Review of firm-level carbon emissions statistics

Magdalena Erdem, Taejin Park and Jose Maria Vidal Pastor

Abstract

Climate change related risks have been increasingly recognised as a major threat to economic growth and financial stability, leading to an increased focus on the exposure of firms to climate change risks in investment and policy decisions. Firms directly exposed to physical and transitional risks from climate change face the immediate challenge of mitigating its impact. Having reliable and relevant data is essential to identify, quantify and tackle such risks. Therefore, an increasing number of firms are reporting data on carbon emissions produced throughout their business activities since the Paris Agreement (December 2015).

While there have been improvements, the frequent data gaps make it challenging to evaluate a firm's GHG performance and exposure to climate risk over time or in comparison to other firms. The lack of mandatory international reporting standards and the differing disclosure policies across firms and countries poses a challenge when building aggregates at the sector, country, or regional levels that accurately represent the impact of industrial activities on the environment.

To address this challenge, a variety of statistical techniques are being used to estimate missing carbon emissions data. However, existing approaches to estimate these emissions have both strengths and weaknesses. This paper aims to review current statistical methods for estimating firm-level carbon emissions, analyse their strengths and limitations, and propose essential statistical considerations to improve the accuracy and reliability of such estimates.

1. Why is environmental data important?

Recent global warming increases and natural disasters arising from climate events have raised awareness of environmental issues. Climate change related risks have been increasingly recognised as a major threat to economic growth and financial stability. Therefore, exposure to the climate change risks has become an important consideration for investment or policy decisions. Firms that are directly exposed to physical and transition risks from climate change face an immediate challenge to mitigate its impact as these risks have impact not only on their risks but also on future expected returns.

To quantify and model the risks that organisations and companies face there is the need to access timely, complete, accurate and reliable environmental statistics. Carbon emissions are the most widely used metric to analyse environmental risk and impact. Emissions are commonly called carbon emissions interchangeably to greenhouse gas emissions (GHG) as carbon dioxide is the gas that produces most emissions from company operations. However, GHGs comprises the seven gases listed in the Kyoto Protocol.[1]

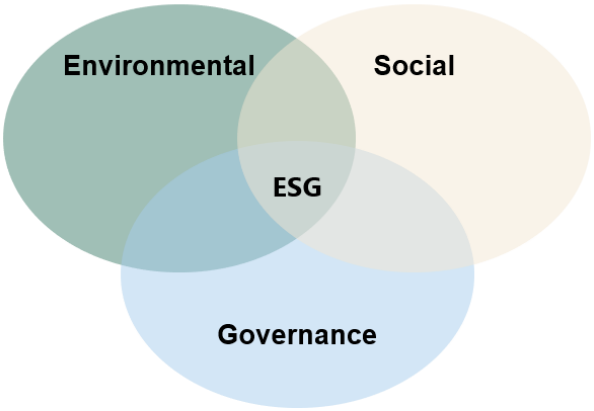
From ESG to carbon emissions

GraphMainSubheading

Graph 1

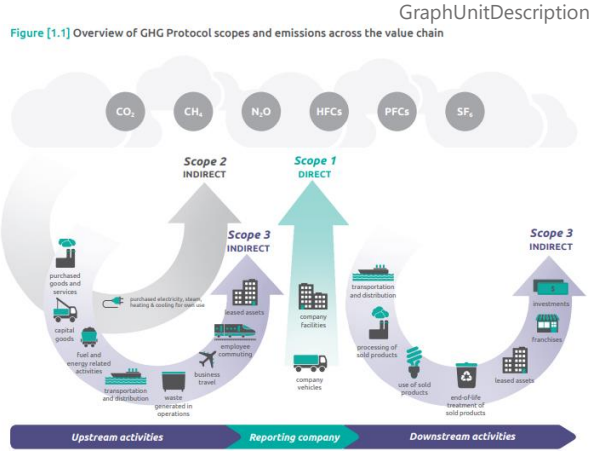
ESG

GraphUnitDescription



GHG emissions

GraphUnitDescription



GraphFootnote

Sources

[1] GHGs are the seven gases listed in the Kyoto Protocol: carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), sulphur hexafluoride (SF_6) and nitrogen trifluoride (NF_3). All GHG emissions are converted into tonnes of carbon dioxide equivalents (CO_2e) using a GHG emissions quantification process. See, for example, the International Financial Institution Framework for a Harmonized Approach to Greenhouse Gas Accounting (unfccc.int/climate-action/sectoral-engagement/ifis-harmonization-of-standards-for-ghg-accounting).

There are multiple objectives that require the availability of good data. To list a few: i) measure the current implications of environmental risks for the economy; ii) forecast future risks to allow adequate risk management practices; iii) adoption of climate policies and input into policy discussions at the local and international levels; iv) increased demand from investors to have access to information on environmental metrics affecting their assets and demanding increased disclosures.

Despite the increasing interest in firm-level microdata on carbon emissions, frequent data gaps make it difficult to conduct analysis at sectoral or country levels. The lack of international reporting standards and the low harmonization of company disclosures present a challenge when building aggregates at the sector, country or regional levels that accurately represent the impact of firm activities on the environment.

There are various ongoing efforts to fill these data gaps from different organisations. In November 2022 the G20 leaders approved[2] a new five-year phase of the Data Gaps Initiative (DGI) which includes climate change as one of the four areas of priority recommendations. The DGI on climate change will work on the development of physical and transition risk indicators that are of high importance to understand both the effects and the risks related to climate change on the economy and the financial system. [Mention ISSB standards and Sustainability Accounting Standards Board]

2. Current landscape of environmental reporting

The status of environmental data varies depending on the specific type of data. Generally, there is a growing amount of environmental data available due to the needs for increased monitoring efforts. However, there are still gaps in data coverage and quality, particularly across regions and sectors. Overall, efforts are being made to improve the collection and reporting of environmental data to better inform decision-making and promote environmental sustainability.

The three main sources from where environmental data comes from are companies own reporting, national statistics and third-party estimates.

Firm reporting has been on an increasing trend in last years, especially since the Paris Agreement in 2015. Despite the greater reporting, disclosures are far from being standardised and harmonized across countries and regions. Public companies tend to report more as they have pressures from investors and public markets. Reporting on environmental metrics requires specialized resources and not all firms have access to them, this is especially applicable in the case of private companies.

National statistics offices and international organizations make great efforts in trying to measure environmental impact. International organizations such as the OECD or the UN put together data from various sources, including national statistics, and create harmonized indicators to help in measuring environmental impacts.[3]

[2] http://www.g20.utoronto.ca/2022/221116-declaration.html

[3] https://unstats.un.org/unsd/envstats/qindicators.cshtml and https://data.oecd.org/environment.htm

Environmental statistics are still in an early stage of development in many countries, and data are often sparse and difficult to obtain.

Traditional financial data providers and specialised database platforms collect environmental data from different sources and work on the standardisation across companies and markets to provide harmonized samples to users. Since there are gaps in the data they also work to try to fill them by providing estimates for the missing data.

As stated above, coverage varies across companies and regions for various reasons. There are differences in firms' reporting between public and private companies because investors require more transparency from public companies and markets demand more information. Data reporting has resources implication as well, which affects the quantity and quality of disclosures. Private companies tend to be smaller in size than public companies and therefore have less resources and incentives to report. There are also differences across regions and countries disclosures. There are no harmonized regulations and requirements in the field, but some countries or regions are putting more emphasis on the topic, so companies within those areas phase a greater scrutiny from stakeholders and report more information than others.

Environmental data can be used for many purposes. Within Economics it allows to answer multiple questions both looking at the past and the future. Companies can calculate the carbon footprint that their operations have on the environment. At the same time, investors and funds can measure the share of the environmental impact that their assets have and make investment decisions based on the outcome. Environmental metrics are also widely used as a factor into credit ratings and credit decisions. Data can be used to take forward looking decisions based on the current risks identified by these metrics. For example, a company can calculate the implications that new policies on carbon can have on future earnings.

3. What are the data gaps?

Environmental statistics are a relatively new field and are still in an early stage of development, which results in data often being difficult to obtain and incomplete. Data gaps arise from different reasons, the following four challenges are important ones worth mentioning, but this is not an extensive list. As mentioned earlier, coverage of environmental data is very different across time, sectors and countries. The availability of data improved significantly after the signing of the Paris Agreement in 2015, however is still very far from being complete. The level of reporting varies a lot between sectors and countries. Sectors that are very exposed to climate risks such as the energy sector face increased pressures to disclose more information. Also, firms in countries with increased probability of climate events happening within their territory tend to report more than other firms. Currently there are various reporting standards and methodologies available to calculate and measure the impact used to report data. This poses challenges to both the companies reporting the impact from their operations and the investors and market participants that need to understand how the metrics are calculated. A clear example of this, is the several standards that are available to account for GHG emissions, which sometimes creates confusion. In

order to address this, and given that there is a strong need for a set of transparent, simple and comparable emissions metrics, public and private stakeholders have launched several initiatives to standardise the measurement, accounting and reporting of GHG emissions. The two most relevant methods for emission calculations are the GHG Protocol and ISO 14064. Both standards are broadly similar and compatible. One difference is that ISO 14064, even though it distinguishes between direct and indirect emissions, does not define scopes. Thus, while Scope 1 emissions in the GHG Protocol correspond to the direct emissions of ISO 14064, Scopes 2 and 3 are summarised as indirect emissions. In addition, ISO 14064 does not provide strict guidelines for the categorisation of indirect emissions and places different requirements on the structure and content of the report.

The short history of the field also translates into data revisions being very common. There are methodological changes occurring often and this results in data being revised for previous period or sometimes making it difficult to compare different reporting vintages.

There are also estimation gaps, which result from not being able to estimate data for the gaps due to the lack of enough data along the history and in the cross section of sectors or countries to build reliable models.

In addition to data reporting gaps, since environmental data disclosure and reporting is not required by harmonized regulations, there is also not external verification by third parties, like it happens with the release of financial statements by companies which are required to be audited.

There are local and international initiatives working toward solutions to fill the gaps in the data and to improve the availability of environmental related metrics for measuring the impact into financial stability. The Data Gap Initiative (DGI) on climate and the NGFS Experts' Network on Data are examples of these initiatives.

[Figure. Stylised facts on data coverage to highlight the data gap problem]

4. How are data gaps currently addressed? And what are the drawbacks in each of the models?

Since climate change has become a factor affecting companies and investors when making financial decisions it is key to have reliable data. As explained in the previous section, environmental data is still far from being complete and there is no regulation that is expected to change this in the near future. Even though there is a substantial number of companies that report on GHG emissions gaps still exist. Statistical methods can be used to overcome the gaps in the data and try to fill the missing information with estimated numbers. This section is going to review the different approaches that have been used for this purpose.

1. Average/mean sector intensity: this method first calculates a carbon intensity metric for all reporting companies available in the sample. Then aggregates the resulting metrics at the sector level and takes either the mean or the median as the industry intensity ratio. Then this ratio is applied to the companies that do not have available data.

This model is a simple approach as there should be enough data available to calculate the intensity ratio for each sector. It uses sector level information which is the highest sectoral level classification, therefore there should be enough companies with data within each of the groups. In addition it does not break down the sample by any other criteria (eg. country). These characteristics are at the same time a drawback of the model as because of its simplicity it loses the opportunity to refine the estimated data using other important factors that also affect environmental data. For example, this approach assigns the same ratio to two companies within two sub-industries of Energy sector while the real GHG emissions might be different. At the same time, the statistical measure chosen, either the mean or the median, is accepted as being representative for all the companies within the sector. There are also other important variables which are not taken into account, such as the company location. Therefore, this approach is easy to compute but lacks precision.

2. Linear model per industry: in this method, one linear model is created for each industry to estimate the GHG emissions of companies within each group. In each linear model, selected features that are particularly relevant to the carbon footprint of companies in that given industry are used to try to find those that best explain the reported carbon emissions. Then the model is applied to the out of sample companies that do not report data on GHG emissions. Common features used in these models are the industry classification, revenue, net fixed assets, energy consumption and number of employees. One common approach when the reported value of a feature is missing is replacing it with the industry average.

   This approach is easy to calculate and implement as it uses a simple linear model for each industry, which allows to customize the variables that are considered for each industry. However, the model is dependent on the data availability for each of the factors considered in addition to the environmental data. It also has the drawback that groups companies by industry, therefore considering that all the firms within a bucket are affected by the same factors. It does not take into account the location of the company or the period dimension, which can both be determinant in the GHG emissions that a company's operations generate.

3. Revenue stream factors: this approach uses a bottom-up analysis with environmental data along with information on company operations, supply chains, and financial performance. It uses a life cycle assessment approach to estimate the environmental impacts of a product or service from the raw material extraction to the end-of-life disposal. Companies' disclosed emissions and revenues are broken down by revenue stream. Then both the emissions and revenues are re-grouped at the sector level, to later aggregate them at the industry level. Finally, estimation factors are calculated for each industry group which can later be applied to the companies that do not have data available.

   The revenue approach makes use of the most granular data available from the operations and supply chain of a company. It needs to have enough detailed data for the sample firms to be able to assign the revenues to the

correct revenue stream by identifying the processes from the raw material extraction till the final use of the products. The strength of this model is that the sector classification is very detailed, but at the same time it creates a main drawback, as it prevents to use other dimensions of the data such as country or time period as it will result in not having enough data points for each bucket.

4. Machine learning method: this model, based on regression trees, is able to produce estimates for companies with enough data available. It uses multiple datasets, such as company location, size, and financial, environmental, social and governance (ESG) data, the breakdown of revenue by industry sectors, and industry-specific company data. The model estimates the data gaps based on the relationships between the variables used as an input.

   This approach uses more complex machine learning techniques to be able to find the relationship between the characteristics that are used as the input to the model. The model creates a relationship for each company which has available data, which allows to highly "customize" the estimation to the specific characteristics of each company. However, this creates drawbacks as the model is highly dependent on the information available for each company. This results in the estimation distribution output being very different for a company that has good data, including environmental, and a company which has not environmental data available.

[Figure. Estimation of two companies with the four models]

Model comparison

Table 1

	Average/mean sector intensity	Linear model per industry	Revenue stream factor model	Machine learning method
Strengths	-Simple to apply and understand -Enough data available	-Simple and customizable to industry -Takes relevant variables into account	-Most granular data used -Detailed revenue stream classification	-Individual company relationships -Customizable
Drawbacks	-Does not consider country and time dimensions -Measure might not be representative	-Dependent on data availability - Does not consider country and time dimensions	- Does not consider country and time dimensions -Highly dependent on data availability per bucket	-Complex to implement -Highly dependent on data

[1] TableNote

Sources: TableNote

[Figure. High uncertainties around the different estimation models]

5. How data gaps can be solved and challenges to overcome?

Current approaches to estimate gaps in environmental data have strengths and can help to calculate missing data for companies that either have incomplete data or do not report data at all. However, there are many challenges to overcome when estimating environmental data and current models do not consider some or many important factors that might be very relevant for the estimation of this type of data. It is challenging to come up with a model that fits all possible scenarios as there are many factors that are import and sometimes difficult to combine. To name a few: sector/industry of operation, company location, time to be estimated, and company characteristics.

The source of the data used is also determinant in the model that can be chosen. There are different levels of granularity which go from using company reported data to national statistics which are normally aggregated at some higher level than the firm itself. Most models use company reported data at some stage, and due to the approaches used they become very dependent on the actual availability of data. At the same time, since methodologies used to calculate environmental metrics are quite unmature and often modified, the models are subject sometimes to significant revisions in the data, which can affect important decisions such as the bucket were a company is placed in the case of some models or how company characteristics affect the GHG emission of a company.[4]

Even if data is available for a significant sample of companies, the data is generally starting only in recent years, therefore models cannot rely on long-history series to make more robust estimations.

The sectoral classification used is very relevant to the final outcome of the estimation. There are different industry classifications, going from very detailed classifications to broad ones. The classification chosen is a trade-off between the amount of data needed and the precision of the estimations.

As stated in previous sections, regional and country differences can be significant as companies in some regions tend to emit more than in others. Therefore, including this dimension in the estimation approach seems appropriate.

6. [Tentative] Proposed model

[Show model performance with a real example/case study]

[4] https://www.ft.com/content/8abd9680-b837-408a-a566-d2a2fe13378e

7. Conclusion