

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

---

# Title

---

*Author:*  
Michael David Hollins

*Supervisor:*  
Dr Ovidiu Serban

Submitted in partial fulfillment of the requirements for the MSc degree in MSc AI of  
Imperial College London

September 2024

## **Abstract**

Your abstract.

---

---

## Acknowledgments

Comment this out if not needed.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background</b>	<b>2</b>
2.1	Historical context . . . . .	2
2.2	Policy context . . . . .	3
2.3	Scope 3 reporting . . . . .	5
2.4	Scope 3 data challenges . . . . .	7
2.4.1	Issues with the data generating process . . . . .	7
2.4.2	Consequential problems for data analysis . . . . .	8
2.5	Attempts to model Scope 3 emissions . . . . .	11
<b>3</b>	<b>Empirical Approach</b>	<b>12</b>
<b>4</b>	<b>Experimental Results</b>	<b>13</b>
<b>5</b>	<b>Conclusion</b>	<b>14</b>



# Chapter 1

## Introduction

Figure 1.1 is an example of a figure.

The logo for Imperial College London, featuring the words "Imperial College" in a dark blue serif font and "London" in a lighter blue sans-serif font below it.

**Figure 1.1:** Imperial College Logo. It's nice blue, and the font is quite stylish. But you can choose a different one if you don't like it.



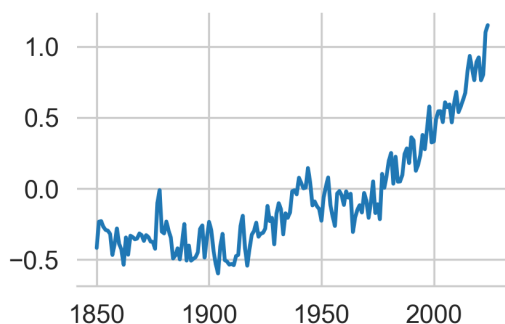
# Chapter 2

## Background

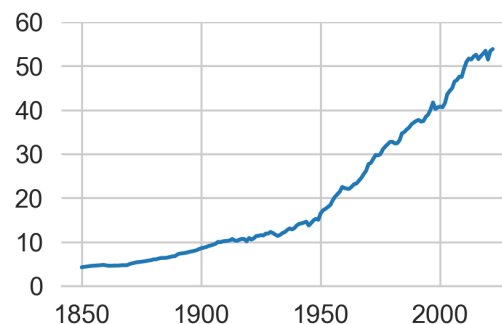
This chapter introduces the background to Scope 3 greenhouse gas (GHG) emissions reporting, standards and data. We begin with the historical and policy context in Sections 2.1 and 2.2 respectively. This serves as a springboard to consider how firms are expected to report their Scope 3 emissions in Section 2.3. Next, we cover the data quality issues inherent to this reporting process in Section 2.4. Finally, we conclude by reviewing various approaches in the literature to tackling these issues for Scope 3 modelling in Section 2.5.

### 2.1 Historical context

Beginning in the 1800s, naturalists were puzzled to find signs of glacial activity in places too warm for glaciers in modern times [? ]. As the century progressed, some scientists discovered that certain gases such as carbon dioxide caused a so-called “greenhouse effect” in their experiments [? ]; therefore, they argued that these greenhouse gases (GHGs) might affect the planet’s temperature through trapping heat in the atmosphere [? ], and so past climates could plausibly have been very different from our own [? ]. Consistent with this theory, during the twentieth century, scientists observed both rising global surface temperatures and steep increases in concentrations of GHGs [? ]. Accordingly, today’s mainstream scientific consensus is that the earth’s recent climate change is mostly driven by human activity: as economic development has increased the combustion of fossil fuels, this has raised GHG levels in the atmosphere and thereby global temperatures [? ? ].

**Figure 2.1:** Global average temperature anomalies and greenhouse gas emissions**(a)** Average global temperature anomaly

Source: Our World in Data [? ]. Global average land-sea temperature anomaly relative to the 1961-1990 average temperature, in degrees Celsius.

**(b)** Global greenhouse gas emissions

Source: Our World in Data [? ]. Greenhouse gas emissions include carbon dioxide, methane and nitrous oxide from all sources, including land-use change. They are measured in billions of tonnes of carbon dioxide-equivalents over a 100-year timescale.

Accompanying the growing weight of scientific evidence was political conviction that something must be done. Consequently, landmark international treaties such as the 1997 Kyoto Protocol [? ] and the 2015 Paris Agreement [? ] legally mandated that countries reduce their GHG emissions. Adding to the scientific and political momentum around climate change, growing public concern for the environment further catalysed domestic support for national targets such as Net Zero [? ].

## 2.2 Policy context

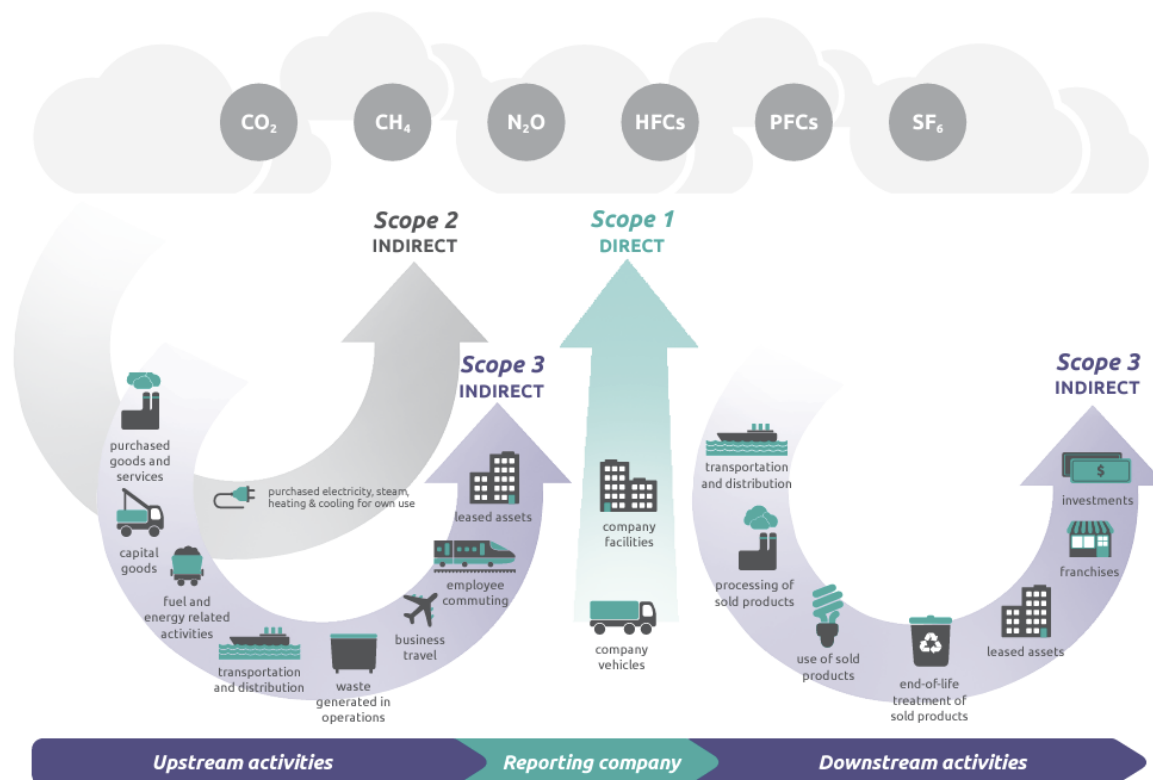
To meet GHG reduction targets, GHG emissions must be accurately and consistently measured. Therefore, as policy makers recognised the need for an international standard for corporate GHG accounting and reporting, in 1998, the GHG Protocol Initiative was launched with the mission to develop GHG reporting standards for businesses [? ]. Since then, the GHG Protocol has increasingly been incorporated into global reporting standards including the Global Reporting Initiative (GRI 305, [? ]), the International Financial Reporting Standards Foundation (IFRS S2, [? ]) and the Taskforce on Climate-related Financial Disclosures (TCFD, [? ]).

Simply put, the GHG Protocol provides a framework for firms to account for their GHG emissions, which are first divided into *direct* and *indirect*. Direct emissions come from sources that are owned or controlled by the company. On the other hand, while indirect emissions are a consequence of activities of the company, they occur at sources owned or controlled outside the company.

To help delineate this further, three “scopes” are defined for GHG accounting and reporting purposes. First, **Scope 1** emissions are **direct** GHG emissions from sources which are owned or controlled by the company. Second, **Scope 2** emissions are the **indirect** GHG emissions from the generation of energy that is purchased, including

electricity, steam, heat and cooling. Finally, **Scope 3** comprises **all other indirect emissions** that occur as a consequence of the company's activities but from sources not owned or controlled by the company. Together, these scopes encompass the total emissions footprint of a company as summarised in Figure 2.2 below.

**Figure 2.2:** Overview of GHG Protocol scopes and emissions across the value chain



Source: [? ], p.6

In accordance with *The Greenhouse Gas Protocol Corporate Accounting and Reporting Standard*, many jurisdictions require mandatory Scope 1 and Scope 2 disclosures for publicly listed firms, including in the EU [? ], the US [? ], the UK [? ] and Japan [? ]. Furthermore, although at the moment Scope 3 reporting is optional in these key financial markets, the regulatory landscape in all of them is shifting away from voluntary efforts and towards mandatory Scope 3 disclosures [? ].

Considering the UK's policy in particular, the UK Government's stated ambition is to introduce mandatory Scope 3 reporting across the entire economy by 2025 in line with the recommendations of the TCFD [? ]. This would require each reporting entity to "comply or explain" their Scope 3 emissions. In other words, if deemed material, each entity must report their Scope 3 GHG emissions in line with the GHG Protocol format. Meanwhile, for listed companies, since 2020 the Financial Conduct Authority (FCA) has phased in mandatory Scope 3 disclosures [? ? ].

## 2.3 Scope 3 reporting

Having set out the origins and development of GHG emission targets, it remains to consider what is required for a firm to comply with the GHG Protocol *Corporate Value Chain (Scope 3) Accounting and Reporting Standard* [? ]. A detailed understanding of this process will help illuminate the issues with data quality which are covered in Section 2.4.

After the firm has divided its GHG emissions into direct and indirect, the first task is to **set the Scope 3 boundary**; in short, this means accounting for all Scope 3 emissions in the value chain and to disclose and justify any exclusions. This includes the gases contained in the Kyoto Protocol: carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) and lastly sulphur hexafluoride (SF<sub>6</sub>). Furthermore, the company is required to account for these emissions in their respective Scope 3 categories as listed in Table 2.1.

**Table 2.1: Scope 3 Categories**

	#	Category	Description
Upstream	1	Purchased goods and services	Extraction, production, and transportation of goods and services purchased or acquired by the reporting company
	2	Capital goods	Extraction, production, and transportation of capital goods purchased or acquired by the reporting company
	3	Fuel- and energy-related activities (not included in Scopes 1 or 2)	Extraction, production, and transportation of fuels and energy purchased or acquired by the reporting company not already accounted for in Scope 1 or Scope 2
	4	Upstream transportation and distribution	Transportation and distribution of products or services purchased by the reporting company between their tier 1 suppliers and own operations, including inbound and outbound logistics and inter-facility transport
	5	Waste generated in operations	Disposal and treatment of waste generated in the reporting company's operations
	6	Business travel	Transportation of employees for business-related activities
	7	Employee commuting	Transportation of employees between their homes and their workplaces
	8	Upstream leased assets	Operation of assets leased by the reporting company not included in Scope 1 or Scope 2
Downstream	9	Downstream transportation and distribution	Transportation and distribution of products sold by the company between the company's operations and the end consumer, including retail and storage
	10	Processing of sold products	Processing of intermediate products sold by downstream companies
	11	Use of sold products	End use of goods and services sold by the reporting company
	12	End-of-life treatment of sold products	Waste disposal and treatment of products sold by the reporting company at the end of their life
	13	Downstream leased assets	Operation of assets owned by the reporting company and leased to other entities not included in Scope 1 and Scope 2
	14	Franchises	Operation of franchises not included in Scope 1 and Scope 2
	15	Investments	Operation of investments (including equity and debt investments and project finance) not included in Scope 1 or Scope 2

Source: Adapted from The GHG Protocol Corporate Value Chain (Scope 3) Accounting and Reporting Standard [? ]

These distinct categories are designed to help firms organise, understand and report on the diversity of Scope 3 activities throughout their value chain. However, while aiming for completeness, it is clear that in some cases, accounting for all of these emissions will not be feasible. Therefore, the GHG Protocol concedes that while some omissions are permitted, “Companies should ensure that the scope 3 inventory appropriately reflects the GHG emissions of the company, and serves the decision-making needs of users, both internal and external to the company. In particular, companies should not exclude any activity that is expected to contribute significantly to the company’s total scope 3 emissions” ([? ], p.60).

After identifying the most likely sources of significant emissions in the value chain, the firm must then **collect data** for quantifying emissions in the Scope 3 categories. There are two main methods to do this. First, where the firm has access to direct emissions data, then they can simply use the direct measurement. In practice, the second method of calculation is much more common, in which the firm uses activity data and emission factors to estimate emissions. In any case, the actual or estimated emissions data is then converted into a carbon dioxide equivalent (CO<sub>2</sub>e) so that Scope 3 emissions can be aggregated and compared regardless of GHG type. If any data remains missing, firms may use proxy data to fill gaps.

Using this data, the firm must **estimate** their Scope 3 emissions for each category. Importantly, companies may use a combination of calculation methods both within and across categories. There are three calculation methods outlined in the Scope 3 Technical Guidance and recognised by the literature [? ? ]. First, there is a process analysis (PA) approach which is bottom-up: it uses primary and secondary production process data and its associated carbon emissions. This has a high level of specificity and focus but may be costly or impossible data to collect. Second, there is input-output analysis (IOA) which is top-down: it uses the input-output tables in the national accounts to allocate the carbon emissions of the economy to smaller units on a sectoral basis. This approach is simple to apply but by providing only averages in a population, it lacks specificity, nuance, and it is difficult to show the results of reduction efforts. Third, there are hybrid approaches which combine PA and IOA which try to combine the strengths of both: under this approach, companies usually begin bottom-up and then fill any remaining gaps top-down [? ].

Finally, to comply with the GHG Protocol Scope 3 Standard [? ], the firm must **report** their Scope 3 emissions with the following requirements:

- Total scope emissions reported separately by each Scope 3 category;
- For each Scope 3 category, total GHG emissions reported in metric tons of CO<sub>2</sub> equivalent, excluding biogenic CO<sub>2</sub> emissions and independent of any GHG trades, such as purchases, sales, or transfers of offsets or allowances;
- A list of Scope 3 categories and activities included in the inventory;
- A list of Scope 3 categories or activities excluded from the inventory, and with justification for their exclusion;
- Once a base year has been established: the year chosen as the Scope 3 base year; the rationale for choosing the base year; the base year's recalculation policy; Scope 3 emissions by category in the base year, consistent with the base year emissions recalculation policy; and appropriate context for any significant emissions changes that triggered base year emissions recalculations;
- For each Scope 3 category, any biogenic CO<sub>2</sub> emissions reported separately;

- For each Scope 3 category, a description of the various types and sources of data, including activity data, emission factors and global warming potential (GWP) values, used to calculate emissions, and a description of the data quality of reported emissions data;
- For each Scope 3 category, a description of the methodologies, the allocation methods, and assumptions used to calculate Scope 3 emissions;
- For each Scope 3 category, the percentage of emissions calculated using data obtained from suppliers or other value chain partners.

These reporting requirements are clearly not trivial. Such is the nature and extent of the requirements that some have concluded that they are practically unworkable in their current form [? ]. The following Section explores the common problems with Scope 3 emissions disclosures that are highlighted in the literature.

## 2.4 Scope 3 data challenges

Early studies exploring the nascent Scope 3 assessments found that the data were largely incomplete and lacking rigour [? ]. Moreover, even after a decade, a report by FTSE-Russell concluded that the data were still characterised by large gaps and poor quality [? ]. Furthermore, it appears that rather than getting better, Scope 3 data is actually becoming less consistent over time [? ]. This suggests that something in the data generating process (as outlined above in Section 2.3) could be driving these inadequate outcomes. This Section describes the main shortcomings of Scope 3 disclosures identified in the literature. Beginning with a review of challenges inherent in the data generating process, we subsequently highlight the problems this causes downstream data analysis.

### 2.4.1 Issues with the data generating process

Klaaßen and Stoll [? ] identify three causes of error in firms' GHG emissions data: reporting inconsistency, boundary incompleteness and activity exclusion.

#### Reporting inconsistency

Firms may disclose their Scope 3 GHG emissions through corporate channels, such as annual reports, or to third parties such as the CDP. However, there is no need for these disclosures to be consistent with each other, as they are written for different purposes and audiences. Accordingly, in line with the expectations of stakeholder theory, studies have found that on average, GHG emissions disclosed in corporate reports are lower than those reported to the CDP [? ? ]. As corporate reports are less prescriptive than the CDP questionnaire, it appears that firms use them to paint a better picture of their GHG emissions to the wider public. Notwithstanding this, different figures do not imply fraudulent disclosure: it may simply reflect the choice of valid but different methodologies [? ]. In any case, the upshot is that even

for a given firm in a given year which self-reports, there is potentially no consistent source of ground truth.

### **Boundary incompleteness**

For the 15 categories of Scope 3 GHG emissions, the GHG Protocol recommends a minimum boundary of what to include. Therefore, for each category companies are encouraged to choose the most appropriate calculation method depending on data quality and quantity. These calculation methods are either top-down (input-output based), bottom-up (processed based), or a hybrid of the two which begins bottom-up and then fills in the gaps top-down. Furthermore, for the most material categories, firms are encouraged to use primary data, i.e. data sourced directly from other firms. However, there is often limited data transparency across the value chain and so this option is commonly unavailable. Where secondary data such as industry reports and production volumes is also missing, boundary incompleteness follows. Some scholars such as Patchell [?] argue that this is an intractable problem due to the complexity of global supply chains and the prohibitive costs involved. This might explain why despite the US Securities and Exchange Commission (SEC) previously intending to introduce compulsory Scope 3 disclosures, a move that some argued was a vital step for improving coverage [? ], it has instead decided against this course of action following consultations [? ].

### **Activity exclusion**

Finally, even when firms report their Scope 3 emissions, they may choose to omit certain relevant categories entirely. Despite the encouragements of standard setters to report Scope 3 emissions that are considered ‘material’ (ISSB) [? ], ‘appropriate’ (TCFD) [? ], or ‘significant’ (GRI) [? ], firms often report what is easier to measure (e.g. business travel) rather than what is important to their carbon footprint. For example, in their 2022 report, CDP found that only 36% of reporting firms disclosed their Purchased Goods and Services (Category 1) emissions, which are relevant for almost every sector [? ]. As this remains a persistent problem, FTSE Russell [? ] has recommended that companies begin by reporting the two most material categories for their sector, which on average should encompass over 80% of their total Scope 3 emissions.

## **2.4.2 Consequential problems for data analysis**

Together, the three problems in Section 2.4.1 prove a major hinderance to analysis.

### **Missing data**

The main reason for the paucity of data comes from the nature of Scope 3 emissions: in contrast to Scopes 1 and 2, they are by construction emissions in the value chain beyond the firm’s direct control or ownership and thereby are much more challenging for a firm to collect. Accordingly, across various datasets, Scope 3 coverage

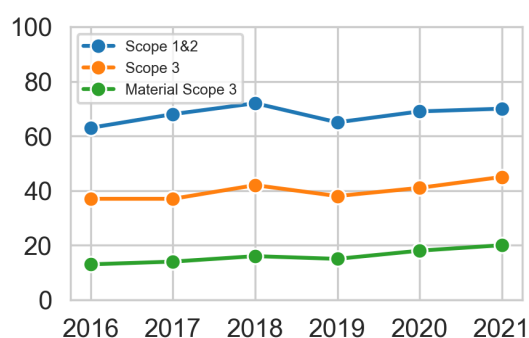
typically exists for only around half of firms.<sup>1</sup> Furthermore, as reporting remains voluntary and the propensity to report depends on firms' characteristics such as size, sector, and reporting jurisdiction, the missing data should be assumed to be 'missing not at random' (MNAR). However, current estimation models used by data providers often do not take this uncertainty into account, leading to misleading prediction intervals and inaccurate predictions [? ]. In this way, missing data is closely related to biased data, the second problem in analysis.

### Biased data

Whereas a firm which is non-reporting is conspicuous by its absence, a more subtle problem is when a firm underreports due to boundary incompleteness and activity exclusion. Without close scrutiny of a firm's methodology, the researcher risks assuming that the reported total Scope 3 emissions is an accurate reflection of its carbon footprint, rather than a reflection of what they choose to measure and report. This problem also further degrades the data generating process; as most firms which self-report underreport their emissions, firms which benchmark themselves to peer averages also underreport as a result [? ]. Both problems of missing data (non-reporting) and biased data (underreporting) are illustrated by Figure 2.3a, which shows that most firms don't report their Scope 3 data (one minus the orange line), and for those that do, they mostly report immaterial figures (gap between orange and green line).

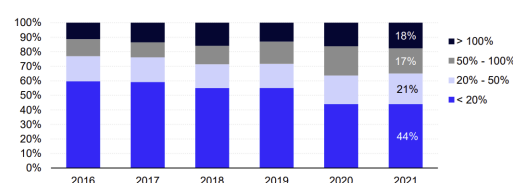
**Figure 2.3:** Global average temperature anomalies and greenhouse gas emissions

**(a)** Percentage of FTSE All-World constituents disclosing Scope 1, 2, 3 emissions



Source: FTSE-Russell [? ]

**(b)** Percentage of reported Scope 3 data within YoY variation thresholds



Source: FTSE-Russell [? ], reporting entities are from the FTSE All-World constituents that disclose Scope 3 emissions

### Volatile data

The relative immaturity of Scope 3 reporting is reflected in the high volatility of the data (see Figure 2.3b). As of 2021, over half of the disclosed values vary by over 20%

<sup>1</sup>45% of FTSE All-World constituents report Scope 3 data as of 2021 [? ]. For firms reporting to the CDP, 41% of disclosed emissions for one or more Scope 3 category in 2022 [? ]. Bloomberg and Refinitiv Eikon appear to have slightly higher proportions [? ].



year-on-year, and over a third vary by at least 50%. Although this is something that could improve over time as reporting practices evolve, any forecasting exercise using this past data will likely be struggle with robustness. Moreover, parameter estimates are unlikely to be statistically significant due to the wide confidence intervals derived from the noisy data.

### **Lack of consistency and comparability**

Besides errors, revisions, and omissions that may arise from the complex task of reporting Scope 3 emissions, there are other reasons why the data may not be comparable across firms or data sources. First, despite what may appear an overly prescriptive approach in Section 2.3, the GHG Protocol is flexible enough to provide a wide latitude in accounting and reporting methodologies. Firms must make a number of choices including their organisational boundaries, which categories to report/deem material, the calculation approach for each category, data sources, how to allocate emissions, whether to report optional extras above the minimum boundaries etc. As long as these choices can be justified according to the five accounting principles (relevance, completeness, consistency, transparency, accuracy) then they are said to conform to Scope 3 reporting standards [? ]. This methodological flexibility opens the possibility for large heterogeneity in firms' accounting and reporting their Scope 3 emissions: firms may be consistent with respect to themselves, but aggregating them into a portfolio risks comparing apples and oranges (see Box 2 in [? ] for an illustration).

Second, the GHG Protocol and Scope 3 Guidance is often complimented with standard setters' frameworks and legal requirements. While these may intend to be aligned to the GHG, invariably they introduce particular interpretations which their signatories are expected to adopt. This could lead to additional inconsistencies, especially between sectors. For example, many financial institutions have adopted the Partnership for Carbon Accounting Financials (PCAF) in an attempt to enhance reporting quality for Category 15 of Scope 3 emissions [? ]. However, as pointed out by Tang et al. [? ], the practical considerations of GHG estimation means that PCAF's framework may provide contradictory guidance, leading to non-standardisation, inaccuracy and inconsistency.

When researchers wish to compare the GHG emissions of many firms, or investors want to evaluate the transition risk of their portfolio, they frequently turn to third party data providers such as Bloomberg, CDP, ISS-Ethix, MSCI, Sustainalytics, Thomas Reuters, and Trucost. The main issue is that data often has to be estimated to fill in the gaps, and unfortunately data providers each use their own methodologies of varying transparency. These different approaches can compound the inconsistency across the data and make it challenging for people to know what the "real" Scope 3 emissions are of a collection of firms. For example, Busch et al. [? ] found that third-party estimated data is less consistent as compared to data stemming directly from corporate reports, that this inconsistency is increasing over time, and that the choice of a particular data provider may significantly affect empirical results. Sur-

prisingly, Nguyen et al. [?] found that divergence even exists between datasets for emissions values that have been reported by firms, having only 68% identical data points between Bloomberg and Refinitiv Eikon. Moreover, Kalesnik et al. [?] argue that these estimates aren't just inconsistent but they're also low quality.

## **2.5 Attempts to model Scope 3 emissions**

Mind the gaps: modelling techniques applied to S1 and 2: still struggle. Data coverage and modelling/literature.

## Chapter 3

# Empirical Approach

The literature suggests that data stemming from corporate reports is more consistent than data generated from third-party estimates [? ].

## **Chapter 4**

# **Experimental Results**

## **Chapter 5**

## **Conclusion**