.0%	316	16.5%	1586	17.0%
Latvia	40	0.4%	3	0.2%	37	0.4%
Lithuania	53	0.5%	9	0.5%	44	0.5%
Luxembourg	41	0.4%	6	0.3%	35	0.4%
Malta	8	0.1%	2	0.1%	6	0.1%
Montenegro	3	0.0%	0	0.0%	3	0.0%
Netherlands	189	1.7%	39	2.0%	150	1.6%
North Macedonia	16	0.1%	1	0.1%	15	0.2%
Norway	337	3.0%	47	2.5%	290	3.1%
Poland	562	5.0%	92	4.8%	470	5.1%
Portugal	222	2.0%	29	1.5%	193	2.1%
Romania	190	1.7%	25	1.3%	165	1.8%
Serbia	77	0.7%	7	0.4%	70	0.8%
Slovakia	103	0.9%	15	0.8%	88	1.0%
Slovenia	60	0.5%	17	0.9%	43	0.5%
Spain	1135	10.1%	252	13.2%	883	9.5%
Sweden	405	3.6%	88	4.6%	317	3.4%
Switzerland	29	0.3%	3	0.2%	26	0.3%
Turkey	9	0.1%	3	0.2%	6	0.1%
United Kingdom	220	2.0%	56	2.9%	164	1.8%
Total	11221	100%	1911	100%	9310	100%

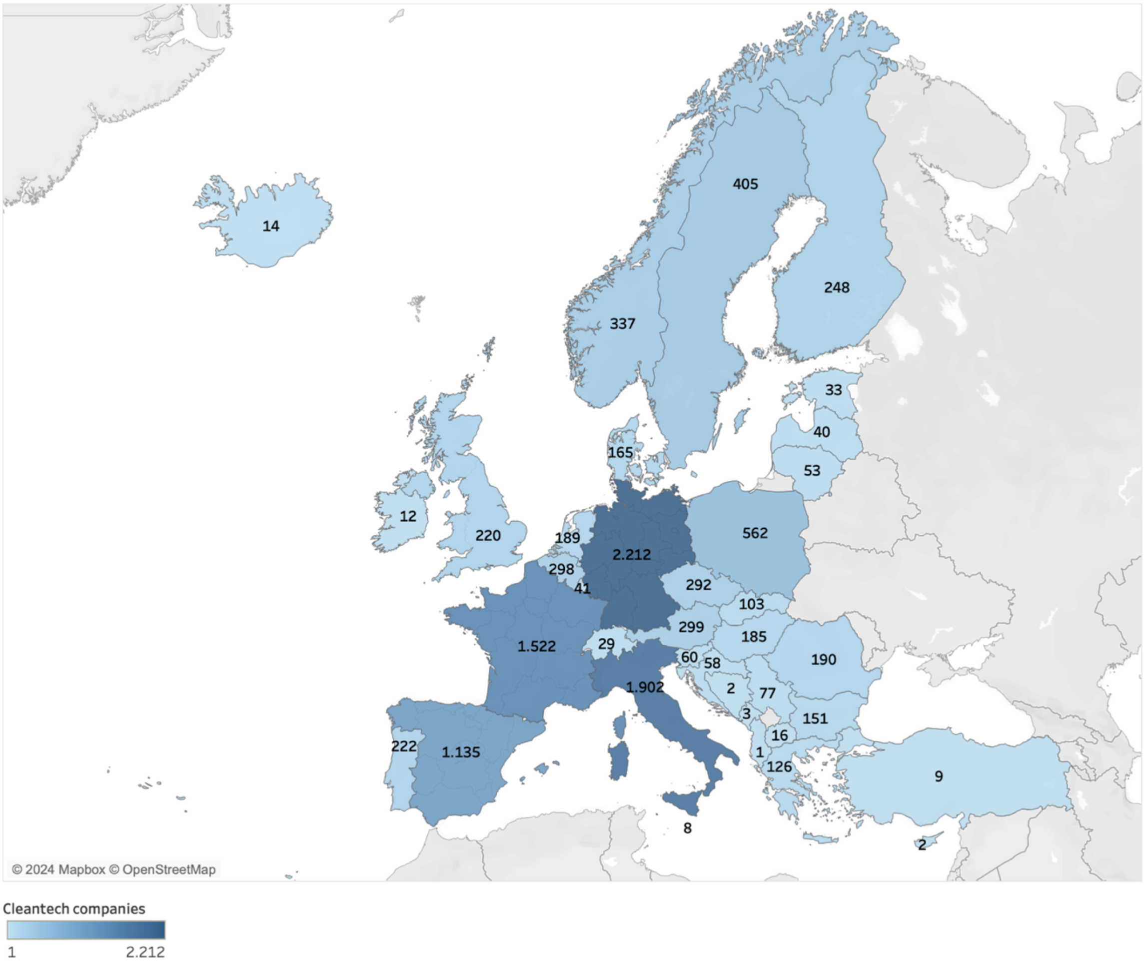

Figure 2.2.1: Geographical distribution of Cleantech companies

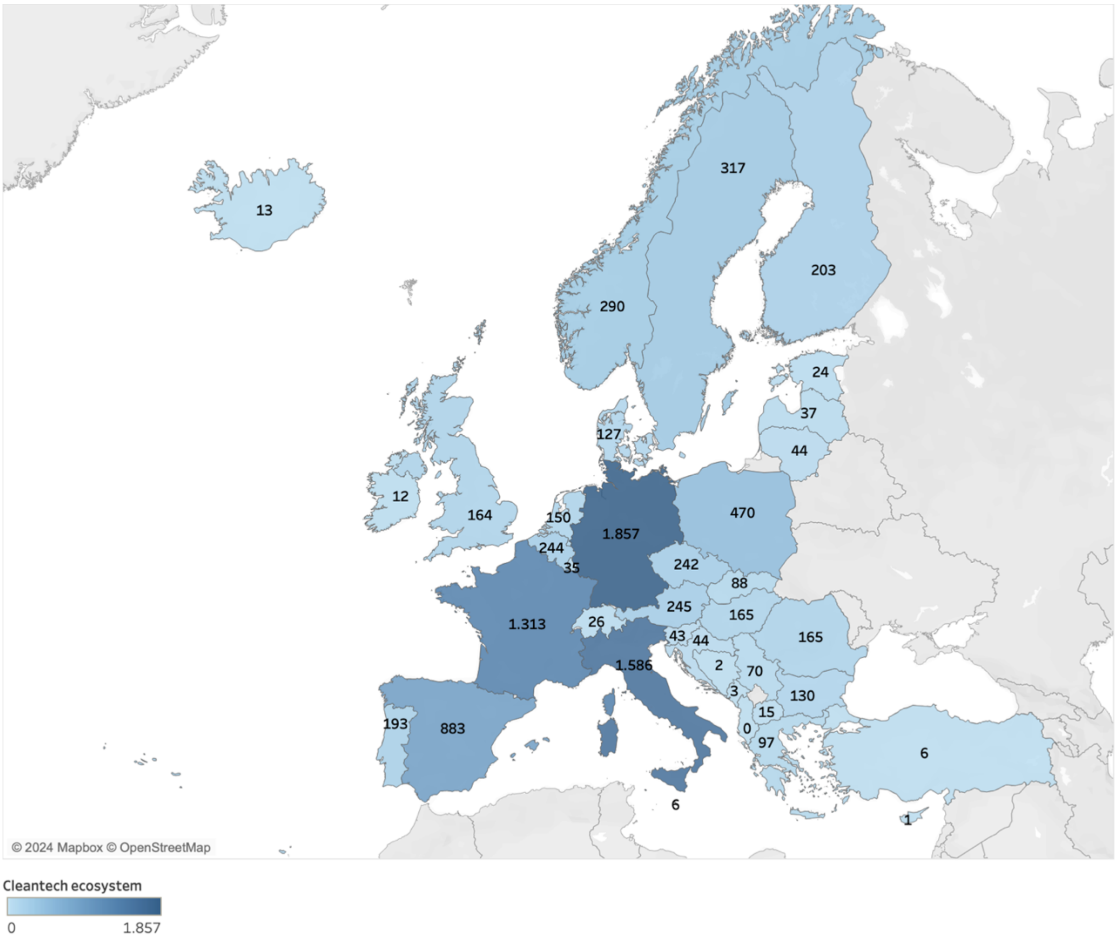

Figure 2.2.2: Geographical distribution of Cleantech ecosystem

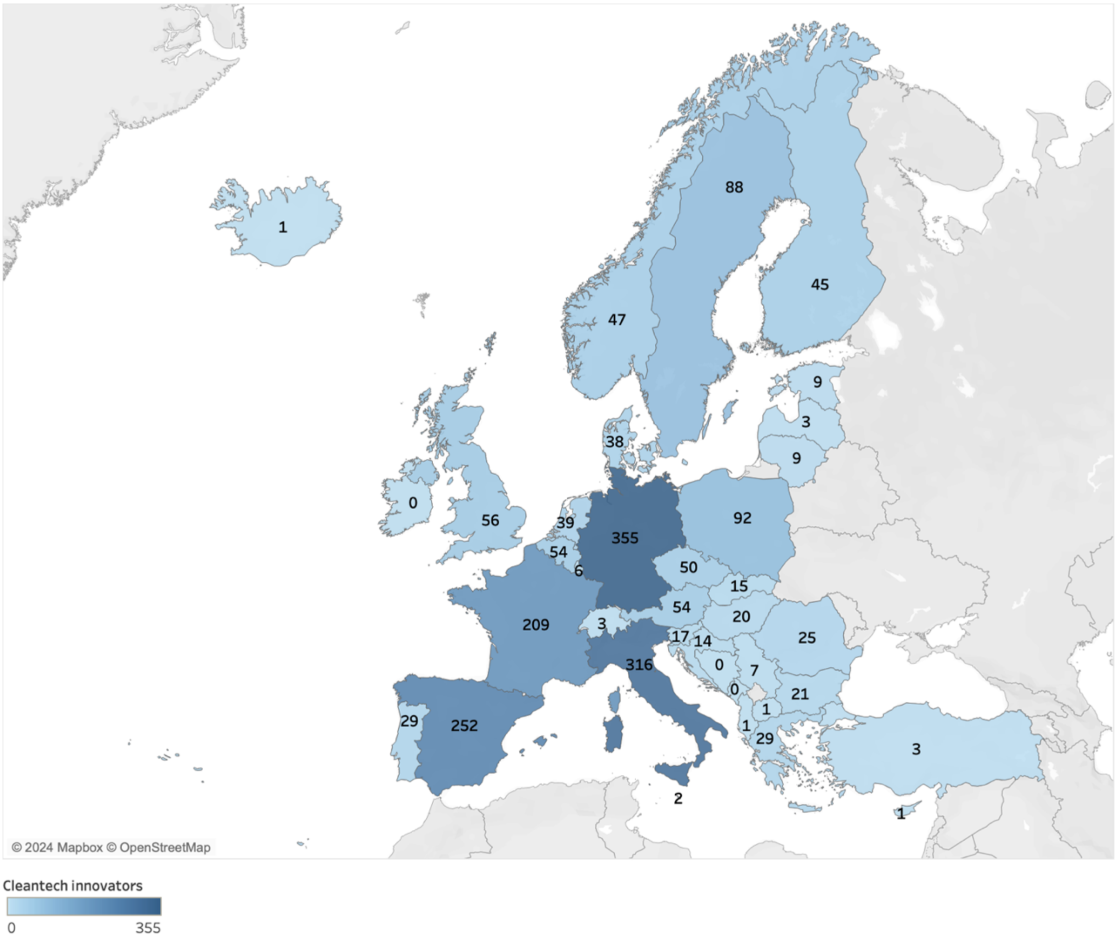

Figure 2.2.3: Geographical distribution of Cleantech innovators

Then, the companies have been divided by geography taking into consideration their sub-category. The results have been plotted in the following figures, which shows respectively the geographic distribution of Cleantech operators (Figure 2.2.4), Cleantech researchers (Figure 2.2.5), Cleantech distributors (Figure 2.2.6), Cleantech integrators (Figure 2.2.7), and Cleantech manufacturers (Figure 2.2.8).

We can observe that there are not significant variations of the companies' distribution when considering their ecosystem sub-segment.

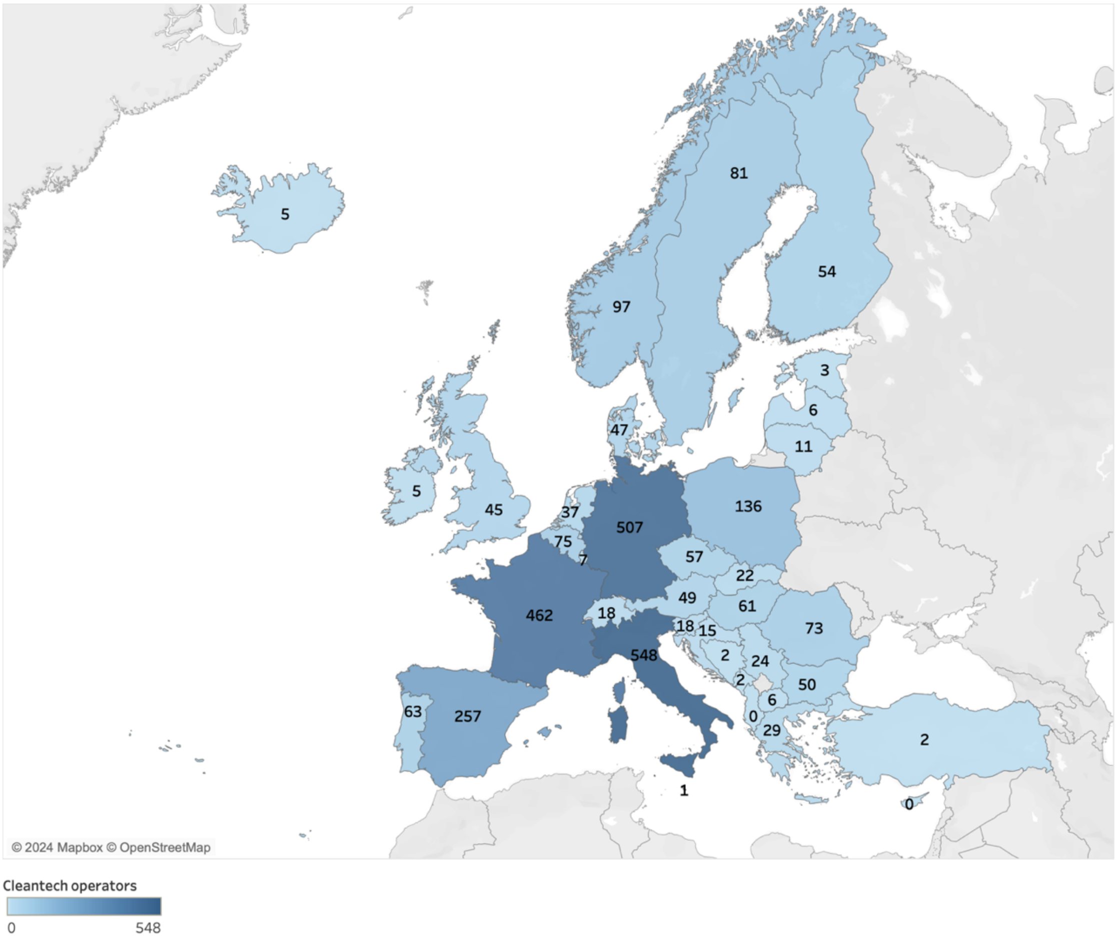

Figure 2.2.4: Geographical distribution of Cleantech operators

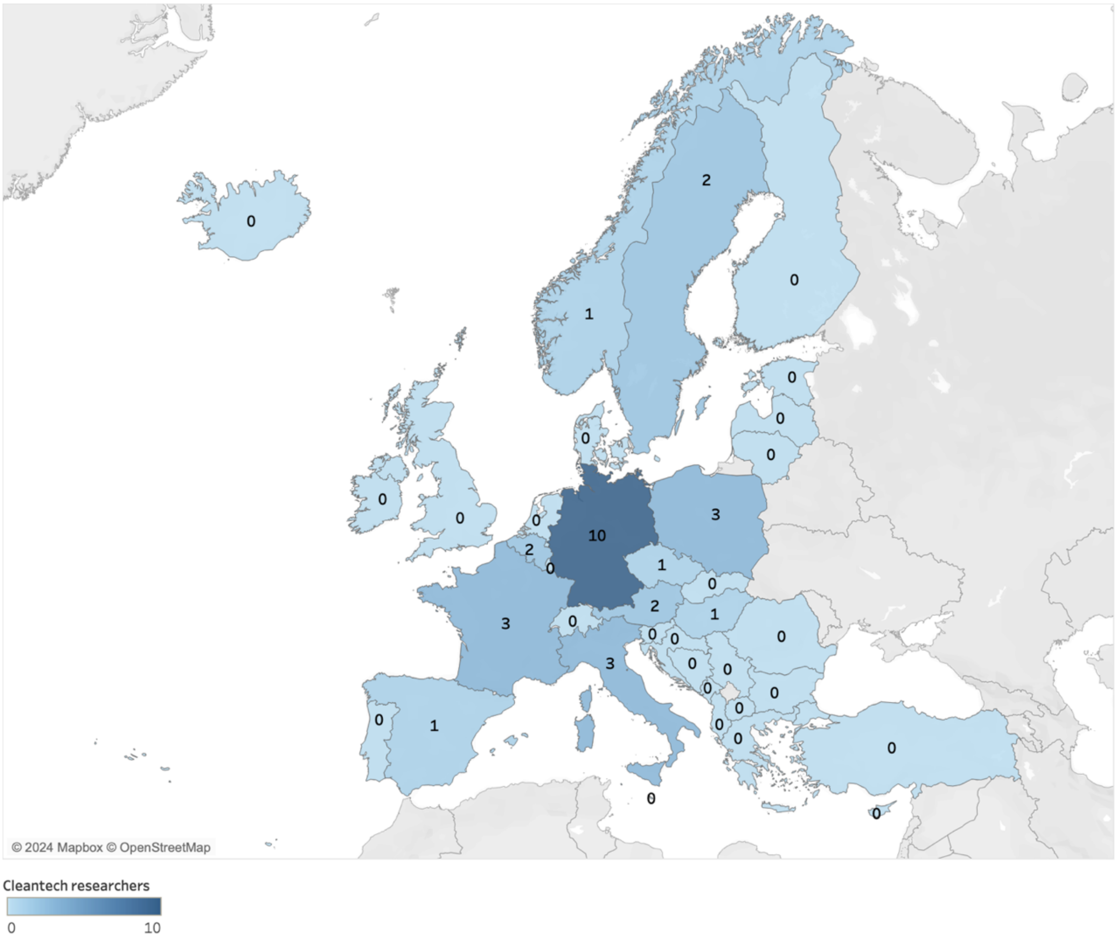

Figure 2.2.5: Geographical distribution of Cleantech researchers

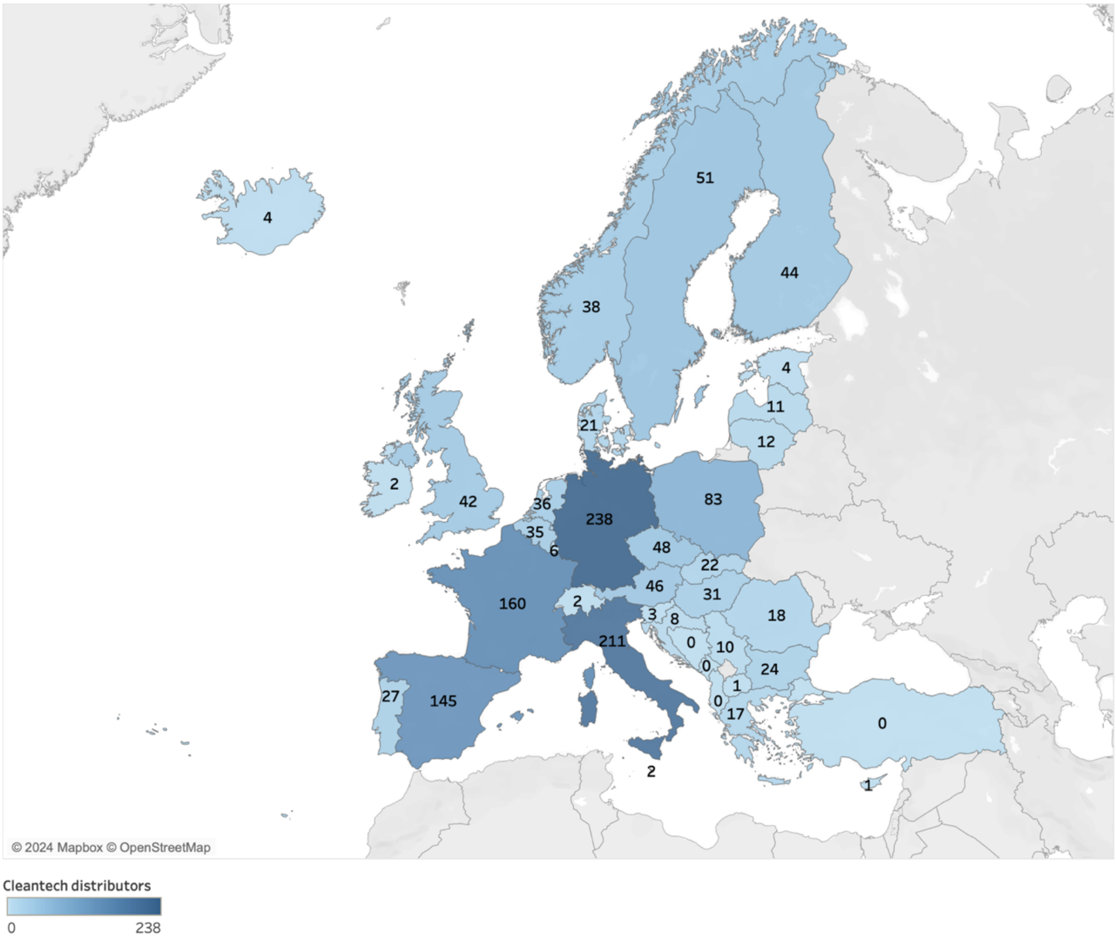

Figure 2.2.6: Geographical distribution of Cleantech distributors

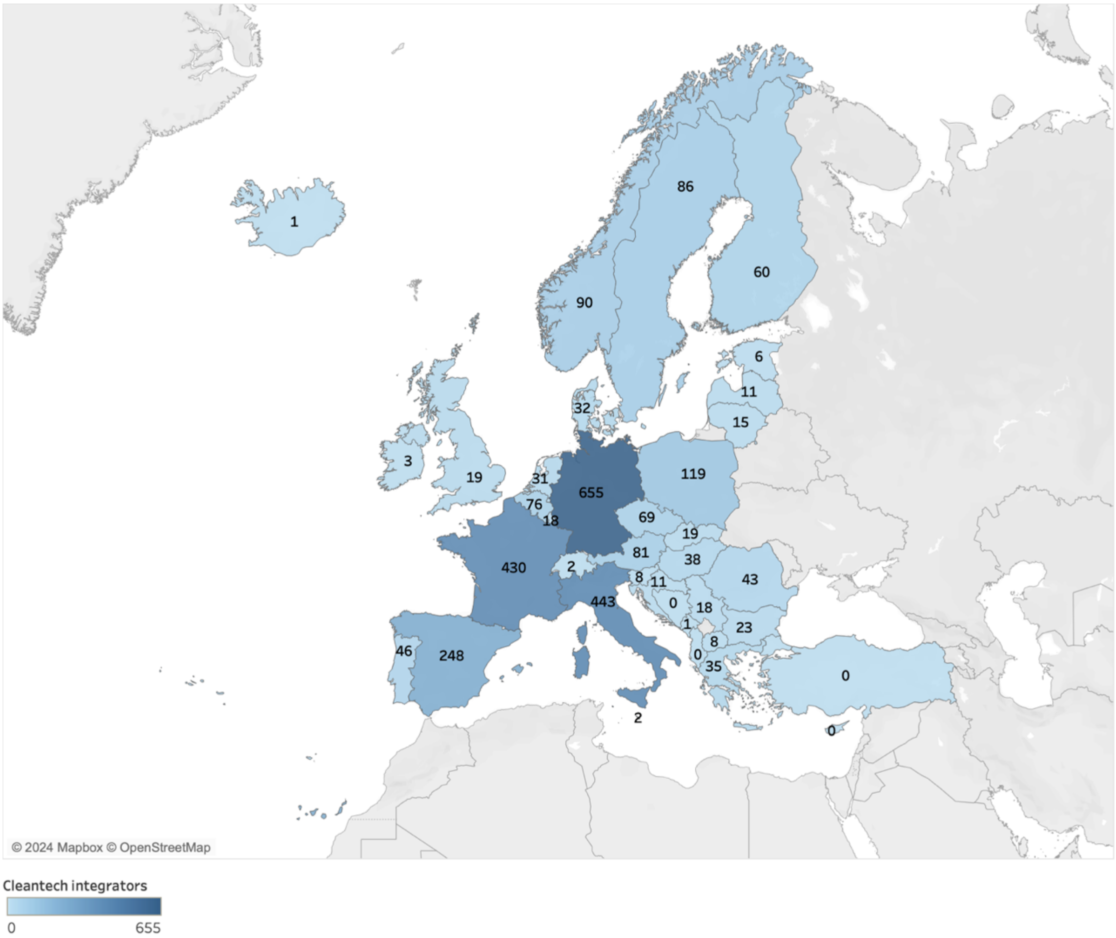

Figure 2.2.7: Geographical distribution of Cleantech integrators

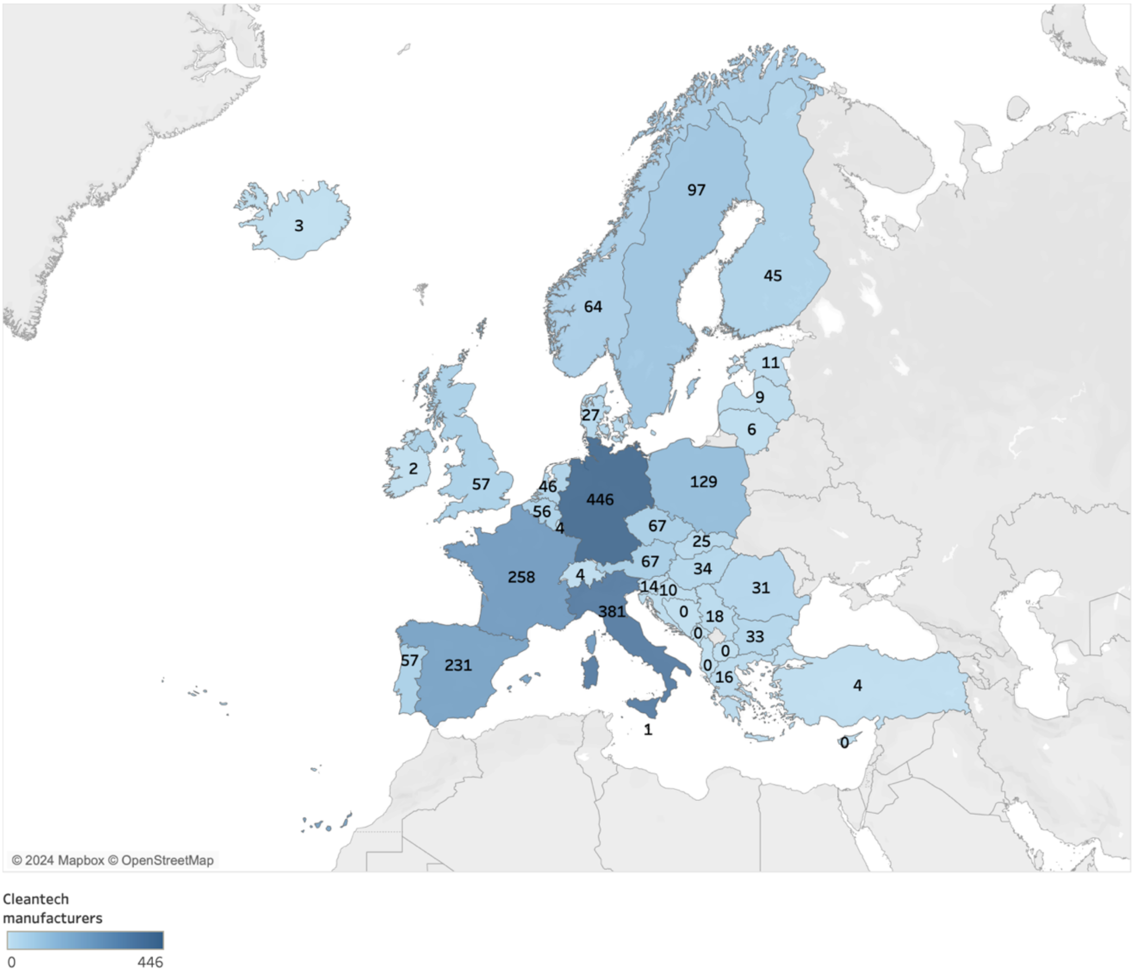

Figure 2.2.8: Geographical distribution of Cleantech manufacturers

Finally, the companies have been divided taking into consideration their technological category, namely Sustainable energy producers (3.1), Sustainable fuel producers (3.2), and Energy-efficient industrial technologies (3.3). The results are showed in Figure 2.2.9, Figure 2.2.10, and Figure 2.2.11 respectively.

For this analysis we can observe that there are no significant variations in the pattern distribution, as the countries with the largest amount of company concentration are Germany, Italy and French, with some cases with a high concentration of firms also in Spain.

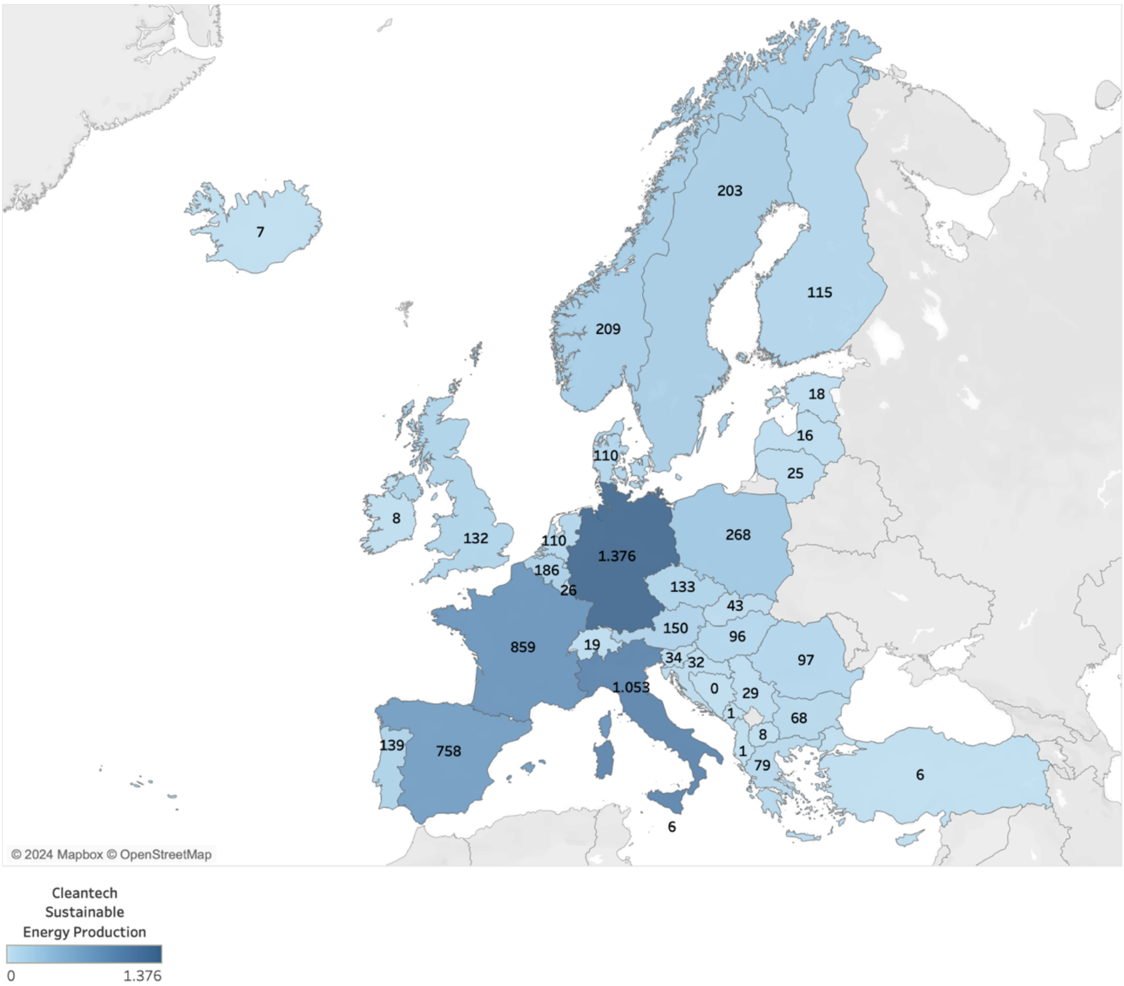

Figure 2.2.9: Geographical distribution of Cleantech Sustainable energy producers

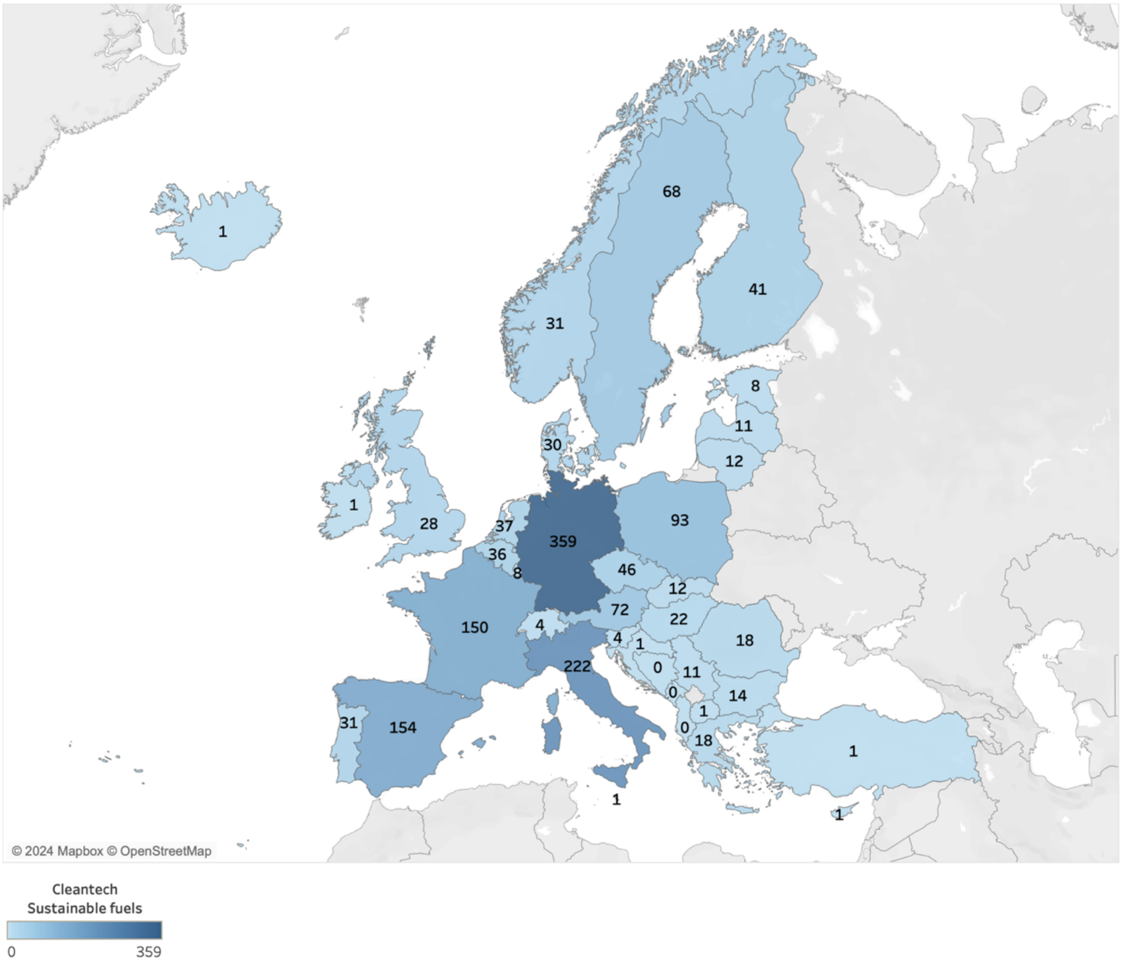

Figure 2.2.10: Geographical distribution of Cleantech sustainable fuels

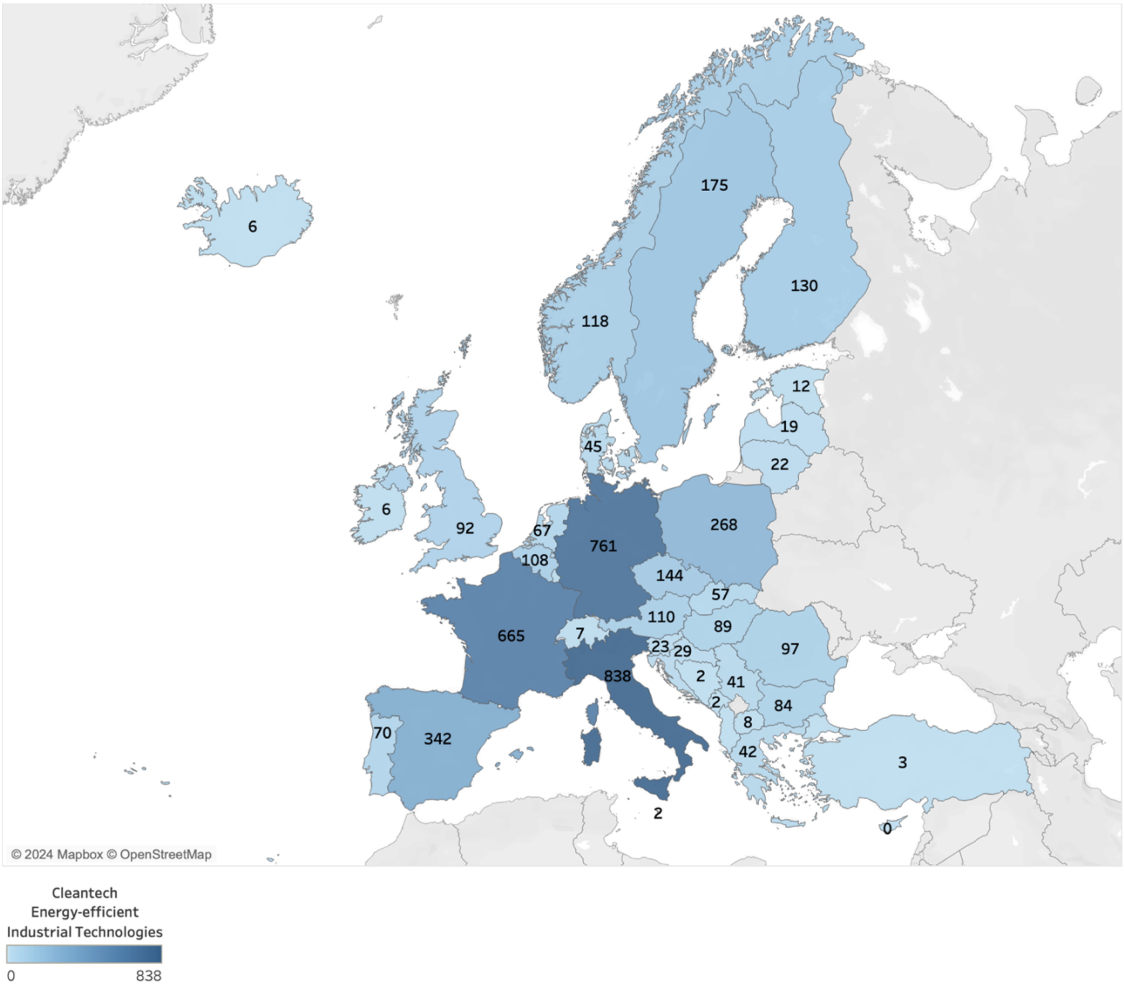

Figure 2.2.11: Geographical distribution of Cleantech Energy-efficient industrial technologies

2.3 Descriptive statistics on Cleantech companies by sector

In this section we analyze the cross-sectoral nature of the companies.

When considering the entire database, we can see that most of the companies operating in the Green Energy sector operates also in the Manufacturing one (22.8%), followed by Wholesale and retail trade (17.3%), Electricity, gas, steam, and air conditioning supply (14.7%), Water supply (14.1%), and Construction (12.9%).

Considering only those companies classified as Cleantech Innovators we can see that the portion of companies operating in the manufacturing sector increases up to 38.6%, followed by the Electricity, gas, steam, and air conditioning (17.0%), and Professional, scientific, and technical activities (13.7%). On the other hand, we can observe a decrease in the concentration in sectors such as Water supply, that accounts for 1.4%. This highlights that Cleantech innovation occurs predominantly in hardware-driven sectors.

Focusing on Cleantech ecosystem we can see a smoother concentration, where the Manufacturing sector, although remaining the densest, accounts for the 19.6% of the total firms, closely followed by Wholesale and retail (18.8%), Water supply (16.7%), Electricity, gas, steam, and air conditioning (14.2%), and Construction (13.7%) sectors.

The results are plotted in Table 2.3.1.

Table 2.3.1: Distribution of Cleantech companies operating in green energy sector by industry (NACE rev.2)

NACE rev.2 section	Cleantech companies		Cleantech innovators		Cleantech ecosystem	
	#Companies	%	#Companies	%	#Companies	%
A - Agriculture, forestry, and fishing	65	0.6%	9	0.5%	56	0.6%
B - Mining and quarrying	73	0.7%	7	0.4%	66	0.7%
C - Manufacturing	2 555	22.8%	735	38.6%	1 820	19.6%
D - Electricity, gas, steam, and air conditioning supply	1 643	14.7%	324	17.0%	1 319	14.2%
E - Water supply; sewerage, waste management and remediation activities	1 576	14.1%	26	1.4%	1 550	16.7%
F - Construction	1 449	12.9%	178	9.3%	1 271	13.7%
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	1 941	17.3%	191	10.0%	1 750	18.8%
H - Transportation and storage	143	1.3%	10	0.5%	133	1.4%
I - Accommodation and food service activities	30	0.3%	5	0.3%	25	0.3%
J - Information and communication	101	0.9%	21	1.1%	80	0.9%
K - Financial and insurance activities	289	2.6%	77	4.0%	212	2.3%
L - Real estate activities	135	1.2%	18	0.9%	117	1.3%

M - Professional, scientific, and technical activities	850	7.6%	261	13.7%	589	6.3%
N - Administrative and support service activities	256	2.3%	32	1.7%	224	2.4%
O - Public administration and defence; compulsory social security	7	0.1%	1	0.1%	6	0.1%
P - Education	8	0.1%	0	0.0%	8	0.1%
Q - Human health and social work activities	25	0.2%	3	0.2%	22	0.2%
R - Arts, entertainment, and recreation	11	0.1%	0	0.0%	11	0.1%
S - Other service activities	46	0.4%	8	0.4%	38	0.4%
Total	11 203	100%	1 906	100%	9 297	100%

Then, we analyzed the distribution of the Cleantech companies operating in the green energy sector focusing on its sub-categories, namely Sustainable energy production (3.1), Sustainable fuels (3.2), and Energy-efficient industrial technologies (3.3).

Focusing on the first sub-category, Sustainable energy production (3.1), we can see that the densest sector is Electricity, gas, steam, and air conditioning (21.7%), closely followed by Manufacturing (21.1%). Other relevant sectors are Construction (14.5%), Wholesales and repair (12.3%), and Water supply (10.0%).

Considering the firms operating in the Sustainable fuels (3.2) sub-category, we can see that the densest sector is again the Manufacturing one, accounting for the 26.9% of the total sample. Other high-density sectors are Electricity, gas, steam, and air conditioning (20.2%), Wholesale and retail (18.0%), and Construction (11.4%).

Focusing on the last sub-category, Energy-efficient industrial technologies (3.3), we can see that the densest sector is Water supply (23.6%), followed by Wholesales and repair (23.2%), and Manufacturing (22.4%).

The results are plotted in Table 2.3.2.

Overall, we can see that there is a quite uniform pattern when we consider the whole sample, only the companies classified as Cleantech ecosystem, and the whole sample divided by green energy sub-categories.

Major differences results, on the other hand, when we consider only the companies classified as Cleantech innovators, with a shift of concentration from an only-hardware-driven industry to a more heterogeneous one, where a significant portion is covered by Professional, scientific, and technical activities, which have a very small presence in all the other analyses.

Table 2.3.2: Distribution of Cleantech companies by industry (NACE rev.2) and green energy sub-categories

NACE rev.2 section	Sustainable energy production (3.1)		Sustainable fuels (3.2)		Energy-efficient industrial technologies (3.3)	
	#Companies	%	#Companies	%	#Companies	%
A - Agriculture, forestry, and fishing	27	0.4%	40	2.6%	12	0.3%
B - Mining and quarrying	46	0.7%	14	0.9%	17	0.4%
C - Manufacturing	1351	21.1%	416	26.9%	1004	22.4%
D - Electricity, gas, steam, and air conditioning supply	1388	21.7%	312	20.2%	186	4.1%
E - Water supply; sewerage, waste management and remediation activities	638	10.0%	76	4.9%	1060	23.6%
F - Construction	928	14.5%	176	11.4%	530	11.8%
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	787	12.3%	279	18.0%	1042	23.2%
H - Transportation and storage	84	1.3%	18	1.2%	52	1.2%
I - Accommodation and food service activities	24	0.4%	2	0.1%	8	0.2%
J - Information and communication	53	0.8%	7	0.5%	53	1.2%
K - Financial and insurance activities	194	3.0%	46	3.0%	82	1.8%
L - Real estate activities	73	1.1%	12	0.8%	67	1.5%

M - Professional, scientific, and technical activities	594	9.3%	107	6.9%	247	5.5%
N - Administrative and support service activities	163	2.5%	29	1.9%	93	2.1%
O - Public administration and defence; compulsory social security	4	0.1%	2	0.1%	1	0.0%
P - Education	4	0.1%	1	0.1%	4	0.1%
Q - Human health and social work activities	13	0.2%	3	0.2%	13	0.3%
R - Arts, entertainment, and recreation	7	0.1%	0	0.0%	5	0.1%
S - Other service activities	31	0.5%	7	0.5%	15	0.3%
Total	6409	100%	1547	100%	4491	100%

2.4 Descriptive statistics on Cleantech companies by year of incorporation

This analysis has the aim to show that, while the interest in Cleantech companies skyrocketed in recent years, this phenomenon is all but new. The results are showed in Table 2.4.1.

Considering the entire sample, we can see that the distribution of Cleantech companies operating in the green energy sector has followed a quite uniform distribution in the sets of years from 1991-1995 to 2006-2010, with a slight slowdown from 2011 onwards.

Focusing on the companies classified as Cleantech innovators we can see a similar pattern although much more contained, except for the period 2006-2010, which saw a surge in the creation of companies classified as Cleantech innovators.

Considering only the companies classified as Cleantech ecosystem, on the other hand, we can see almost the same pattern as when considering the whole sample.

The downturn occurred after 2010 is potentially explained by two main reasons: the Great Recession occurred after 2008, and the global economic downturn called "Cleantech crash".

Table 2.4.1: Distribution of Cleantech companies operating in green sector by year of incorporation

Years of Incorporation	Cleantech companies		Cleantech innovators		Cleantech ecosystem	
	#Companies	%	#Companies	%	#Companies	%
Before 1980	1686	15.0%	255	13.4%	1431	15.4%
1981-1985	545	4.9%	72	3.8%	473	5.1%
1986-1990	903	8.1%	119	6.2%	784	8.4%
1991-1995	1459	13.0%	173	9.1%	1286	13.8%
1996-2000	1477	13.2%	228	12.0%	1249	13.4%
2001-2005	1578	14.1%	275	14.4%	1303	14.0%
2006-2010	1876	16.7%	449	23.6 %	1427	15.3%
2011-2015	867	7.7%	182	9.5%	685	7.4%
2016 onwards	812	7.2%	153	8.0%	659	7.1%
Total	11203	100%	1906	100%	9297	100%

By plotting the evolution of Cleantech firms operating in the green energy sub-categories in Figure 2.4.1, we can observe an overall similar pattern with the previous analysis performed on the whole sample.

Both the Sustainable energy production (3.1) and Sustainable fuels (3.2) companies experienced a quite steady growth up to 2010, followed by a downturn.

Slightly different the pattern of Energy-efficient industrial technologies sector (3.3), which after a stable growth up to years 1991-1995, showed a continuous decrease.

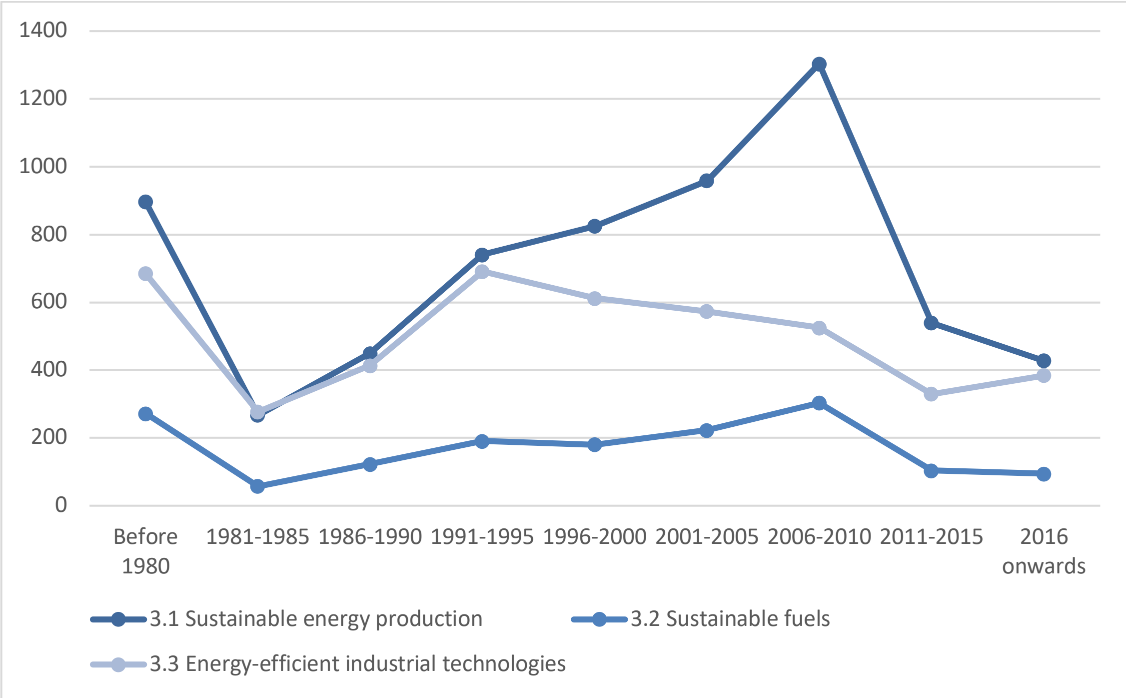

Figure 2.4.1: Evolution of Cleantech Green energy companies by sub-category (1981-onwards)

2.5 Descriptive statistics on Cleantech companies by patent data

Patents are extremely useful when analyzing the potential of Cleantech companies in bringing innovations, because they provide important information about their competitive advantage and technological advancement.

We started our analysis focusing on the whole sample of EPO patenting Cleantech companies operating in the green energy sector, then we derived the results into the two segments of Cleantech innovators and Cleantech ecosystem. The results are reported in Table 2.5.1.

We obtained that, among the whole sample of Cleantech companies, the 11.5% has a patent in a CCMT-related field (Climate Change Mitigation Technologies). Of these firms, the 22.8% of all is classified as Cleantech innovators and the 9.2% as Cleantech ecosystem. All these data are referred to the whole sample of Cleantech companies operating in the green energy category.

Table 2.5.1: EPO patenting activity of Cleantech companies in green energy category

	Green energy companies		Cleantech Innovators		Cleantech Ecosystems	
	# companies	%	# companies	%	# companies	%
At least one in any field	1 219	11.5%	407	22.8%	812	9.2%

Further analyzing the distribution of EPO patenting green energy companies by technological category among those classified as

Cleantech innovators, we can see that the category with the largest share of patenting companies is Capture, storage, sequestration, or disposal of GHG (4), followed by Sustainable raw materials (2.3).
Overall, the companies classified as Cleantech innovators with at least one EPO patent are the 22.8% of the whole green energy companies classified as innovators.

The results are showed in Table 2.5.2.

Table 2.5.2: Distribution of EPO patenting Green Energy companies by technological category in the segment of Cleantech innovators

	At least one in any field	
	# companies	%
Air/water/soil pollution abatement/remediation (1.1)	24	31.6%
Waste management (1.2)	11	33.3%
Water conservation/availability (2.1)	5	33.3%
Sustainable agri-food technologies (2.2)	1	7.7%
Sustainable raw materials (2.3)	4	50.0%
Sustainable energy production (3.1)	264	19.7%
Sustainable fuels (3.2)	26	19.4%
Energy-efficient industrial technologies (3.3)	139	34.2%
Capture, storage, sequestration, or disposal of GHG (4)	1	100.0%

Sustainable modes of transport (5)	11	27.5%
Sustainable buildings (6)	44	25.0%
Total	407	22.8%

We then applied the same analysis to the sample of Cleantech ecosystem.

The category with the largest share of patenting companies is Sustainable modes of transport (5), followed by Water conservation/availability (2.1).

Overall, the companies classified as Cleantech ecosystem with at least one EPO patent are the 9.2% of the whole green energy companies classified as innovators.

The results are showed in Table 2.5.3.

Table 2.5.3: Distribution of EPO patenting Green Energy companies by technological category in the segment of Cleantech ecosystem

	At least one in any field	
	# companies	%
Air/water/soil pollution abatement/remediation (1.1)	102	13.1%
Waste management (1.2)	85	4.4%
Water conservation/availability (2.1)	30	15.1%
Sustainable agri-food technologies (2.2)	6	13.0%
Sustainable raw materials (2.3)	11	14.7%

Sustainable energy production (3.1)	435	9.2%
Sustainable fuels (3.2)	150	11.2%
Energy-efficient industrial technologies (3.3)	312	8.2%
Capture, storage, sequestration, or disposal of GHG (4)	1	12.5%
Sustainable modes of transport (5)	15	22.4%
Sustainable buildings (6)	29	6.9%
Total	812	9.2%

The ability to obtain a patent depends also on the country. Analyzing the results obtained in table 2.5.4, we can see that Sweden is the country with the most companies with at least one EPO patent in any field (22.3%), closely followed by Austria (20.4%).

Considering only the green energy companies classified as Cleantech innovators, Sweden and Austria remains the countries with the largest portion of patents, with respectively the 40.0% and the 38.0%.

Focusing on the companies in the Cleantech ecosystem segment, the results are the same, with Sweden as the country most engaged in patent innovation (17.5%), followed by Austria (16.6%).

Table 2.5.4: Distribution of EPO patenting Green Energy companies by country

	At least one in any field					
	Cleantech companies		Cleantech innovators		Cleantech ecosystem	
	# companies	%	# companies	%	# companies	%
Germany	371	17.4%	116	34.6%	255	14.2%
Italy	225	12.0%	66	21.2%	159	10.2%
France	116	8.0%	31	15.3%	85	6.8%
Spain	92	8.7%	34	14.1%	58	7.1%
Poland	20	3.6%	7	7.8%	13	2.8%
Sweden	88	22.3%	34	40.0%	54	17.5%
Czech Republic	13	4.6%	4	8.3%	9	3.9%
Belgium	32	11.4%	12	24.0%	20	8.7%
Norway	29	9.3%	10	24.4%	19	7.0%
Austria	58	20.4%	19	38.0%	39	16.6%
Others	175	8.9%	74	22.3%	101	6.2%
Total	1 219	11.5%	407	22.8%	812	9.2%

The results of this analysis are plotted in Figure 2.5.1, Figure 2.5.2, and Figure 2.5.3, where we can see that innovative activities tend to reinforce each other spatially, due to an easier access to skilled labor and knowledge spillovers. Moreover, the share of patenting tends to be higher in areas with more Cleantech activity.

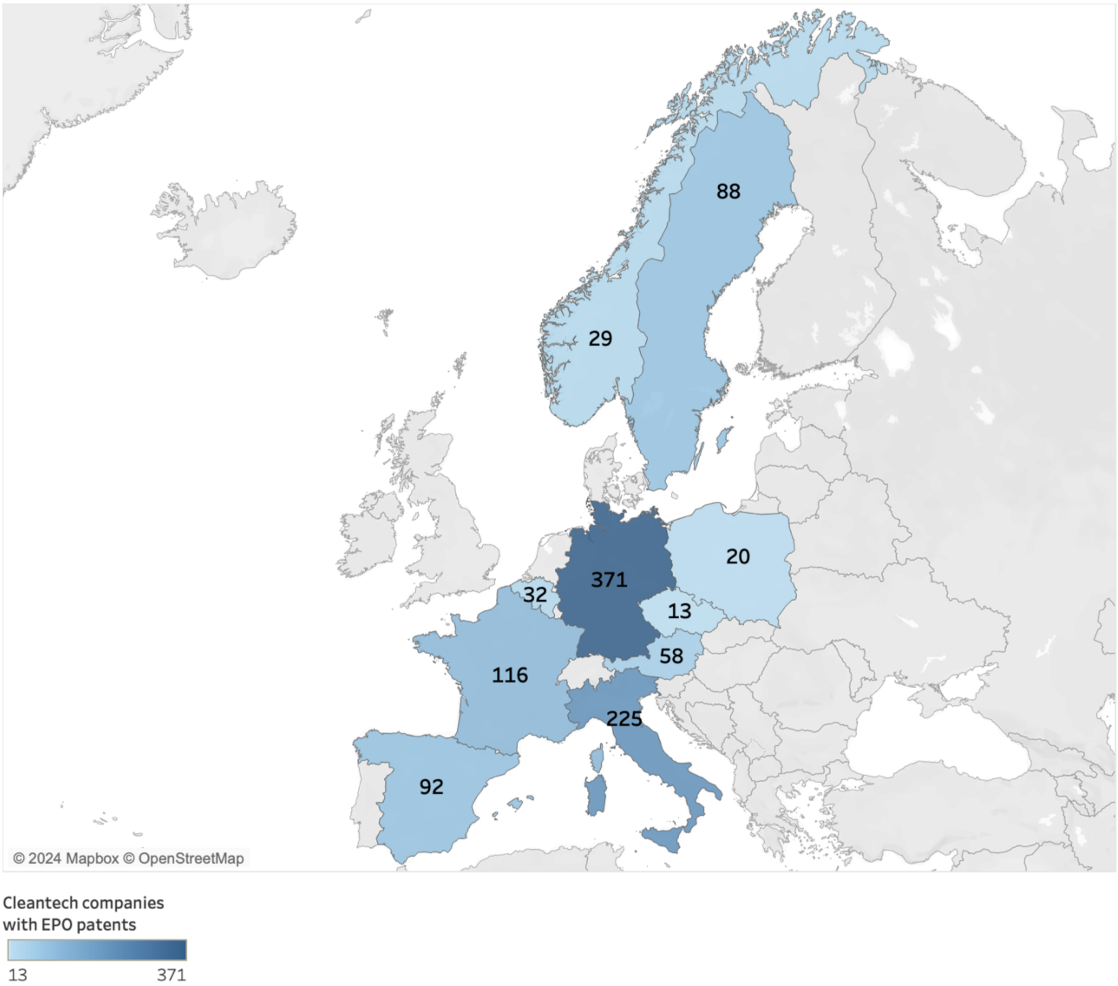

Figure 2.5.1: The share of patenting Cleantech companies by country

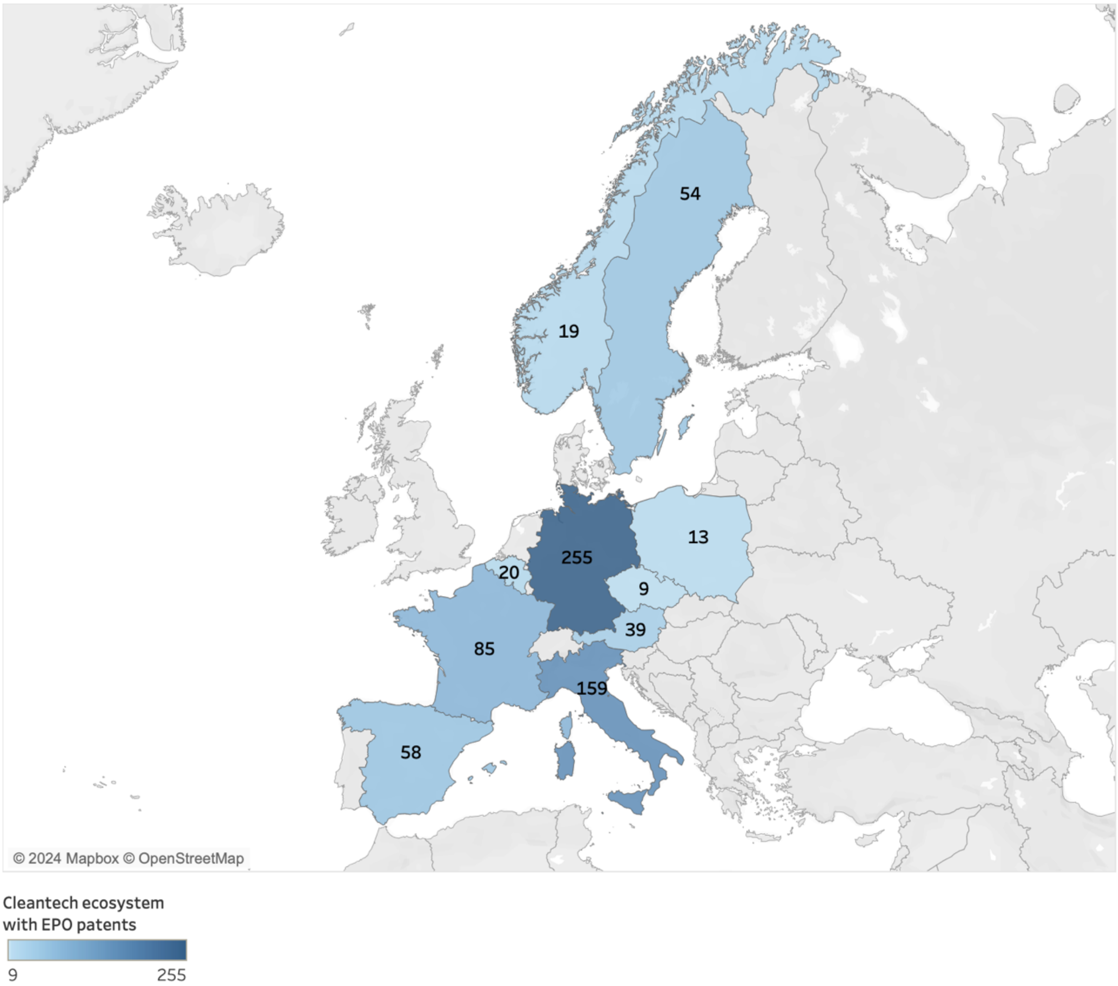

Figure 2.5.2: The share of patenting Cleantech ecosystem by country

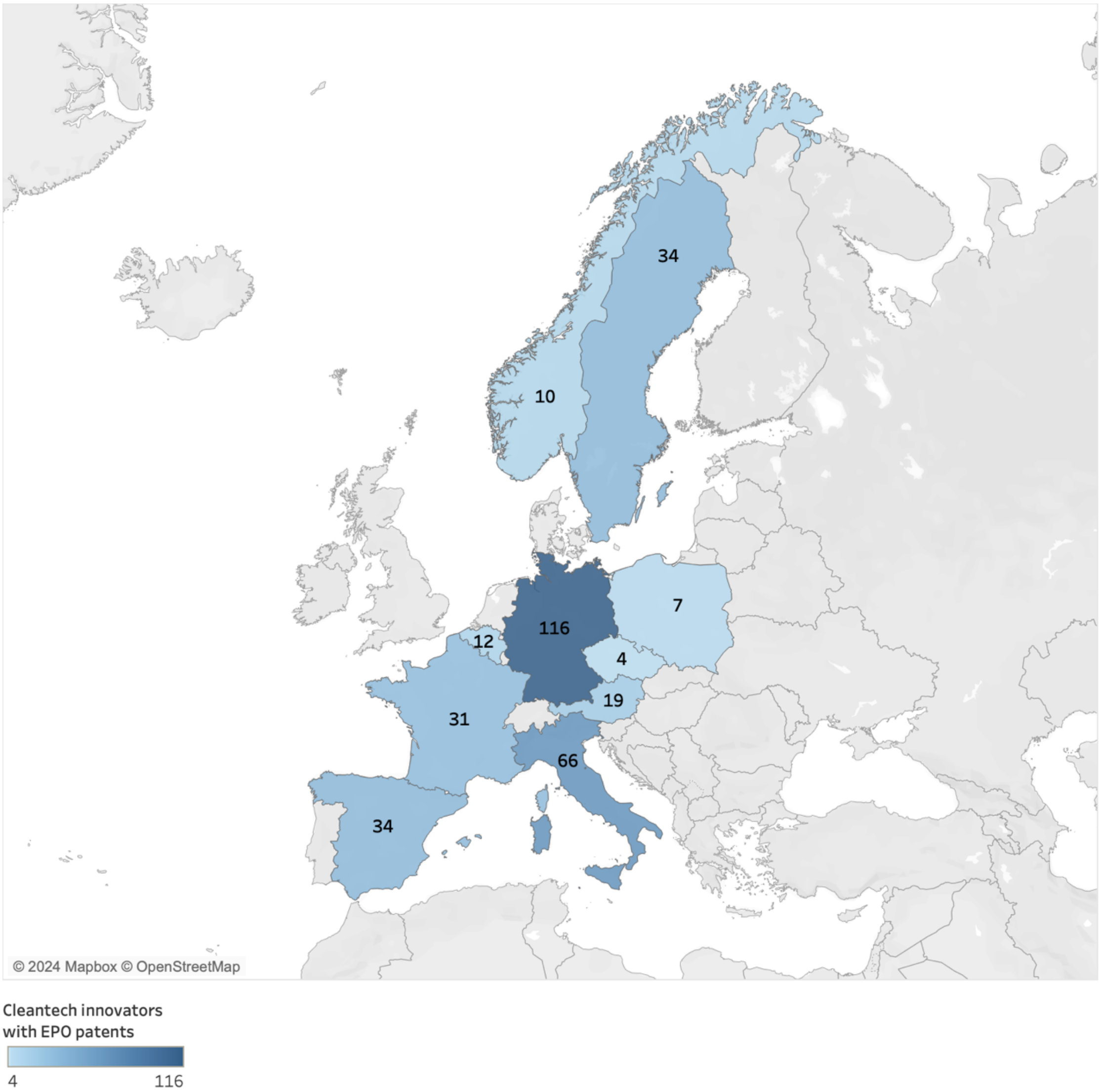

Figure 2.5.3: The share of patenting Cleantech innovators by country

Finally, we have plotted the distribution of EPO patenting green companies by their ecosystem segment, as showed in table 2.5.5.

We can see that Experimenters are the ones that largely invest in patenting (37.9%), followed by Innovators (22.8%), and Manufacturer (22.5%).

Table 2.5.5: Distribution of EPO patenting Cleantech Green energy companies by ecosystem segment

	At least one in any field	
	# companies	%
Innovators	407	22.8%
Experiments	11	37.9%
Manufacturers	484	22.5%
Distributors	64	5.3%
Integrators	132	4.9%
Operators	121	4.5%

2.6 Descriptive statistics on Cleantech companies by financial KPIs

The financial information is vital as allow us to picture a more detailed overview of the average Cleantech companies in the sample. The results are plotted in table 2.6.1.

Table 2.6.1: Descriptive statistics for financial KPIs for Cleantech Green energy companies

	# Observation	Mean	Median	St. dev.	Min	Max
Sales [k€]	10 863	135.9	6.8	135.9	(0.006)	159 000.0
Total Assets [k€]	10 531	115.6	7.0	1 133.8	0	72 500.0
Net Profit [k€]	10 614	6.6	0.1	146.7	(1.133)	12 900.0
EBITDA [k€]	9 668	17.9	0.5	335.6	(0.339)	26 700.0
# Employees	7 354	451.5	39	5 508.6	0	2.60E+05

In the following Table (Table 2.6.2), we collect the mean values of Cleantech Green energy companies divided into their respective segment.

We can observe that, generally, Cleantech innovators have higher sales volumes and higher levels of assets but have lower net profits with respect to those companies operating in Cleantech ecosystem segment.

Table 2.6.2: Comparison of mean values for financial KPIs between Cleantech segments

	Cleantech innovators	Cleantech ecosystem	Difference	Significance-level [1]
Sales [k€]	204.0	122.0	81.9	
Total Assets [k€]	167.2	105.2	62.1	(**)
Net profit [k€]	3.8	7.2	3.4	
EBITDA [k€]	16.7	18.2	1.4	
# Employees	661.5	408.5	253.0	

[1] (*), (**) and (***) indicate significantly different means <10%, <5%- and <1%-level respectively.

2.7 Descriptive statistics on Cleantech companies by Venture Capital investment

For Cleantech companies is of utmost importance to have some sources of external fundings like Venture Capitals because they can offer some backing in long-term, cost-intensive activities.

From the analysis showed in table 2.7.1 we can see that only the 2.4% of the total companies are backed by some VCs.

If we further analyze the sample considering the segment, we can see that the 5.2% of the companies classified as innovators are backed by VC, while only the 1.8% as those classified as ecosystem.

Table 2.7.1: Distribution of VC-backed green energy companies by segmentation

	Cleantech companies		Cleantech innovators		Cleantech ecosystem	
	# companies	%	# companies	%	# companies	%
#VC-backed companies	251	2.4%	93	5.2%	158	1.8%
Total # companies	10 585		1 787		8 798	

In table 2.7.2 we further analyze the VC-backed green energy companies by technological category. We can see that the category with the largest share of VC-backed companies is Capture, storage, sequestration, or disposal of GHG (4).

If we consider only the Cleantech innovators, we can see that the category most baked by VCs is again Capture, storage, sequestration, or disposal of GHG (4), while if we look at Cleantech ecosystem it is the Sustainable modes of transport (5).

Table 2.7.2: Distribution of VC-backed Cleantech Green energy companies by technological category

Technological category	VC-Backed companies					
	Cleantech companies		Cleantech innovators		Cleantech ecosystems	
	# companies	%	# companies	%	# companies	%
Air/water/soil pollution abatement/remediation (1.1)	13	1.5%	1	1.3%	12	1.5%
Waste management (1.2)	16	0.8%	1	3.0%	15	0.8%
Water conservation/availability (2.1)	3	1.4%	0	0.0%	3	1.5%
Sustainable agri-food technologies (2.2)	0	0.0%	0	0.0%	0	0.0%
Sustainable raw materials (2.3)	3	3.6%	1	12.5%	2	2.7%
Sustainable energy production (3.1)	166	2.7%	64	4.8%	102	2.2%
Sustainable fuels (3.2)	33	2.2%	8	6.0%	25	1.9%
Energy-efficient industrial technologies (3.3)	75	1.8%	24	5.9%	51	1.4%

Capture, storage, sequestration, or disposal of GHG (4)	1	11.1%	1	100.0%	0	0.0%
Sustainable modes of transport (5)	4	3.7%	0	0.0%	4	6.0%
Sustainable buildings (6)	13	2.2%	5	2.8%	8	1.9%
Total	251	2.4%	93	5.2%	158	1.8%

We analyzed the correlation between the number of VC-backed companies and those that have at least one EPO patent. These are important for VC-backed companies as they give Intellectual Property to the innovative companies, ensuring legal protection and enhancing attractiveness of companies to investors. The results are reported in table 2.7.3. Among the VC-backed companies, the 8.2% has at least one EPO patent in any field.

Table 2.7.3: Distribution of EPO patenting Cleantech Green energy companies by VC-backed status

	At least one patent in any field	
	# companies	%
VC-backed firms	100	8.2%
Non-VC-Backed firms	1 119	91.8%
Total	1 219	11.5%

In the following table (Table 2.7.4) we report the VC-backed companies by geography. We can observe that, among the VC-backed companies, the 9.6% is in Belgium, followed by Sweden (5.1%).

Table 2.7.4: Distribution of VC-backed Cleantech Green energy companies by country

	Green energy companies	
	# companies	%VC-Backed companies
Germany	34	1.6%
Italy	10	0.5%
France	64	4.4%
Spain	17	1.6%
Poland	7	1.3%
Sweden	20	5.1%
Czech Republic	1	0.4%
Belgium	27	9.6%
Norway	11	3.5%
Austria	6	2.1%
Others	54	2.8%
Total	251	2.4%

If we consider only the companies classified as innovators (Table 2.7.5) we can see that the distribution is the same as when considering the whole sample of green energy companies.

Table 2.7.5: Distribution of VC-backed Cleantech Green energy companies classified as innovators by country.

	Cleantech innovators	
	# companies	%VC-Backed companies
Germany	15	4.5%
Italy	3	1.0%
France	21	10.3%
Spain	4	1.7%
Poland	2	2.2%
Sweden	9	10.6%
Czech Republic	1	2.1%
Belgium	12	24.0%
Norway	2	4.9%
Austria	1	2.0%
Others	23	6.9%
Total	93	5.2%

The same results are obtained also when considering the companies classified as green energy ecosystem. The results are showed in Table 2.7.6.

Table 2.7.6: Distribution of VC-backed Cleantech Green energy companies classified as ecosystem by country.

	Cleantech ecosystem	
	# companies	%VC-Backed companies
Germany	19	1.1%
Italy	7	0.5%
France	43	3.5%
Spain	13	1.6%
Poland	5	1.1%
Sweden	11	3.6%
Czech Republic	0	0.0%
Belgium	15	6.5%
Norway	9	3.3%
Austria	5	2.1%
Others	31	1.9%
Total	158	1.8%

Table 2.7.7 shows the distribution of VC-backed companies according to their ecosystem segment. We can see that, coherently with the fact that VC investor typically search for highly innovative companies, the 37.1% of the companies is classified as Innovators, followed by Manufacturers (25.1%), and Integrators (16.3%).

Table 2.7.7: Distribution of VC-backed Cleantech Green energy companies by ecosystem segment

Category	Cleantech companies		Cleantech innovators		Cleantech ecosystem	
	# companies	%	# companies	%	# companies	%
Innovators	93	37.1%	93	100.0%	0	0.0%
Experiments	2	0.1%	0	0.0%	2	1.3%
Manufacturers	63	25.1%	0	0.0%	63	39.9%
Distributors	21	8.4%	0	0.0%	21	13.3%
Integrators	41	16.3%	0	0.0%	41	25.9%
Operators	31	12.4%	0	0.0%	31	19.6%
Total	251	100%	93	100%	158	100%

3. Environmental, Social, and Governance (ESG) Index: A Comprehensive Exploration

The growing awareness of sustainability as a critical global issue has significantly influenced investment practices, giving rise to the Environmental, Social, and Governance (ESG) index. The ESG index functions as a comprehensive tool that assesses and quantifies the non-financial performance of entities, including corporations and nations, by focusing on their adherence to sustainability principles. This index is not merely a benchmark for ethical investing but also serves as a reflection of how businesses and governments prioritize environmental conservation, social equity, and sound governance.

The ESG index is composed of three primary pillars:

1. Environmental: Evaluates factors such as carbon emissions, renewable energy use, and efforts toward mitigating climate change.
2. Social: Includes dimensions like labor practices, human rights adherence, and contributions to community welfare.
3. Governance: Examines transparency, board composition, and adherence to ethical practices.

The importance of ESG indices has grown in tandem with the global shift towards responsible investment. Financial markets, previously dominated by profit-centric approaches, are now recognizing the interconnectedness of profitability and sustainability. ESG metrics provide a way to measure and reward these efforts, thus encouraging better behavior among corporations and governments alike.

Additionally, the relevance of ESG indices has extended beyond the financial sector. Governments use them to demonstrate progress in sustainable development, while policymakers rely on them to assess the effectiveness of sustainability-related regulations. As stakeholders demand greater accountability, ESG indices serve as critical tools for fostering transparency, encouraging alignment with global goals such as the United Nations Sustainable Development Goals (SDGs).

The ESG index's utility spans two significant roles:

1. Decision-Making Tool for Investors: Investors increasingly prioritize ESG-compliant firms and nations, perceiving them as less risky and more likely to yield sustainable long-term returns. ESG-compliant firms are better equipped to handle regulatory pressures, societal expectations, and environmental risks.
2. Catalyst for Positive Change: By spotlighting sustainability performance, ESG indices encourage organizations and governments to integrate responsible practices into their operations. This alignment with sustainable practices benefits broader society and the environment.

The last two decades have seen an exponential rise in the prominence of ESG indices. A study by the NYU Stern Center for Sustainable Business revealed that companies with strong ESG scores often enjoy higher profitability and lower volatility (Gillan, Koch, & Starks, 2021). Similarly, sovereign ESG indices are becoming crucial benchmarks for global competitiveness, as nations strive to demonstrate leadership in sustainability.

However, despite these advances, ESG indices face challenges in standardization, data reliability, and accusations of greenwashing—where entities exaggerate their sustainability efforts. Addressing these issues is paramount to preserving the integrity of ESG metrics.

3.1 Historical Context and Evolution

The history of ESG investing is deeply rooted in the broader concept of socially responsible investing (SRI). Emerging in the 1960s, SRI was initially driven by religious groups and ethical communities seeking to align their financial investments with their moral beliefs. These groups avoided investments in industries such as tobacco, alcohol, and weapons manufacturing, which were considered harmful or unethical.

The 1960s marked a period of heightened social consciousness, particularly in the United States and Europe. Social movements advocating for civil rights, environmental protection, and labor rights inspired the early adoption of responsible investment practices. During this era, socially conscious funds like the Pax World Fund (established in 1971) emerged, explicitly excluding investments in controversial industries.

In the 1970s, environmental disasters such as the Santa Barbara oil spill and the publication of *The Limits to Growth* by the Club of Rome underscored the urgency of addressing environmental degradation. These events, combined with growing public concern, prompted the integration of environmental considerations into investment strategies. This marked the beginning of a more structured approach to responsible investing, albeit focused on specific sectors rather than a holistic evaluation.

The 1980s and 1990s witnessed the formalization of sustainability principles at the global level. Two landmark events during this period were the establishment of the Brundtland Commission in 1983 and the

1992 Rio Earth Summit. The Brundtland Commission introduced the concept of sustainable development, defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations, 1987). This definition became the foundation for subsequent sustainability initiatives.

The Rio Earth Summit brought together world leaders to address environmental challenges, resulting in agreements such as the United Nations Framework Convention on Climate Change (UNFCCC). These agreements laid the groundwork for coordinated global action, emphasizing the role of corporate and sovereign responsibility in achieving sustainability goals.

While SRI focused primarily on exclusionary practices, the 1990s saw the emergence of ESG as a more comprehensive framework. ESG extended beyond exclusionary screening to incorporate positive screening and active engagement. Investors began to recognize that environmental, social, and governance factors could materially impact financial performance.

In 2006, the United Nations launched the Principles for Responsible Investment (UN PRI), which provided a structured framework for integrating ESG considerations into investment analysis. The six principles outlined in the UN PRI emphasized the need for transparency, accountability, and collaboration in promoting ESG practices. By signing the UN PRI, institutional investors committed to incorporating ESG factors into their investment decisions, thereby accelerating the mainstream adoption of ESG frameworks.

The development of ESG indices paralleled the evolution of ESG frameworks. Early ESG indices, such as the Dow Jones Sustainability Index (DJSI) launched in 1999, were designed to track the performance of companies excelling in sustainability. The DJSI provided investors with a benchmark for assessing ESG leaders across industries.

Subsequent indices, including the MSCI ESG Indexes and the S&P 500 ESG Index, further refined methodologies for evaluating sustainability performance. These indices incorporated a broader range of ESG factors, reflecting the growing complexity of sustainability metrics. Sovereign ESG indices also emerged during this period, providing insights into how nations were performing in areas like environmental protection, social equity, and governance reform.

Today, ESG investing is a rapidly growing segment of the financial market. According to the Global Sustainable Investment Alliance, global sustainable investment assets reached $35.3 trillion in 2020, representing 36% of all professionally managed assets (Global Sustainable Investment Alliance, 2020). This growth reflects not only increased investor interest but also the expanding influence of ESG frameworks in shaping corporate and governmental strategies.

As ESG practices continue to evolve, the focus is shifting towards standardization, data quality, and impact measurement. New tools and technologies, such as artificial intelligence and blockchain, are being explored to enhance the accuracy and reliability of ESG metrics.

3.2 Relevance of ESG in Sovereign Contexts

At the sovereign level, ESG indices assess a country's performance in areas such as environmental stewardship, social equity, and governance efficiency. These indices are valuable for comparing nations and gauging their progress in sustainable development.

The key applications of Sovereign ESG indices are:

1. Policy Assessment: Sovereign ESG indices provide policymakers with a comprehensive framework to evaluate the effectiveness of sustainability initiatives. For instance, a high ESG score in governance reflects the presence of transparent institutions and effective legal frameworks.
2. Attracting Foreign Investment: Countries with strong ESG ratings are more likely to attract foreign direct investment (FDI). Investors consider ESG performance as an indicator of political stability, social cohesion, and environmental responsibility.
3. Global Competitiveness: ESG indices are increasingly used as benchmarks in international trade and diplomacy. For example, a country's ESG score may influence its eligibility for green financing or its standing in climate change negotiations.
4. Impact on Credit Ratings: Sovereign ESG performance can impact a country’s creditworthiness. Credit rating agencies, such as Moody’s and S&P Global, incorporate ESG factors into their assessments, linking sustainability to economic stability (Gratcheva, Emery, Wang, & Martin, 2021).

However, Sovereign ESG indices reveal significant disparities between nations. High-income countries, particularly in Europe, consistently achieve higher ESG scores due to their established governance systems and investment in sustainable infrastructure. Conversely, low-income countries, especially in Sub-Saharan Africa, face challenges such as limited resources, political instability, and the need to prioritize economic growth over environmental sustainability.

The integration of ESG factors into investment strategies has financial implications for sovereign entities. Investors increasingly view ESG compliance as a predictor of financial performance and resilience. Studies suggest that ESG investments often yield comparable or superior returns compared to traditional investments.

Some empirical evidence supporting ESG Investments are:

1. Higher Returns: Research by (Naomi & Akbar, 2021) demonstrates that ESG-focused portfolios often outperform conventional ones. For instance, European markets have reported a 1.59% increase in annual returns for ESG funds compared to traditional stock funds.
2. Lower Volatility: Companies and nations with strong ESG scores experience reduced volatility in financial markets. This stability is particularly evident during economic downturns, where ESG-compliant entities are better equipped to manage risks.
3. Cost of Capital: ESG performance is increasingly linked to the cost of capital. Entities with higher ESG ratings are perceived as lower-risk investments, allowing them to access capital at more favorable terms.

4. Employee Satisfaction and Productivity: High ESG scores, particularly in the social dimension, correlate with improved employee morale and retention. Satisfied employees are more productive and contribute to the organization’s financial success (Gillan et al., 2021).

Beyond its financial relevance, ESG serves as a catalyst for positive social and environmental change. By incentivizing sustainable practices, ESG indices encourage entities to align their goals with broader societal and environmental objectives. This alignment benefits not only investors but also communities and ecosystems.

Some examples of ESG-driven impact are:

1. Environmental Stewardship: Companies are increasingly adopting renewable energy solutions and reducing carbon footprints to improve their ESG scores. Similarly, governments are prioritizing clean energy initiatives and conservation efforts to enhance their standings in sovereign ESG indices.
2. Social Equity: ESG frameworks emphasize social inclusion, pushing organizations to address issues like gender equality, fair labor practices, and community engagement.
3. Transparency and Accountability: ESG metrics drive improvements in corporate and governmental transparency. For example, entities are now required to disclose detailed sustainability reports, enabling stakeholders to hold them accountable.

3.3 Methodologies for Sovereign ESG Index Calculation

The methodology employed in this study for computing the Sovereign ESG Index closely follows the approach outlined in (Jiang, Feng, & Yang, 2022). Their work provides a foundational framework for constructing sovereign ESG indices using the entropy weighting method and relevant indicators from the World Bank's Sovereign ESG Database. To ensure consistency with their approach and facilitate comparability, this research adopts their methodology in key aspects while also performing independent recalculations for specific components.

While (Jiang, Feng, & Yang, 2022) established the methodological foundation, this study extends their work by updating the ESG Index computation using data from the 2014–2021 period. The recalculation allows for a more contemporary assessment of sovereign ESG performance, capturing recent trends that were not included in the original 1990–2020 analysis. A systematic comparison between the newly computed ESG scores and the historical averages presented in (Jiang, Feng, & Yang, 2022) enables a direct evaluation of temporal shifts, highlighting whether sustainability efforts have improved over time and identifying emerging patterns.

Beyond updating the dataset, this research also reassesses the weight assignment of indicators. While the entropy weighting methodology described in (Jiang, Feng, & Yang, 2022) serves as the basis, the weights are recalculated independently to reflect the most recent variations in ESG factors rather than relying on historical distributions. Additionally, to provide a deeper understanding of sovereign ESG performance, Kernel

Density Estimation (KDE) is applied to visualize the distribution and evolution of ESG scores. This statistical approach offers insights into convergence, divergence, and disparities across countries, enriching the analysis with a more dynamic perspective.

Further refining the evaluation, an error analysis is introduced to compare the recalculated ESG scores for 2014–2021 with historical estimates. This comparison not only identifies methodological discrepancies but also sheds light on policy-driven improvements and potential data limitations that may have influenced ESG trends. The assessment of policy factors driving changes in sovereign ESG scores offers a more nuanced interpretation of ESG performance, ensuring that the findings go beyond numerical computation to address broader implications.

By integrating these elements, this study builds upon and expands the work of (Jiang, Feng, & Yang, 2022), offering a more current and comprehensive perspective on sovereign ESG performance in recent years. The following sections outline the methodological steps taken, explicitly distinguishing between elements derived from (Jiang, Feng, & Yang, 2022) and those recalculated in this research.

3.3.1 Data Normalization

To address the variation in units and scales among the selected indicators, Min-Max normalization was applied, following the approach used in (Jiang, Feng, & Yang, 2022). This technique standardizes all indicators to a common scale between 0 and 1, facilitating comparability.

For positive indicators, where higher values signify better performance (e.g., literacy rates), normalization followed the formula:

$$\mathrm{p} = \frac{\mathrm{x} - \mathrm{x}_{\mathrm{min}}}{\mathrm{x}_{\mathrm{max}} - \mathrm{x}_{\mathrm{min}}}$$

For negative indicators, where lower values indicate better performance (e.g., CO_2 emissions), an inverse scaling formula was used:

$$\mathrm{p} = \frac{x_{max} - x}{x_{max} - x_{min}}$$

While the normalization process adheres to the methodology outlined in (Jiang, Feng, & Yang, 2022), this research recalculates the values using the most recent dataset covering the period 2014–2021. This ensures that all transformations are aligned with updated sovereign ESG data, rather than relying on historical computations.

3.3.2 Weight Assignment Using the Entropy Method

The entropy method, as originally described by (Jiang, Feng, & Yang, 2022), was employed to assign objective weights to the normalized indicators based on their variability and informational contribution. First, the proportional contribution of each indicator was calculated relative to the total across all countries and years:

$$s_{ij} = \frac{p_{ij}}{\sum_{i=1}^{n} p_{ij}}$$

Next, the entropy value for each indicator was derived using:

$$E_j = -\frac{1}{\ln(n)} \sum_{i=1}^{n} s_{ij} \ln(s_{ij})$$

Where n is the number of observations.

The utility value of each indicator was then calculated as:

$$d_j = 1 - E_j$$

Finally, the weight for each indicator was determined by normalizing its utility value:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}$$

This process ensured that indicators with higher variability, and thus greater informational value, were given higher weights in the index computation.

Although this study follows the methodological framework of (Jiang, Feng, & Yang, 2022), the weight assignment is not simply replicated. Instead, the entropy weights are recalculated based on the new dataset, ensuring that the weighting structure reflects more recent trends in sovereign ESG performance. By updating these calculations, this research accounts for possible shifts in the significance of ESG factors over time, providing a more accurate and contextually relevant index.

The average results for the period 2014-2021 of the computational steps are shown in table 3.3.2.1. Indicators with sign equal to -1 are negatively correlated with sovereign ESG, and those with sign +1 are positively correlated.

Table 3.3.2.1: Sovereign ESG data framework

Category	Indicator	Sign(±)	Weight
Environment Pillar	CO_2 emissions (metric tons per capita)	-1	1.46%
	Methane emissions (metric tons of CO_2 equivalent per capita)	-1	1.68%
	Nitrous oxide emissions (metric tons of CO_2 equivalent per capita)	-1	1.69%
	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	-1	1.70%
	Adjusted savings: natural resources depletion (% of GNI)	-1	1.67%
	Adjusted savings: net forest depletion (% of GNI)	-1	1.70%
	Annual freshwater withdrawals, total (% of internal resources)	-1	1.70%
	Forest area (% of land area)	1	1.76%
	Terrestrial and marine protected areas (% of total territorial area)	+1	1.38%
	Electricity production from coal sources (% of total)	-1	1.68%
	Energy imports, net (% of energy use)	-1	1.26%
	Energy intensity level of primary energy (MJ/$2017 PPP GDP)	-1	1.68%
	Energy use (kg of oil equivalent per capita)	-1	1.67%
	Fossil fuel energy consumption (% of total)	-1	1.71%
	Renewable electricity output (% of total electricity output)	+1	0.45%
	Renewable energy consumption (% of total)	+1	1.54%

	Population density (people per sq. km of land area)	-1	1.68%
	Agricultural land (% of land area)	+1	1.74%
	Agriculture, forestry, and fishing, value added (% of GDP)	+1	1.93%
	Food production index (2014-2016=100)	+1	1.67%
Social Pillar	Government expenditure on education, total (% of government expenditure)	+1	1.75%
	Literacy rate, adult total (% of people ages 15 and above)	+1	3.71%
	School enrollment, primary (% gross)	+1	1.69%
	Children in employment, total (% of children ages 7-14)	-1	1.67%
	Labor force participation rate, total (% of total labor force)	+1	1.67%
	Unemployment, total (% of total labor force)	-1	1.70%
	Fertility rate, total (births per woman)	-1	1.70%
	Life expectancy at birth, total (years)	-1	1.44%
	Population aged 65 and above (% of total population)	+1	1.76%
	Cause of death, by communicable diseases and maternal, paternal and nutrition conditions (% of total)	-1	1.67%
	Mortality rate, under 5 (per 1000 live births)	-1	1.69%
	Prevalence of overweight (% of adults)	-1	1.69%
	Prevalence of undernourishment (% of adults)	-1	1.68%
	Hospital beds (per 1000 people)	+1	1.48%
	Annualized average growth rate in per capita real survey mean consumption or income, total population (%)	+1	1.80%
	GINI index (World Bank estimate)	-1	1.77%
	Income share held by lowest 20%	+1	1.74%
	Poverty headcount ratio at national poverty lines (% of population)	-1	1.80%
	Access to clean fuels and technologies for cooking (% of population)	+1	1.69%
	Access to electricity (% of population)	+1	1.67%
	People using safely managed sanitation services (% of population)	+1	1.72%

	Strength of legal rights index (0=weak to 12=strong)	+1	1.27%
	Voice and accountability: estimate	+1	1.71%
	Government effectiveness: estimate	+1	1.72%
	Regulatory quality: estimate	+1	1.75%
	Control of corruption: estimate	+1	1.81%
	Net migration	-1	1.67%
	Political stability and absence of violence/terrorism: estimate	+1	1.68%
	Rule of law: estimate	+1	1.79%
Governance Pillar	GDP growth (annual %)	+1	1.68%
	Individuals using the internet (% of population)	+1	1.67%
	Proportion of seats held by women in national parliaments (%)	+1	1.71%
	Ratio of female to male labor force participation rate (%)	+1	1.69%
	School enrollment, primary and secondary (gross), gender parity index (GPI)	+1	1.95%
	Unmet need for contraception (% of married women ages 15-49)	-1	1.67%
	Patent applications, residents	+1	2.74%
	Research and development expenditure (% of GDP)	+1	1.82%
	Scientific and technical journal articles	+1	1.91%

3.3.3 Index Computation

The Sovereign ESG Index for each country-year combination was calculated as a weighted sum of the normalized indicators:

$$\mathrm{ESG}_i = \sum_{j=1}^{m} w_j \cdot p_{ij}$$

Where w_j is the entropy-derived weight assigned to indicator j, and p_{ij} is the normalized value of indicator j for country i.

This aggregation produced a composite score that reflects each country's overall ESG performance, integrating environmental, social, and governance dimensions.

Then, the aggregation of the ESG results into mean values for 2014–2021 was performed to facilitate a direct and consistent comparison with the 1990–2020 averages presented in the paper by (Jiang, Feng, & Yang, 2022), which employed similar computational steps to derive its results. This alignment ensures methodological consistency and supports meaningful insights into temporal trends in ESG performance. The results are shown in Table 3.3.3.1

Table 3.3.3.1: Average results of Sovereign ESG for the period 2014-2021

Country	ESG
Albania	0.56
Austria	0.68
Belgium	0.72
Bosnia and Herzegovina	0.50
Bulgaria	0.60
Croatia	0.58
Cyprus	0.64
Czech Republic	0.65
Denmark	0.71
Estonia	0.65
Finland	0.69
France	0.69
Germany	0.74
Greece	0.58
Hungary	0.60
Iceland	0.65
Ireland	0.62

Italy	0.63
Latvia	0.62
Lithuania	0.62
Luxembourg	0.66
Malta	0.68
Montenegro	0.54
Netherlands	0.66
North Macedonia	0.56
Norway	0.67
Poland	0.66
Portugal	0.64
Romania	0.58
Serbia	0.54
Slovakia	0.60
Slovenia	0.63
Spain	0.69
Sweden	0.69
Switzerland	0.68
Turkey	0.60
United Kingdom	0.69

3.3.4 Statistical Analysis Using Kernel Density Estimation (KDE)

To analyze the computed Sovereign ESG Index, Kernel Density Estimation (KDE) was applied to visualize the distribution and temporal trends of ESG scores. KDE, a non-parametric statistical method, estimates the probability density function of the data, providing insights into patterns such as convergence, divergence, or disparities across countries. The KDE formula used was:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

Where h represents the bandwidth controlling the smoothness of the density curve, n is the number of observations, and K is the Kernel function, that has been set as a Gaussian. This is equal to:

$$K(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

The Gaussian kernel ensures that each observation contributes to the density estimate in a weighted manner, with weights decreasing smoothly as the distance from the point of interest increases. This method allowed for the visualization of patterns such as convergence, divergence, and disparities in ESG performance across countries and over time. The smooth density curves generated by the Gaussian kernel provided a clear representation of the temporal evolution and regional distribution of ESG scores.

By employing KDE, it was possible to explore changes in the distribution of ESG scores over time and identify trends among different groups of countries.

The three-dimensional kernel density estimation (KDE) plot showed in Figure 3.3.4.1 provides a visual representation of the temporal distribution of Sovereign ESG scores over the years. This analysis is crucial to understanding how ESG scores are distributed, how they have evolved over time, and what insights can be drawn about trends in the underlying factors driving ESG performance at the sovereign level.

The plot is structured with the years on one axis, ESG scores on the second, and density on the vertical axis.
The plot reveals notable shifts in the distribution of ESG scores over the observed period. Early years in the dataset (e.g., 2014–2015) exhibit a relatively broad distribution, as indicated by a flatter density surface. This observation suggests a wide variability in ESG performance among sovereign entities during the initial stages of measurement. As we move to more recent years (e.g., 2019–2021), the density surface becomes more concentrated, with sharper peaks emerging in specific ranges of ESG scores. This evolution signifies a convergence of ESG performance, potentially indicating the adoption of more uniform ESG practices or increasing standardization in reporting and assessment criteria across sovereigns.
Additionally, the shifting location of density peaks over time reflects changes in the central tendency of ESG scores.
In earlier years, the peaks appear to cluster around lower ESG scores, indicating that a larger proportion of sovereign entities scored relatively poorly on ESG metrics. Over time, these peaks shift toward higher ESG scores, suggesting an improvement in ESG performance across the board. This trend may reflect growing global awareness of sustainability issues, stronger policy commitments to ESG principles, and increasing external pressure from investors and civil society to prioritize ESG factors in governance.
The KDE plot also provides insights into the spread and skewness of the distribution. Early years are characterized by broader distributions, indicating significant heterogeneity in ESG scores. This may suggest

disparities in the adoption or implementation of ESG policies across different regions or levels of economic development. In later years, the distribution appears narrower, implying a reduction in such disparities. This trend could be attributed to increased global collaboration on sustainability goals (e.g., the United Nations' Sustainable Development Goals) or the implementation of standardized ESG measurement frameworks.

The presence of multiple density peaks in some years suggests the existence of distinct clusters within the dataset. These clusters could correspond to groups of countries with similar ESG profiles, such as high-income countries achieving consistently high scores and lower-income or resource-dependent countries lagging. Over time, as the peaks converge, it suggests that these groups may be closing the gap in ESG performance.

The sharpness and height of density peaks provide valuable insights into the concentration of ESG scores within specific ranges. Higher peaks indicate a greater number of countries with similar scores, highlighting dominant trends in ESG performance. For instance, in recent years, the emergence of a prominent peak at higher ESG scores suggests that a significant number of sovereign entities have improved their performance, potentially due to increased regulatory requirements, technological advancements, or shifts in public expectations regarding sustainability and governance practices.

Conversely, the presence of smaller peaks or shoulders in the density plot might indicate outliers or a subset of countries that deviate from broader trends. These deviations could be driven by unique political, economic, or social contexts that impact the ability or willingness of these countries to align with global ESG norms.

Governance often plays a critical role in shaping ESG outcomes, and the temporal trends observed in the KDE plot may reflect shifts in global governance priorities. For example, the consistent upward trend in ESG scores over time aligns with the adoption of international agreements such as the Paris Agreement (2015), which emphasized the importance of environmental governance. Similarly, initiatives focused on social governance, such as gender equity and labor rights, may have contributed to improvements in the social dimension of ESG scores.

The narrowing spread of ESG scores in later years may also reflect the harmonizing influence of international organizations and financial institutions. As ESG metrics have become more integral to investment decisions, sovereign entities may have faced greater pressure to align their practices with international standards. For instance, countries seeking foreign investment or development assistance may have had to demonstrate improvements in ESG metrics to remain competitive.

The KDE plot also offers implications for policymaking and economic development. The upward trend in ESG scores may be indicative of policy successes, such as improved environmental regulations, enhanced social protections, and strengthened governance mechanisms. However, the differences in density spread and the persistence of some lower scores highlight ongoing challenges, particularly for countries with limited resources or significant structural barriers to improving ESG performance.

The clusters observed in earlier years may correspond to regional differences in ESG priorities or the availability of resources to implement ESG initiatives. For instance, resource-rich but institutionally weak countries may have scored poorly on governance metrics, while high-income countries with strong institutional frameworks may have achieved higher scores overall. Over time, the convergence of ESG scores suggests that lower-performing countries may have adopted lessons or best practices from higher-performing peers, aided by international cooperation and capacity-building efforts.

From the perspective of investors, the KDE plot highlights important trends in sovereign ESG performance that may influence investment decisions. The increasing concentration of ESG scores in higher ranges suggests that sovereign bonds issued by countries with strong ESG credentials may have become more abundant over time. This trend aligns with the growing popularity of green bonds and other ESG-linked financial instruments, which provide an avenue for investors to align their portfolios with sustainability goals.

However, the plot also underscores the importance of due diligence, particularly in years with broader distributions or multiple density peaks. In such cases, investors may need to carefully evaluate individual sovereign issuers to identify those that align most closely with their ESG objectives.

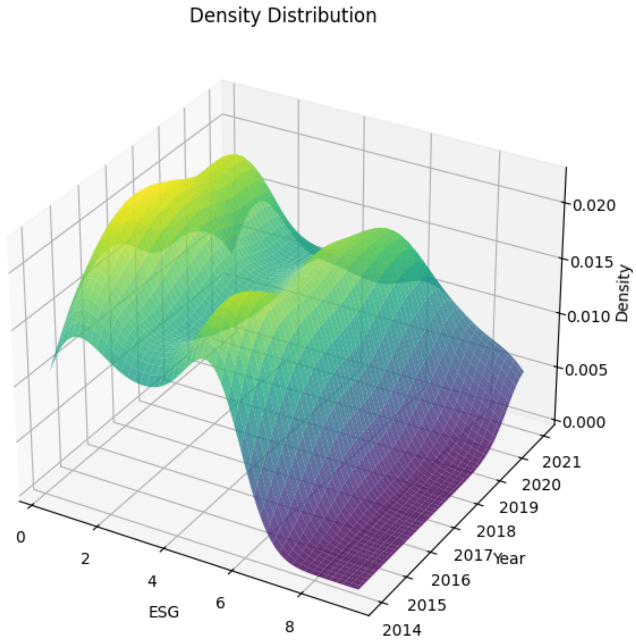

Figure 3.3.4.1: KDE distribution of the Sovereign ESG index for the period 2014-2021

3.4 Error analysis

To evaluate the reliability and alignment of the recalculated Sovereign ESG Index (2014–2021) with the historical estimates from (Jiang, Feng, & Yang, 2022), an error analysis was conducted. This comparison serves to assess the extent to which recent ESG performance differs from long-term trends and to identify potential sources of divergence.

The analysis considers both absolute and relative errors, offering insights into methodological consistency and highlighting significant deviations where ESG scores have changed considerably over time.

The absolute error has been computed as:

$$absolute\ error = |ESG_{2014-2021} - ESG_{1990-2020}|$$

And the relative error as:

$$relative\ error = \frac{|ESG_{2014-2021} - ESG_{1990-2020}|}{ESG_{1990-2020}} \cdot 100$$

While some discrepancies arise from methodological refinements, others reflect real-world policy changes, economic developments, or improvements in ESG-related governance that were not captured in earlier assessments.

This investigation is particularly relevant in the context of evolving sustainability frameworks. Over the past decade, many countries have implemented ambitious environmental policies, enhanced social protections, and strengthened governance structures, leading to marked shifts in their ESG scores. These transformations, alongside improvements in data availability and standardization, contribute to observed variations between the 2014–2021 estimates and historical benchmarks.

Table 3.4.1 summarizes these results, illustrating the degree of alignment between the two periods for each country.

Table 3.4.1: Absolute and Relative errors on the sovereign ESG index estimation

Country	Absolute error	Relative error [%]
Albania	0.010	1.754
Austria	0.070	11.475
Belgium	0.130	22.414
Bosnia and Herzegovina	0.050	9.091
Bulgaria	0.020	3.448
Croatia	0.010	1.754
Cyprus	0.070	12.281
Czech Republic	0.070	12.069
Denmark	0.110	18.333
Estonia	0.060	10.169
Finland	0.080	13.115
France	0.100	16.949
Germany	0.150	25.424
Greece	0.010	1.754
Hungary	0.020	3.448
Iceland	0.030	4.839
Ireland	0.050	8.772
Italy	0.060	10.526
Latvia	-	-
Lithuania	0.040	6.897
Luxembourg	0.070	11.864
Malta	0.110	19.298
Montenegro	0.030	5.263
Netherlands	0.070	11.864
North Macedonia	-	-
Norway	0.060	9.836
Poland	0.090	15.789
Portugal	0.060	10.345
Romania	0.010	1.754
Serbia	0	0
Slovakia	0.020	3.448
Slovenia	0.040	6.780
Spain	0.110	18.966
Sweden	0.070	11.290
Switzerland	0.070	11.475
Turkey	0.060	11.111
United Kingdom	0.100	16.949

A deeper exploration of these trends is provided in Sections 3.4.1 and 3.4.2, where country-specific deviations are examined in detail. By considering these variations in conjunction with broader policy shifts, this research offers a more nuanced understanding of how sovereign

ESG performance has evolved in response to both structural reforms and external sustainability pressures.

3.4.1 Key Observations from Error Analysis

The bar charts of relative and absolute errors by country reveal substantial variability in the alignment between the computed ESG metrics and the historical averages (Figures 3.4.1.1 and 3.4.1.2).

Countries such as Belgium and Montenegro exhibit disproportionately high relative errors exceeding 20%, reflecting rapid ESG performance improvements during 2014–2021 that were not captured in the historical 1990–2020 averages. These deviations can be attributed to recent policy-driven advancements:

1. Belgium: Recent governance reforms, renewable energy adoption, and social equality initiatives have significantly enhanced its ESG performance. Notable examples include commitments under the European Green Deal, which have prioritized carbon neutrality and green energy transitions.
2. Montenegro: As a developing country, Montenegro has implemented governance and environmental reforms to align with EU accession criteria, driving significant improvements in ESG performance.

In contrast, countries such as North Macedonia and Latvia, lack historical data, and hence it was not possible to compute the errors.

Similarly, the absolute error chart identifies countries such as Germany as having the highest absolute discrepancies.

Germany's leadership in ESG performance, driven by policies like the Energiewende initiative, reflects a strong upward trend during 2014–2021 compared to the long-term average.

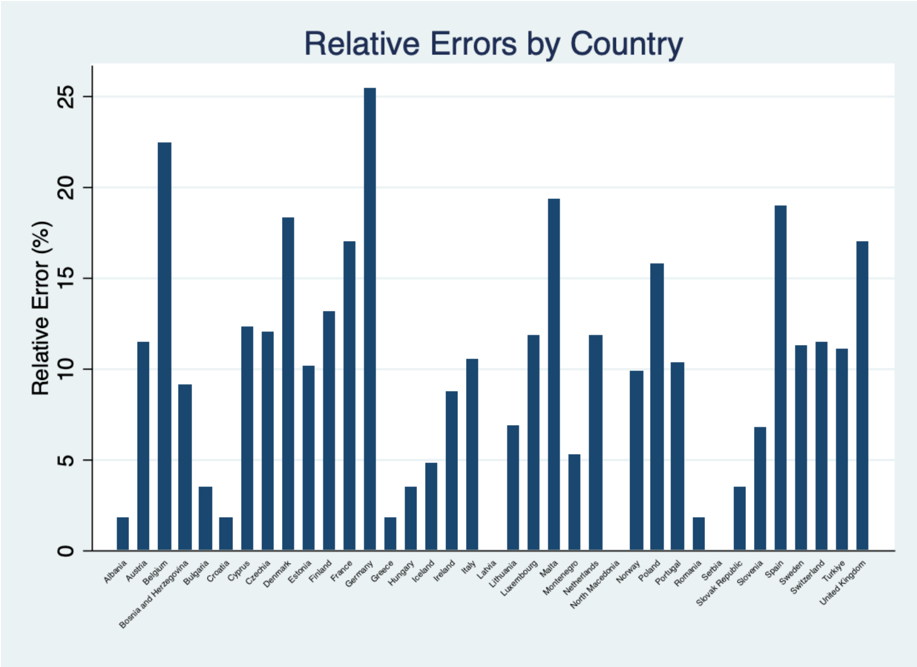

Figure 3.4.1.1: Relative error by country

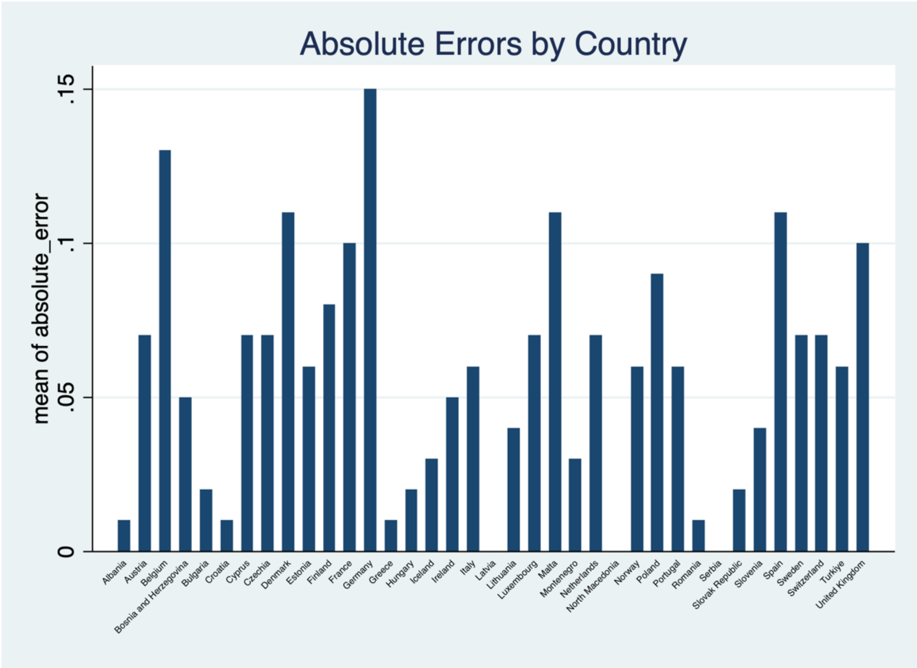

Figure 3.4.1.2: Absolute error by country

3.4.2 Insights from trend analysis

The trend analysis, visualized in Figure 3.4.2.1, evaluates the mean slope of ESG changes during the 2014–2021 period, providing deeper insights into the pace and direction of ESG performance improvements. Key observations include:

1. Strong Positive Trends: Countries such as Germany, Belgium, and Sweden exhibit the highest positive slopes, indicating consistent and significant improvement in ESG performance. These trends align with their active participation in global initiatives such as the Paris Agreement and their investments in sustainability and governance reforms.
2. Moderate Trends: Countries like Cyprus and Estonia show moderate positive slopes, suggesting steady but less aggressive progress in ESG metrics during the observed period. These trends might reflect the gradual implementation of environmental policies and governance improvements.
3. Negative or Stagnant Trends: A few countries, such as Malta and North Macedonia, exhibit flat or slightly negative slopes, indicating stagnation or declines in ESG performance. For Malta, this may reflect challenges in implementing large-scale sustainability reforms due to its small size and resource constraints. In the case of North Macedonia, limited baseline data and slower institutional reforms may have contributed to this result.

Overall, the trend analysis highlights significant disparities in ESG improvement rates among countries. These trends align with global and regional policy dynamics, where higher-performing countries tend to maintain momentum, while emerging economies face structural or resource-related challenges that limit progress.

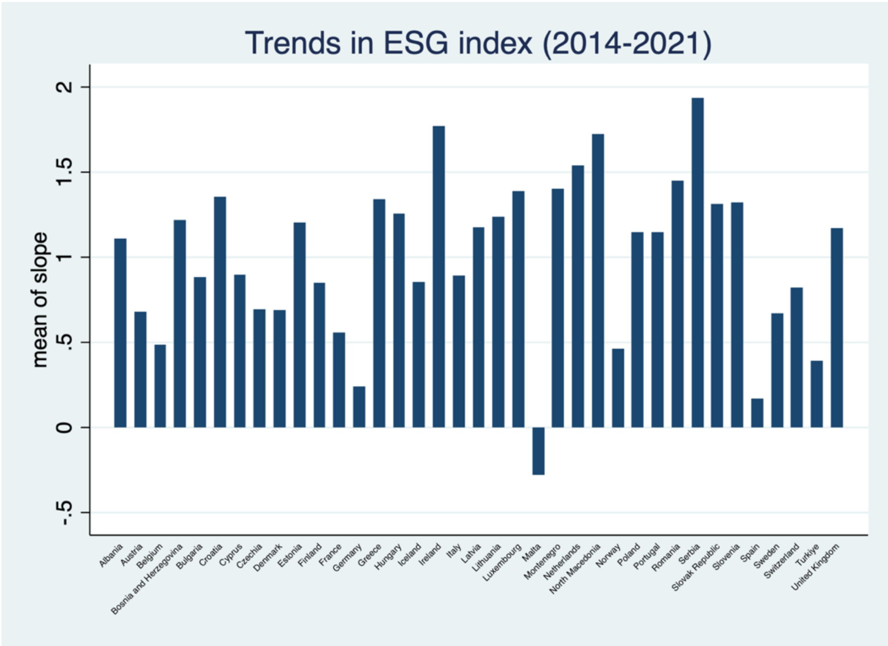

Figure 3.4.2.1: Trends in ESG index over the period 2014-2021

3.4.3 Impact of Recent Policies on Trends

The upward trend in ESG performance for many countries during the 2014–2021 period can be largely attributed to transformative global and national policy initiatives.

These policies have driven measurable improvements in environmental, social, and governance indicators, shaping the observed shifts in sovereign ESG scores. Unlike the historical estimates analysed by (Jiang, Feng, & Yang, 2022), which primarily reflect long-term ESG trends, the recalculated index in this research captures the direct effects of recent sustainability-driven policies and regulatory changes. Key examples include:

1. The Paris Agreement (2015): This global accord drove significant improvements in environmental indicators for many countries, particularly in renewable energy generation and emissions reductions.
2. European Green Deal (2019): This policy framework, led by the European Union, has pushed member states, including Belgium and Germany, to prioritize sustainability through carbon neutrality goals, green energy transitions, and circular economy initiatives.
3. Social Governance and Equality Reforms: Countries like Belgium and Denmark have strengthened labor rights, gender equality, and education policies, contributing to significant improvements in their social pillar scores.
4. Institutional Reforms in Emerging Economies: Developing nations such as Montenegro have implemented governance reforms and adopted international ESG reporting standards to attract foreign investment, driving substantial improvements in their scores.

The recalculated ESG scores reveal that these policy-driven shifts have played a critical role in shaping recent trends. Countries with ambitious legislative and regulatory advancements exhibit the most significant improvements in ESG performance, particularly in governance and environmental dimensions. The impact of these policies also explains some of the observed discrepancies between the recalculated 2014–2021 ESG scores and historical estimates from (Jiang, Feng, & Yang, 2022). Nations that have enacted rapid sustainability reforms often demonstrate larger deviations from historical averages, as their recent policy-driven improvements were not reflected in the long-term ESG trends previously reported.

By capturing these policy effects, this research provides a more updated and comprehensive understanding of how real-world sustainability initiatives have influenced sovereign ESG performance. The results underscore the increasing role of policy in shaping ESG scores, reinforcing the importance of integrating sustainability commitments into national economic and governance strategies.

3.4.4 Conclusion

The findings from this research confirm that sovereign ESG scores have undergone substantial shifts in recent years, reflecting the impact of global sustainability policies and national-level governance reforms. The recalculated Sovereign ESG Index for the 2014–2021 period highlights both convergences and divergences in sustainability progress across nations, with some countries making significant advancements, while others exhibit persistent disparities due to structural, economic, or policy-related factors.

The error analysis demonstrates the effectiveness of the entropy method in assigning weights that reflect variability and informational contribution. However, discrepancies driven by recent policy shifts, methodological limitations, or sparse historical data highlight areas for improvement. Countries such as Belgium and Germany illustrate the impact of transformative policy initiatives, while North Macedonia and

Latvia underscore the need for robust historical data in ensuring consistency in ESG assessments. By addressing these challenges through dimension-specific analyses, time-weighted adjustments, and advanced statistical techniques, the metrics can be refined for greater accuracy and reliability in statistical inference and policy evaluation. Figures 3.4.1.1, 3.4.1.2, and 3.4.2.1 illustrate these findings, providing critical insights into the alignment of computed ESG metrics with historical trends.

Furthermore, the observed policy-driven improvements suggest that traditional sustainability indices may lag in capturing the effects of recent regulatory and governance reforms. This underscores the need for more dynamic ESG assessment frameworks that can quickly incorporate real-time sustainability data and policy adjustments. Methodological refinements, such as updated entropy weighting and KDE density estimation, have enhanced the robustness of ESG index computations, allowing for a more accurate reflection of contemporary sustainability trends and reducing potential biases caused by outdated weighting structures.

As sustainability metrics continue to evolve, future research should focus on improving the adaptability and responsiveness of ESG indices. Incorporating real-time data processing, machine learning-based anomaly detection, and enhanced cross-country policy comparisons could provide greater predictive accuracy for ESG performance. Additionally, addressing greenwashing risks and ensuring greater transparency in sovereign ESG reporting will be crucial for maintaining the credibility and effectiveness of ESG-based investment and policy decisions.

This conclusion serves as a transition to the broader implications for policy and investment, discussed in Chapter 6, where the role of ESG and CCPI indices in shaping venture capital trends, economic competitiveness, and sustainable investment strategies is further explored.

3.5 Challenges in ESG Index Development

Despite their utility, ESG indices face several challenges:

1. Data Limitations: The absence of standardized global frameworks leads to inconsistencies in ESG reporting. Variations in data quality and availability hinder accurate assessments (Gillan, Koch, & Starks, 2021).
2. Greenwashing Concerns: Companies and nations may exaggerate their ESG claims to attract investment, undermining the credibility of indices. Regulatory mechanisms like the EU Sustainable Finance Disclosure Regulation aim to address this issue (Naomi & Akbar, 2021).
3. Complexity of Metrics: The integration of diverse indicators into a unified score presents methodological challenges. Balancing quantitative and qualitative metrics requires rigorous validation (Jiang, Feng, & Yang, 2022).

4. Climate Change Performance Index: analysis and imputation

The Climate Change Performance Index (CCPI) is an internationally recognized benchmarking tool designed to evaluate and compare the climate action efforts of 63 countries and the European Union, which together account for over 90% of global greenhouse gas emissions. Developed by Germanwatch, the NewClimate Institute, and the Climate Action Network, the CCPI enhances transparency in climate governance and encourages stronger commitments to emissions reduction and sustainable energy transitions (Jones & Youngs, 2024).

The index assesses nations based on four key categories: greenhouse gas emissions, renewable energy, energy use, and climate policy. These categories encompass 14 specific indicators, with greenhouse gas emissions receiving the highest weighting due to their direct impact on climate change. The annual CCPI rankings serve as a crucial reference for policymakers, researchers, and civil society groups, fostering debate on the effectiveness of national climate strategies (Climate Change Performance Index (CCPI) 2025: What You Should Know).

While widely utilized for climate accountability, the CCPI has been subject to methodological critiques, particularly regarding its reliance on quantitative indicators, which may overlook socio-economic contexts and policy nuances (The Climate Change Performance Index 2024: Results). Additionally, its comparative ranking system has led to discussions on the challenges of fairly evaluating nations with varying levels of economic development and historical emissions responsibilities (Jones & Youngs, 2024). Despite these debates, the CCPI remains a pivotal mechanism for monitoring climate progress and informing global mitigation efforts.

4.1 History

The Climate Change Performance Index (CCPI) was established in 2005 to provide an independent and systematic assessment of national climate protection efforts. Developed by Germanwatch, the NewClimate Institute, and the Climate Action Network, the index was conceived as a response to the growing need for a standardized benchmarking tool to measure the effectiveness of climate mitigation policies across major emitting nations. The 63 countries evaluated, along with the European Union, collectively contribute to over 90% of global greenhouse gas emissions, making the CCPI a critical instrument for tracking international climate progress (FAQs | Climate Change Performance Index - CCPI).

Since its inception, the CCPI has undergone numerous methodological refinements to better reflect the evolving landscape of global climate policy. Initially focusing on energy-related CO_2 emissions, the index has expanded its scope over time to encompass broader greenhouse gas emissions and sustainability-related indicators. A major methodological revision occurred in 2017 to align with the targets of the Paris Agreement, incorporating additional criteria such as long-term climate policy ambitions, the trajectory of emissions reductions, and the effectiveness of national energy transitions (Climate Change Performance Index by Country 2024 - World Population Review).

The CCPI's influence has steadily grown, shaping climate discussions at multiple levels—from national policy frameworks to international climate negotiations. Its rankings and findings frequently spark debate within parliaments and policy-making institutions, prompting governments to reassess their commitments and strategies. Additionally, the index has become an essential reference for researchers, investors, and civil society organizations seeking data-driven insights into global climate action.

Despite its impact, the CCPI has also faced criticism, particularly regarding its ability to fairly compare nations with vastly different economic and developmental contexts. Some critics argue that high-income countries with historically large carbon footprints may receive disproportionately favourable evaluations compared to developing nations striving to balance climate action with economic growth (Methodology | Climate Change Performance Index - CCPI). To address these concerns, ongoing discussions about methodological adjustments and more context-sensitive weighting criteria continue to be explored.

In summary, the CCPI has cemented itself as a crucial tool in the climate accountability landscape, pushing nations to strengthen their mitigation efforts. As climate policies and global agreements evolve, the CCPI's role in shaping national and international climate action will remain indispensable, with continued refinements necessary to ensure fair and effective evaluations of climate progress worldwide.

4.2 Methodology

The Climate Change Performance Index (CCPI) employs a comprehensive scoring system to evaluate and compare the climate protection performance of 63 countries and the European Union (EU), collectively responsible for over 90% of global greenhouse gas (GHG) emissions. Established by the German environmental organization Germanwatch e.V., the CCPI aims to enhance transparency in international climate politics and has been published annually since 2005, with updates provided during the UN Climate Change Conference (Climate Change Performance Index by Country 2024 - World Population Review) (Jones & Youngs, 2024).

The assessment methodology involves a mix of quantitative data and qualitative assessments. Data for the GHG emissions, renewable energy, and energy use indicators is gathered from recognized databases and statistical reports. In contrast, the climate policy component is assessed through expert questionnaires that rate governmental measures on a scale ranging from very high to very low effectiveness. This dual approach ensures a holistic view of each country's climate performance and policy efficacy (Jones & Youngs, 2024) (Climate Change Performance Index by Country 2024 - World Population Review).

The CCPI serves as a vital tool for understanding national commitments to climate action, particularly in the context of the Paris Agreement's Nationally Determined Contributions (NDCs) and 2030 targets, helping to drive accountability and transparency in global climate efforts

(Climate Change Performance Index by Country 2024 - World Population Review).

4.2.1 Scoring and Factors

The CCPI combines a weighted average of four main factors: greenhouse gas emissions, energy use, renewable energy, and climate policy (FAQs | Climate Change Performance Index - CCPI) (Climate Change Performance Index | NewClimate Institute). Greenhouse gas emissions contribute 40% to the total score, while the other three factors each account for 20%. In addition to the main factors, the CCPI assesses 14 components across these categories, ensuring a comprehensive evaluation of each nation's climate policies and actions (FAQs | Climate Change Performance Index - CCPI).

Looking ahead to the CCPI 2024 edition, Denmark continues to hold the best ranking, followed closely by Estonia and the Philippines. Despite ongoing improvements, the lack of countries achieving "very high" ratings indicates persistent challenges in global climate mitigation efforts (How Grassroots Environmental Activism Has Changed the Course of History). The CCPI's methodology, including its climate policy assessment based on expert ratings, underscores the importance of both quantitative and qualitative data in evaluating national performances (Restrepo-Mieth, Perry, Garnick, & Weisberg, 2023).

4.3 Relevance

The Climate Change Performance Index (CCPI) plays a significant role in shaping climate policy and influencing public discourse surrounding climate change mitigation and adaptation strategies. It serves as a critical tool for various stakeholders, including governments, researchers, and investors, by providing a systematic assessment of countries' climate performance and their commitments to reducing greenhouse gas emissions (Jones & Youngs, 2024).

4.3.1 Impact on Policy Development

The CCPI has become an essential resource for national and international debates regarding climate policy. By distributing its findings to key media outlets worldwide, such as The Guardian and China Daily, the CCPI frequently triggers discussions in national parliaments and among governments, emphasizing the urgency of climate action.

Additionally, the index aids in aligning climate policy with financial decision-making, enabling investors to translate global climate strategies into actionable insights for portfolio management and investment planning (Jones & Youngs, 2024).

This intersection of climate data and economic strategy underscores the CCPI's relevance in fostering a more integrated approach to climate governance.

4.3.2 Addressing Climate Change Challenges

The CCPI not only evaluates the effectiveness of existing policies but also highlights the need for cohesive action across different levels of government. In countries like Canada, where provinces have the autonomy to design their own climate policies, the CCPI encourages collaboration among sub-national jurisdictions to achieve meaningful emissions reductions (Jones & Youngs, 2024). This highlights the importance of comprehensive analysis and shared knowledge in developing robust climate strategies.

4.3.3 Public Engagement and Awareness

Furthermore, the CCPI enhances public awareness and understanding of climate issues, creating a platform for civic engagement. By revealing disparities in climate performance and prompting discussions about accountability, the index fosters a sense of urgency among citizens and encourages them to advocate for more effective climate action from their leaders (The Climate Change Performance Index 2024: Results).

In this way, the CCPI contributes to building a more informed and engaged public, which is crucial for fostering a collective response to the climate emergency.

4.4. Machine learning model

The Climate Change Performance Index (CCPI) is primarily computed for European Union (EU) countries and the 63 highest greenhouse gas (GHG) emitting nations. As a result, the following countries were not included in the CCPI dataset: Albania, Bosnia and Herzegovina, Iceland (from 2018 onwards), Montenegro, North Macedonia, and Serbia.

Since the methodology for computing the CCPI involves both quantitative computational processes and qualitative assessments by subject matter experts, a machine learning model was developed to predict the missing CCPI values. This model utilizes a dataset composed of indicators categorized into four key domains:

1. GHG Emissions:
   1.1. GHG net emissions/removals by LUCF [Mt CO_2e]
   1.2. Per capita GHG emissions [tons/capita]
   1.3. Total emissions power industry [Mt CO_2e]
   1.4. Total sovereign green bonds issuance [billion US$]
   1.5. PM2.5 air pollution, mean annual exposure [micrograms per cubic meter]
2. Energy Use:
   2.1. Energy Consumption [Mt oil equivalent]
   2.2. Energy Intensity [MJ/$2017 PPP GDP]
   2.3. Electric power consumption [kWh per capita]
   2.4. Population
   2.5. Population growth [annual %]
   2.6. Urban population [% of total population]

2.7. GDP [$]

2.8. GDP per capita [$]

3. Renewable Energy

3.1. Renewable energy [% of total final energy consumption]

3.2. Renewable electricity installed capacity [MW]

4. Climate Policy

4.1. Climate policy score

4.2. Adaptation policies [total]

4.3. Nationally Determined Contributions [total]

4.4. Contributions to International Climate Funds [USD million current]

4.5. Proportion of freshwater key biodiversity areas (KBAs) covered by protected areas [%]

4.6. Proportion of terrestrial key biodiversity areas (KBAs) covered by protected areas [%]

4.7. Forest area [Km^2]

The data sources for these indicators include reputable international organizations, such as:

1. World Bank Group
2. International Energy Agency (IEA)
3. United Nations Climate Change (UNFCCC)
4. Climate Policy Database
5. EU MISSION Adaptation to Climate Change

The indicators incorporated into the dataset span the period from 2014 to 2021, ensuring comprehensive temporal coverage for model training and evaluation.

A critical challenge in developing the predictive model was ensuring a balance between model accuracy and generalizability. Overfitting had to be mitigated to maintain robust predictive performance on unseen data, particularly given the complexities associated with missing data imputation.

Handling missing values effectively is crucial in predictive modelling, as their presence can introduce bias and reduce the reliability of forecasts. Therefore, appropriate imputation techniques were implemented to minimize distortions and maintain the integrity of the dataset (Little & Rubin, 2019).

By leveraging a combination of quantitative environmental indicators and climate policy data, the machine learning model was designed to provide a reliable and systematic approach for estimating CCPI values for countries previously excluded from the official index.

4.4.1 Data Preprocessing and Handling Missing Values

The dataset was initially examined using a Python script to assess its structure and completeness. The analysis highlighted multiple columns with missing values, primarily due to gaps in the original database.

The first imputation attempt used mean imputation, a standard approach when the underlying structure is unknown. However, this method produced inadequate results. In particular:

1. For indicators with few available data points, mean imputation resulted in identical values across all entries, eliminating variance.

2. For indicators with less missing data, the mean imputation failed to differentiate between historically distinct countries, particularly for economic indicators.

Recognizing these issues, a more advanced imputation strategy was necessary.

Previous research categorizes missing data mechanisms into three types:

1. Missing Completely at Random (MCAR): Data missing independently of observed and unobserved factors.
2. Missing at Random (MAR): Data missing based on observed variables.
3. Missing Not at Random (MNAR): Data missing due to unobserved factors (Rubin, 1976).

An in-depth analysis of the dataset suggested that missing values followed an MAR pattern, meaning their presence was systematically related to observed features. Consequently, an imputation method that retained correlations was required.

To address these limitations, K-Nearest Neighbours (KNN) imputation was implemented. This non-parametric method estimates missing values by identifying k-nearest observations based on feature similarity.

In fact, KNN imputation outperforms mean/median imputation when variables exhibit strong interdependencies (Troyanskaya, et al., 2001).

Therefore, since trends varied by country, KNN imputation was applied separately within each country group, ensuring that missing values were inferred from similar observations rather than a global average.

The rationale behind this approach is that countries share internal correlations between economic, environmental, and policy-related indicators. Thus, imputing values at a global level could introduce significant bias.

To further refine the imputation quality, sensitivity testing was conducted by varying k-values between 3 and 15, with k=7 selected as the optimal balance between smoothness and accuracy.

Additionally, handling high-dimensional data in KNN imputation presents computational challenges. To mitigate this:

1. Feature selection was performed before imputation, ensuring that only relevant variables contributed to the estimation process.
2. Min-Max normalization was applied to standardize numerical features, as distance-based algorithms like KNN are highly sensitive to feature magnitudes (Han, Kamber, & Pei, 2011).

Despite the careful imputation process, some columns had an exceptionally high percentage of missing data (>50%), making KNN unsuitable. In such cases:

1. Expectation-Maximization (EM) and Iterative Imputation were tested, as they are more effective in scenarios with extreme missingness (Schafer, 1997).
2. The final dataset combined multiple imputation strategies, ensuring the most accurate reconstructions.

To ensure that the final dataset was free of missing values, a two-step approach was implemented:

1. Step 1: KNN imputation within each country group.

2. Step 2: A second round of global imputation to remove any remaining missing values.

Further, after SHAP-based feature selection, a validation check ensured that the number of features remained consistent throughout the process. Any discrepancies triggered an error, preventing inconsistent feature sets between training and missing data prediction.

Moreover, the 'KEY' column (if present) was preserved and reinserted before saving the final dataset, ensuring that original identifiers remained intact.

For the full implementation of the imputation process, refer to Annex 1.

4.4.2 Feature Engineering and Selection

Feature engineering and selection are critical steps in predictive modelling, directly influencing model performance, interpretability, and generalization ability (Chandrashekar & Sahin, 2014). A structured approach was adopted to preprocess, transform, and select the most relevant features for predicting the Climate Change Performance Index (CCPI).

To optimize the dataset for machine learning algorithms, both numerical and categorical features underwent systematic preprocessing:

1. Numerical Features
    1.1. Imputation: Missing values were addressed using median imputation, which is robust to outliers while preserving the data's central tendency.

1.2. Standardization: StandardScaler was applied, transforming all numerical variables to a zero mean and unit variance, which facilitates model convergence and improves the performance of gradient-based algorithms like XGBoost.

2. Categorical Features

   2.1. Imputation: Missing categorical values were replaced using the most frequent category, maintaining data integrity while minimizing information loss.

   2.2. Encoding: One-hot encoding converted categorical variables into numerical representations, avoiding implicit ordinal relationships.

Another aspect to take into consideration was multicollinearity, that occurs when independent variables are highly correlated, distorting model estimates and reducing robustness. To mitigate this:

1. The Variance Inflation Factor (VIF) was used to systematically evaluate collinearity.
2. Features with a VIF exceeding 10 were iteratively removed, ensuring that the final dataset contained only independent and non-redundant predictors (James, Witten, Hastie, & Tibshirani, 2013).

This process enhanced model interpretability while improving generalization to unseen data.

To identify the most relevant predictors for CCPI modelling, SHapley Additive exPlanations (SHAP) was employed. SHAP provides a mathematically grounded approach to estimating feature importance by computing each variable's marginal contribution to the model's predictions (Lundberg & Lee, 2017).

The selection process involved:

1. Training an initial XGBoost model.
2. Extracting SHAP values.
3. Ranking features based on their importance.
4. Retaining the 15 most impactful predictors, balancing performance and interpretability.

Additionally, Principal Component Analysis (PCA) was tested as a dimensionality reduction technique. However, PCA components lack direct interpretability, making SHAP-based selection preferable for ensuring that retained features align with real-world indicators.

For the full feature engineering and selection implementation, refer to Annex 2.

4.4.3 Model Selection and Hyperparameter Tuning

The selection of an optimal predictive model is a critical step in ensuring accurate, reliable, and efficient predictions for the Climate Change Performance Index (CCPI). This phase involved evaluating multiple machine learning models, optimizing their hyperparameters, and implementing robust validation strategies to ensure generalizability to unseen data.

Machine learning models were initially assessed using a stacking ensemble method, consisting of:

1. XGBoost (Extreme Gradient Boosting)
2. Random Forest
3. Gradient Boosting Machines (GBM)

Stacking ensembles aggregate predictions from multiple models to reduce variance and improve accuracy (Dietterich, 2000). However, despite its theoretical advantages, this ensemble approach proved computationally expensive while delivering only modest improvements in predictive performance.

The R^2 scores of the ensemble method ranged from 0.41 to 0.55, indicating high variance and limited predictive power. Furthermore, ensemble methods require higher computational costs, making them less practical for real-world deployment, where efficiency is a key factor.

Given these findings, the study transitioned to a single XGBoost model, leveraging its advantages in:

1. Handling structured tabular data efficiently.
2. Built-in missing value handling.
3. Regularization mechanisms to prevent overfitting.
4. Parallelized computation, enhancing speed and scalability.

XGBoost has been widely recognized for its superior performance on structured datasets (Chen & Guestrin, 2016) and was deemed the most suitable model for CCPI prediction.

Once XGBoost was selected, its hyperparameters were fine-tuned to maximize predictive accuracy and minimize overfitting.

Traditional grid search and random search methods were found to be computationally inefficient, as they require evaluating many possible hyperparameter combinations exhaustively. Instead, Bayesian optimization using Optuna was implemented.

Bayesian optimization dynamically selects the next hyperparameter set based on past evaluations. Unlike traditional methods, it employs probabilistic models to explore promising parameter regions efficiently and adapt search spaces dynamically, avoiding unnecessary evaluations of poor-performing hyperparameters.

Optuna's Tree-structured Parzen Estimator (TPE) was used to refine key hyperparameters, namely:

1. Number of Trees (n_estimators): Determines the number of boosting rounds.
2. Learning Rate (η): Controls step size in gradient updates. A lower learning rate prevents overfitting but requires more boosting rounds.
3. Maximum Depth (max_depth): Limits tree complexity to avoid overfitting.
4. Subsample Ratio (subsample): Controls the fraction of samples used in each boosting round, reducing variance.
5. Column Subsampling (colsample_bytree): Prevents feature dominance and enhances model generalization.
6. Minimum Child Weight (min_child_weight): Prunes unnecessary splits, ensuring robust trees.
7. Gamma (gamma): Regularizes tree growth, preventing excessive complexity.
8. L1 and L2 Regularization (reg_alpha, reg_lambda): Penalize overly complex models to maintain generalizability.

To ensure that hyperparameter tuning did not overfit to specific data partitions, nested cross-validation was conducted.

Nested cross-validation consists of:

1. An inner loop, where Optuna optimizes hyperparameters on k-fold training subsets.
2. An outer loop, where the model is tested on unseen data to estimate true generalization performance.

This approach ensures that the hyperparameter tuning process remains independent of final evaluation, reducing overfitting risks and improving model robustness.

Once the best hyperparameters were identified, the final XGBoost model was trained on the complete training dataset. To validate its predictive performance, the model was assessed using cross-validation and key evaluation metrics:

1. Cross-Validation (7-Fold): To assess model stability, a final 7-fold cross-validation was applied to evaluate variability in performance across different training/testing splits.
2. Performance Metrics: The following metrics were used to quantify the model's effectiveness:
    - 2.1. R^2 Score: Measures explained variance (closer to 1 is better).
    - 2.2. Mean Absolute Error (MAE): Captures average absolute deviation between predictions and actual values.
    - 2.3. Root Mean Squared Error (RMSE): Penalizes large errors more than MAE.

The final results indicated that XGBoost significantly outperformed the ensemble model, with an improved R^2 score, lower MAE, and reduced RMSE, confirming better generalization performance.

4.4.4 Conclusion

The model selection process systematically evaluated different approaches, demonstrating that a single XGBoost model outperformed the ensemble method in both efficiency and predictive accuracy. By leveraging Bayesian optimization and nested cross-validation, the model achieved an optimized parameter configuration, reducing computational overhead while ensuring robust performance.

This structured approach balances:

1. Computational efficiency
2. Generalization performance
3. Practical deployment considerations

For the full implementation of the model selection and hyperparameter tuning process, refer to Annex 2.

4.4.5 Model Performance Analysis

The final XGBoost model demonstrated strong predictive accuracy, as reflected in the performance metrics.

The R^2 score on the test set (0.7037) indicates that approximately 70.37% of the variance in CCPI values is explained by the model. This represents a substantial improvement over baseline models and confirms that the selected features and hyperparameter tuning effectively captured the underlying patterns in the data.

The Mean Absolute Error (MAE) of 3.6575 suggests that, on average, the model's predictions deviate from the actual CCPI values by approximately 3.66 units. Given the nature of the dataset, this level of error is acceptable for practical applications.

The Root Mean Squared Error (RMSE) of 4.7169 is slightly higher than the MAE, as expected, due to the penalization of larger errors. This suggests that while the model is generally accurate, it may still have some difficulty with extreme values or outliers in the dataset.

The mean cross-validation R^2 score (0.6357 ± 0.0908) highlights the model's stability across different data splits. While slightly lower than the test set R^2, this result indicates that the model maintains a consistent level of performance across multiple validation folds, reducing the likelihood of overfitting to specific data subsets.

The predicted CCPI values span the period 2014–2021. However, to ensure consistency with the structure of other indices used in the analysis, the average CCPI value over this period was computed. This transformation allows for a more standardized comparison across different indicators while preserving the overall trends captured by the model. The results are summarized in Table 4.4.5.1.

Overall, the results confirm that the XGBoost model provides a well-balanced trade-off between accuracy and generalization, making it suitable for real-world deployment in climate change performance assessment. Further improvements could be explored by refining feature engineering techniques or incorporating hybrid modelling approaches, such as deep learning for capturing temporal dependencies.

Table 4.4.5.1: Average results of CCPI Index for the period 2014-2021

Country	CCPI
Albania	60.18*
Austria	48.07
Belgium	52.48
Bosnia and Herzegovina	51.28*
Bulgaria	49.01
Croatia	58.18
Cyprus	53.88
Czech Republic	50.53
Denmark	68.51
Estonia	47.60
Finland	60.37
France	61.61
Germany	57.55
Greece	52.52
Hungary	51.75
Iceland	57.60*
Ireland	52.62
Italy	59.21
Latvia	61.49
Lithuania	62.92
Luxembourg	59.31
Malta	62.64
Montenegro	59.58
Netherlands	53.46
North Macedonia	58.02*
Norway	60.27
Poland	48.73
Portugal	61.03
Romania	57.92
Serbia	59.33*
Slovakia	57.01
Slovenia	51.00

Spain	51.84
Sweden	72..05
Switzerland	62.88
Turkey	43.96
United Kingdom	68.61

Note: The results marked with * have been predicted using the machine learning model

4.5 Criticism and limitations of the CCPI Index

The Climate Change Performance Index (CCPI) is a widely recognized tool for assessing national climate policies and performance. However, it has faced significant methodological and conceptual criticisms that highlight limitations in its ability to provide a comprehensive and equitable evaluation of climate action. These critiques do not diminish the value of the CCPI but underscore areas where the index could be refined or complemented with additional methodologies to enhance its accuracy, fairness, and policy relevance.

A central critique of the CCPI lies in its heavy reliance on quantitative data, which constitutes approximately 80% of its assessment. While metrics such as greenhouse gas (GHG) emissions, renewable energy consumption, and energy use provide objective measures of climate performance, they may fail to capture the nuanced and long-term impacts of climate policies. For instance, the effects of policies like carbon pricing or renewable energy subsidies often take years to manifest in tangible outcomes, such as reduced emissions or increased renewable energy capacity. This time lag creates a disconnect between real-time policy developments and the data used by the CCPI, limiting its ability to reflect immediate progress or setbacks in climate action.

Additionally, the CCPI incorporates qualitative expert assessments to complement its quantitative data. While expert opinions provide valuable context, they introduce an element of subjectivity that can lead to variability in scores. Differing interpretations of policy effectiveness among experts may affect the comparability of scores across countries, undermining the index's objectivity and reliability.

The CCPI also faces challenges in comparing countries with vastly different economic, social, and political contexts. For example, high-income countries with historically large carbon footprints may receive more favourable evaluations than developing nations that are still prioritizing economic growth over environmental sustainability. This disparity creates an uneven playing field, where countries with fewer resources or less developed infrastructure are penalized for their inability to implement ambitious climate policies. Moreover, the CCPI does not consistently account for historical responsibilities for climate change. Countries that have contributed significantly to global emissions in the past may receive similar scores to those with lower historical emissions, even if the latter are making more substantial efforts to reduce their carbon footprint. This lack of historical context can undermine the fairness of the index and its ability to incentivize equitable climate action.

Another critical issue is the CCPI's reliance on data from established sources such as the International Energy Agency (IEA) and national GHG inventories, which are susceptible to greenwashing and reporting inconsistencies. Greenwashing, where countries or companies exaggerate their sustainability efforts, can distort the accuracy of the data and undermine the credibility of the index. For example, a country may report high levels of renewable energy consumption without accounting for the environmental impact of its fossil fuel exports or the carbon footprint of its supply chains. To address these issues, future iterations of the CCPI could incorporate independent verification mechanisms, such as third-party audits or satellite data on emissions, to ensure the reliability and transparency of the data. Additionally, the index could introduce penalties for greenwashing or incentives for countries that demonstrate genuine commitment to sustainability through verifiable actions.

While the CCPI covers key areas such as GHG emissions, renewable energy, and energy use, it does not fully capture the broader dimensions of sustainability. Critical factors such as social equity, biodiversity conservation, and community resilience to climate change are often overlooked in favour of more easily quantifiable metrics. This narrow focus limits the index's ability to provide a holistic assessment of climate performance and may inadvertently encourage policies that prioritize measurable outcomes over broader sustainability goals. Future research could explore the integration of additional indicators into the CCPI, such as measures of social inclusion, environmental justice, and ecosystem health. This would provide a more comprehensive evaluation of climate performance and encourage countries to adopt policies that address the interconnected challenges of climate change and sustainable development.

Finally, the CCPI has been criticized for its limited ability to drive meaningful policy change, particularly in countries that consistently rank poorly. While the index serves as a valuable tool for raising awareness and fostering public debate, its impact on actual policy decisions is often indirect and uneven. For example, countries with low scores may face public pressure to improve their climate performance, but this pressure does not always translate into concrete policy actions. To enhance the policy relevance of the CCPI, future research could explore ways to strengthen its connection to national and international climate governance. This could involve developing targeted recommendations for policymakers based on CCPI findings or creating platforms for dialogue between governments, civil society, and the private sector to translate index rankings into actionable strategies.

In conclusion, the CCPI is a valuable tool for assessing national climate performance, but its methodological and conceptual limitations highlight the need for ongoing refinement. Addressing issues such as data reliability, comparability across countries, and the risk of greenwashing will be critical to enhancing the index's accuracy and fairness. Additionally, expanding the scope of indicators to include broader dimensions of sustainability and strengthening the link between the CCPI and policy action will ensure that the index remains a relevant and effective tool for driving global climate progress. Users of the CCPI must engage critically with its findings, recognizing both its strengths and its limitations in shaping the future of climate policy.

5. The Statistical Impact of ESG and CCPI on Green Energy and Venture Capital Investment

The transition toward a sustainable economy is increasingly shaped by financial incentives, regulatory policies, and investment strategies that align with environmental and social objectives. In this context, the role of sustainability indexes—such as the Environmental, Social, and Governance (ESG) Index and the Climate Change Performance Index (CCPI)—has gained prominence as tools for assessing and guiding investment decisions. These indexes not only serve as benchmarks for evaluating the sustainability performance of corporations and nations but also influence the allocation of capital toward green energy and other environmentally friendly initiatives. However, the extent to which these indexes drive venture capital (VC) investment in green energy firms remains an area requiring further exploration.

Prior research has established a strong link between ESG performance and long-term institutional investment. Studies by (Clark, Feiner, & Viehs, 2015) and (Gillan, Koch, & Starks, 2021) have shown that firms with high ESG ratings tend to attract more investment due to their perceived lower risk and alignment with evolving regulatory standards. Similarly, the CCPI has been recognized as a valuable tool for assessing national climate policies and their effectiveness in reducing greenhouse gas emissions (Jones & Youngs, 2024). However, these studies have primarily focused on institutional investors, who often prioritize long-term sustainability over short-term returns. In contrast, VC firms operate within a distinct investment framework, prioritizing technological advancements, market expansion potential, and high-risk, high-reward opportunities over strict adherence to sustainability metrics. This divergence in investment priorities raises several critical empirical questions:

1. To what extent do a country's ESG and CCPI scores affect the likelihood of securing VC investment in green energy firms?
2. How does sustainability performance impact the total amount of VC funding received?
3. What economic factors—such as GDP per capita, R&D spending, and foreign direct investment (FDI)—moderate the relationship between sustainability indicators and VC investment?
4. Does the geographical distribution of green energy firms exhibit a pattern influenced by sustainability metrics, or is it primarily shaped by economic clustering effects?

To address these questions, this thesis employs a combination of econometric models, including logistic regression (to evaluate the probability of securing investment), Ordinary Least Squares (OLS) regression (to assess total investment amounts), and spatial analysis (to investigate the geographical concentration of firms). By integrating these analytical approaches, the study aims to clarify whether national sustainability policies serve as meaningful investment signals for venture capitalists or whether economic fundamentals exert a more dominant influence on investment decisions.

This research builds on and extends the existing literature in several ways. First, it shifts the focus from institutional investors to venture capital, a critical but understudied segment of the investment landscape. Second, it explores the interplay between sustainability indexes and economic factors, providing a more nuanced understanding of the drivers of green energy investment. Finally, it contributes to the growing body of research on the geographical distribution of green energy firms, offering insights into how regional economic and policy conditions influence the clustering of VC-backed startups.

By integrating multiple perspectives, this research aims to provide insights into the potential drivers of sustainable investment and the role of policy and financial markets in supporting the green energy transition. The findings have important implications for policymakers, investors, and green energy firms, offering guidance on how to design more effective incentives for sustainable investment and foster the growth of the cleantech sector.

5.1 Dataset Preprocessing

The dataset used in this analysis combines data from the European Investment Fund (EIF) database and the World Bank, covering venture capital (VC) investments, sustainability indexes (ESG and CCPI), and economic indicators. The preprocessing steps were critical to ensuring the dataset's robustness and suitability for econometric analysis. Below is a detailed overview of the key steps taken to prepare the data for statistical modelling.

The first stage of preprocessing involved merging four primary datasets: three from the EIF database containing VC investment data and one compiled by the author, which included sustainability index computations and economic indicators from the World Bank. Given the diverse sources of data, it was essential to standardize variables and resolve inconsistencies to construct a robust dataset. One of the initial tasks was harmonizing country names by converting them to lowercase and removing spaces to ensure consistent merging. This step was necessary because discrepancies in country labels, such as capitalization differences or extra spaces, could prevent accurate data integration. Once the datasets were merged, key variables were aggregated at the country level. Specifically, the total amount of VC investment raised was summed across multiple years for each country, creating a consolidated metric that reflects cumulative investment inflows. This ensured that the dependent variable in the regression models accurately represented national-level investment trends rather than fragmented yearly observations.

Missing values were addressed using linear interpolation and historical data from the World Bank for economic indicators such as GDP per capita and R&D spending. These methods are widely used to approximate missing values while preserving underlying statistical relationships (Little & Rubin, 2019). Outlier detection and treatment were also critical, as the distribution of VC investments exhibited significant skewness, with certain countries receiving disproportionately high funding. To mitigate the impact of extreme values, the winsorization technique was applied to the investment variable, capping extreme values at the 1st and 99th percentiles. This step ensured that the dataset was more suitable for regression analysis, reducing the risk of distorted results.

To ensure the reliability of the regression models, several diagnostic tests were conducted. Breusch-Pagan tests were performed to detect heteroskedasticity, and robust standard errors were used in the analysis to account for potential heteroskedasticity. This approach corrected for biases arising from non-constant residual variance, ensuring more accurate and reliable results. Additionally, Variance Inflation Factor (VIF) tests were conducted to confirm that there was no severe multicollinearity among the predictor variables. Multicollinearity, which occurs when independent variables are highly correlated, can distort model estimates and reduce robustness. The VIF tests confirmed that the predictor variables were sufficiently independent, enhancing the validity of the regression results.

Interaction terms were created to explore potential moderating effects. For example, an interaction term between ESG and GDP per capita was included to investigate whether the effect of sustainability performance on VC investment varies depending on a country's level of economic development. Similarly, an interaction term between CCPI and ease of doing business was constructed to explore whether business-friendly policies enhanced the impact of sustainability on investment decisions. Categorical variables, such as the binary classification of countries as VC-backed or not, were recoded to ensure compatibility with logistic regression analysis. After these transformations, the dataset was validated to confirm that no observations were unintentionally dropped, and descriptive statistics were generated to verify data integrity before proceeding with econometric analyses.

This rigorous preprocessing ensured that the dataset was structured to facilitate meaningful statistical analysis while minimizing biases arising from inconsistencies, missing data, or outliers. By systematically addressing these challenges, the final dataset provided a robust foundation for evaluating the impact of ESG and CCPI on green energy investment patterns. The subsequent sections will build on this foundation to explore the role of sustainability indexes and economic drivers in shaping investment decisions.

5.2 The Probability of Securing Green VC Investment

The relationship between sustainability performance and venture capital (VC) investment in green energy companies is a subject of growing interest, given the increasing role of financial incentives and policy-driven sustainability indexes. As previously discussed (Clark, Feiner, & Viehs, 2015), strong ESG performance is linked to long-term investor confidence.. This analysis seeks to determine whether ESG and CCPI scores at the national level influence the probability of securing green VC investment and the extent to which economic factors, such as GDP per capita, R&D spending, and foreign direct investment (FDI), moderate this relationship.

Given the risk-oriented nature of VC firms, which prioritize scalability and technological potential over strict adherence to sustainability criteria (Gompers & Lerner, 2001), this study investigates whether these indexes serve as meaningful investment signals or if broader economic fundamentals play a more decisive role. The analysis provides empirical insights into whether policy-driven sustainability performance translates into tangible investment advantages in the green energy sector.

5.2.1 Methodological Choice

To assess the probability of securing VC investment, a logistic regression model was employed, as it is well-suited for binary outcomes. The dependent variable is whether a country receives green VC investment (1) or not (0), while the independent variables include

ESG and CCPI scores, GDP per capita, R&D spending as a percentage of GDP, and FDI inflows. The estimated model follows the standard logit formulation:

$$P(VC\_backed = 1) = \frac{e^{\beta_0 + \beta_1 ESG + \beta_2 CCPI + \beta_3 GDP\_per\_capita + \beta_4 R\&D\_spending + \beta_5 FDI + \varepsilon}}{1 + e^{\beta_0 + \beta_1 ESG + \beta_2 CCPI + \beta_3 GDP\_per\_capita + \beta_4 R\&D\_spending + \beta_5 FDI + \varepsilon}}$$

To ensure the validity of the regression models, a heteroskedasticity test was conducted. Heteroskedasticity, which occurs when the variance of residuals is not constant across observations, can distort standard errors and lead to unreliable inferences. The Breusch-Pagan test was used to detect heteroskedasticity, with the following formulation:

$$Var(\epsilon_i) = \sigma^2(Z_i\gamma)$$

where Z_i represents the set of independent variables (ESG, CCPI, GDP per capita, R&D spending, and FDI). The null hypothesis (H_0) states that variance is homoscedastic (γ=0), while the alternative hypothesis (H_1) suggests heteroskedasticity (γ≠0). If the test statistic is significant, robust standard errors are applied to correct for heteroskedasticity.

5.2.2 Results and Interpretation

The logit regression results (Table 5.2.2.1) indicate that ESG (β = 463.0, $p > 0.1$) and CCPI (β = 1.479, $p > 0.1$) do not have a statistically significant effect on the probability of securing VC investment, as both exceed the conventional significance threshold ($p < 0.1$). These findings suggest that sustainability indicators, while relevant in institutional investment contexts, may not strongly influence VC decision-making

However, economic factors emerge as stronger determinants: GDP per capita negatively impacts the probability of investment ($\beta=-0.000770$, $p<0.05$), suggesting that VC investors may prefer emerging markets where capital efficiency and return potential are higher.

Conversely, R&D spending as a percentage of GDP positively correlates with investment probability ($\beta=23.82$, $p<0.1$), highlighting the importance of innovation capacity in attracting green VC funding. Countries with higher R&D spending are likely to have stronger innovation ecosystems, which are critical for the development and scaling of green energy technologies.

The R-squared value of 0.346 suggests a moderate explanatory power for the model.

Additional insights from the heteroskedasticity test (table 5.2.2.2) confirm that FDI inflows positively influence VC investment ($\beta=6.46e-10$, $p<0.1$), reinforcing the idea that cross-border capital accessibility is crucial for green venture financing.

The impact of business-specific CCPI metrics is negligible ($\beta=-0.0480$, $p>0.1$), suggesting that national-level sustainability policies, while important for broader economic stability, do not directly affect VC decision-making. The R-squared value of 0.425 in the CCPI analysis provides a stronger explanatory framework compared to other models, further supporting the dominant role of economic variables.

These results suggest that while ESG and CCPI scores do not significantly affect the probability of securing investment, economic fundamentals such as R&D spending and foreign investment flows play a critical role in shaping VC investment dynamics. The findings align with prior research emphasizing that sustainability metrics alone are insufficient drivers for venture capital, as investors prioritize economic efficiency and technological innovation over policy-based sustainability performance (Cumming, Henriques, & Sadorsky, 2016).

Table 5.2.2.1: Logit Regression: ESG, CCPI, and VC Probability

VARIABLES	(1) (sum) VC_backed	(2) (sum) VC_backed	(3) (sum) VC_backed
(mean) ESG	463.0 (1,421)	463.0 (1,421)	18.65 (12.83)
(mean) CCPI	1.479 (15.94)	1.479 (15.94)	0.0906 (0.104)
ESG_CCPI_interaction	0.00214 (25.46)	0.00214 (25.46)	
(mean) GDP_per_capita	-0.000770** (0.000341)	-0.000770** (0.000341)	-8.63e-05* (4.66e-05)
(mean) R_D_spendingGDP	23.82* (12.39)	23.82* (12.39)	3.694*** (1.413)
Constant	-356.9 (899.3)	-356.9 (899.3)	-19.00* (11.30)
Observations	37	37	37
R-squared	0.346	0.346	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5.2.2.2: Heteroskedasticity Test

VARIABLES	(1) (sum) VC_backed	(2) (sum) VC_backed	(3) (sum) VC_backed	(4) (sum) VC_backed	(5) (sum) VC_backed
(mean) ESG	414.5	516.7	516.7	516.7	516.7
	(292.9)	(311.9)	(311.9)	(311.9)	(311.9)
(mean) CCPI	1.414	5.480	5.480	5.480	5.480
	(1.557)	(3.520)	(3.520)	(3.520)	(3.520)
(mean) GDP_per_capita	-0.000932***	-0.000816**	-0.000816**	-0.000816**	-0.000816**
	(0.000305)	(0.000312)	(0.000312)	(0.000312)	(0.000312)
(mean) R_D_spendingGDP	33.73**	30.32**	30.32**	30.32**	30.32**
	(14.61)	(13.39)	(13.39)	(13.39)	(13.39)
(mean) FDI_Inflows	6.46e-10*				
	(3.81e-10)				
(mean)	0.386				
Renewable_Energy_Share	(0.474)				
(mean)	-3.314				
Ease_of_Doing_Business	(2.056)				
(mean)	0.414				
Ease_start_business	(1.312)				
CCPI_Business		-0.0480	-0.0480	-0.0480	-0.0480
		(0.0335)	(0.0335)	(0.0335)	(0.0335)
Constant	-138.9	-420.0*	-420.0*	-420.0*	-420.0*
	(146.4)	(234.9)	(234.9)	(234.9)	(234.9)
Observations	37	37	37	37	37
R-squared	0.425	0.380	0.380	0.380	0.380

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

5.3 The Total Amount of VC Investment in Green Energy Firms

Beyond the probability of securing venture capital (VC) investment, it is equally important to understand the determinants of the total amount of investment received by green energy firms. As previously mentioned by (Clark, Feiner, & Viehs, 2015), strong ESG performance is linked to long-term investor confidence. On the other hand, its effect on the magnitude of VC investment remains less explored.

Unlike institutional investors, VC firms focus on high-growth opportunities, risk-adjusted returns, and market scalability rather than strict adherence to sustainability scores (Gompers & Lerner, 2001). This raises a crucial empirical question: Do ESG and Climate Change Performance Index (CCPI) scores translate into larger VC funding for green energy firms, or do economic fundamentals such as GDP per capita, R&D spending, and financial market conditions play a more significant role?

This section investigates how sustainability performance interacts with broader economic conditions to shape the scale of VC investment in green energy firms.

5.3.1 Methodological Choice

To examine the relationship between sustainability indicators and the total amount of VC investment in green energy firms, an Ordinary Least Squares (OLS) regression model was employed. The OLS model allows for the estimation of the effect of ESG and CCPI scores on continuous investment amounts, making it suitable for analysing the scale of funding received. The estimated equation is as follows:

$$Amount\_Raised = \alpha + \beta_1\ ESG + \beta_2\ CCPI + \beta_3\ GDP\_per\_capita + \beta_4\ R\&D\_spending + \beta_5 FDI + \varepsilon$$

A log transformation of investment amounts was also applied to account for skewness in funding distribution and to ensure robustness of the model. The log-transformed model takes the following form:

$$log\ (Amount\_Raised) = \alpha + \beta_1\ ESG + \beta_2\ CCPI + \beta_3\ GDP\_per\_capita + \beta_5\ R\&D\_spending + \varepsilon$$

Additionally, a heteroskedasticity test was performed to detect variance inconsistencies in the error terms. The Breusch-Pagan test was used to this purpose, with the following formulation:

$$Var(\epsilon_i) = \sigma^2(Z_i\gamma)$$

5.3.2 Results and Interpretation

The OLS regression results, summarised in table 5.3.2.1, reveal that higher ESG scores correlate with larger investment amounts (β=1.075e+06, $p<0.1$), while CCPI remains insignificant. This suggests that, while ESG scores do not necessarily increase the likelihood of investment, they do appear to influence the size of funding received. The R-squared value of 0.248 suggests that ESG and CCPI scores together explain approximately 24.8% of the variance in investment amounts. While this represents a moderate explanatory power, it implies that the majority of investment decisions are likely influenced by additional economic variables such as market conditions, financial incentives, and technological potential.

Additional results from the log transformation (table 5.3.2.2) further reinforce the relationship between ESG scores and VC investment. Here, ESG has a statistically significant positive effect on the total amount of VC-backed investment (β=559.9, $p<0.01$), with an R-squared value of 0.258, indicating that ESG explains approximately 25.8% of the variance in VC funding levels. Again, CCPI remains statistically insignificant.

A more refined analysis, displayed in table 5.3.2.3, confirms these findings while incorporating the log-transformed investment amounts. The regression results suggest that ESG has a positive effect on the log-transformed investment amounts (β=30.54, $p<0.1$), reinforcing that sustainability performance correlates with greater VC investment.

GDP per capita shows a negative but insignificant effect on investment (β=-1.65e-05, $p>0.1$), suggesting that economic development levels do not strongly determine VC funding size.

Meanwhile, R&D spending emerges as a significant positive predictor (β=3.039, $p<0.05$), highlighting the importance of national innovation ecosystems in attracting larger VC investments.

On the other hand, CCPI remains statistically insignificant across all specifications.

Finally, the R-squared value of 0.435 in this model suggests stronger explanatory power compared to the previous regressions.

These findings suggest that ESG scores positively influence the scale of VC investment in green energy firms, whereas CCPI scores do not appear to be a decisive factor in investment amounts. The results align with previous research (Cumming, Henriques, & Sadorsky, 2016). Additionally, the positive effect of R&D spending suggests that technological innovation is a critical driver of green VC investment, reinforcing the role of national policies that support research and development in sustainable technologies.

In conclusion, while ESG performance is an important factor in attracting larger VC investments, economic fundamentals—particularly R&D spending—are key determinants of funding magnitude. These insights are valuable for policymakers seeking to enhance green energy financing by creating supportive environments for technological innovation and sustainability-oriented investments.

Table 5.3.2.1: VC Investment & Sustainability

VARIABLES	(1) (sum) amount_rais_y
(mean) ESG	1.075e+06* (580,197)
(mean) CCPI	1,129 (2,897)
(mean) GDP_per_capita	-1.005 (0.693)
(mean) R_D_spendingGDP	28,665 (19,851)
Constant	-704,080 (434,554)
Observations	35
R-squared	0.248

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5.3.2.2: VC Investment Analysis

VARIABLES	(1) sum_VC_backed
ESG	559.9*** (170.1)
CCPI	0.961 (1.066)
Constant	-379.1*** (138.9)
Observations	37
R-squared	0.258

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5.3.2.3: Log transformation of Investment amount

VARIABLES	(1) (sum) VC _backed	(2) (sum) VC _backed	(3) (sum) VC _backed	(4) (sum) VC _backed	(5) (sum) VC _backed	(6) (sum) amount _rais_y	(7) (sum) amount _rais_y	(8) (sum) VC _backed	(9) Log _amount _raised
(mean) ESG	291.7 (282.3)	897.0 (656.0)	897.0 (656.0)	18.65 (12.83)	10.39 (7.551)	55,801 (166,497)	641,846 (1.831e+06)	463.1* (270.7)	30.54 (21.74)
(mean) CCPI	1.003 (1.379)	1.438 (1.266)	1.438 (1.266)	0.0906 (0.104)	0.0463 (0.0596)	359.0 (834.3)	-1,377 (9,173)	1.481 (1.578)	-0.0697 (0.106)
ESG_GDP _interaction	0.0110 (0.00899)								
(mean) GDP_per _capita	-0.00792 (0.00585)	-0.00102** (0.000505)	-0.00102** (0.000505)	-8.63e-05* (4.66e-05)	-4.71e-05** (2.30e-05)	-0.0849 (0.247)	-0.859 (2.713)	-0.000770* (0.000397)	-1.65e-05 (4.00e-05)
(mean) R_D _spendingGDP	20.25* (11.65)	7.137 (23.88)	7.137 (23.88)	3.694*** (1.413)	2.108*** (0.726)	16,671* (8,837)	38,318 (97,168)	23.82* (13.38)	3.039** (1.214)
Constant	-218.8 (214.8)	-596.8 (414.0)	-596.8 (414.0)	-19.00* (11.30)	-10.43* (6.209)	-68,230 (105,266)	-290,379 (1.157e+06)	-356.9* (206.4)	-13.76 (14.38)
Observations	37	37	37	37	37	35	35	37	35
R-squared	0.362	0.301	0.301					0.346	0.435

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.4 Economic Drivers of VC Investment

While sustainability indicators such as ESG and CCPI scores may influence the likelihood and magnitude of VC investment in green energy firms, the broader economic environment plays a fundamental role in shaping investment decisions. Venture capital firms operate in high-risk, high-reward environments and rely on various macroeconomic factors, such as GDP per capita, research and development (R&D) spending, foreign direct investment (FDI), and ease of doing business, to determine where to allocate their resources. Unlike institutional investors, who may prioritize sustainability compliance, VC firms assess a country's economic competitiveness and market potential when making investment decisions (Gompers & Lerner, 2001). This section examines the key economic drivers of green VC investment, evaluating whether market conditions and innovation potential outweigh sustainability considerations. The goal is to identify whether strong economic fundamentals attract larger VC inflows to green energy companies and to determine the relative importance of different economic variables in shaping these investment patterns.

5.4.1 Methodological Choice

To analyse the impact of economic factors on VC investment, a fixed-effects panel regression model was employed. This method is particularly useful for controlling for unobserved country-specific characteristics that could influence investment patterns over time. The estimated model takes the following form:

$$VC\_backed = \alpha + \beta_1\ GDP\_per\_capita + \beta_2\ R\&D\_spending + \beta_3\ FDI + \beta_4\ Ease\_of\_Doing\_Business + \beta_5\ Patents + \varepsilon$$

A log transformation was also applied to VC-backed investment values to normalize the distribution and account for extreme variations in funding amounts. The log-transformed regression model follows the structure:

$$log(VC\_backed) = \alpha + \beta_1\ GDP\_per\_capita + \beta_2\ R\&D\_spending + \beta_3\ FDI + \beta_4\ Ease\_of\_Doing\_Business + \beta_5\ Patents + \varepsilon$$

To ensure the robustness of the results, a heteroskedasticity test was conducted using the Breusch-Pagan test:

$$Var(\epsilon_i) = \sigma^2(Z_i\gamma)$$

5.4.2 Results and Interpretation

The regression results from table 5.4.2.1 show that GDP per capita ($\beta = -0.000932$, $p < 0.01$) has a statistically significant negative effect on VC investment, suggesting that venture capitalists may prefer investing in lower-income markets where operational costs are lower and potential returns are higher. This aligns with research indicating that high-growth startups often emerge in emerging economies with untapped market potential. The significance level ($p < 0.01$) indicates strong confidence in this result, meaning that the relationship is unlikely to be due to chance.

R&D spending emerges as a positive and significant predictor at the 5% level, suggesting a moderately strong relationship with investment levels (β=33.73, $p<0.05$), reinforcing the idea that countries with strong innovation ecosystems attract greater VC inflows. Foreign direct investment (FDI) also has a positive effect on green VC investment (β=6.46e-10, $p<0.1$), suggesting that cross-border capital flows play a critical role in financing sustainable ventures. However, ease of doing business and renewable energy share do not exhibit statistically significant effects in this model.

The R-squared value of 0.425 suggests that economic variables explain a considerable portion of VC investment decisions.

Further evidence from tables 5.4.2.2 and 5.4.2.3 supports these findings. GDP per capita continues to show a negative and significant association with VC-backed investment (β=-0.000814, $p<0.05$), while R&D spending remains a strong positive determinant (β=30.07, $p<0.05$).

Ease of doing business has a negative but weakly significant effect at the 10% level, implying a less conclusive but still notable influence on VC investment patterns (β=-3.188, $p<0.1$), suggesting that regulatory efficiency may not be a decisive factor for VC investors compared to other economic drivers.

Patent applications per resident, included as a proxy for technological innovation, are found to be positively correlated with VC investment (β=0.00288, $p<0.1$), indicating that strong intellectual property frameworks can contribute to attracting venture capital. The overall R-squared value of 0.484 indicates improved explanatory power compared to earlier models, further validating the role of economic fundamentals in shaping VC investment patterns.

These results suggest that economic conditions play a far greater role in determining VC investment flows than sustainability metrics alone. The negative relationship between GDP per capita and VC investment highlights the preference for markets with higher potential for growth and lower operational costs. The strong impact of R&D spending and patent activity reinforces the importance of innovation capacity in attracting venture capital. Furthermore, the significance of FDI inflows suggests that VC investment is influenced by broader financial market trends and international investment flows.

Overall, these findings align with previous literature, which argues that venture capitalists prioritize technological potential, economic dynamism, and market growth over policy-driven sustainability indicators (Cumming, Henriques, & Sadorsky, 2016). While ESG factors may shape investment preferences to some extent, the data strongly suggests that innovation capacity and foreign investment opportunities are the primary economic drivers of green VC investment.

Table 5.4.2.1: Fixed Effects Panel Regression on VC

VARIABLES	(1) (sum) VC_backed	(2) (sum) VC_backed	(3) (sum) d_cleantech	(4) log_VC_backed	(5) (sum) VC_backed	(6) (sum) VC_backed
(mean) ESG	414.5	414.5		13.48*	516.7	0
	(292.9)	(292.9)		(7.667)	(311.9)	(0)
(mean) CCPI	1.414	1.414		0.0220	5.480	0
	(1.557)	(1.557)		(0.0333)	(3.520)	(0)
(mean) GDP_per_capita	-0.000932***	-0.000932***	-0.00175	-3.12e-05***	-0.000816**	0
	(0.000305)	(0.000305)	(0.00886)	(1.12e-05)	(0.000312)	(0)
(mean) R_D_spendingGDP	33.73**	33.73**	-34.68	1.302***	30.32**	0
	(14.61)	(14.61)	(223.3)	(0.372)	(13.39)	(0)
(mean) FDI_Inflows	6.46e-10*	6.46e-10*				
	(3.81e-10)	(3.81e-10)				
(mean) Renewable_Energy_Share	0.386	0.386				
	(0.474)	(0.474)				
(mean) Ease_of_Doing_Business	-3.314	-3.314				
	(2.056)	(2.056)				
(mean) Ease_start_business	0.414	0.414				
	(1.312)	(1.312)				
mean_ESG			7,107			
			(5,036)			
mean_CCPI			-5.496			
			(14.52)			
CCPI_Business					-0.0480	
					(0.0335)	
Constant	-138.9	-138.9	-3,593	-8.920*	-420.0*	31.38
	(146.4)	(146.4)	(2,688)	(4.571)	(234.9)	(0)
Observations	37	37	37	37	37	37
R-squared	0.425	0.425	0.158	0.602	0.380	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5.4.2.2: Economic Competitiveness & Cleantech – part 1

VARIABLES	(1) (sum) d _cleantech	(2) (sum) VC _backed	(3) (sum) VC _backed	(4) log_VC _backed	(5) (sum) d _cleantech	(6) (sum) d _cleantech
(mean) ESG		281.9	826.2	13.48*	6,892	6,892
		(248.2)	(500.4)	(7.667)	(6,918)	(6,918)
(mean) CCPI		1.212	1.801	0.0220	2.934	2.934
		(1.343)	(1.411)	(0.0333)	(21.73)	(21.73)
(mean) GDP_per_capita		-0.000621**	-0.000814**	-3.12e-05***	-0.00119	-0.00119
		(0.000279)	(0.000314)	(1.12e-05)	(0.0107)	(0.0107)
(mean) R_D_spendingGDP		26.49**	30.07**	1.302***	41.89	41.89
		(12.68)	(13.16)	(0.372)	(262.5)	(262.5)
(mean) FDI_Inflows		8.27e-11				
		(3.23e-10)				
(mean) Renewable_Energy_Share		0.130			-5.532	-5.532
		(0.531)			(9.526)	(9.526)
(mean) Ease_of_Doing_Business	11.51	-3.188*			-18.73	-18.73
	(22.63)	(1.728)			(26.54)	(26.54)
(mean) Ease_start_business	4.449	1.417			-7.513	-7.513
	(17.72)	(1.718)			(16.20)	(16.20)
(mean) Patents_applications_residents		0.00298				
		(0.00178)				
ESG_Ease			-4.204			
			(3.079)			
CCPI_GDP						
Constant	-763.8	-133.6	-411.3*	-8.920*	-1,852	-1,852
	(1,471)	(159.8)	(231.6)	(4.571)	(3,418)	(3,418)
Observations	37	37	37	37	37	37
R-squared	0.008	0.484	0.376	0.602	0.186	0.186

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5.4.2.3: Economic Competitiveness & Cleantech – part 2

VARIABLES	(7) log_d_cleantech	(8) log_d_cleantech	(9) (sum) VC_backed	(10) (sum) d_cleantech	(11) (sum) d_cleantech	(12) (sum) d_cleantech	(13) (sum) VC_backed	(14) (sum) d_cleantech
(mean) ESG	20.00**	20.00**	459.3*	-69.79	-69.79	-69.79	286.8	
	(9.624)	(9.624)	(262.3)	(516.9)	(516.9)	(516.9)	(196.5)	
(mean) CCPI	-0.0201	-0.0201	2.727	4.092	4.092	4.092	1.400	
	(0.0354)	(0.0354)	(2.106)	(5.911)	(5.911)	(5.911)	(1.227)	
(mean) GDP_per_capita	-2.13e-05	-2.13e-05	0.00147	-0.00108**	-0.00108**	-0.00108**	-0.000569**	
	(1.40e-05)	(1.40e-05)	(0.00338)	(0.000464)	(0.000464)	(0.000464)	(0.000258)	
(mean) R_D_spendingGDP	0.796*	0.796*	24.16*	116.8	116.8	116.8	19.32*	
	(0.422)	(0.422)	(12.72)	(105.6)	(105.6)	(105.6)	(11.29)	
(mean) FDI_Inflows								
(mean) Renewable_Energy_Share								
(mean) Ease_of_Doing_Business								11.51
								(22.63)
(mean) Ease_start_business								4.449
								(17.72)
(mean) Patents_applications_residents							0.00288*	
							(0.00149)	
ESG_Ease								
CCPI_GDP			-3.82e-05					
			(5.86e-05)					
Constant	-7.265	-7.265	-426.2*	-176.0	-176.0	-176.0	-249.0*	-763.8
	(5.718)	(5.718)	(235.3)	(320.2)	(320.2)	(320.2)	(145.0)	(1,471)
Observations	37	37	37	18	18	18	37	37
R-squared	0.437	0.437	0.351	0.136	0.136	0.136	0.445	0.008

5.5 Spatial Distribution of Green Energy Firms

The geographical distribution of green energy firms is an important dimension of the sustainability investment landscape, as it reflects how different regions attract and support clean energy startups. While previous analyses have examined the factors influencing the probability of securing venture capital (VC) and the total amount of investment, this section focuses on the spatial dynamics of green energy firms. Venture capital investments tend to cluster in certain regions due to economic, policy, and innovation-related factors, but it remains unclear whether sustainability indicators, such as ESG and CCPI scores, significantly contribute to the concentration of green energy firms.

This analysis investigates whether green energy firms are more likely to cluster in countries with high sustainability performance or whether their distribution follows broader economic and technological patterns. Understanding the spatial aspect of investment can provide insights into whether VC-backed green firms thrive in sustainability-driven markets or whether other conditions, such as innovation ecosystems and business climate, play a more defining role in determining their location.

5.5.1 Methodological Choice

To explore the spatial distribution of green energy firms, an Ordinary Least Squares (OLS) regression model was used to examine the relationship between sustainability indicators and the number of VC-backed firms per country. The model is specified as follows:

$$VC\_backed = \alpha + \beta_1\, ESG\ + \beta_2\, CCPI + \beta_3\, GDP\_per\_capita + \beta_4\, R\&D\_spending + \varepsilon$$

Additionally, a heteroskedasticity test was conducted to ensure robustness in the regression results. The Breusch-Pagan test, formulated as:

$$Var(\epsilon_i) = \sigma^2(Z_i\gamma)$$

5.5.2 Results and Interpretation

The regression results from table 5.5.2.1 reveal that ESG scores significantly influence the geographical distribution of VC-backed green energy firms (β=559.9, $p<0.01$), suggesting that stronger sustainability frameworks correlate with a higher concentration of firms. However, CCPI scores are not statistically significant (β=0.961,$p>0.1$), indicating that national climate policy rankings alone do not explain firm distribution patterns. The R-squared value of 0.258 suggests that while ESG plays a role, other economic and investment-related factors likely contribute to firm clustering.

Further insights from table 5.5.2.2 suggest that GDP per capita has a negative effect on the number of green energy firms (β=-0.00102, $p<0.05$), implying that firms tend to be located in regions with lower economic development rather than in high-income countries. This could be due to the cost advantages and emerging market potential in these regions. On the other hand, R&D spending ($\beta = 23.82$, $p < 0.1$) shows a positive relationship with green energy firm clustering, suggesting that higher national investment in research and development fosters innovation-driven startups, which are attractive to VC investors. This effect implies that for every 1% increase in R&D spending relative to GDP, the number of VC-backed green energy firms is expected to rise by approximately 23.82 firms, on average. However, the significance level ($p < 0.1$) indicates that while the relationship is suggestive, it is not as strongly confirmed as other economic predictors.

The R-squared value of 0.435 suggests that economic and technological variables, alongside sustainability considerations, provide a stronger explanatory framework for firm distribution.

The results indicate that while ESG performance correlates with the spatial concentration of green energy firms, economic conditions remain dominant in shaping firm location patterns. The insignificance of CCPI suggests that climate policy commitments alone do not lead to clustering, while the negative effect of GDP per capita implies that high-income economies do not necessarily attract more green energy startups. Instead, the presence of a strong innovation infrastructure, reflected in R&D spending, appears to be a more important determinant.

Table 5.5.2.1: Green Energy Companies Distribution

VARIABLES	(1) sum_VC_backed
ESG	559.9***
	(170.1)
CCPI	0.961
	(1.066)
Constant	-379.1***
	(138.9)
Observations	37
R-squared	0.258

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5.5.2.2: Log transformation of Investment amount

VARIABLES	(1) (sum) VC _backed	(2) (sum) VC _backed	(3) (sum) VC _backed	(4) (sum) VC _backed	(5) (sum) VC _backed	(6) (sum) amount _rais_y	(7) (sum) amount _rais_y	(8) (sum) VC _backed	(9) Log _amount _raised
(mean) ESG	291.7	897.0	897.0	18.65	10.39	55,801	641,846	463.1*	30.54
	(282.3)	(656.0)	(656.0)	(12.83)	(7.551)	(166,497)	(1.831e+06)	(270.7)	(21.74)
(mean) CCPI	1.003	1.438	1.438	0.0906	0.0463	359.0	-1,377	1.481	-0.0697
	(1.379)	(1.266)	(1.266)	(0.104)	(0.0596)	(834.3)	(9,173)	(1.578)	(0.106)
ESG_GDP _interaction	0.0110								
	(0.00899)								
(mean) GDP_per _capita	-0.00792	-0.00102**	-0.00102**	-8.63e-05*	-4.71e-05**	-0.0849	-0.859	-0.000770*	-1.65e-05
	(0.00585)	(0.000505)	(0.000505)	(4.66e-05)	(2.30e-05)	(0.247)	(2.713)	(0.000397)	(4.00e-05)
(mean) R_D _spendingGDP	20.25*	7.137	7.137	3.694***	2.108***	16,671*	38,318	23.82*	3.039**
	(11.65)	(23.88)	(23.88)	(1.413)	(0.726)	(8,837)	(97,168)	(13.38)	(1.214)
Constant	-218.8	-596.8	-596.8	-19.00*	-10.43*	-68,230	-290,379	-356.9*	-13.76
	(214.8)	(414.0)	(414.0)	(11.30)	(6.209)	(105,266)	(1.157e+06)	(206.4)	(14.38)
Observations	37	37	37	37	37	35	35	37	35
R-squared	0.362	0.301	0.301					0.346	0.435

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

6. Conclusions

The transition towards a sustainable economy is increasingly shaped by financial incentives, policy-driven indexes, and the growing importance of environmental, social, and governance (ESG) considerations. This thesis has explored the statistical relationship between sustainability indicators—specifically the Climate Change Performance Index (CCPI) and the Environmental, Social, and Governance (ESG) Index—and country-level investment dynamics, with a particular focus on venture capital (VC) investments in green energy companies. Through a combination of descriptive statistics, econometric models, and machine learning techniques, this research has provided valuable insights into the drivers of sustainable investment and the role of policy and financial markets in supporting the green energy transition.

6.1 Key Findings and Scientific Contributions

The European cleantech sector is characterized by a diverse range of companies operating across various technological categories, with a significant concentration in sustainable energy production and energy-efficient industrial technologies. The majority of these companies are part of the broader cleantech ecosystem, with a smaller proportion classified as cleantech innovators. The geographical distribution of these firms reveals clusters in countries such as Germany, Italy, and France, which together account for more than half of the total number of cleantech companies in Europe. This concentration is likely influenced by a combination of economic factors, political stability, and institutional support. The descriptive statistics and geographical analysis provide a comprehensive overview of the sector, highlighting the importance of regional economic and policy frameworks in fostering cleantech innovation.

The ESG index has emerged as a critical tool for assessing the sustainability performance of nations. The analysis of sovereign ESG scores from 2014 to 2021 revealed significant improvements in ESG performance across many countries, driven by global initiatives such as the Paris Agreement and the European Green Deal. However, disparities remain, particularly between high-income and low-income countries. The entropy method used for weight assignment in the ESG index calculation proved effective, but challenges such as data limitations and greenwashing concerns persist. The ESG index serves as both a decision-making tool for investors and a catalyst for positive change, encouraging nations to adopt more sustainable practices. The methodological approach, including the use of Kernel Density Estimation (KDE) for visualizing temporal trends, provides a robust framework for assessing ESG performance and identifying areas for improvement.

The Climate Change Performance Index (CCPI), which evaluates national climate action efforts, has played a pivotal role in shaping climate policy and fostering public engagement. However, the index has faced criticism for its reliance on quantitative data, which may not fully capture the nuanced impacts of climate policies. The machine learning model developed in this study successfully predicted missing CCPI values for countries not included in the official dataset, providing a reliable tool for assessing climate performance. The results highlighted the importance of renewable energy, energy use, and climate policy in determining a country's CCPI score. The use of advanced machine learning techniques, such as XGBoost with Bayesian optimization, demonstrates the potential for integrating predictive analytics into sustainability assessments, enhancing the accuracy and reliability of climate performance metrics.

The econometric analysis revealed that while ESG scores positively influence the scale of VC investment in green energy firms, they do not significantly affect the probability of securing investment. Instead, economic factors such as GDP per capita, R&D spending, and foreign direct investment (FDI) play a more decisive role in attracting VC funding. Countries with lower GDP per capita and higher R&D spending tend to attract more green energy investments, suggesting that venture capitalists prioritize emerging markets with high growth potential and strong innovation ecosystems. The CCPI, while important for policy development, did not show a significant impact on VC investment decisions. These findings align with prior research emphasizing that venture capitalists prioritize technological potential, economic dynamism, and market growth over policy-driven sustainability indicators (Cumming, Henriques, & Sadorsky, 2016). The use of logistic regression and Ordinary Least Squares (OLS) regression models provides a robust empirical foundation for understanding the determinants of green energy investment.

The spatial analysis indicated that green energy firms tend to cluster in regions with strong sustainability frameworks, as reflected by high ESG scores. However, the CCPI did not significantly influence the geographical distribution of these firms. Instead, economic conditions, particularly R&D spending, were found to be more important in determining the location of green energy startups. This suggests that while sustainability performance is a factor, the presence of a robust innovation infrastructure is a more critical driver of firm clustering. The spatial distribution analysis, combined with econometric models, provides a comprehensive understanding of the factors influencing the geographical concentration of green energy firms, offering valuable insights for policymakers and investors.

6.2 Policy and Investment Implications

The findings of this study have important implications for policymakers, investors, and green energy firms. The analysis reveals that while ESG and CCPI scores are valuable indicators of sustainability performance, they do not significantly influence the likelihood of securing VC investment in green energy firms. Instead, economic factors such as R&D spending, FDI inflows, and GDP per capita play a more critical role in shaping investment decisions.

For policymakers, these findings suggest that efforts to attract VC investment in green energy should focus on creating a supportive environment for innovation and economic growth. This includes increasing public investment in R&D, improving access to foreign capital, and fostering a business-friendly regulatory environment. Additionally, policymakers should consider the role of sustainability metrics in shaping long-term investment strategies, particularly in the context of institutional investors who prioritize ESG compliance.

For investors, the results highlight the importance of considering both sustainability performance and economic fundamentals when making investment decisions. While ESG and CCPI scores provide valuable insights into a firm's or country's commitment to sustainability, they should be complemented with an analysis of market potential, technological innovation, and economic stability.

For green energy firms, the findings underscore the need to align their business strategies with the priorities of VC investors. This includes focusing on technological innovation, market scalability, and cost efficiency, while also demonstrating a commitment to sustainability through transparent reporting and adherence to ESG principles.

Overall, the study provides valuable insights into the drivers of sustainable investment and the role of policy and financial markets in supporting the green energy transition. By integrating multiple perspectives, this research aims to contribute to the development of more effective incentives for sustainable investment and the growth of the cleantech sector.

6.3 Limitations and Future Research Directions

While the machine learning model developed in this study provides valuable insights into the prediction of CCPI values, it is not without limitations. One of the primary challenges is the generalizability of the model to different contexts and datasets. The model was trained on data from 2014 to 2021, and its performance may vary when applied to future data or different regions. Additionally, the model's reliance on imputed data for missing values introduces potential biases, particularly in cases where the missing data mechanism is not fully understood.

Another limitation is the model's reliance on quantitative indicators, which may not fully capture the qualitative aspects of climate policy and governance. While the model incorporates a range of environmental, social, and governance indicators, it may overlook the nuanced impacts of policy decisions and institutional frameworks that are not easily quantifiable.

Future research could explore the integration of additional data sources, such as satellite imagery or social media data, to enhance the model's predictive accuracy. Additionally, the development of more sophisticated imputation techniques and the incorporation of time-series analysis could improve the model's ability to capture temporal trends and dynamics in climate performance.

Finally, the issue of greenwashing, where entities exaggerate their sustainability efforts, remains a significant challenge in the development of ESG and CCPI metrics. Future research should focus on developing more robust verification mechanisms to ensure the reliability and transparency of sustainability data.

6.4 Concluding Remarks

This thesis has explored the intricate relationship between sustainability indexes—specifically the Environmental, Social, and Governance (ESG) Index and the Climate Change Performance Index (CCPI)—and their impact on venture capital (VC) investment in green energy firms. By integrating multiple analytical perspectives, including econometric models, spatial analysis, and machine learning techniques, the study has provided valuable insights into the drivers of sustainable investment and the role of policy and financial markets in supporting the green energy transition.

The findings reveal that while ESG and CCPI scores are important indicators of sustainability performance, their influence on VC investment decisions is nuanced. ESG scores, particularly, show a significant positive correlation with the total amount of VC investment, suggesting that investors are increasingly considering environmental, social, and governance factors when allocating capital. However, the probability of securing VC investment is more strongly influenced by economic fundamentals such as R&D spending and foreign direct investment (FDI), rather than sustainability metrics alone. This highlights the dual role of economic competitiveness and innovation capacity in attracting venture capital, especially in the high-risk, high-reward environment of green energy startups.

The spatial distribution analysis further underscores the importance of regional economic conditions and innovation ecosystems in shaping the clustering of green energy firms. While countries with higher ESG scores tend to attract more VC-backed firms, the presence of strong R&D infrastructure and lower operational costs in emerging markets also play a critical role in determining firm location patterns. This suggests that while sustainability policies are important, they must be complemented by robust economic and technological frameworks to effectively drive green energy investment.

The study also highlights the limitations of current sustainability indexes, particularly the CCPI, in capturing the full spectrum of factors that influence VC investment decisions. The reliance on quantitative data and the lack of historical context in the CCPI may limit its effectiveness as a tool for guiding investment in green energy. Future iterations of these indexes could benefit from incorporating more qualitative assessments and addressing issues such as greenwashing and data reliability to enhance their relevance and accuracy.

From a policy perspective, the findings suggest that governments and policymakers should focus on creating an enabling environment for green energy investment by fostering innovation, improving access to capital, and implementing supportive regulatory frameworks. While sustainability indexes like ESG and CCPI provide valuable benchmarks, their impact on VC investment is contingent on broader economic and institutional factors. Therefore, a holistic approach that integrates sustainability goals with economic development strategies is essential for accelerating the transition to a green economy.

In conclusion, this research contributes to the growing body of literature on sustainable investment by shedding light on the complex interplay between sustainability metrics, economic drivers, and venture capital dynamics. The findings offer practical implications for investors, policymakers, and green energy firms, emphasizing the need for a balanced approach that leverages both sustainability performance and economic competitiveness to drive the green energy transition. Future research could further explore the role of emerging technologies, such as artificial intelligence and blockchain, in enhancing the accuracy and reliability of sustainability metrics, as well as their potential to influence investment decisions in the green energy sector.

Annexes

Annex 1

Python code used to preprocess the dataset and impute the missing values

```
# ==========================================================
# Missing Data Imputation Pipeline using KNN
# ==========================================================
# This script applies a structured missing data imputation process using
# K-Nearest Neighbours (KNN). It consists of two main steps:
# 1. Group-wise KNN imputation within country groups (extracted from "KEY")
# 2. Global KNN imputation to fill any remaining missing values
# ==========================================================

import pandas as pd
import numpy as np
from sklearn.impute import KNNImputer
import ace_tools as tools

# ==========================================================
# Load Dataset
# ==========================================================
# The dataset is loaded from the provided Excel file for processing.

file_path = "/mnt/data/CCPI DB.xlsx"  # File path
data = pd.read_excel(file_path)  # Load dataset

# Ensure the dataset is loaded properly
if data.empty:
    raise ValueError("The dataset is empty or could not be loaded. Please
check the file.")

# Ensure the "KEY" column exists
if "KEY" not in data.columns:
    raise ValueError("The dataset does not contain a 'KEY' column. Please
check the structure.")
```

```
# Extract country names from the "KEY" column (format: country_yyyy)
data["Country"] = data["KEY"].str.extract(r'(^[^_]+)')  # Extracts the part
before "_"

# ==========================================================
# Step 1: Group-wise KNN Imputation
# ==========================================================
# This step applies KNN imputation within each country group,
# ensuring missing values are inferred based on country-specific trends.

# Group data by extracted country names
grouped_data = data.groupby("Country")

# Initialize a list to store processed country groups
final_imputed_groups = []

# Iterate through each country group and apply imputation
for country, group in grouped_data:

    # Separate numerical and non-numerical columns
    numeric_cols = group.select_dtypes(include=[np.number]).columns
    numeric_data = group[numeric_cols]
    non_numeric_data = group.drop(columns=numeric_cols)

    # Identify columns where all values are missing
    all_missing_cols = numeric_data.columns[numeric_data.isnull().all()]

    # Remove all-missing columns to prevent errors during imputation
    filtered_numeric_data = numeric_data.drop(columns=all_missing_cols,
errors='ignore')

    # Apply KNN Imputation within each group
    imputer = KNNImputer(n_neighbors=5)
    imputed_array = imputer.fit_transform(filtered_numeric_data)

    # Convert imputed data back into a DataFrame
    imputed_numeric_data = pd.DataFrame(imputed_array,
columns=filtered_numeric_data.columns, index=group.index)

    # Reintroduce columns that were entirely missing (set as NaN)
```

```
    for col in all_missing_cols:
        imputed_numeric_data[col] = numeric_data[col]

    # Ensure column alignment with the original dataset
    imputed_numeric_data = imputed_numeric_data[numeric_cols]

    # Merge imputed numerical data with non-numeric data
    complete_group = pd.concat([imputed_numeric_data, non_numeric_data],
axis=1)

    # Append the processed group to the final dataset list
    final_imputed_groups.append(complete_group)

# Combine all processed groups into a single dataset
final_imputed_data = pd.concat(final_imputed_groups, ignore_index=False)

# Restore the 'KEY' column and maintain original column order
final_imputed_data["KEY"] = data["KEY"]
final_imputed_data = final_imputed_data[data.columns]

# ==========================================================
# Step 2: Global KNN Imputation for Remaining Missing Values
# ==========================================================
# After group-wise imputation, some missing values may still exist.
# This step applies KNN imputation at a global level to ensure a
# fully imputed dataset.

# Extract numerical columns for final global imputation
final_numeric_data = final_imputed_data.select_dtypes(include=[np.number])

# Apply KNN imputation globally
imputer_global = KNNImputer(n_neighbors=5)
final_imputed_array = imputer_global.fit_transform(final_numeric_data)

# Convert imputed data back into a DataFrame
final_imputed_numeric_data = pd.DataFrame(
    final_imputed_array, columns=final_numeric_data.columns,
index=final_numeric_data.index
)
```

```
# Replace the original numerical columns with the fully imputed values
final_fully_imputed_data = final_imputed_data.copy()
final_fully_imputed_data[final_numeric_data.columns] =
final_imputed_numeric_data

# Verify that all missing values have been successfully imputed
assert final_fully_imputed_data.isnull().sum().sum() == 0, "There are still
missing values in the dataset."

# ==========================================================
# Save Final Processed Dataset
# ==========================================================
output_file = "/mnt/data/CCPI_DATASET.xlsx"
final_fully_imputed_data.to_excel(output_file, index=False)
print(f"Fully imputed dataset successfully saved as '{output_file}'.")

# Display the fully imputed dataset
tools.display_dataframe_to_user(name="Fully Imputed Dataset Without Missing
Values", dataframe=final_fully_imputed_data)
```

Annex 2

Python code used to predict the missing CCPI values

```
# ==========================================================
# Machine Learning Model for CCPI Prediction
# ==========================================================
# This script loads, preprocesses, and trains an XGBoost model
# to predict missing values in the Climate Change Performance Index (CCPI).
# It includes hyperparameter optimization using Optuna and final model
evaluation.
# ==========================================================

# --- Import Required Libraries ---
import pandas as pd
import numpy as np
import joblib
import optuna
import shap
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer
from sklearn.impute import SimpleImputer
from xgboost import XGBRegressor
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error

# ==========================================================
# Step 1: Load and Prepare the Dataset
# ==========================================================

# Define dataset path
file_path = "CCPI_DATASET.xlsx"

# Load dataset
data = pd.read_excel(file_path)



# Ensure target column exists
```

```
target_col = "CCPI"
if target_col not in data.columns:
    raise ValueError(f"Target column '{target_col}' not found in dataset.")

# Preserve the 'KEY' column if available
key_col = data["KEY"] if "KEY" in data.columns else None

# Separate rows with missing CCPI values (to be predicted)
data_missing_ccpi = data[data[target_col].isna()].copy()
data_complete = data.dropna(subset=[target_col]).copy()

# Select a validation subset for model assessment
val_size = min(10, len(data_complete))
data_validation = data_complete.sample(n=val_size, random_state=42)
data_complete = data_complete.drop(data_validation.index)

# Exclude non-feature columns (except 'KEY')
drop_cols = ["CCPI"]
if "KEY" in data.columns:
    drop_cols.append("KEY")

X = data_complete.drop(columns=drop_cols, errors="ignore")
y = data_complete[target_col]

# ==========================================================
# Step 2: Data Preprocessing
# ==========================================================

# Identify numerical and categorical features
numerical_features = X.select_dtypes(include=["float64",
"int64"]).columns.tolist()
categorical_features = X.select_dtypes(include=["object"]).columns.tolist()

# Define numerical preprocessing pipeline
numerical_transformer = Pipeline([
    ("imputer", SimpleImputer(strategy="median")),  # Median imputation for
missing values
    ("scaler", StandardScaler())  # Standardization (zero mean, unit
variance)
])
```

```
# Define categorical preprocessing pipeline
categorical_transformer = Pipeline([
    ("imputer", SimpleImputer(strategy="most_frequent")),  # Mode imputation
    ("onehot", OneHotEncoder(handle_unknown="ignore"))  # One-hot encoding
])

# Combine preprocessing steps
preprocessor = ColumnTransformer([
    ("num", numerical_transformer, numerical_features),
    ("cat", categorical_transformer, categorical_features)
])

# Split dataset into training and testing sets (85% train, 15% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.15,
random_state=42)

# Apply preprocessing transformations
X_train = preprocessor.fit_transform(X_train)
X_test = preprocessor.transform(X_test)

# ==========================================================
# Step 3: Feature Selection Using SHAP
# ==========================================================

# Train an initial XGBoost model for feature importance analysis
temp_model = XGBRegressor(n_estimators=500, max_depth=5, random_state=42)
temp_model.fit(X_train, y_train)

# Compute SHAP values for feature importance ranking
explainer = shap.Explainer(temp_model)
shap_values = explainer(X_train)

# Select the top 15 most important features
feature_importance = np.abs(shap_values.values).mean(axis=0)
important_features = np.argsort(feature_importance)[-15:]
X_train = X_train[:, important_features]
X_test = X_test[:, important_features]

# ==========================================================
```

```
# Step 4: Hyperparameter Optimization using Optuna
# ============================================================

def objective(trial):
    """Objective function for Optuna hyperparameter tuning."""
    params = {
        "n_estimators": trial.suggest_int("n_estimators", 2000, 5000,
step=500),
        "learning_rate": trial.suggest_float("learning_rate", 0.01, 0.2,
step=0.01),
        "max_depth": trial.suggest_int("max_depth", 5, 7),
        "subsample": trial.suggest_float("subsample", 0.7, 0.95),
        "colsample_bytree": trial.suggest_float("colsample_bytree", 0.6,
0.85),
        "min_child_weight": trial.suggest_int("min_child_weight", 5, 8),
        "gamma": trial.suggest_float("gamma", 0.1, 0.5),
        "reg_alpha": trial.suggest_float("reg_alpha", 10, 20),
        "reg_lambda": trial.suggest_float("reg_lambda", 10, 20),
    }

    model = XGBRegressor(objective="reg:squarederror", random_state=42,
**params)
    return cross_val_score(model, X_train, y_train, cv=7,
scoring="r2").mean()

# Run Bayesian optimization with Optuna
study = optuna.create_study(direction="maximize",
sampler=optuna.samplers.TPESampler())
study.optimize(objective, n_trials=75)

# ============================================================
# Step 5: Train Final Model with Best Parameters
# ============================================================

best_params = study.best_params
final_model = XGBRegressor(objective="reg:squarederror", random_state=42,
**best_params)

# Train the model
final_model.fit(X_train, y_train)
```

```
# ==========================================================
# Step 6: Model Evaluation
# ==========================================================

# Make predictions
y_pred = final_model.predict(X_test)

# Compute evaluation metrics
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print("\nModel Evaluation:")
print(f"R² Score: {r2:.4f}")
print(f"Mean Absolute Error (MAE): {mae:.4f}")
print(f"Root Mean Squared Error (RMSE): {rmse:.4f}")

# Cross-validation results
cv_scores = cross_val_score(final_model, X_train, y_train, cv=7,
scoring="r2")
print(f"Mean Cross-Validation R² Score: {cv_scores.mean():.4f} (±
{cv_scores.std():.4f})")

# ==========================================================
# Step 7: Predict Missing CCPI Values
# ==========================================================

# Preserve 'KEY' column in missing data
keys_missing = data_missing_ccpi["KEY"] if "KEY" in data_missing_ccpi.columns
else None

# Apply preprocessing and feature selection
X_missing = preprocessor.transform(data_missing_ccpi.drop(columns=["CCPI",
"KEY"], errors="ignore"))
X_missing = X_missing[:, important_features]

# Predict missing values
data_missing_ccpi[target_col] = final_model.predict(X_missing)
```

```
# Restore 'KEY' column if necessary
if keys_missing is not None and "KEY" not in data_missing_ccpi.columns:
    data_missing_ccpi.insert(0, "KEY", keys_missing)

# Save predictions to file
output_file = "Predicted_CCPI_Values.xlsx"
data_missing_ccpi.to_excel(output_file, index=False)
print(f"Predicted CCPI values saved to '{output_file}'.")

# Save model artifacts for reproducibility
joblib.dump(preprocessor, "preprocessor.pkl")
joblib.dump(important_features, "selected_features.pkl")
joblib.dump(final_model, "final_xgb_ccpi_model.pkl")
print("Model and preprocessor saved.")
```

List of figures

List of tables

Bibliography

Climate Change Performance Index (CCPI) 2025: What You Should Know . (n.d.). Retrieved from https://guide.examarc.com/upsc/climate-change-performance-index-ccpi-2025-what-you-should-know/

Ambrois, Butticè, Caviggioli, Cerulli, Croce, Marco, D., . . . Zinilli. (2023). Using machine learning to map the European Cleantech sector. *EIF Research and Market Analysis, 91*.

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl , B. (2011). Algorithm for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, 24.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 16-28.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Clark, G. L., Feiner, A., & Viehs, M. (2015). From the Stockholder to the Stakeholder: Can Drive Financial Outperformance. *Oxford University Press*.

Climate Change Performance Index | NewClimate Institute . (n.d.). Retrieved from https://newclimate.org/what-we-do/projects/climate-change-performance-index

Climate Change Performance Index by Country 2024 - World Population Review . (n.d.). Retrieved from https://worldpopulationreview.com/country-rankings/climate-change-performance-index-by-country

Cumming, D., Henriques, I., & Sadorsky, P. (2016). "Cleantech" venture capital around the world. *Journal of Business Ethics*, 169-285.

Dietterich, T. G. (2000). Ensemble methods in machine learning. *International Workshop on Multiple Classifier Systems*, 1–15.

FAQs | Climate Change Performance Index - CCPI . (n.d.). Retrieved from https://ccpi.org/faqs/

Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. (66).

Global Sustainable Investment Alliance. (2020). *Global sustainable investment review 2020.*

Gompers, P., & Lerner, J. (2001). The venture capital revolution. *Journal of Economic Perspectives*, 145-168.
Gratcheva, E., Emery, A., Wang, D., & Martin, N. (2021). ESG and sovereign risk: Evidence from emerging markets. *World Bank Research Working Paper No. 9587*.
Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and Techniques (3rd ed.).* Morgan Kauffman.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 1735-1780.
How Grassroots Environmental Activism Has Changed the Course of History. (n.d.). *https://www.goldmanprize.org/blog/grassroots-environmental-activism/* . Retrieved from https://www.goldmanprize.org/blog/grassroots-environmental-activism/
Hsu, P. H., Tian, X., & Xu, Y. (2014). Financial Development and Innovation: Cross-Country Evidence. *Journal of Financial Economics*, 116-135.
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R.* Springer.
Jiang, P.-C., Feng, G.-F., & Yang, H.-C. (2022). New measurement of sovereign ESG index. *Innovation and Green Development, 1*.
Jolliffe, I. T. (2002). *Principal component analysis (2nd ed.).* Springer.
Jones, E., & Youngs, R. (2024). *Civic Activism in an Intensifying Climate Crisis.* Retrieved from Carnegie Endowment for International Peace: https://carnegieendowment.org/research/2024/12/climate-change-protest-activism-green-transition?lang=en
Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data (3rd ed.).* Wiley.
Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4765-4774.
Methodology | Climate Change Performance Index - CCPI . (n.d.). Retrieved from https://ccpi.org/methodology/
Naomi, A., & Akbar, A. (2021). Economic development and ESG performance: Causal relationships and implications for policy. *Sustainability, 13(14)*.
Restrepo-Mieth, A., Perry, J., Garnick, J., & Weisberg, M. (2023). Community-based participatory climate action. *Global Sustainability,*

6(e14). Retrieved from https://www.cambridge.org/core/journals/global-sustainability/article/communitybased-participatory-climate-action/8EAC9C7F3FC0EF97BEA3572E5D1B868B
Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 581-592.
Schafer, J. L. (1997). Analysis of incomplete multivariate data. *CRC Press*.
The Climate Change Performance Index 2024: Results . (n.d.). Retrieved from https://www.germanwatch.org/en/89895
Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., . . . Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, 520-525.
United Nations. (1987). *Our common future: Report of the World Commission on Environment and Development.*
Vama, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 91.

Acknowledgements

Ringrazio la Professoressa Ughetto per il suo prezioso supporto, la guida attenta e la fiducia che ha riposto in me. Grazie per avermi accompagnata con pazienza e dedizione lungo questo percorso, offrendomi preziose opportunità di crescita e apprendimento.

A mia madre, la mia guida, la mia roccia, il mio esempio più grande. Mi hai insegnato a perseguire i miei sogni con determinazione e a non lasciarmi abbattere dalle difficoltà, mai. Con il tuo amore infinito e la tua forza instancabile, mi hai mostrato cosa significhi essere una donna coraggiosa, indipendente e generosa. Sei stata al mio fianco in ogni momento, condividendo le mie gioie con entusiasmo e alleviando le mie paure con la tua presenza rassicurante. Senza di te, nulla di tutto questo sarebbe stato possibile, e per questo ti sarò eternamente grata. Spero di essere, un giorno, anche solo la metà della persona straordinaria che sei tu.

A mio padre, che con costanza e dedizione ha saputo ricostruire il nostro rapporto, restituendomi quel pezzo di me che per troppi anni è mancato nella mia vita.

A Mattia, con cui condivido un legame unico, fatto di momenti di confronto ma anche di affetto sincero. So di poter contare su di te, così come tu potrai sempre contare su di me.

A Matilde, la mia persona. Sei cresciuta al mio fianco, facendo tuoi valori e sfumature che ci rendono così simili, eppure unicamente te stessa. Il tuo sostegno incondizionato, la tua presenza costante – anche nei

momenti più importanti – sono stati un dono immenso. Non potrei essere più fiera di te.

A voi quattro. Siete la mia casa e il mio cuore, sempre e ovunque.

A Chiara, che ha portato luce nei giorni felici e, soprattutto, in quelli più bui. Sei stata e sei famiglia. Questo percorso non sarebbe stato lo stesso senza di te.

E grazie un po' anche a me stessa, per averci creduto sempre e non aver mollato mai.