ESG Factors and Climate Change for Credit Analysis and Rating

ESG-CREDIT.EU DATABASE

Table of contents

I. INTRODUCTION TO THE DATABASE

In the last decades the public opinion was driven by a growing awareness on climate and environmental related topics. In addition, governments understood the relevance of social issues, after having tackled several financial crisis. As a direct consequence, corporates' behaviour has been affected more and more by these aspects over time. The significance of the theme is nowadays undeniable; in fact, new specialized entities were created to assess the environmental, social and governance behaviour of the companies: they are known today as ESG Rating Agencies.

Focusing on this phenomenon, the ESG-Credit.eu project was developed to provide a methodological support for the construction and use of credit ratings, which will include ESG and climate change related factors. In order to pursue this intention, the project was created based on seven steps that go from the analysis of the associated literature to the design of a final methodology for the integration of ESG and climate change factors in credit ratings. One of these stages, a crucial one, is referred to the construction of a comprehensive "ESG database", which aims to collect high-quality and sufficiently long time series data to allow to carry out relevant empirical analyses.

The database is composed by three main blocks:

- <u>Firm</u>, which contains information about:
  - Market data, gathered from Thomson Reuters Eikon.
  - Credit ratings, provided by S&P, Moody's, and Fitch.
  - ESG ratings and measures, collected from several providers with the intention of having different points of view, trying to avoid the relevant and well-known problem of the divergence among ESG rating methodologies.
- <u>State Variables</u>, related with financial data (market indices, volatility indices, …) and macroeconomic data.
- <u>Climate Change Factors</u>, a block in which data are collected with the support of ENERGYA, a European Research Council project.

The ambitious aim of the program is to collect these type of data to enable the conduction of analysis using both different models and approaches. For this purpose, it was required to include detailed data on credit risk as well as ESG and climate risk criteria across several sectors and geographical locations. Thus, one key aspect of this project is the data accuracy and quality, and in particular the ability to rely on a sufficiently long sample of historical data.

Looking at the database, there was initially collected data with respect to the first block defined above. Firm level data can be divided into different sublevels; specifically, in the download and data management process, these information was partitioned into six main groups, described as follow:

- Market Data, a block containing information about volume, price, total return, and market value.
- Financial Data, which involves both general data, derived from income statement and balance sheet, and more specific data, resulting from cash flow statement and financial ratios.
- Credit Ratings, a sub-level built relying on the creditworthiness assessment published by the big three Credit Rating Agencies (Moody's, Fitch, S&P).
- ESG Ratings, field obtained collecting valuations issued by different ESG rating agencies and data providers.

- ESG Data, which helps to explain how companies' behaviour has evolved over time and how this could affect their creditworthiness.
- CDS Data, the last sub-level, which allows to better explain the credit standing of the companies.

The present paper was drafted with the purpose to explain the composition of the database, exploring all its features, and clarifying how it was developed. As stated by the project general purpose, the database is designed in an object-oriented framework, a paradigm where each block represents a structure that contains both data and procedures. This framework is particularly convenient given its modularity and flexibility, which easily allows ad hoc interventions such as extensions by the inclusion of new type of data.

The development of the database was performed with the use of Python as this enabled to carried out a more efficient process. Especially, the download process required an automatization because of the wide amount of data required.

II. SAMPLE OF FIRMS AND METRICS OF INTEREST

The construction of this extensive database required, first of all, the construction of a proper sample of firms, reasonably large to provide a comprehensive set of companies. The size of the selected universe is a crucial point because it may help avoid incurring in several biases mentioned in the literature, allowing so to develop accurate analysis. Moreover, the project deals with the challenge of finding an inclusive set of measures, needed to perform an appropriate investigation of the influence of ESG factors on companies in terms of market metrics, financial metrics, and creditworthiness.

1. Universe of firms

Due to the importance and the great ambition of the research project's purpose, it was required to create the largest, most comprehensive possible set of listed companies headquartered in the European Union. There were considered the 27 countries belonging to the EU, relying on the list provided by the European Union itself by its website[1]. To those 27 countries there was added also the United Kingdom, due to its relevance. Thus, the updated database contains firms belonging to the 27 EU countries and to United Kingdom.

After the identification of the countries to consider in the construction of the database, the main problem arising was linked to the research of the listed companies existing in each state. To overcome this issue, there were exploited different available providers of data; as a result, the decision was to rely on *Thomson Reuters* as a source of information. In particular, the universe of firms was developed using *Datastream*. This is an historical financial database developed by Thomson Reuters, which collects over 35 million individual instruments or indicators across all major asset classes, including 8.5 million active economic indicators. Furthermore, it features 65 years of data, across 175 countries[2].

In addition to several types of financial and macroeconomic data, Datastream also provides some interesting and valuable services, such as the *Worldscope Global Database*, a relevant source of detailed financial and profile data about public companies. The objective of this database is indeed to provide the most comprehensive, accurate and timely data on publicly quoted companies belonging to the specified geographical area; it started from a universe of about 4,000 firms in 1987, counting, at March 2007, a list of over 51,100 companies[3]. It is worth noting that Worldscope provides not only information relative to active firms, but it yields data about extinct or inactive companies, that is those which have merged, liquidated, or became privately held, too. Therefore, history for these corporations remains on the database.

The database was built using the aforementioned platform since it provides, among others, readily available country lists encompassing all listed entities in each country of interest. So, in this context, there were used the ready-made lists provided by Worldscope for every country within the European Union, getting for each one all available lists present in a specific file, which was published by Thomson Reuters itself[4]. In fact, for larger equity markets there are multiple lists, considering that each of them cannot contain more than 1,000 companies. Hence, in similar circumstances, the overall universe of listed firms was divided in different smaller sub-lists.

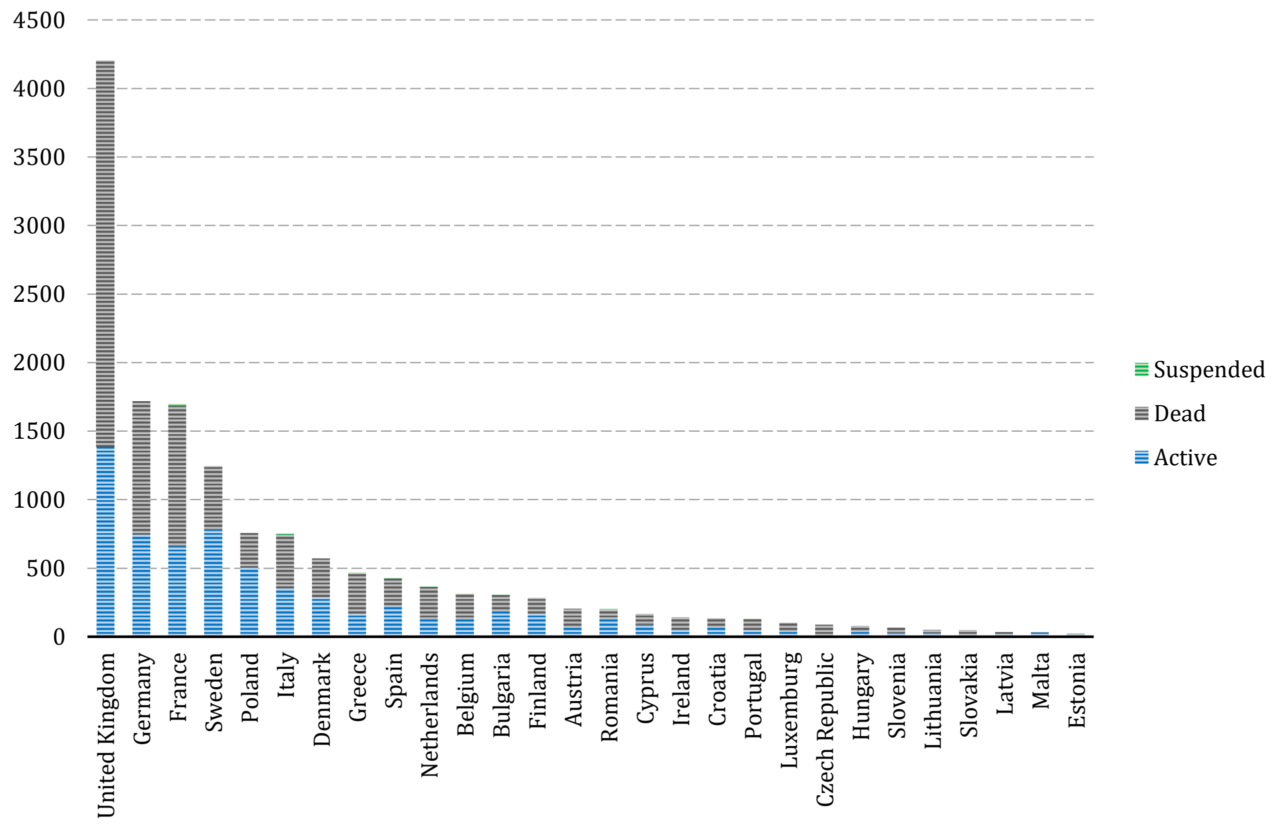
[1] *Source: Countries | European Union (europa.eu)*
[2] *Source: Datastream Macroeconomic Analysis | Refinitiv*
[3] *Source: https://www.tilburguniversity.edu/sites/default/files/download/WorldScopeDatatypeDefinitionsGuide_2.pdf*
[4] *Datastream Worldscope list (columbia.edu)*

Thus, the creation of the database began downloading all the existing lists for each of the 28 countries considered. Such step generated a list of ISIN, allowing to dispose of a quite huge amount of unique company identification numbers; specifically, the process led to a result of a comprehensive list of *15,988* firms. It is worth notice that the list we are speaking about was obtained operating a screening from the overall universe of companies provided by Worldscope lists. Precisely, any present duplicated firm was deleted relying on a check based on the ISIN codes. In addition, amid the remaining list, there was developed a process to firstly detect and secondly delete the companies located in countries different from the ones of the European Union.

As stated before, Worldscope contains information of dead and inactive companies and, therefore, lists include both listed and no longer listed firms. The decision to preserve the latter in the final list was taken thinking about a possible survivorship bias which could arise in the regression analysis, with a direct effect on the accuracy of the results. Anyway, from the remaining *15,988* companies, there was performed a last selection, aimed to remove from the list all the companies dead before 2000. The decision was taken since the database collects information starting from *31/12/1999*; hence, there would be no data for those firms. So, the final list, which was used as basis for the construction of the entire database, includes *14,626* companies.

The final sample of firms considered in database development is presented in Figure 1, which breaks down the overall list among the different EU countries, giving indications about the distribution between dead, suspended, and active companies. More specific information can be found in *Appendix A*.

Figure 1 – Country breakdown of companies' universe.



Going more in deep, as of October 2021, the total amount of dead firms is 8,010, and consists in the 54.77% of the complete list. The selected universe of firms was stored in an Excel file, which brings together all the firms located and listed in a European Union country or in United Kingdom. This file plays a key role in the construction of the database as it can be considered the starting point for all the subsequent steps.

For each company the file contains a wide range of *“static”* information; actually, it should be considered as a descriptive, static map. It was built starting from the collection of *14,626* ISINs, obtained from Worldscope and representing the selected firms’ universe. After the development of the list, the ISIN codes were used as an input to gather detailed information about each company. To achieve this purpose, *Thomson Reuters Eikon*, *Datastream, Bloomberg* and *FactSet* were used as data providers to inhabit the static map, since they provide different type of information.

Finally, the overall information collected was organized to allow a more comprehensive interpretation of the map. Thus, it was derived a static map structured in four different macro-areas; the structure implemented is summarized by Table 1 – Structure of the static map used to develop the database.

Table 1 – Structure of the static map used to develop the database.

IDENTIFICATION			**CLASSIFICATION**		
Field	*Source*	*Coverage*	*Field*	*Source*	*Coverage*
Exchange	Datastream	100.00 %	General Industry Class.	Datastream	99.62 %
ISIN Code	Datastream	100.00 %	Industry Group	Datastream	99.80 %
Bloomberg Ticker	Bloomberg	79.50 %	GICS Sector Name	Eikon	44.20 %
RIC	Eikon	100.00 %	GICS Industry Name	Eikon	44.21 %
Worldscope ID	Datastream	98.84 %	GICS Sub-Industry Name	Eikon	44.19 %
Organisation PermID	Datastream	94.75 %	TRBC Economic Sector	Datastream	96.73 %
SEDOL Parent Code	Datastream	98,11 %	TRBC Business Sector	Datastream	96.74 %
			TRBC Industry Group	Datastream	96.76 %
			TRBC Industry	Datastream	96.77 %
			NACE Classification	Eikon	96.62 %
			NACE (4-digit)	Eikon	96.62 %
DESCRIPTIVE			**CDS**		
Field	*Source*	*Coverage*	*Field*	*Source*	*Coverage*
Country	Eikon	100.00 %	CDS Spread 5y Ticker	Datastream	0.88 %
Currency	FactSet	96.87 %	Has CDS Quote	Eikon	8.66 %
Name (with status spec.)	Datastream	100.00 %	Primary CDS Quote	Eikon	3.18 %
Company Name	Eikon	99.96 %	Primary Country of Risk	Eikon	53.12 %
City of Headquarters	Eikon	86.48 %			
Latitude	Eikon	68.54 %			
Longitude	Eikon	68.54 %			
Dead Date	Datastream	54.77 %			
Inactive Date	Datastream	54.37 %			
Is Active (flag)	Eikon	100.00 %			
Is Delisted (flag)	Eikon	100.00 %			
Equity Status	Datastream	100.00 %			
Operating MIC	Eikon	77.53%			

* Coverage (%) is computed as the ratio between number of companies with available information and the total amount of firms (14,626).

The distinction between active, inactive, and dead companies is very clear in the static map because the status is properly identified by a specific field, as shown by Table 1. Datastream is very accurate because it is able to provide the name of the entity, containing the status (whether it was delisted or dead) and the date of the eventually occurred delisting.[5] For more information about the decomposition of dead firms by period, see *Appendix B*.

Unfortunately, the database is not able to provide a complete information for each company because of both Thomson Reuters Datastream and Eikon availability of data. Hence, the file lacks in information for some data, especially for what concerns the CDS macro area. This situation may be due to the fact that very few companies have a quoted CDS in the market. Specifically, in the developed database there are only 480 firms with a quoted CDS, relying on the availability offered by Thomson Reuters for this type of data.

2. Universe of Metrics

After the identification of the universe of firms where to rely on in the construction of the database, the following, natural step consisted in the development of a comprehensive set of metrics, wide enough to ensure the performance of complete and correct analysis. So, it was required to discover proper measures, both financial and non-financial, for each company in the list. In compliance with the target of the project, the attention was principally focused on indicators related to creditworthiness and environmental, social and governance aspects. On the financial side, both strictly financial indicators, derived from income statements and balance sheets, and market data, linked to company stocks, were also considered in the development of the list of metrics.

As what was done for the universe of firms, also in this case there was built an Excel, summary file to easily map the chosen metrics. This, next to the static map, constitutes the basis from which the database is built. With the aim of finding an appropriate set of measures, there were exploited different data providers; Therefore, the metrics research process was implemented through, among others, Thomson Reuters Eikon, Bloomberg, MSCI and CDP. We use also other providers, such as NRG and Standard Ethics, specialized in the assessment of ESG features.

The table summarizing the metrics' universe exhibits for every measure a detailed list of fields, in order to ease the interpretation of the indicator. In particular, for each metric there were developed ten fields which help the specification of it. More precisely, they allow to know the unit of measure by which the indicator is expressed, and the data provider who furnishes it, in addition to the code needed to retrieve the specific metric. Furthermore, the map specifies the frequency by which the measure is updated by the provider. Finally, another relevant field involves the coverage; it expresses, as a percentage, the number of companies for which the metric is available versus the overall number of companies in the database.

Actually, the overall list counts *1125 metrics*. Due to this large volume of indicators, they were divided in groups, depending on the topic they are referred on. The distribution was developed identifying the following four categories:

I. *Market Data*. It is the lowest populated group since it includes only few measures (7), that are the ones directly related to information regarding the stocks of the companies. For this type of

[5] *Source: https://www.ifm.unibe.ch/e39710/e39716/e274093/files495520/DATASTREAM.pdf*

metrics, the database relies on information provided by Thomson Reuters Eikon; due to their nature, these data are updated (and available) in a daily frequency.

II. *Financial Data*. The category collects 80 indicators; they are mainly referred to information derived by Balance Sheet, Income, and Cash Flow statements. Moreover, information is gathered also looking at financial ratios. As for market data, financial measures are collected from Thomson Reuters Eikon. However, in this case the information is provided in a yearly fashion with respect to financial statements, and in a quarterly basis for financial ratios. Indeed, even if listed companies are required to provide a quarterly based financial disclosure, Thomson Reuters Eikon does not allow to have so granular information.

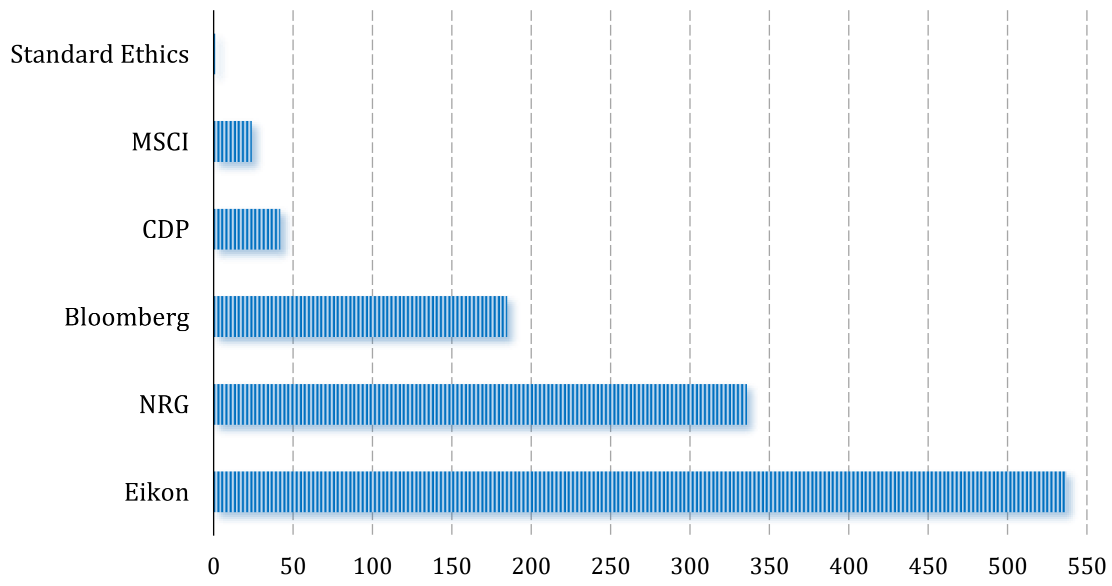
III. *ESG & Credit Ratings*. This group is composed by 88 measures; only 11 among these directly refers to credit ratings, while the other are all focused on ESG aspects. Specifically, the majority of ESG related indicators are expressed as scores and are collected from an extensive sample of ESG rating agencies and data providers. Due to the different source of information, data was collected either in a monthly, weekly, or daily frequency.

IV. *ESG Factors.* This represents the most relevant part of the metrics' map. Indeed, it is composed by the considerable amount of 945 indicators, all directly concerning ESG themes. Due to the large number of measures and the use of several providers, data was collected either in monthly or in yearly frequency, depending on the availability granted by them. To collect these measures, we relied on information published by six different providers.

The decomposition of the entire list of indicators into four distinct area was developed thinking at the possible database structure; the chosen partitioning method turned out to be extremely efficient in easing the subsequent data download process.

The overall list currently counts *1120 metrics*, as stated before; they are collected from a wide variety of data provider. The most used was Thomson Reuters Eikon since it provides both financial and non-financial information and grants an enormous capacity of download with respect to Bloomberg. Anyway, to have a comprehensive set of ESG measures, the collection relied also on different data providers. The contribution of each is graphically summarized by Figure 2.

Figure 2 – Breakdown of the 1125 metrics between the different providers.



According to Figure 2, the most used provider is represented by Thomson Reuters Eikon, mainly because of its high capacity of data download. Furthermore, the choice to mainly rely on Thomson Reuter Eikon was taken looking at the availability and the variety of data the different providers guarantee.

Next to Eikon, another widely used provider considered is NRG[6], a database created in 2016 from a team of market professionals and academic researchers on the field of corporate governance. The database offers accurate data about corporate governance of listed companies around the world, retrieving information from annual reports, SEC filings, corporate governance reports, etc. NRG Metrics are divided between six different datasets: Corporate Governance, Ownership Structure, Directors and Officers, Family Firms, Compensation and Audit.

Another relevant role is played by Bloomberg. This provider offers very precise information but has an enormous disadvantage: it provides a very low data capacity limit, which is set on a monthly basis. Hence, you must pay attention on the download process. In our case, the provider was used only to download the most relevant ESG metrics, not to collect financial information.

Then, after this three big players, other providers were used in the development of the database. Such providers are CDP, Standard Ethics, and MSCI. Differently from the first three, they play a marginal role. In particular, Standard Ethics was used only for the ESG rating it provides; in fact, it is the first independent Sustainability rating agency with a standard methodology and a proprietary algorithm following the applicant-pay model. So, the information provided by those last three providers is very relevant for the purpose of the project.

The distribution of the different providers among the four different groups of metrics is expressed by Table 2.

Table 2 – Decomposition of metrics into different providers per metrics group.

Provider	Market Data	Financial Data	ESG & Credit Ratings	ESG Factors	Total
Eikon	7	80	6	439	*532*
Bloomberg			50	135	*185*
CDP			7	35	*42*
MSCI			24		*24*
Standard Ethics			1		*1*
NRG				336	*336*
Total	*7*	*80*	*88*	*945*	***1120***

Hence, the map of metrics is quite helpful if you would explore the indicators used in the construction of the database since it accurately illustrates the meaning of each measure, exhibiting also all the possible useful information. For these reasons, the construction process of the map was performed with the intention of simplifying the most possible the research of a metric when you are interested in it.

[6] *https://nrgmetrics.com/.*

III. DATABASE STRUCTURE

The two maps described in the previous section (*Static Map, Map Metrics*) were used as the starting point for the construction of the database. So, having defined the list with all the companies of interest for the 27 European Union countries plus United Kingdom and the universe of indicators of interest for each of them, it was then possible to think about the most suitable database structure with respect to available information.

Looking at both the static and the metrics map, the intuition consisted in developing a unique, large database and to divide then it into five smaller sub-databases. The division was manly based on the macro-areas identified by the map collecting the 1120 indicators. In addition to the group decomposition developed in the metrics' map, it was necessary to introduce another cluster, referred to CDS data. Furthermore, the division in sub-groups was necessary since it allows to take into account that different type of information is published in different frequencies. Thus, the following five sub-level databases were identified:

1. *Market Data*
2. *Financial Data*
3. *ESG & Credit Ratings*
4. *ESG Data*
5. *CDS*

The final structure of the database is expressed graphically, as a diagram, by Figure 3.

Figure 3 – Database diagram.

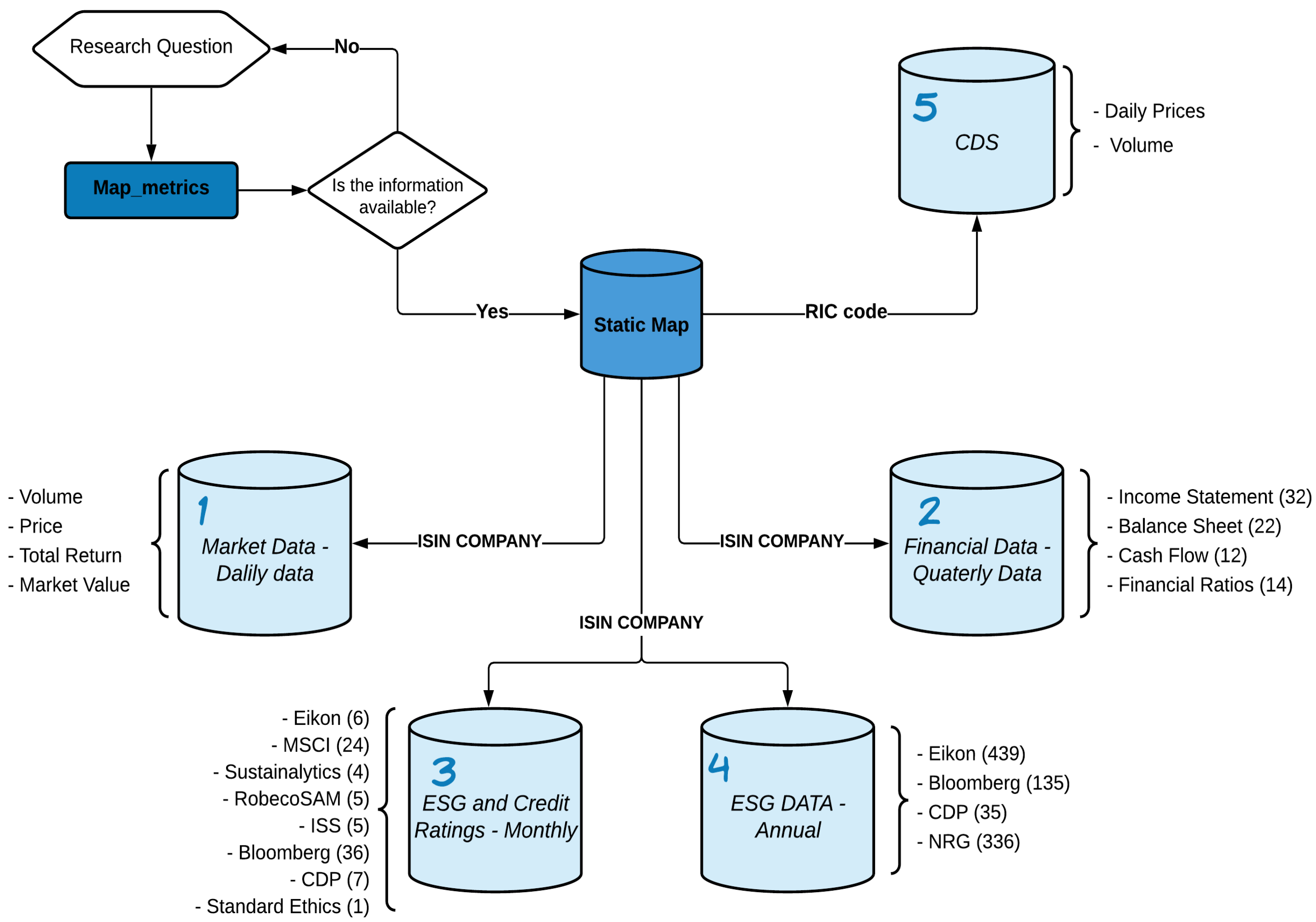


Figure 3 is really helpful because provides a very detailed and concise explanation of the database, giving information about providers of data and data frequency. Actually, as you can easily see in the figure, each sub-level database contains information given in a specific frequency, depending on the availability of the data provided by the platform used. In fact, for example, market data changes in an intraday frequency while the largest sample of ESG data is disclosed only annually by the firms. The databases were numbered; this was important in the download process because it was performed relying on the numeration of the databases.

The construction of the database was developed through a massive download process. The latter was performed generating specific scripts on Python, which allow to automatize the process, making it faster and more efficient. The scripts were different, in order to follow faithfully the structure of the database. So, there is a script for each sub-database. After the download process, the data management step followed, again with specific Python's scripts. Basically, there were developed two scripts for each sub-database: one relative to the download process, and one related to the data management process. The latter allows to export the data downloaded to files in a **.csv** format. The download process ended with the creation of a series of files, each one containing different data present in the metrics map, that were exported in a particular format; in fact, Python, thanks to the *pickle* module, allows to save on disk any Python object. This is possible because it serializes the object of interest, converting it into a byte stream, with the idea that this character stream contains all the information necessary to reconstruct the object in another Python script. Hence, whenever you want to load the data in the pickle format, you only need to de-serialized the byte stream.

Thanks to the data management process, it was possible to export the data collected through the creation of five **.csv** files for each company analysed, one for every sub-level database. In addition, there were created four files, one for each of the first four sub-databases, in each download performed; these files were built to summarize the major information about the download. They contain the list of companies involved in the download process, identified with the ISIN code. Each of those Excel files contains three sheets: the first, "*info_metrics*", shows, for each firm, if the metrics downloaded were available (1) or not (0); the second, "*info_start*", expresses the starting date of the series downloaded for each firm; the latter, "*info_end*", expresses instead the end point of the series for each firm.

Summarizing, the database was built basing on the two maps described above, the static map and the metrics map. So, the database includes *14,626* companies, all headquarters and operating in one of the 27 countries of European Union or in the United Kingdom. For each firm, there were collected *1120* measures, the ones indicated by the map of the metrics, everyone characterized by a specific frequency.

With regard to the time horizon, there was decided to start with the collection of data from *01/01/2000* and to gather information until *30/10/2021*. The choice was developed following the aim of the project, that consist in the creation of the most possible comprehensive database of financial and non-financial data for the companies listed in the 27 countries of EU plus United Kingdom. In addition, the wide time horizon ensures the possibility to dispose of an available, reliable, and sufficiently long sample of historical data.

In the sub-sections that follow, there is going to be developed a precise description of the five sub-levels databases built, which together constitute the overall database.

1. Market Data

All the market data of interest were provided by Thomson Reuters Eikon platform and are collected and available, due to their nature, in a daily frequency. In fact, because of their volatility, they would be available even in an intraday frequency. The database in question is composed by the following measures:

- *volume of trading.*
- *high, low, close, open prices.*
- *total return* (expressed as percentage).
- *market capitalization.*

All the prices are expressed in the currency linked to the country where the firm is listed in. Looking at the market capitalization, instead, it was downloaded in euro for all the firms in the universe. For each of the 14,626 companies, the same time series was created in order to have a standardised layout. The series consists in the complete set of business days within the pre-specified time horizon.

Market information was the most available; in fact, the average coverage of the metrics collected is about 90 %.

2. Financial Data

As done for the first sub-database, financial data were collected entirely relying on Thomson Reuters Eikon. Anyway, in this case it was needed to face a trade-off in the download process between the two main sources of data: Bloomberg and Eikon. Indeed, Bloomberg provides more accurate information with respect to Eikon. Specifically, Bloomberg data are provided in a quarterly frequency, where Eikon only allows to work with annual information. Even so, there was noticed that the coverage of data furnished by Eikon was almost double with respect to the one offered by Bloomberg, for the considered universe of firms (~80% vs. ~45%).

Differently from market data, financial information was stored in different frequencies, depending on the type of data. In this case, within the sub-database there can be identified four different groups of information:

- *Balance Sheet*, which includes 22 indicators.
- *Income Statement*, that collects 32 measures.
- *Cash Flow*, a group containing 12 different metrics.
- *Financial Ratios*, the last group, made by 14 indicators.

For the first three areas, the information was downloaded in a yearly frequency since, after a check, it was discovered that the quarterly frequency created some problems linked to the alignment of data between companies in term of dates. Even though the decision was to utilize the yearly frequency, the information can be considered accurate because for each firm Thomson Reuters Eikon returns the effective date of publication of the financial statements.

With respect to the financial ratios area, it was possible to directly applicate the quarterly frequency because the ratios change in a quite frequent way and are computed by the data provider itself. Another important aspect to underline consists in the currency; with the purpose of ease the analysis, all the information was collected in euro, independently from the country where the firms were located.

Hence, while market data were all collected in a daily frequency, in this case we had data collected both in a quarterly and in a yearly fashion. For this reason, it was necessary to create a monthly frequency time series for each company in order to normalize the different time series, because different firms publish the financial statements in different dates over the year. So, for each firm there is a monthly time series where information is stored only in the months where it was released. No filling or manipulation to data was performed; so, you will find only a month with data for each year. This decision was taken in order to allow people to manage the data in the best way needed for the analysis.

3. ESG & Credit Ratings

The sub-database was built focusing on ratings, both on ESG aspects and on creditworthiness. In this case, information was provided mainly by Bloomberg and MSCI. Only information on creditworthiness was gathered by Thomson Reuters Eikon. The frequency of the data varies a lot, depending on the way providers publishes each type of measure. So, it was collected information stored in a daily, weekly, monthly, and quarterly frequencies. Also, metrics are provided in a wide variety of unit of measures, depending on the nature of the analysed measure.

The overall number of measures composing the database is *88*. Only six of those metrics are referred to credit rating, while the remaining 82 are related to ESG scores assessed by ESG rating agencies. There were considered several ESG rating agencies in order to mitigate problems related to the lack of comparability and the non-transparency of the assessment processes. Some of the agencies considered in this type of analysis were Refinitiv, Sustainalytics, ISS, MSCI, RobecoSAM, Standard Ethics, and CDP.

Due to the divergence of frequencies, after the download process, the data management process had to tackle this issue. The problem was overcome standardising the time series, creating a monthly frequency series for each company. Differently from the financial data, in this case it was performed a fill of the next 11 empty cells, in order to have fill the information for each cell in the year.

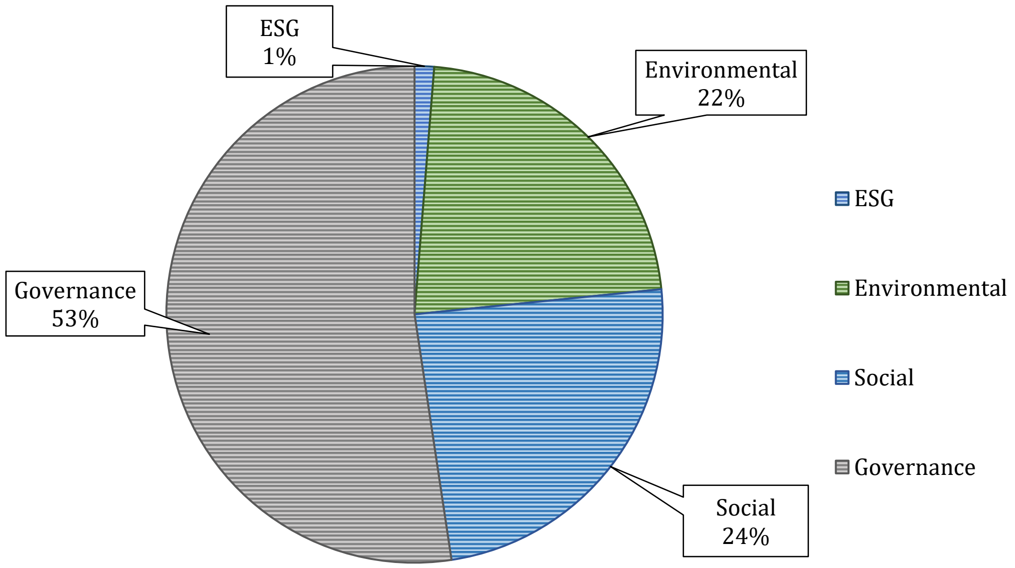
4. ESG Data

This is certainly the most relevant part of the database considering that it collects the type of information for which the project was developed, and it is crucial to fulfilling its aim. So, in the process of researching adequate measures, the purpose was to identify the highest possible number of indicators in order to build a comprehensive set of data, necessary to perform several and appropriate analyses, avoiding the largest possible type of biases.

Over the abovementioned reasons, data were extracted from three different sources of information: Bloomberg, Thomson Reuters Eikon, NRG, and CDP. The overall number of measures collected is *945*. The whole number of metrics is furnished in a yearly fashion.

The metrics map is a crucial instrument and allows to understand the frequencies by which each metrics is provided, also giving information about the provider of the specific indicator. The distribution of the 945 indicators among the different ESG pillars for the sub-level database under analysis is graphically expressed by Figure 4, which is useful to easily and immediately interpret this information.

Figure 4 – Metrics breakdown between ESG pillars.



Due to the importance of this 4th database, it is interesting to investigate the role of each provider in the provision of the overall list of measures. So, a more analytical breakdown of the indicators between the four different data providers used is shown by Table 3.

Table 3 – Measures breakdown between data providers.

ESG pillar	Eikon	Bloomberg	CDP	NRG	Total
ESG	3	9			*12*
Environmental	125	49	35		*209*
Social	181	48			*229*
Governance	130	29		336	*495*
Totale	*439*	*135*	*35*	*336*	***945***

As can be seen in Table 3, CDP only provides data about environmental aspects, while NRG furnishes exclusively governance information. This can be explained by the fact that these companies are specialized in single and precise ESG pillars. For instance, CDP is a not-for-profit charity that runs the global disclosure system for investors, companies, cities, states and regions to manage their environmental impacts, and it is considered as the standard of environmental reporting with the richest and most comprehensive dataset on corporate and city action[7].

5. CDS Data

This is the last sub-database, as expressed by the diagram that represents graphically the whole database, presented by Figure 3. This database collects information about Credit Default Swaps for the entire universe of 14,626 companies. This type of data was gathered, as for market data, relying on the platform Eikon, developed by Thomson Reuters.

[7] In addition, each year, CDP takes the information supplied in its annual reporting process and scores companies and cities based on their journey through disclosure and towards environmental leadership. Source: *https://www.cdp.net/en*

Since CDS information is comparable to market one, data were stored in a daily frequency. Unfortunately, with regard to this category of data it was difficult to find a proper set of measures to explore in deep the dynamics of prices; in fact, Eikon provides only few metrics. The most interesting, collected measures used in the construction of the database under analysis can be summarised as follow:

- *low price*
- *high price*
- *open price*

Hence, differently from market data, Eikon does not provide information with respect to closing prices; as a consequence, it was not possible to derive any type of information about volumes of trading.

CDS are specific instruments, and it is not easy to find information about them as it is for equities. For this reason, from the overall sample of 14,626 firms, only *306* unique identifiers of CDS were found. To sum up, the complete CDS database is composed only by CDS data on *306* companies.

In this case the data management was very quick, similar to the one performed for market data, and finished with the exportation of the database, creating for each company with available data a daily time series.

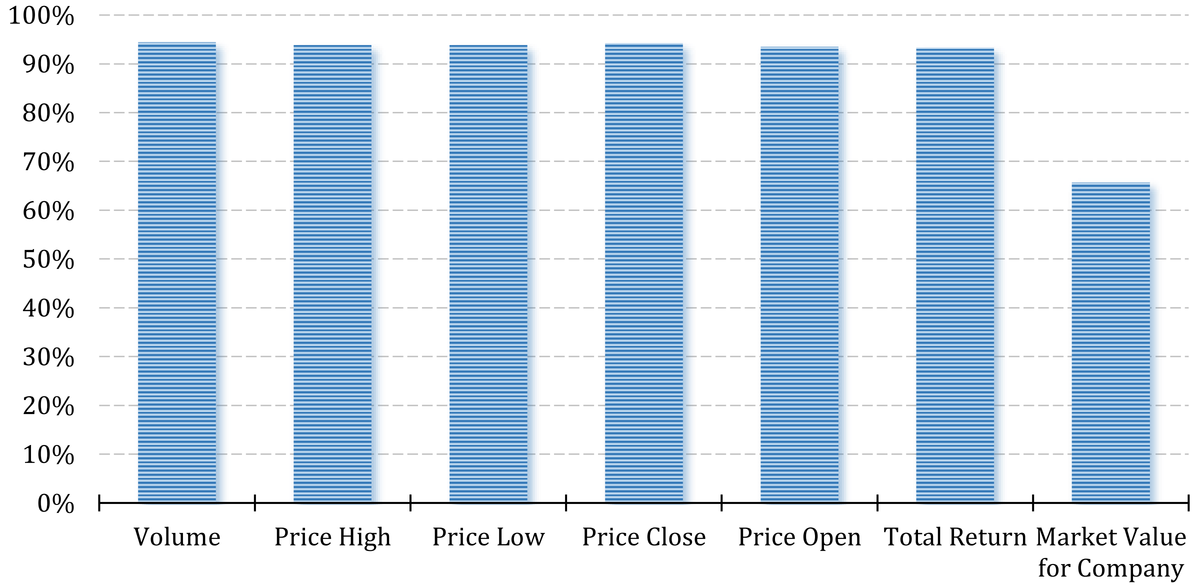
IV. DATABASE STATISTICS

This section was developed to introduce some important descriptive statistics linked to the database. In particular, the fourth sub-level database, the one referred to ESG factors, was deeply explored and analysed. So, the aim of the section is to find the main features of the metrics for the sample of firms under study. Each sub-database will be examined, either with descriptive analysis or with an analysis on the coverage.

1. Market Data

As stated before, the database related to market data is composed by only 7 fields, which are linked to the stocks of the listed companies composing the entire sample. This sub-database was developed relying on information collected from Thomson Reuters Eikon, since it allowed to have a great coverage. The coverage for each field is summarised in by Figure 5.

Figure 5 - Coverage of market data metrics for the 14,626 firms.



The coverage was computed for both this and the other databases as the ratio between the number of companies with at least one data and the total amount of companies in the universe of interest (14,626).

In this case, it was not developed a descriptive analysis since there is a wide heterogeneity in market information; in fact, the price of the stocks may differ a lot, depending on the policy adopted by each company. In addition, the volume of the stocks traded is influenced by a great heterogeneity of factors.

2. Financial Data

As for the market data database, there would have been not useful to investigate about descriptive statistics in the sample for financial data, due to the heterogeneity of the sample in terms of both size and revenues. Differently from the first sub-database, here there were developed four different sections, each one referred to a specific financial aspect: Balance Sheet, Income Statement, Cash Flow, and Financial Ratios.

Also in this case, a graph was created to graphically express the coverage of the different metrics; the metrics were grouped by the aspect they are linked to. This graph is expressed by

Figure 6.

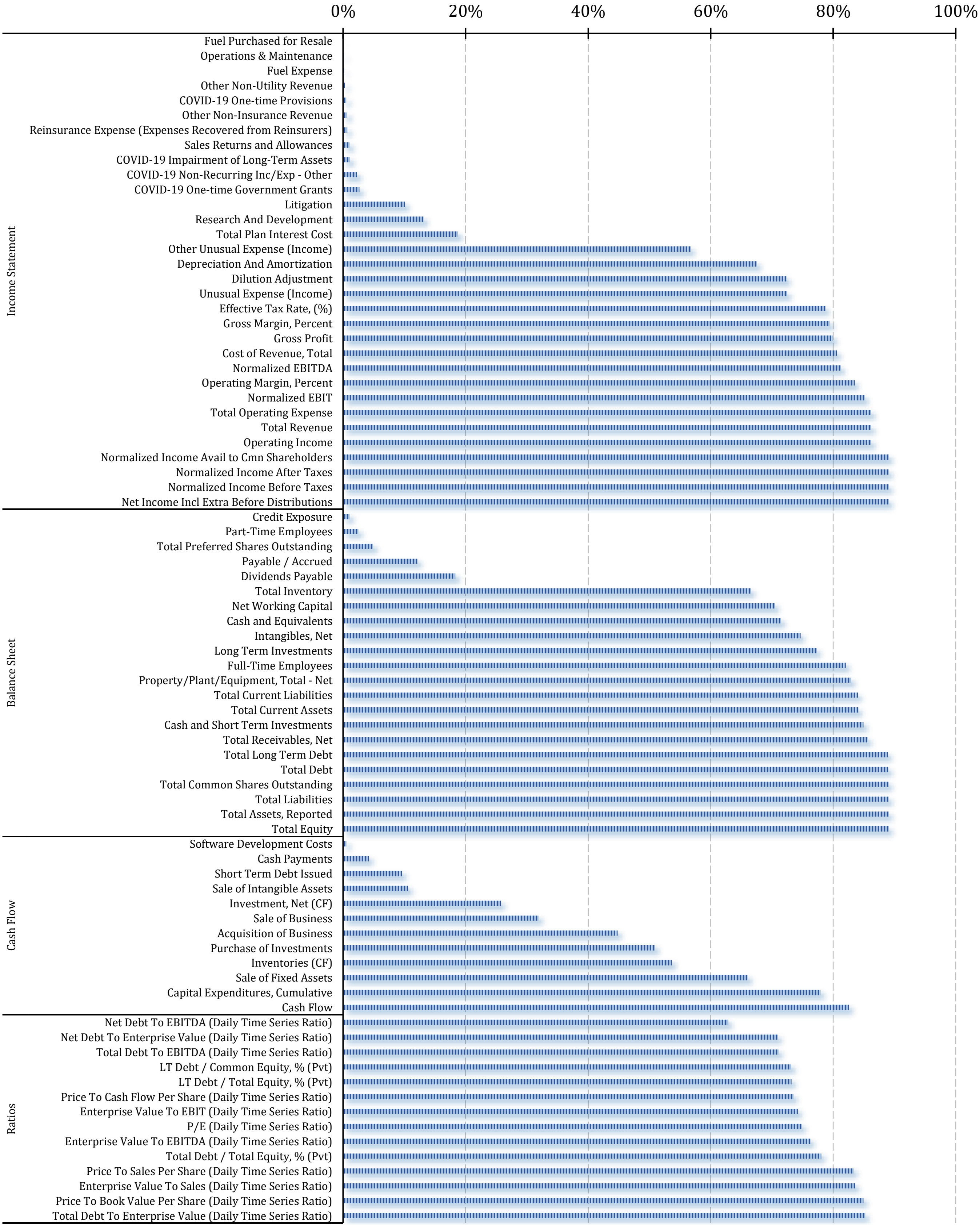


Figure 6 - Coverage of financial data metrics for the 14,626 firms.

In the above figure, you can see how most metrics show a considerable coverage, of about 80%. The indicators for which the coverage is not so high are the ones linked to particular aspects, such as COVID-

19. Examples of fields with about zero coverage are fuel expenses or software development costs, metrics difficult to measure for the companies and specially not compulsory to disclose.

Anyway, the most relevant metrics have a coverage which is very good focusing on the possibility to develop a proper analysis. In fact, the most common metrics and ratios are available for the majority of the firms in the sample; this allows the possibility to have a huge number of possible indicators to use for different kind of analysis and investigations.

3. ESG & Credit Ratings

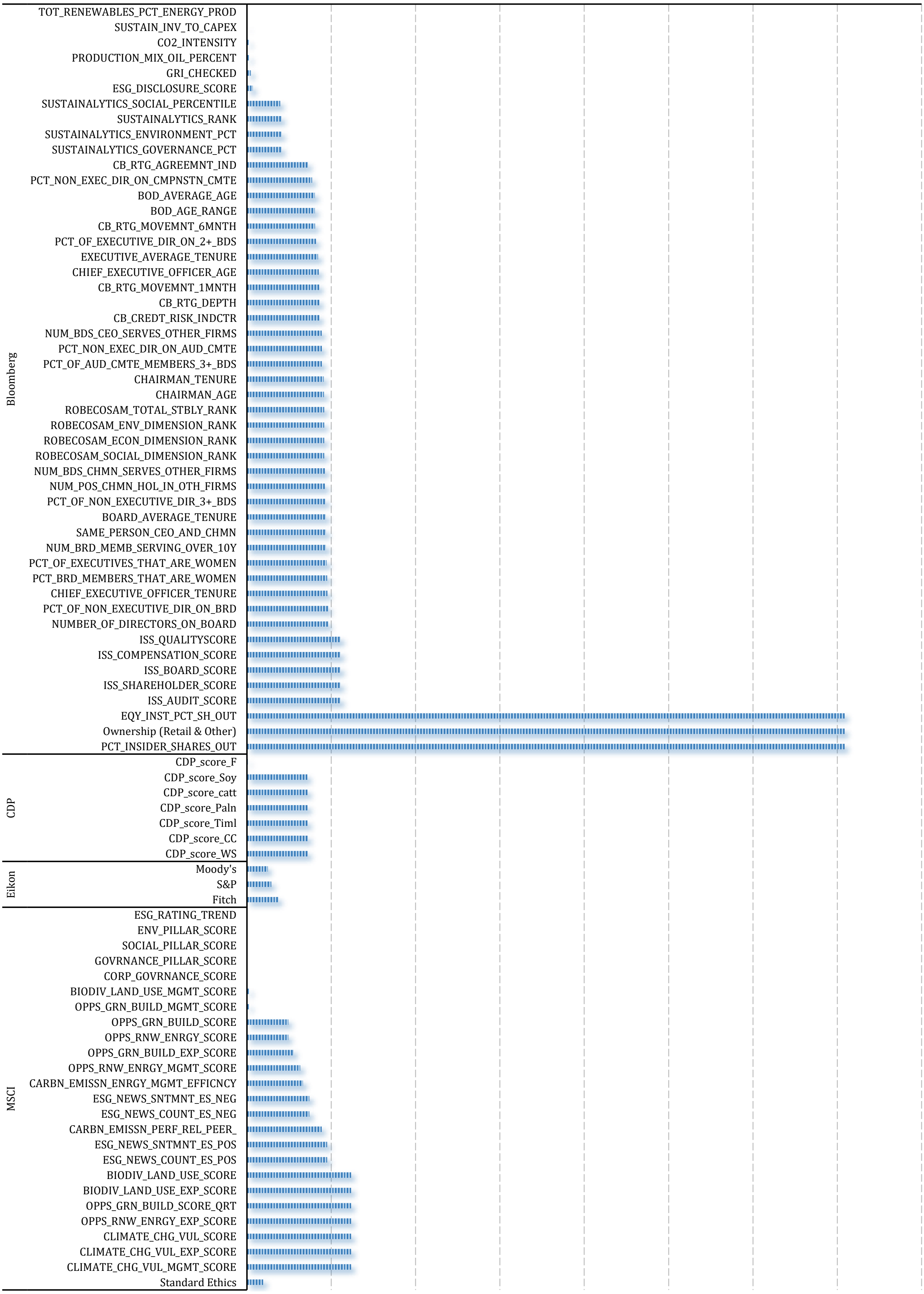
This sub-database is focused on both ESG and Credit ratings, collecting an overall number of 88 metrics. In this case, there were collected various scores, referred not only on the general ESG general aspect, but on different aspects, even more specific than the single E, S, and G pillars.

In this case, not only Thomson Reuters Eikon was used as data provider of reference in the download process. On the contrary, also MSCI, Bloomberg, CDP and Standard Ethics were used, as discussed in the previous chapter. This database is interesting since collects information that summarises ESG aspects of the companies in the sample, differently from the fourth database, which collects granular information.

As done in the previous two sections, also in this section there was developed a graphical analysis about the coverage of each metrics in the third database. This information is exhibited in Figure 7.

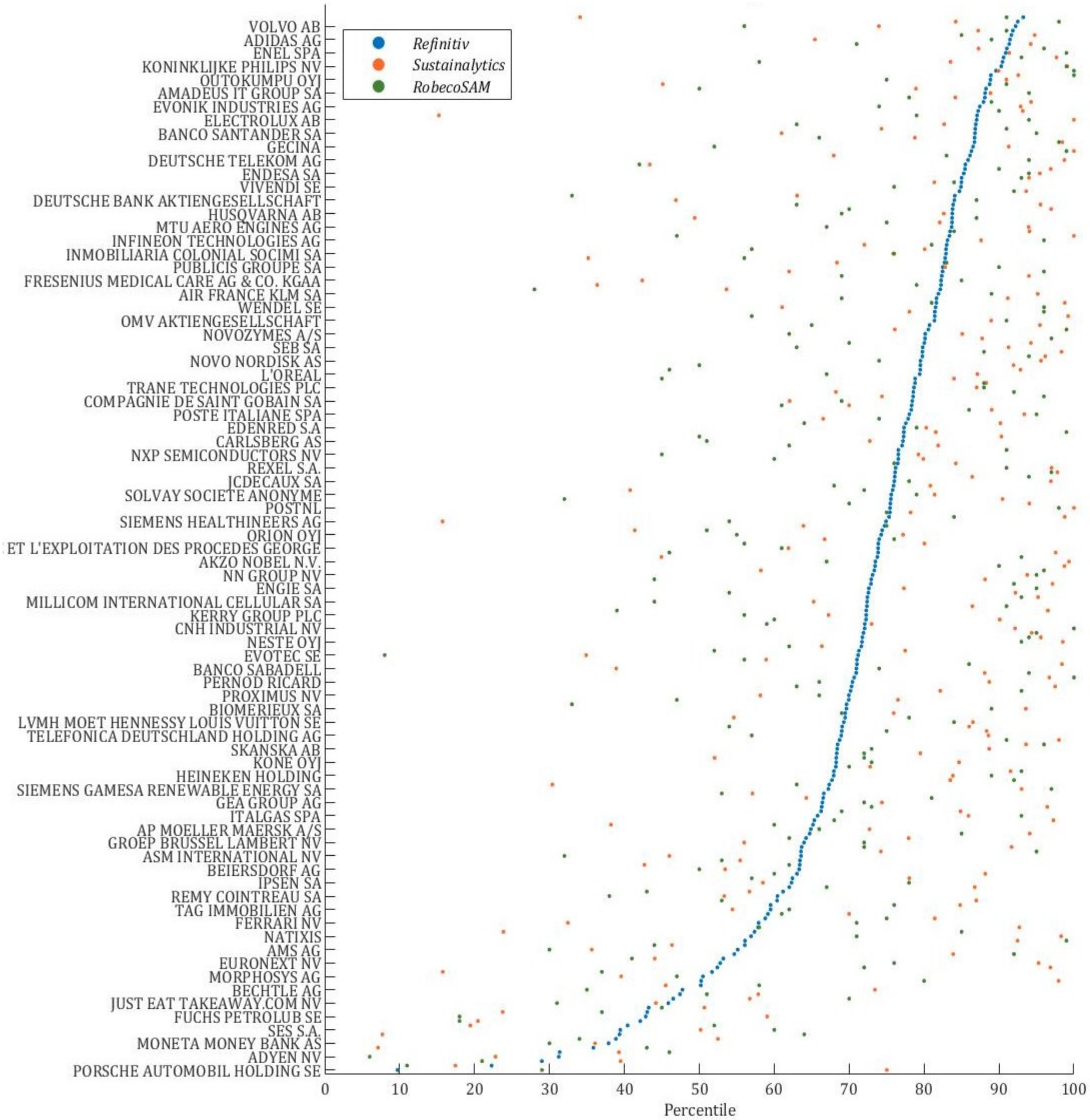
Figure 7 is very interesting and allows to see at some important features of the data in the database. First of all, the coverage is extremely low compared with the average one shown by the both the financial and the market data. You can see that the metrics derived from Bloomberg have an interesting coverage on the average, which is around the level of 10% (1,463 companies). Bloomberg provides data of ISS, Sustainalytics, and RobecoSAM, three important ESG rating agencies; those ESG ratings are very relevant to investigate the misalignment expressed by many academic papers in the last years. Then, Bloomberg offers some metrics developed by itself, referred to both environmental and governance pillars, not to the social one. Another provider considered in the construction of the database is CDP; in particular, we considered the different environmental scores published by it, linked to a letteral classification. The only CDP metric with a low coverage, near to zero, is the one referred to the forest treatment. As you can guess, this is a not so relevant aspect in the environmental assessment of a firm behaviour. In database three, we rely on Thomson Reuters Eikon only with respect to the collection of credit ratings. In fact, the platform allows to gather information on creditworthiness assessments published by the big three (Moody's, S&P, and Fitch). The coverage is very similar for each rating agency considered. Then, we have used information from MSCI. In this case, you have to take into account that the information is only cross-sectional and refer to the last year disclosure (2020). In particular, strangely it was not possible to collect relevant information about ESG scores assessed by the rating agency. Perhaps, more information could be retrieved directly with a subscription to the MSCI platform.

Figure 7 – Coverage of ESG & Credit Ratings information for the 14,626 firms in the sample.



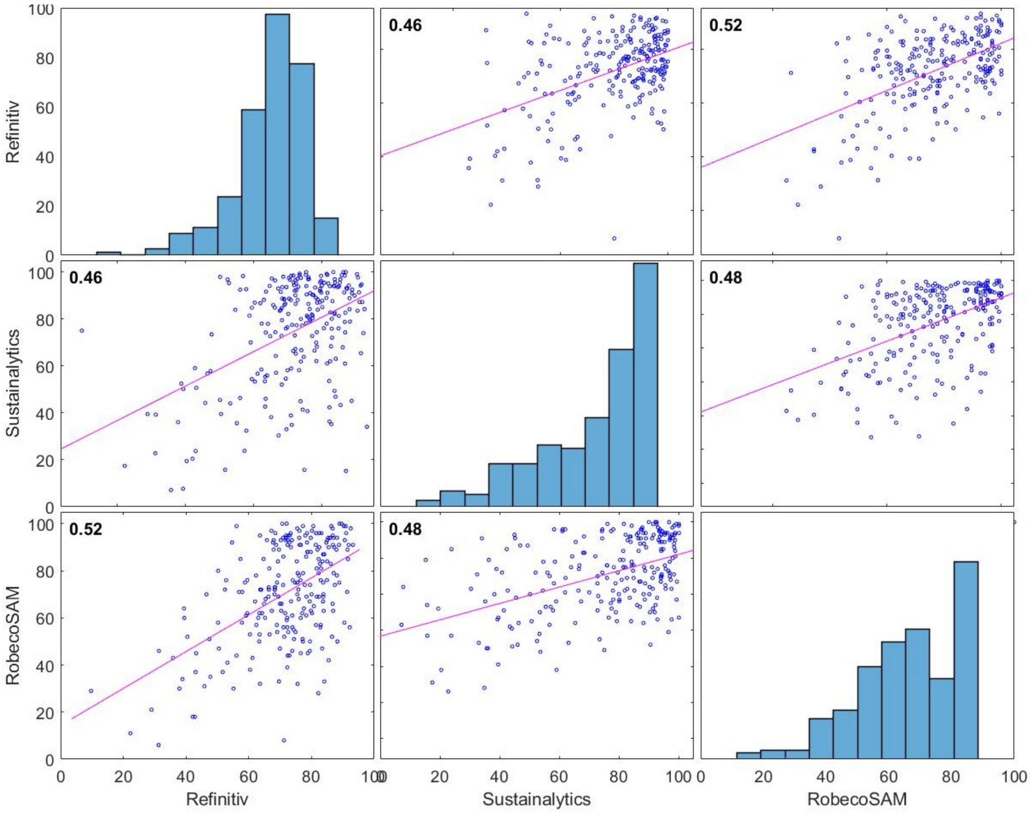
As stated before, the ESG ratings are extremely useful because allow to highlight the observable divergence studied by different researchers. The possibility to dispose of ESG ratings provided by different agencies is quite interesting. In fact, the inadequate regulation and the divergence of assessment methodologies was easy to verify since we have the ratings for a common sample of firms. In particular, we developed the graphical representation of this divergence for a common sample of 271 companies, relying on the ratings published by Refinitiv, Sustainalytics, and RobecoSAM. The comparison was implemented in a straightforward way since all the ratings are expressed in percentiles. Figure 8 exhibits the divergence between ratings and relies on data published in 2020.

Figure 8 - ESG scores divergence.



In addition to this graphical analysis, a simple computation of the correlation between ratings was performed. In fact, we were interested in find a simple, analytical measure which would be able to summarise the convergence/divergence between ESG rating agencies considered in a very intuitive way. The correlation matrix was also represented graphically; this graphical representation is exhibited by Figure 9.

Figure 9 - Correlation matrix of the three ESG rating agencies considered.



As you can see, the ratings provided by the three agencies do not strongly agree. In fact, the correlation matrix plot is very interesting and insightful since it gives the opportunity to better understand the divergence of the ratings shown in Figure 8. A key aspect would consist in identifying the key drivers guiding the divergence of the ratings. Anyway, this aspect is very difficult to implement and must be developed through precise analysis. investigations.

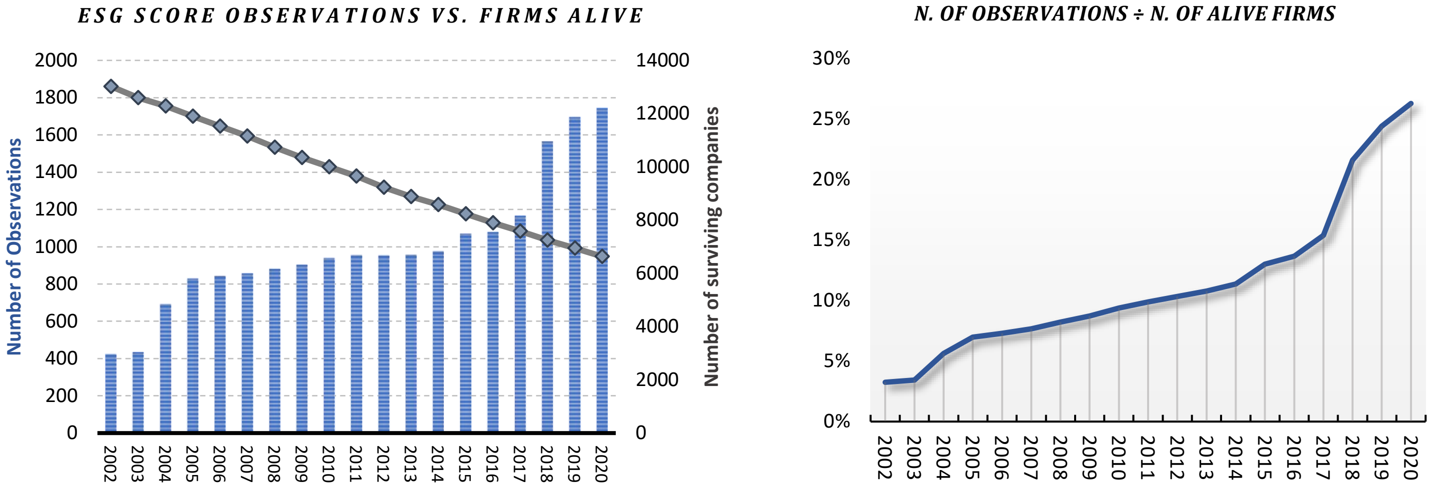
4. ESG Data

Looking at this database, there is a wide variety of metrics, precisely 945; in fact, in this case it is not possible to summarise the coverage of each metrics through a unique graph, as done in the previous sections. Anyway, some relevant aspect may be pointed out. In fact, the database includes, in addition to ESG ratings provided by Refinitiv, also granular metrics, such as Scope1, Scope2, and Scope3 emissions, that are the emissions related to CO2.

Differently from the information expressed in market and financial sub-databases, the data involved in the third and fourth databases is very poor in coverage for the first ten years of the time horizon covered by the database. In fact, non-financial information has started to be disclosed only in the recent years, due to the lack of regulation. In order to display this phenomena, there was developed a graph summarising the fact. This was performed taking as reference the ESG rating metric provided by Refinitiv. The graph was developed considering that during the time horizon, a lot of companied died, as show previously. Indeed, starting from a universe of 14,626 companies, only 6,370 are currently active.

The analysis expressed just before is summarised in Figure 10.

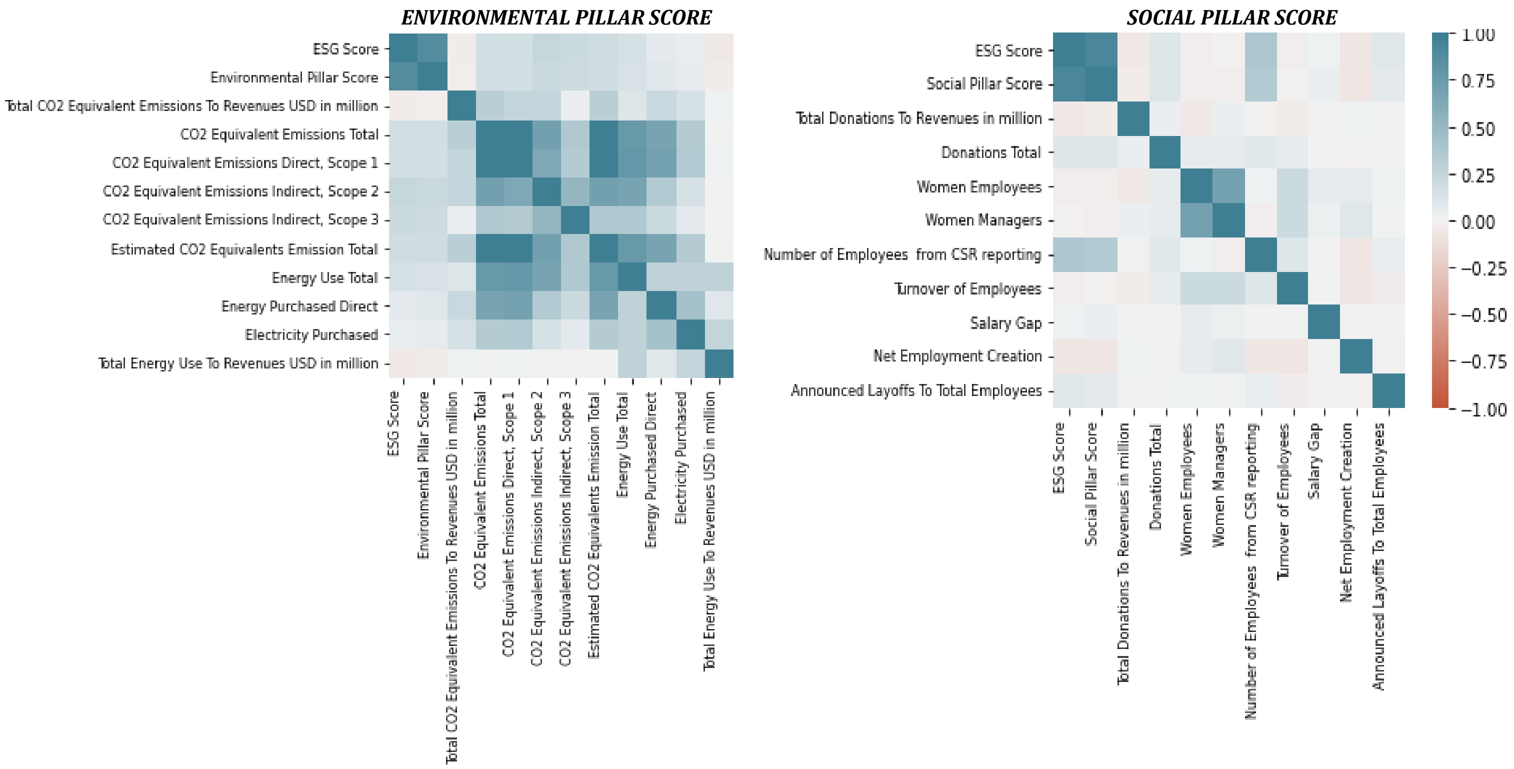
Figure 10 - Description of the data availability for Refinitiv's ESG ratings.

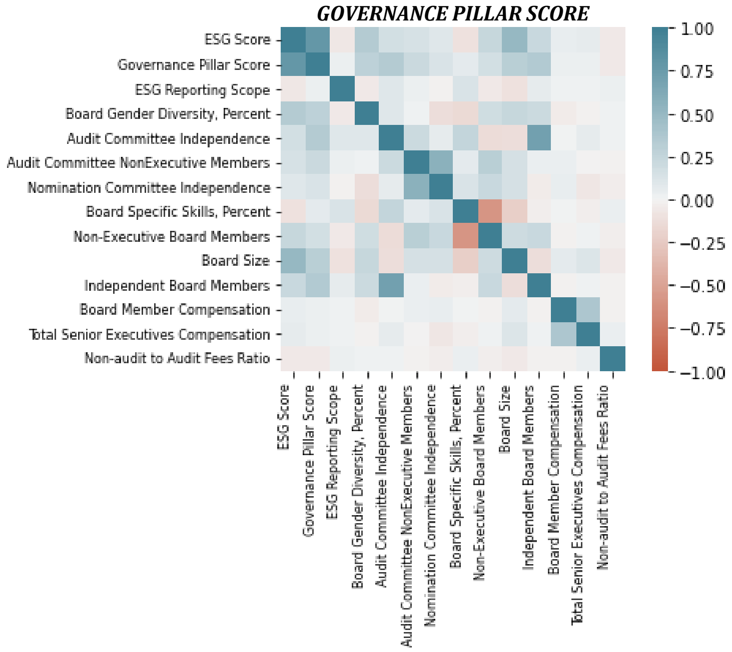


This figure is interesting and evidently shows how non-financial, ESG information, has grown in disclosure in the last three years. This grown is so evident and coincides with the development of the international regulation about non-financial aspects and their disclosure. In addition, the focus on those aspects has been relevant in a wide variety of conferences.

After this brief analysis related to the coverage, another interesting analysis was performed based on the metrics provided by Refinitiv. In fact, the internal correlation between pillar scores and granual indicators was developed through a hitmap, which is able to graphical summarise a correlation plot in a intuitive way. In particular, a graph for each ESG pillar was developed. The three graphs are enclosed in Figure 11.

Figure 11 - Internal correlation of Refinitiv scores by pillar.





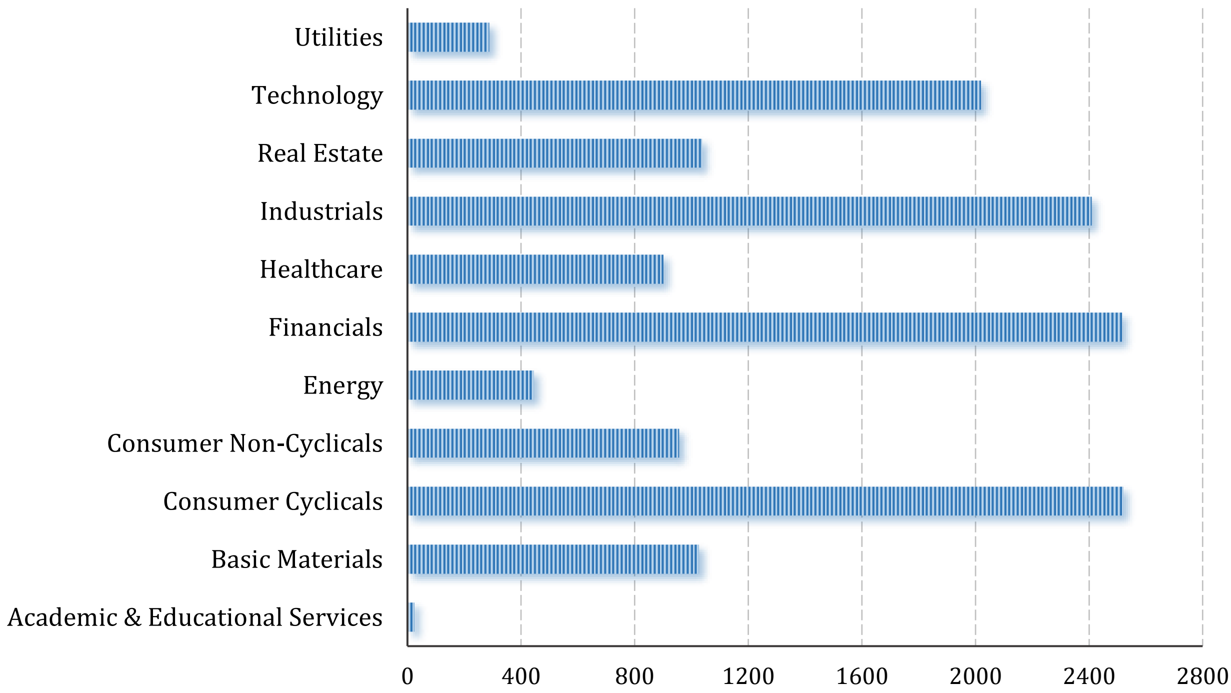
As you can see, the internal correlation is very scarce. In fact, there is a low correlation between each pillar score and the metrics referred to that E, S, or G aspect.

APPENDIX A – Main features of the selected universe of companies.

Table A.1. Analytical country breakdown of the overall sample of companies.

Country	Active	Dead	Suspended	*Total*
Austria	69	137		*206*
Belgium	126	187	1	*314*
Bulgaria	184	121	1	*306*
Croatia	70	66		*136*
Cyprus	86	80	2	*168*
Czech Republic	12	76		*88*
Denmark	287	284		*571*
Estonia	17	5		*22*
Finland	165	119		*284*
France	660	1023	15	*1698*
Germany	745	974		*1719*
Greece	164	297	9	*470*
Hungary	34	44		*78*
Ireland	48	94		*142*
Italy	352	392	8	*752*
Latvia	16	19		*35*
Lithuania	28	24		*52*
Luxemburg	38	65		*103*
Malta	29	2		*31*
Netherlands	128	236	1	*365*
Poland	503	256		*759*
Portugal	47	83	1	*131*
Romania	130	66	7	*203*
Slovakia	14	34		*48*
Slovenia	28	40		*68*
Spain	222	203	4	*429*
Sweden	780	467		*1247*
United Kingdom	1388	2813		*4201*
Total	***6370***	***8207***	***49***	***14626***

Table A.2. Analytical and graphical sector breakdown based on TRBC Classification[8].



TRBC Economic Sector	N. of firms
Academic & Educational Services	*24*
Basic Materials	*1027*
Consumer Cyclicals	*2521*
Consumer Non-Cyclicals	*957*
Energy	*445*
Financials	*2517*
Healthcare	*902*
Industrials	*2409*
Real Estate	*1038*
Technology	*2020*
Utilities	*288*

[8] Thomson Reuters Business Classification (TRBC) is a market-based classification scheme, similar to GICS. It classifies companies on the basis of degree of impact on markets, rather than establishment-based classification systems, such as NAICS.

APPENDIX B – Specific information about the 8,010 dead firms.

Figure B.1. Dead firms' breakdown between different years.

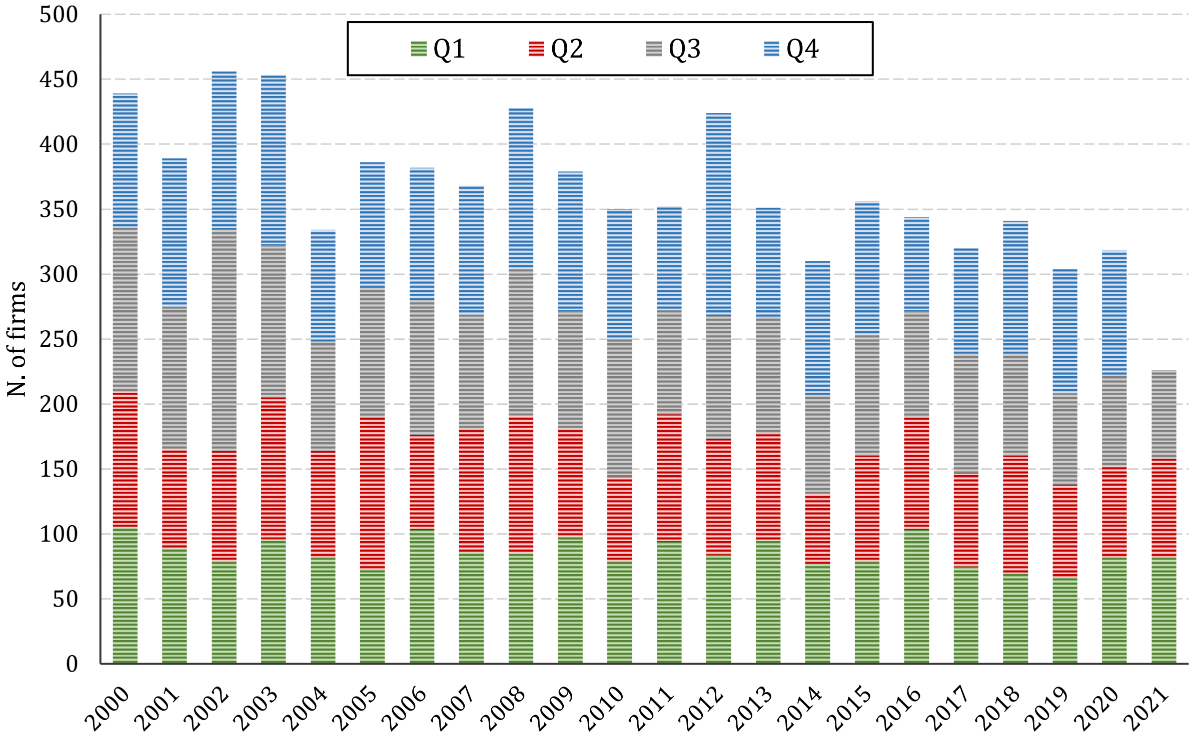


Table B.1. Analytical decomposition of dead firms between different periods.

Year	Q1	Q2	Q3	Q4	Total
2000	105	104	127	103	*439*
2001	89	76	110	114	*389*
2002	79	85	170	122	*456*
2003	96	109	117	131	*453*
2004	82	82	84	86	*334*
2005	73	117	99	97	*386*
2006	103	73	104	102	*382*
2007	86	94	89	99	*368*
2008	85	106	113	124	*428*
2009	98	83	90	108	*379*
2010	79	65	106	100	*350*
2011	94	98	80	80	*352*
2012	84	89	95	156	*424*
2013	95	82	90	84	*351*
2014	77	54	76	103	*310*
2015	79	81	93	103	*356*

2016	103	86	82	73	*344*
2017	75	72	92	81	*320*
2018	70	90	79	102	*341*
2019	67	71	70	96	*304*
2020	82	70	70	96	*318*
2021	82	76	68		*226*
Total	**1883**	**1863**	**2104**	**2160**	***8010***