# Leveraging financial transaction data for real-time emission estimation in small and medium-sized enterprises

Alec Phillpotts

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Earth and Environment

December 2025

# Front Matter

## Declaration

The candidate confirms that this work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and other authors to this work is explicitly indicated below. The candidate confirms that appropriate credit is given within the thesis where the reference has been made to the work of others.

The work in **Chapter 2** of this thesis is published in Journal of Industrial Ecology as:

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2025. Bridging the SME Reporting Gap: A New Model for Predicting Scope 1 and 2 Emissions. *Journal of Industrial Ecology*. https://doi.org/10.1111/jiec.70106.

The conceptualisation for this paper was done by all authors. All analysis was done by me, with feedback from the co-authors. I produced all the figures and tables in the article. I wrote the manuscript, which was improved by suggestions and comments from all the co-authors, the anonymous reviewers, and editors of Journal of Industrial Ecology.

The work in **Chapter 3** of this thesis is published in Journal of Industrial Ecology as:

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2026. Completing the SME carbon profile: scalable prediction of scope 3 emissions. *Journal of Industrial Ecology*. https://doi.org/10.1007/s44498-026-00003-5.

The conceptualisation for this paper was done by all authors. All analysis was done by me, with feedback from the co-authors. I produced all the figures and tables in the article. I wrote the manuscript, which was improved by suggestions and comments from all the co-authors, the anonymous reviewers, and editors of Journal of Industrial Ecology.

The work in **Chapter 4** of this thesis is in manuscript form, ready to submit:

Phillpotts, A., Leake, D., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Owen, A., *in prep*. Towards Timelier Embodied Emission Factors: Investigating methods to enable real-time EEMRIO applications.

The conceptualisation for this paper was done by all authors. All analysis was done by me, with feedback from the co-authors. I produced all the figures and tables in the article. I wrote the manuscript, which was improved by suggestions and comments from all the co-authors.

# Acknowledgements

This contents of this thesis were made possible through the continued guidance and encouragement of many people. I would like to thank my academic supervisors, Dr Anne Owen and Dr Jonathan Norman, for their support across three years, your expert knowledge, encouragement to maximise opportunities and supervision for navigating this doctoral research process has been key to the output of this thesis. To my industry supervisors, Dr David Leake and Dr Anna Trendl, thank you for providing me with this opportunity, and adding the supervision of my PhD to your already extraordinary workload - I am immensely grateful. I'd also like to extend a sincere thank you to all the other key individuals both within Lloyds Banking Group and outside for enabling this research to happen; Norbert Jobst, Jonathan Gathergood, Jasmine Wells, Trystan Davies, Andrew McCulloch, Neil Oliver, Ranil Boteju and Jaideep Mann. Also, to the Lloyds Behavioural Science Team, past and present, thank you for making the last three years such a valuable learning experience.

# Thesis Abstract

Timely and accurate emissions data is an essential first step to the understanding of business environmental impacts, as national economies look to shift away from reliance on emission generating activities. Current measurement approaches, however, represent a significant practical obstacle for small and medium-sized enterprises (SMEs). Without widespread SME emission measurement, a gap in data occurs. This has created a dual challenge: many SMEs remain disengaged from emissions measurement due to limited capacity and low reporting expectations, while external stakeholders such as large firms, financial institutions and policymakers lack the reliable emissions data for SMEs needed to inform analysis and decision-making. This thesis investigates how financial transaction data (FTD) and environmentally extended multi-regional input–output (EEMRIO) analysis can enable low burden, spend-based measurement approaches to address these issues through efficient, scalable, standardised, transparent, and timely emissions estimates.

The first two empirical chapters develop the processes required to transform large scale FTD into emissions estimates for SMEs and evaluate whether the resulting data can be meaningfully modelled across firms and industries. These processes are introduced for estimating SME Scope 1 and 2 emissions from energy and fuel spending, ensuring compatibility with established activity-based emissions factors, before the approach is extended to upstream components of Scope 3 using EEMRIO multipliers. A set of hierarchical regression models is then constructed to identify the minimal inputs needed to predict SME emissions accurately. The findings show that emissions can be estimated reliably using just a small number of firm-level variables, demonstrating that modelled approaches can be used to produce accessible emissions insights, even in the absence of detailed transactional data. These models provide a practical mechanism for producing SME benchmark emissions profiles, informed by extensive microdata, in turn allowing for external stakeholders to approximate the unreported emissions of SMEs.

The final empirical chapter then addresses the multi-year publication lag of EEMRIO databases by evaluating three nowcasting methods that vary in data requirement and underlying assumptions. Applied to the UKMRIO for a set of known years (2018–2022) the

methods achieve broadly similar accuracy, with the analysis showing that simple inflation-based updates provide a strong basis for real-time estimation, performing comparably to more complex approaches designed to adjust MRIO structures.

Taken together, this thesis advances methodological tools for producing SME emissions estimates and for updating EEMRIO-based conversion factors in lagged periods. It demonstrates how FTD can help fill the SME emissions data gap, how emissions can be predicted in low-data contexts, and how EEMRIO systems can be aligned with real-time economic conditions. These contributions support the wider ambition of creating emissions data that is both operationally practical and broadly accessible, thereby enabling more inclusive participation in climate policies.

# Table of Contents

# Table of Figures



# Table of Tables

# Table of Appendices

# List of Abbreviations

| | |
|---|---|
| ABS | Annual Business Survey |
| AE | Absolute Error |
| AIC | Akaike Information Criterion |
| APE | Absolute Percentage Error |
| bp | Basic Prices |
| CCC | Climate Change Committee (formerly Committee on Climate Change) |
| CDP | Carbon Disclosure Project |
| CPI | Consumer Price Index |
| EEMRIO | Environmentally Extended Multi-Regional Input–Output |
| FIs | Financial Institutions |
| FTD | Financial Transaction Data |
| GDP | Gross Domestic Product |
| GHG | Greenhouse Gas |
| IMF | International Monetary Fund |
| IO | Input-Output |
| IOA | Input-Output Analysis |
| IOT | Input-Output Tables |
| kWh | Kilowatt-Hours |
| LBG | Lloyds Banking Group |
| LCFS | Living Cost and Food Survey |
| ln | Natural Logarithm |
| MRIO | Multi-Regional Input–Output |
| OLS | Ordinary Least Squares |
| ONS | Office for National Statistics |
| PCAF | Partnership for Carbon Accounting Financials |
| pp | Purchasers Prices |
| PPI | Producer Price Indices |
| RQ | Research Question |
| RSQ | R-Squared |
| SECR | Streamlined Energy and Carbon Reporting |
| SIC | Standard Industrial Classification |
| SME | Small- and Medium-Sized Enterprise |
| SMRIO | Symmetric Multi-Regional Input–Output |
| SNAC | Single-Country National Accounts Consistent |
| SPIN | Scenario-Based Projection of the International Trade Network |
| STD | Standard Deviation |
| SUT-MRIO | Supply and Use Table Multi-Regional Input–Output |
| SUTs | Supply and Use Tables |
| UK | United Kingdom |
| WEO | World Economic Outlook |
| WTT | Well-to-Tank |

# Chapter 1. Introduction

## 1.1 Overview

### 1.1.1 Thesis Context

Over the course of the twentieth century, anthropogenic emissions became an increasing focus of concern in relation to changes in the Earth's atmosphere. Human activity is now understood to be the dominant driver of atmospheric change (IPCC, 2021), and annual average global surface temperatures continue to rise, with 2023 approaching the Paris Agreement's upper limit of 1.5°C above pre-industrial levels (UNFCCC, 2015; WMO, 2024). Changes to the atmosphere are contributing to more frequent and intense extreme weather events, which in turn create significant physical and economic risks and have far-reaching social and environmental impacts (Cologna et al., 2025; IPCC, 2023; Newman and Noy, 2023). To address this, many countries have pledged to reach net zero greenhouse gas (GHG) emissions in some shape or form by the mid-century (Fankhauser et al., 2022).

Whilst small and medium-sized enterprises (SMEs) may not be the principal source of GHG emissions, given their vast numbers, they make a crucial contribution to meeting any national net zero ambitions (OECD, 2023). In the United Kingdom (UK), SMEs make up 99.9% of the business population (BEIS, 2022), yet their contribution to total GHG emissions is crudely estimated (BBB, 2021), whilst the emissions embodied through their trade is unknown. This is largely due to the focus of existing emission disclosure frameworks on larger firms, leaving SMEs outside their scope, creating a significant data gap in relation to SMEs. With net zero targets fast approaching, this situation is insufficient. Measuring and monitoring environmental impacts represents a critical first step in the transition to net zero, supporting the development of awareness, knowledge, and the setting of evidence-based reduction strategies.

### 1.1.2 Thesis Contribution

This thesis focuses on this challenge, making several individual contributions. First, through a partnership with Lloyds Banking Group (LBG), one of the major retail banks in the

UK, this research has unprecedented access to proprietary financial transaction data (FTD) for a large sample of UK SMEs, enabling entirely novel analyses and large-scale insights on an otherwise challenging to observe segment of the economy. Second, all data structures and methods are aligned with the GHG Protocol, the global standard for corporate GHG accounting, making this UK focused study a transferable case study for all other economies with significant digital transaction activity. Third, by working on real-time FTD, this work addresses one of the most significant and frequently cited limitation associated with emissions data, it's timeliness. Finally, an accompanying R package is provided to operationalise the methods detailed here, enabling practitioners to reproduce estimates and apply to their own use cases.

While the UK is the sole country of analysis due to the availability of FTD, it has various benefits for being so. Firstly, as previously stated, SMEs make up 99.9% of the UK's physical business population. This, combined with the UKs legally binding climate targets of a 68% reduction by 2030 (relative to 1990), and net zero by 2050 (DESNZ and BIES, 2022a), make SME engagement crucial. As interim carbon budgets tighten, practical ways to engage SMEs in measurement and emission mitigation will only grow in importance. Secondly, the single-country focus of this work follows best practice, as emissions factors are derived from official national statistical aligned databases, enhancing robustness of results (Edens et al., 2015; Tukker et al., 2018). Lastly, as in most countries, in the UK, SMEs remain outside of mandatory emissions disclosure regulations (HM Government, 2019), creating both a regulatory and data gap. Alternative emissions data sources for SMEs are scarce, underscoring the need for scalable, low-burden measurement approaches. This work examines the role of FTD as a rational and efficient means of developing scalable models or policy frameworks to support SMEs in contributing to decreasing emissions.

Legislative amendments in 2019 established the UK as the first major economy to set a binding target to cease its net contribution to global GHG emissions by 2050 (BEIS, 2021). In recent years, the UK's territorial emissions have seen a steady decrease, with emissions decreasing to 50% of 1990 levels by 2024. The primary driver of this decline has been the reduction in emission intensity within the electricity supply sector, which alone accounts for approximately half of all national emission reductions since 2008 (CCC, 2025). As the UK electricity grid advances towards decarbonisation, the potential for further reductions in this sector diminishes, requiring greater emission reductions from other sources. Business and industry account for 25% of UK territorial emissions, with approximately half of these from SMEs (Hutton, 2020).

Alongside territorial emission targets, which capture direct emissions generated within the UK, the UK also tracks and officially publishes consumption-based emissions, which reflect the global supply-chain impacts of goods and services consumed domestically. These figures are particularly important in the UK context, where reductions in territorial emissions have coincided with increases in emissions embodied in imports. The emissions linked to the UK's consumption emissions rose by 41% between 2001 and its peak in 2007. These emissions have since fallen to 740 MtCO2e, remaining 23% above 2001 levels (DEFRA, 2025a; Owen and Kilian, 2025).

The introduction of this thesis continues with an outline of the literature surrounding this PhD research, followed by a presentation of both the research aims and objectives, and data and methods contained within this thesis.

## 1.2 An Introduction to Emission Accounting

### 1.2.1 National Emission Inventories

When measuring national emissions inventories, different accounting perspectives can be applied, each requiring distinct methodologies and producing different estimates. Territorial emissions account for all GHG emissions occurring within a nation's geographic boundaries, including offshore areas under its jurisdiction. They underpin national net zero targets and are reported under the Kyoto Protocol (UNFCCC, 1997). Production-based emissions follow the residency principle, attributing emissions to resident households and enterprises regardless of where those emissions physically occur. These inventories also contain the bunker fuels from aviation and shipping. Production-based emissions align with the System of National Accounts, enabling coherent integration of emissions and economic statistics. Whilst these allocation principles differ, both metrics quantify the emissions directly produced from the energy use, fuel combustion, and processes associated with nations or their residents.

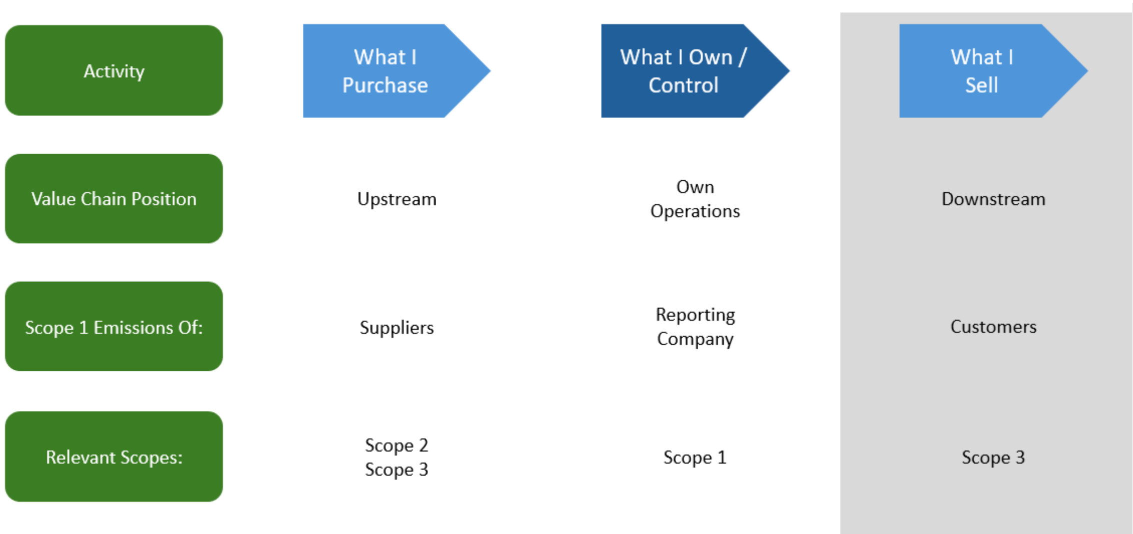
Limitations inherent to these accounting principles have led to increasing calls for the switching to, or amalgamation with other approaches (Afionis and Sakai, 2022; Peters, 2008). Of particular concern is the concept of carbon leakages through imports, where emissions are generated abroad to meet a nation's consumption demand. In these cases, domestic emission reductions may be offset by increased reliance on carbon-intensive goods produced elsewhere, meaning emissions are effectively shifted rather than reduced. This undermines the integrity of national climate targets and obscures the responsibility of high-consumption economies, as global emissions may remain unchanged or increase, even when domestic inventories fall (Davis and Caldeira, 2010; Franzen and Mader, 2018; Hertwich and Peters, 2009; Wiedmann et al., 2010). The globalization of supply chains and geographically uneven nature of consumption has amplified this limitation, underscoring the need to quantify emissions occurring outside of national borders but attributable to domestic demand (Hertwich and Peters, 2009; Owen et al., 2014). Estimates show that approximately one quarter of global CO2 emissions are embodied in international trade. In 2019 emissions associated with international trade amounted to 7.9 GtCO2 (22% of the global total), a 1.4Gt CO2 increase since 1995 (Maquet et al., 2024). A country's consumption-based emissions adjust for this, reallocating emissions from the point of production to the consumer of the final good or service.

When quantifying emissions, differences in the allocation of responsibilities across measurement approaches may stem from variations in operational control, the scope of policy, or considerations of fairness and equity.

### 1.2.2 Corporate Emission Accounting

While national inventories focus on how emissions are generated and allocated on an economy level, corporate emission accounting operates at a smaller scale, requiring individual firms to understand, measure and report the emissions associated with their own activities and value chains. The first GHG Protocol (WRI & WBCSD, 2001) was considered a major step towards enabling coherent emission accounting at the corporation level (Lenzen and Murray, 2010). The protocol has since had subsequent revisions, and now defines accounting principles, system boundaries and reporting standards, thus providing a widespread framework of emission accounting. The concepts of producer, consumer, and shared responsibility are embedded within the GHG Protocol, and operationalized through the practice of carbon footprint analysis (Wiedmann and Minx, 2008). To measure the complete carbon footprint of a corporation, the GHG Protocol defines clear, differentiated "Scopes" that capture all emissions associated with a business's operations and broader value chain. The scopes encompass emissions arising from activities related to a business's purchases, its owned or controlled operations, and the goods and services it provides. Figure 1.1 illustrates where each scope of emission occurs within an individual firm's value chain, indicating that one firm's Scope 2 and 3 emissions are the allocated Scope 1 emissions of another (Hertwich and Wood, 2018).

*Figure 1.1 - "Spheres of corporate responsibility" included within the corporate carbon footprint. Adapted from Lenzen and Murray (2010). Grey section represents the portion of the footprint not represented within FTD, and thus outside the scope of this study.*



1.2.2.1 Scope 1

GHG emissions produced from operations owned or controlled by a company are classified as Scope 1 emissions. These are direct emissions and can include stationary combustion, mobile combustion, and process-related emissions (Lenzen and Murray, 2010). These emissions appear in the central column of Figure 1.1 and could arise from direct on-site activities such as a gas boiler for heating or using a combustion engine vehicle. These emissions can be substantial in businesses with large operational complexity and scale, or those that use energy-intensive industrial processes. Calculating Scope 1 emissions can therefore be data intensive: it requires the identification of emission sources, collection of detailed activity and usage data, and application of appropriate conversion factors.

Businesses may consume various fuels in different units; natural gas may be reported in thermal or volumetric units, while liquid fuels are measured in gallons, barrels, or cubic metres. Calculating emissions from mobile combustion introduces further complexity, as calculations may need to account for both distance travelled and vehicle characteristics. To ensure consistency, fuel characteristics such as the gross calorific value should also be incorporated. Once each source has been identified, and usage data has been collected, data conversion requires accurate selection of emission factors from official datasets. In the UK, government departments publish official usage conversion factors, with the condensed file alone containing over 1,000 potential Scope 1 factors for firms to choose from. In practice, errors in matching activity data to the correct units or conversion factors frequently result in significant misreporting (Financial Reporting Council, 2021).

1.2.2.2 Scope 2

GHG emissions from the generation of purchased electricity, steam, and heating or cooling consumed by a company are classified as Scope 2 emissions. These are indirect emissions, arising outside the organisation's direct operational control, are considered upstream as they occur before the point of purchase. In Figure 1.1, Scope 2 emissions are in the first column, and mainly include electricity purchases to operate lighting, machinery, and other equipment. Compared with Scope 1, Scope 2 measurement should be more straightforward, as consumption is usually reported directly on utility bills in consistent units such as kilowatt-hours (kWh). Measuring Scope 2 emissions also involves applying conversion factors, with two distinct accounting approaches recognised under the GHG Protocol. The location-based method calculates emissions using average grid emission factors for the region where the electricity is consumed, reflecting the fuel mix of the electricity grid. Alternatively, the market-based method, incorporates the specific energy contract of the business, enabling firms to account for specific attributes of the electricity they procure. Entirely renewable tariffs could then be designated an emission factor of 0.

Although Scope 2 reporting is less complex in terms of activity data requirements, the choice between location- and market-based methods has important implications for transparency, comparability.

#### 1.2.2.3 Scope 3

Scope 3 emissions encompass all other indirect GHG emissions that occur within a company's value chain but outside of its direct operations. These emissions present the greatest measurement challenges and whilst not produced as a direct consequence of a firm's operations, these emissions can often account for the largest share of a firm's carbon footprint (Hertwich and Wood, 2018). The GHG Protocol identifies 15 distinct categories of Scope 3 emissions, spanning both upstream and downstream activities. Upstream emissions are linked to how a business spends its money. These include categories such as purchased goods and services, capital goods, fuel- and energy-related activities not captured in Scope 1 or 2, upstream transportation and distribution, waste generated in operations, business travel, employee commuting, and leased assets. Downstream emissions, by contrast, relate to the goods and services a business sell. These cover downstream transportation and distribution, processing and use of sold products, end-of-life treatment of products, as well as downstream leased assets, franchises, and investments. A business's bank account can proxy for much of the upstream component of its Scope 3 emissions since many relevant purchases flow directly through financial transactions. The main exception is employee commuting, which does not typically appear in business spending records, whilst downstream emissions are also generally not observable in this way.

Scope 3 emission measurement is complex relative to Scope 1 and 2 emissions (Blanco et al., 2016; Cheema-Fox et al., 2021). Their calculation necessitates subjective decisions on system boundaries, whilst requiring extensive external data on supplier locations, production techniques, and knowledge on various other factors within global supply chains (Acquaye et al., 2014). This introduces significant complexity and practical obstacles: precise external data may be hard to source or unavailable, firms might then rely on secondary data or industry averages which may require a level of technical expertise to acquire, and transparency in reporting is required. Each Scope 3 category demands different assumptions, system boundaries, and conversion factors, making Scope 3 the most resource-intensive and uncertain element of corporate emissions accounting.

A further complication is the issue of double counting. Under the GHG scopes, the Scopes 2 and 3 emissions of firms are the Scope 1 emissions of others, accounting for them both implies a double counting. Rather than a flaw, this overlap is inherent to value-chain accounting as several actors influence and can therefore reduce the same underlying emissions. In this sense, double counting signals shared responsibility and highlights more opportunities for emission reductions (Hertwich and Wood, 2018).

### 1.2.3 Uses of Corporate Emission Data

Once a firm measures its emissions, it typically reports them either voluntarily or in response to mandatory disclosure requirements. These requirements vary by jurisdiction and are typically defined using firm-size thresholds (for example turnover or employee numbers) and, in some cases, levels of energy-consumption. Voluntary reporting is often done through platforms like Carbon Disclosure Project (CDP), or in company sustainability reports, while regulatory disclosure is done so in the requested format. In the UK, mandatory disclosures are included in annual reports filed with Companies House, making them publicly accessible but not compiled into a centralised emissions database. Instead, the data sits within individual, annual company filings. Increasingly, providers have sought to collect and standardize emissions data from these sources, often supplementing gaps with estimation models. Key providers include: the CDP, Trucost, Morgan Stanley Capital International (MSCI), Sustainalytics, Refinitiv.

In practice, these datasets are often utilised by external stakeholders of business emission data, forming the basis for financed emissions or climate-related risk calculations (Nguyen et al., 2021; PCAF, 2022), whilst they also have varied utility in research. For instance Banerjee et al. (2025) use the Refinitiv database to study the impact of increased political risk on firm-level emissions, whilst Okimoto and Takaoka (2024) use the Trucot database to examine the relationship between emissions and firm default risk. The CDP database has also been used to examine the effect of carbon emissions on firm value (Matsumura et al., 2014) and understand the relevance and significance of Scope 3 categories (Buchenau et al., 2025).

### 1.2.4 Predicting Corporate Emission Data

There is a growing body of literature that uses corporate emissions datasets to extrapolate or estimate emissions for firms that do not report them, thereby filling the data gap that arises when only a subset of businesses produce emissions measurements. Goldhammer et al., (2017) use the joint Scope 1 and 2 emissions of 93 European companies sourced from the CDP, Corporate Sustainability Reports, and the Thomson Reuters database to produce a regression-based specification using variables on a firm's industry, size, capital intensity, and centrality of production to externally estimate the emissions of a firm, highlighting the benefits of enhanced efficiency for external stakeholders. Griffin et al. (2017) develop a similar approach using the reported emissions of 1,657 companies held by the CDP to enhance the sample size for their study on relevance of emission disclosers for investors. Nguyen et al. (2021), build on this, finding prediction error reductions can be achieved from applying machine learning techniques on a larger sample of 13,435 firm-year observations for 2,289 firms (sourced from Thomson Reuters). Further examples of machine learning based models can be found (Assael et al., 2023; Heurtebize et al., 2022a), which aim to maximise variable inputs to provide high accuracy emission estimates for the large set of companies that do not report emissions.

External prediction methods for Scope 1 and 2 emissions offer utility to a wide range of stakeholders. Policymakers use such estimates to design and evaluate environmental policies that can drive behavioural and operational change, ensuring progress toward the

UK's legally binding Carbon Budgets. For financial institutions (FIs), robust emissions data are critical for developing, assessing, and implementing financed emissions strategies, and for aligning lending and investment portfolios with climate targets. Under Task Force on Climate-related Financial Disclosures (TFCFD)-aligned disclosure frameworks, FIs are expected to measure and disclose their financed emissions, with the Partnership for Carbon Accounting Financials (PCAF, 2022) providing the methodological standard most commonly used for this purpose. As a result, the availability and accuracy of corporate emissions data are essential. Moreover, increased Scope 3 reporting has heightened the importance of accurate Scope 1 emissions data from suppliers, as larger firms seek to quantify emissions across their value chains. This creates pressure for reliable emissions data even among smaller businesses that do not currently disclose such information, a trend likely to strengthen further as Scope 3 targets become more widespread and are reinforced through investor and market pressures.

Scope 3 prediction models have also been developed, but their utility for external stakeholders is more limited. Because Scope 3 disclosure remains largely voluntary, available training data are incomplete and sparse. Nguyen et al., (2023) employ various emissions datasets to train machine learning models for Scope 3 emissions, but find that error reductions are modest regardless of the method used. Similarly, Serafeim and Velez Caicedo (2022) apply an adaptive boosting algorithm to the CDP dataset, using financial statement variables to predict Scope 3 emissions, whilst Wang and Ye (2025) also utilise machine learning techniques in a Taiwan-focused Scope 3 estimation study.

### 1.2.4 Limitations of Using Corporate Emission Data

Models that predict firm-level emissions using publicly disclosed data inevitably face several well-documented limitations. Self-reported figures frequently contain inconsistencies, such as rounding errors and unit misclassifications (Bajic et al., 2023; Financial Reporting Council, 2021). Data reliability is further undermined by discrepancies across providers: Busch et al. (2022) report similarity scores can be as low as 16% for Scope 3 emissions between data providers, while Goldhammer et al. (2017) also found the persistence of missing values and conflicting entries across providers. These variations raise significant reliability and comparability concerns, which are worsened by inconsistent or non-standardised emissions calculations (Dragomir, 2012; Talbot and Boiral, 2013). Moreover, reliance on mandatory and voluntary disclosures restricts sample sizes; a limitation often compensated for by the pooling of multi-year and multi-region observations (Assael et al., 2023), introducing further heterogeneity into the underlying data. Such challenges have direct implications for prediction models as inconsistent training data limits model accuracy and validity, whilst the sparse coverage of reporting firms hampers the ability of a model to truly generalise across firms, or sectors.

Importantly, reliance on self-disclosed information limits coverage of SMEs, which fall outside mandatory reporting requirements and often lack the resources to disclose emissions voluntarily. These firms stand to benefit most from accessible and efficient emissions estimation methods (Assael et al., 2023; Goldhammer et al., 2017), as they represent the largest source of the current data gap.

## 1.3 The SME Challenge

### 1.3.1 Status of SMEs in Decarbonising

The vast numbers of SMEs operating within the UK economy makes their engagement crucial in achieving the UKs net zero ambitions. Unlike larger firms, SMEs are exempt from formal emissions reporting requirements under frameworks such as the UK's Streamlined Energy and Carbon Reporting (SECR) scheme, which mandates the disclosure of emissions as part of broader extra-financial reporting (Meath et al., 2016). SECR eligibility cover companies who meet at least two of the following criteria: turnover of £36m or more; balance sheet assets of £18m or more; 250 employees or more. Companies consuming less than 40,000 kWh of energy per year are exempt (HM Government, 2019). The omission of SMEs has led to a situation where the vast majority of SMEs do not know how to measure their carbon footprint (BBB, 2021), and many are yet to set any reduction targets (Mazhar et al., 2022). As a key starting point to decarbonisation, this is a worrying situation.

Extending reporting obligations to SMEs presents a complex challenge. While SMEs have a significant environmental impact collectively, their individual contributions are often minimal. In addition, SMEs are overly exposed to economic disruptions, energy prices and other external changes given their lower resources. With an estimated 60% of small businesses already failing within their first three years of operation (May, 2019), effective policy should be designed with this vulnerability in mind, ensuring that additional requirements do not overburden firms.

The true contributions of SMEs to total emissions are unknown, with estimates calculated through linking SME shares of labour, turnover or value added to an industry emission intensity value. Estimates range between 43-53% of total UK business emissions, and between 29-36% of UK territorial emissions (BBB, 2021). This estimation process is crude, and rely on an assumed constant relationship between emissions and the specified economic variable (European Commission, 2022), illustrating how the limited data availability forces the use of simplified approximation methods.

### 1.3.2 Challenges to SME Decarbonisation

In addition to economic challenges, SMEs face numerous internal and external barriers to pursuing decarbonisation that are unique to firms of their size (Alipour and Rahimpour, 2020; Oyewole et al., 2024). These challenges are well documented in the literature, with Caldera et al (2019) describing six categories; a lack of financial resources; a lack of time; a lack of knowledge; risks associated with implementing a new sustainable practice; current policies and regulations; and existing organizational culture. In this study, beyond financial constraints, the most frequently cited barriers among SMEs were their lack of knowledge, skills, awareness, and time.

The theme of knowledge and awareness barriers continues in other studies. Mazhar et al (2022) found that 79% of SMEs do not actively measure their emissions, and 41% do not have any carbon reduction targets. These findings are reflected by the British Business Bank, who found only 6% of surveyed SMEs have attempted to measure their carbon footprint in the last 5 years, and only 3% had set reduction targets (BBB, 2021). A lack of sustainability knowledge is a direct impediment to applying sustainable practices such as emissions measurement (Bakos et al., 2020).

SMEs also have limited capacity to dedicate time and resources to staff training and education (Yacob et al., 2018). With many SMEs operating with ten or fewer employees, responsibilities for decarbonisation are typically spread across individuals whose primary roles lie elsewhere. Engaging external consultants can provide expertise, but the associated costs are high and such services rarely rank as a priority for SMEs. As a result, significant gaps in understanding remain, including uncertainty about the primary sources of operational emissions and the extent to which upstream suppliers contribute to a firm's overall footprint (Oluleye et al., 2025).

The lack of personnel resources to dedicate to sustainability have also led to SMEs being described as 'data poor' when it comes to understanding energy (Janda et al., 2014), with little or no energy tracking (Conway, 2015). Linked to this, actions to correct for data limitations, or decarbonise are further limited due to practical constraints such as being tenants of a premises (Mazhar et al., 2022). Central to overcoming these barriers is the improved access to more data for decision-making, alongside stronger, more efficient support mechanisms tailored to SMEs.

## 1.4 Financial Institutions and Financial Transaction Data

### 1.4.1 Financial Institutions as Data Holders

In contemporary economies, financial institutions (FIs) play a foundational role. Among them, retail banks are particularly central, facilitating capital flows and everyday transactions. The relationships fostered through this role extends to both individuals and businesses, whilst larger FIs hold important lobbying powers with policy makers (Lambert and Igan, 2019). Retail banking markets are generally dominated by a handful of large institutions that collectively capture most of the market share. In 2015, the four largest current account providers in Great Britain (RBS Group, Lloyds Banking Group, Barclays, and HSBC Group) held approximately 70% of all main personal current accounts and 83% of all business current accounts (Competition Markets Authority, 2016). Most financial data is therefore held by a small set of dominant institutions, positioning FI's well to influence and shape the sustainability transition.

### 1.4.2 Financial Transaction Data (FTD)

Of the data amassed by FIs, FTD stands out for use in emission estimation. It is intrinsically accumulated with every payment made through a debit or credit card, and when combined with an effective categorisation system, can be used to gain spending behaviour insights. Amongst others, academic uses cases of FTD include economic responses to income changes (Boudt et al., 2022), investigating markers of degenerative health or financial abuse (Trendl et al., 2025), and as a data source for household emission estimation (Trendl et al., 2023).

The concept of linking transactions to emissions factors is not new, with various so called 'carbon calculators' available for individual or business use (Barendregt et al., 2020). In the private sphere, companies are developing apps that provide individuals and businesses with transaction-linked emissions estimates (Cogo, 2022; Svalna, 2022). However, the use of this data asset to address the SME regulatory and data gap remains underexplored, despite its considerable potential. Leveraging FTD could address the key barriers to SME decarbonisation, by supporting the development of practical tools that place less reliance on self-reporting, reducing the burden on already resource-constrained SMEs.

To assess the advantages and limitations of using FTD, it is useful to compare key data characteristics with the current alternative approaches to emission estimation such as surveys or activity data. Table 1.1 categorises the characteristics of FTD under three headings: Availability, Quality, and Detail. In the table, a tick indicates a strength, a cross indicates a weakness, whilst the appearance of both denotes that the characteristic may be either, depending on the specific subtype of data.

*Table 1.1 - Overview of the strengths and weaknesses of carbon measurement sources, from the perspective of researchers. Adapted from Trendl et al. (2023).*

| Criteria | Surveys | Activity Data | FTD |
|---|---|---|---|
| **Data Availability** | | | |
| Data Access | ☑ ☒ | ☒ | ☒ |
| Big data | ☒ | ☒ | ☑ |
| Scalability | ☒ | ☒ | ☑ |
| Representative sample | ☑ | ☒ | ☑ |
| Longitudinal view | ☒ | ☒ | ☑ |
| Frequency of data points | ☒ | ☒ | ☑ |
| Demographic variables | ☑ | ☒ | ☑ |
| **Data Quality** | | | |
| Completeness | ☑ | ☒ | ☒ |
| Reliability | ☒ | ☒ | ☑ |
| Comparability | ☑ | ☒ | ☑ |
| **Data Detail** | | | |
| Granularity | ☑ | ☑ | ☒ |
| Physical Volumes | ☒ | ☑ ☒ | ☒ |
| Differentiation by brand or product | ☒ | ☒ | ☒ |

1.4.2.1 Data availability

In terms of public accessibility, generally only surveys are available at scale. In the UK, the Living Cost and Food Survey (LCFS), is the most comprehensive, official survey of annual household spend. The LCFS is openly available via the UK Data Service and frequently used to generate household emission insights (Betts-Davies et al., 2025; Kilian et al., 2023; Owen and Büchs, 2024). The business equivalent of the LCFS, the Annual Business Survey (ABS), is not publicly available, with limited access granted on a project-by-project basis. On the other hand, neither physical activity data nor FTD are made widely available. FTD is subject to data proprietary constraints and privacy protections, whilst physical activity data can only be sourced and shared by the user. This represents a major barrier to the FTD use case in climate policy contexts, and one that this thesis seeks to address.

When access is possible, FTD stands out for its data availability strengths. As an inherent dataset in modern economies, an emissions measurement could be constructed for any current account without relying on any active participation. Those within a sample are selected purely on the grounds of having an active account, reducing any self-selection bias (West et al., 2016), whilst samples of FTD from dominant FIs can be assumed to be broadly reflective of the wider population. FTD has the potential to be collected and stored across multiple years, allowing for the analysis of change over time, whilst its frequency positions it as the only real-time data asset.

#### 1.4.2.2 Data quality

Whilst surveys do provide a fuller and comprehensive picture of spending, assuming no user error, a bank account may not reflect the same complete picture. This is because individuals and businesses are not constrained to a single account, with 22% of UK adults having more than one financial service provider (Competition Markets Authority, 2016). The comparable number for businesses is not widely reported or known. Where multi-banked accounts exist, the FTD of one provider may underestimate true expenditure, and subsequent emissions. To reduce the impact of multi-banked firms, analyses using FTD should aim to minimise their inclusion during the sample selection process.

In terms of its reliability and comparability traits, FTD excels. It is collected objectively, eliminating risks associated with human errors or misreporting. Whilst calculations are standardised and transparent, ensuring consistency across observations. This is a key advantage not always present by self-reported emissions based on activity data.

#### 1.4.2.3 Data detail

FTD, however, remains limited in the level of detail required for precise emissions estimation. Transactions are recorded and classified at the merchant-level, rather than at the receipt or item level, meaning the specific products, quantities, and brands purchased are not observed (Trendl et al., 2023). As a result, variation in carbon intensity across goods and suppliers is obscured. The absence of physical quantity information in FTD also necessitates reliance on the assumption that physical consumption scales linearly with monetary value (Lenzen, 2000). In contrast, physical usage data captures both fuel type and actual consumption levels, whilst surveys, can elicit product-level details unobserved by FTD.

## 1.5 Research Aims and Objectives

This thesis takes a pragmatic approach to extending corporate emissions measurement and reporting to SMEs, with an overarching aim of exploring FTD as a means to deliver real-time, standardised estimates of SME emissions for the purposes of increased SME engagement, and external stakeholder calculations. To address this, the following research questions (RQ) are asked:

**RQ1**: How can FTD be used to produce models that capture variation in SME emissions across firms and industries?

**RQ2**: What level of accuracy can be retained whilst minimising user input requirement to produce SME emission estimates?

**RQ3**: What processes are required to ensure FTD is applied to timely emission conversion factors, thus enabling real-time estimates?

These RQs are set out below, along with brief explanations of the aims and objectives they address and an indication of where each is covered within the thesis chapters.

### 1.5.1 RQ1: How can FTD be used to produce models that capture variation in SME emissions across firms and industries?

Central to this thesis is the construction of SME emission profiles from FTD. While expenditure-based methods are acknowledged in the GHG Protocol, there is no single, standardised procedure for firm-level implementation, and expenditure-based emissions multipliers, compatible with the Protocol, are not officially published in the UK. Accordingly, this thesis develops and present a standardised, transparent, and robust process to produce expenditure-based emission estimates that are consistent with the Protocol's scope definitions.

This process is then applied to the FTD of the sample of SMEs for which this thesis has access to, producing a large sample of emissions data for an otherwise unobserved segment of the business population. This thesis then aims to address the principal limitation of using FTD for emissions estimation, its proprietary nature and subsequent inaccessibility. Whilst work on developing an expenditure-based process offers value in itself, utility is contained by the data owners (FIs) that store this data for a large-scale sample. Without their active participation, the benefit is FTD data is restricted. To address this, Chapters 2 and 3 develop models, trained on FTD emissions estimates, that can be used to produce predictions of emissions, with firm characteristics. This provides a means of accessing FTD insights, in the absence of access to the vast microdata itself.

### 1.5.2 RQ2: What level of accuracy can be retained whilst minimising user input requirement to produce SME emission estimates?

The models produced in this thesis use a hierarchical regression specification, where industry and firm level variables are used to try and capture the variation observed within the data. To assess model performance relative to practical feasibility, variables are added sequentially, allowing simple baseline models to be compared with more complex iterations.

The second research question examines the extent to which these models can be simplified without substantially sacrificing accuracy. This reflects a practical requirement for real-world application, where the aim is to provide SMEs with a low burden tool. The chosen specification is evaluated through multiple rounds of testing to assess its accuracy and stability when applied to out-of-sample data. These tests draw on several independent data sources, including withheld sample data (FTD-based estimates), predictions derived from industry-average intensities (mirroring current financed-emission methods like PCAF), and, where available, self-reported emissions from SECR disclosures. Together, these exercises assess how well the model can perform under realistic deployment conditions.

### 1.5.3 RQ3: What processes are required to ensure FTD is applied to timely emission conversion factors, thus enabling real-time estimates?

The final RQ aims to overcome a central limitation of EEMRIO database use, their multi-year publication lag. Addressing this limitation enables EEMRIO compatibility with FTD and harnesses its real-time resolution. To answer this RQ, high-frequency indicators (price indexes, output and trade data) are used to nowcast EEMRIO base tables to produce timelier embodied-emission conversion factors. Methods are designed based on the existing literature, and nowcasted EEMRIOs are benchmarked against databases for known years to assess accuracy, stability, and practical utility. This chapter aims to quantify the associated uncertainty of nowcasting approaches, with the aim of producing a transparent, reproducible approach that practitioners, policymakers, and researchers can operationalise to enable expenditure-based emissions estimation and other real-time applications, whilst retaining EEMRIO consistency.

## 1.6 Data and Methods

This section introduces the key data and methods used throughout this thesis. Further, more detailed, accounts of data and methodology applications are provided in the three empirical chapters.

### 1.6.1 Financial Transaction Data- Lloyds Banking Group

To facilitate this research project, access was granted to the otherwise proprietary FTD of SMEs banking with Lloyds Banking Group (LBG), one of the major retail banks in the UK, with a client base of over 900,000 commercial customers and 28 million retail customers. Insights derived from this banking data were then used to produce anonymised profiles of SMEs at scale. For this study, data is collected for the single year 2021. The reason for this, is at the time of analysis the most recent environmental data for the UK was available up to 2021 (production- and consumption-based accounts) (DEFRA, 2025b). In practice, firms report emissions to coincide with their financial reporting requirements, meaning year ends vary. Here, the calendar year is used as start and end dates to match public environmental datasets. The intention of this work is to serve as a case study, under the understanding that additional years could easily be modelled, with annual data updates to transactions, and emission conversion factors.

The first characteristic included in each SME profile is the firm's annual turnover, which serves as an indicator of the scale of its operations. As retail banks do not typically hold full accountancy records for all commercial clients, turnover is estimated from credit transactions recorded in the bank account. To improve the accuracy of this estimate, selected credit lines are excluded, including inter-account transfers, refunds, and HMRC rebates. This approach reflects standard banking practice and is internally validated.

The second required characteristic is industry classification, which is recorded by relationship managers when a business establishes its relationship with LBG. This follows the Standard Industrial Classification (SIC) system and is recorded at both the two- and three-digit levels. A firms SIC code identifies its' primary business activities, and groups it with similar businesses.

Finally, annual firm expenditure is compiled from business account transactions. Each transaction is automatically classified using a transaction classification system with a detailed hierarchy of 624 recipient-based categories. This schema is a custom-built, expanded version of LBG's standard internal classification system, developed to provide greater disaggregation for emissions accounting, particularly in key areas such as energy, fuels, and wider business and industry spending. Although the classification system is not itself part of the SIC structure, it is designed to be compatible with it, as merchants are assigned both an SIC section and division alongside a transaction classification code. Owing to the greater level of detail in the transaction classification system, multiple transaction codes may correspond to a single SIC code. This classification provides the foundation for linking transaction data to appropriate categories of emissions and relevant emission factors.

Upon opening an account, LBG customers consent to their data to be used for research purposes (Phillpotts et al., 2025; Trendl et al., 2023, 2025; Wells et al., 2025b). In line with LBG's data privacy requirements, all firm-level data are anonymized, and only the minimum necessary information is utilised. To further ensure anonymity, all insights produced throughout this thesis are reported at the aggregate level. Further information on sample sizes and characteristics is provided in Chapters 2 and 3, along with their respective appendices

### 1.6.2 Public Accounts – Economic and Environmental Datasets

In addition to FTD, various public datasets are required throughout this thesis to assist with the estimation and modelling of emissions, as well as the nowcasting of EEMRIO databases. Table 1.2 summarises the public datasets used, listing dataset title and which chapters they are used in.

*Table 1.2 - Summary of public datasets used throughout this thesis.*

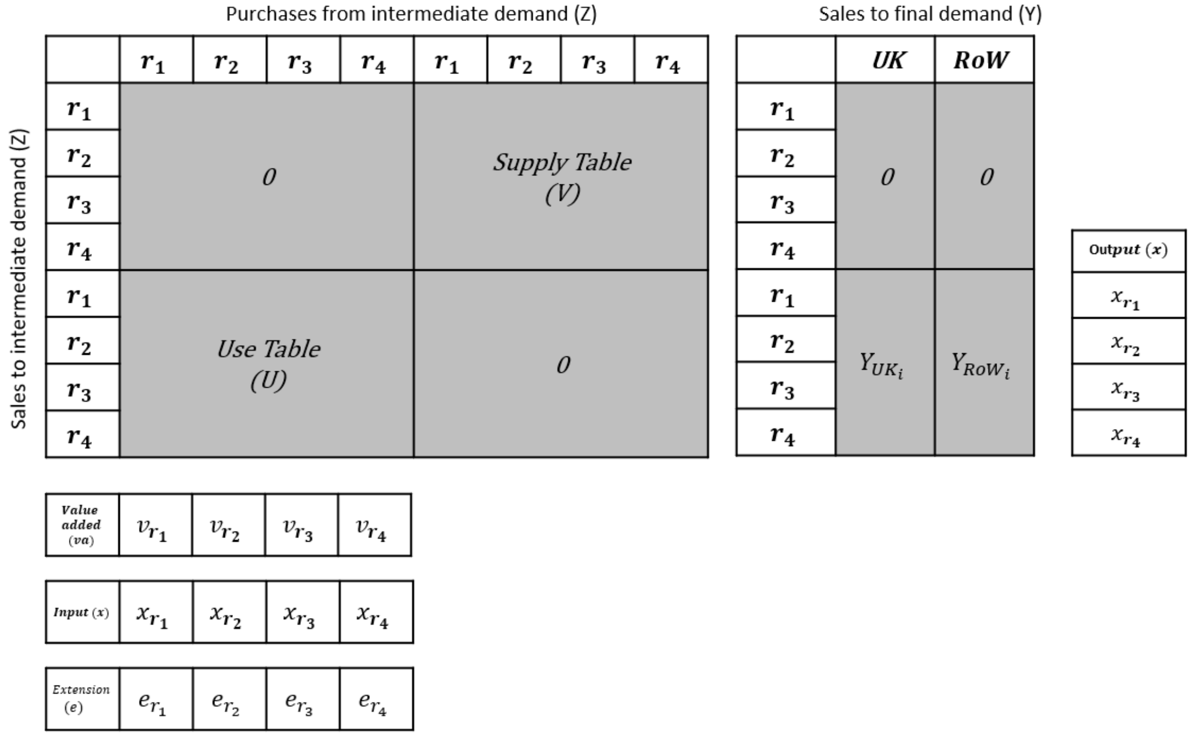
| Dataset | Role | Reference | Chapter |
|---|---|---|---|
| Energy intensity: Annual Business Survey 2019 and Annual Purchases Survey 2018 | Typical firm gas-to-electricity consumption ratio, used to differentiate utility purchases. | (ONS, 2021) | 2, 3 |
| UK gas and electricity prices in the non-domestic sector | Data on consumption bands and energy unit prices are collected | (DESNZ, 2024) | 2, 3 |
| UK greenhouse gas reporting: conversion factors | Usage-based conversion factors for fuels and energy consumption are collected | (BEIS and DEFRA, 2022) | 2, 3 |
| Digest of UK Energy Statistics | Data on the UK's petrol–diesel consumption split is collected | (DESNZ and BIES, 2022b) | 2, 3 |
| Business population estimates for the UK and regions 2022 | Firm-turnover distributions by industry are calculated | (BEIS, 2022) | 2, 3 |
| UK Supply and Use Tables | Data on sectoral output and sectoral electricity purchases is collected | (ONS, 2022) | 2, 3, 4 |
| UK Environmental Accounts | Data on sectoral emissions is collected | (ONS, 2025) | 2, 3, 4 |
| International Monetary Fund's World Economic Outlook Database | Data is sourced to calculate real GDP, imports, and export growth rates for all regions in the UKMRIO | (IMF, 2022) | 4 |
| Producer Price Indices | Annual, industry specific inflation indices are collected | (ONS, 2025a) | 4 |
| Contributions to monthly GDP | Data is collected to calculate the contributions different sectors make the annual GDP change | (ONS, 2025b) | 4 |

### 1.6.3 UKMRIO

Throughout this thesis, the UKMRIO database is used (Owen and Kilian, 2025). This is an EEMRIO database developed at the University of Leeds with the primary purpose of producing and publishing the UK's consumption-based emissions accounts (Barrett et al., 2013). To maintain an official status, the database is constructed with an emphasis on preserving original data released by the Office for National Statistics (ONS), making it a Single-country National Accounts Consistent (SNAC) MRIO, a framework specifically designed for credible, robust country-specific analysis (Edens et al., 2015; Tukker et al., 2018).

In the UKMRIO, data is structured into Supply and Use tables (SUTs) (Figure 1.2), splitting products and services into 112 sectors. The UK's SUTs are released with a relatively prompt lag of 22 months following the reference period, typically in October each year (ONS, 2022). Following their release, the UKMRIO is constructed, and subsequent consumption-based emission inventories are then calculated and published in June, approximately 30 months after the reference year (Owen and Kilian, 2025). For the years used in this thesis, the international element of the UKMRIO is extracted from FIGARO, an MRIO database produced by Eurostat (2021). This database is also available in a SUT structure for 64 products and industries, and is therefore manipulated into the same regional, industry and product structure of the UK's SUTs. A comprehensive account of the UKMRIO construction methodology is provided in the official documentation published on the UK Government website, released alongside the annual UK consumption-based accounts (DEFRA, 2025b; Owen and Kilian, 2025).

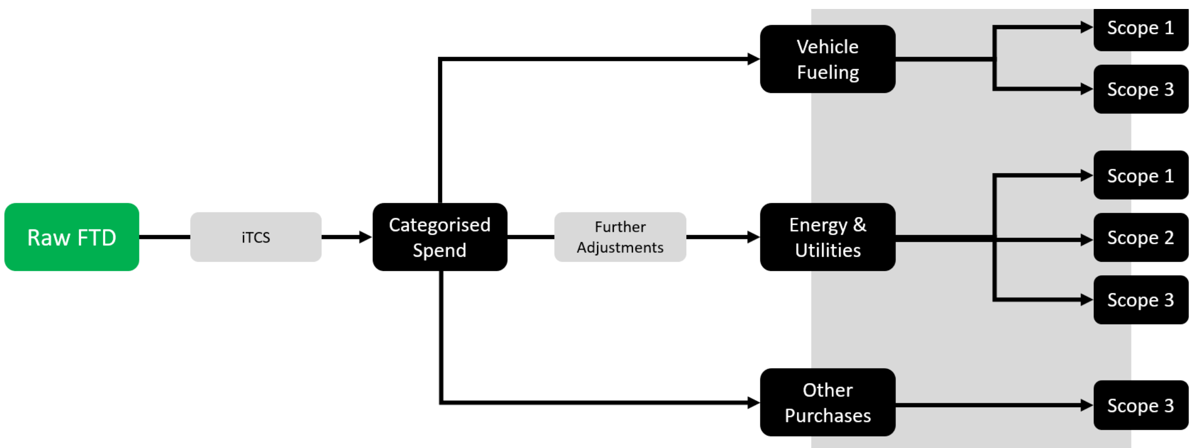
*Figure 1.2 - UKMRIO structure for a 4-region example.*




## 1.6.4 Spend-based Emission Estimation

The process of producing emission estimates from FTD forms the basis for much of the work conducted in this thesis. Figure 1.3 illustrates a simplified flow chart of this process, with the subsequent subsections further detailing the conversion process and factor used by spend type.

*Figure 1.3 - Overview of spend-based emission estimation. Black rectangles indicate derived data, green denotes raw data, and grey highlights specific processes. Here, iTCS refers to the internal Transaction Classification System to categories spend.*



### 1.6.4.1 Calculating energy- and fuel-related emissions

*Spending adjustments*

Spend relating to Scope 1 and 2 emissions are categorised under “Energy & Utilities” and “Vehicle Fuelling”. The former refers to 4 sub-categories of expenditure, containing energy providers selling electricity, natural gas, and heating oil, whilst the latter refers to diesel and petrol purchased for mobile combustion. With these categories, the Scope 1 and 2 emissions generated from a firm’s energy-use are captured, mirroring what is required of larger firms under the SECR framework.

Two adjustments to spending classified under ‘Energy & Utilities’ are necessary to address limitations arising from the granularity of FTD and the proportionality assumption underlying emissions estimation in this important category. First, due to many UK utility firms providing electricity, gas, and heating oil to consumers, it is a necessary step to disaggregate spending into Scope 1 (gas and oil) and Scope 2 (electricity). To do this, external data signally the expected split between spend on electricity and other fuels in the UK is sourced. To estimate the split between Scope 1 and Scope 2, industry data is used to determine the typical gas-to-electricity ratio of firms in each industry (ONS, 2021), and applied to those firms known to be selling various energy types.

Second, transactions made with energy companies also require additional adjustment steps to account for discounts given to consumers of high energy spend. Businesses receive a cheaper unit price for energy, as energy consumption increases. As a result, spend is adjusted upwards for higher consumption levels to reflect actual consumption levels (and thus emissions), accounting for the lower unit price. Whilst spend is also adjusted downwards for lower consumption levels to reflect the above average unit price charged to smaller consumers. Adjustment bands are calculated from UK statistics on consumption bands and energy unit prices (DESNZ, 2024).

*Emission conversion*

Next, a set of conversion factors are constructed to estimate the emissions associated with a unit purchase of different energy and fuel types. As official conversion factors are provided in terms of physical usage rather than expenditure, these must be derived in a manner that ensures consistency and comparability with usage-based factors. Official emissions factors per physical unit (BEIS and DEFRA, 2022) are divided by the average price paid by non-domestic consumers per unit (DESNZ, 2024) to derive emissions per pound spent. This calculation is shown in Equation 1.1.

$$Emissions\ Factor_{\mathrm{e},£} = \frac{Unit\ factor_{e,u}}{Unit\ price_{\ u,£}} \tag{1.1}$$

Where e = kgCO2e

$u$ = chosen unit of energy or fuel (see tables 1.3 and 1.4)

As spending within “Vehicle Fuelling” reflects the consumption of both petrol and diesel, a single conversion factor is calculated through a weighted average based on the UK’s consumption split between petrol and diesel (DESNZ and BIES, 2022b) and the unit price difference (DESNZ, 2024). For gas and oil, an average conversion factor is assigned given the near zero difference in individual factors. Finally, for electricity, the conversion factor is based on the UK grid’s emission intensity (location-based method). This process and resulting factors are presented in Table 1.3.

*Table 1.3 - Calculation of direct emission conversion factors. Units are fuel-specific: petrol, diesel, and oil are measured in litres, while gas and electricity are measured in kilowatt-hours.*

| | Scope 1 | | | | Scope 2 |
|---|---|---|---|---|---|
| | Petrol | Diesel | Gas | Oil | Electricity |
| Emissions per Unit (kg CO2e / Unit) | 2.19 | 2.51 | 0.18 | 2.54 | 0.21 |
| Average Unit Price (Unit / £) | 1.31 | 1.35 | 0.03 | 0.43 | 0.15 |
| Expenditure factors (kg CO2e / £) | 1.67 | 1.86 | 5.96 | 5.94 | 1.41 |
| Consumption Split | 68% | 32% | | | |
| **Conversion Factors (kg CO2e / £)** | **1.802** | | **5.949** | | **1.408** |

The same spend categories can also be used the calculate the upstream emissions from fuel and energy consumption, Category 3 defined in the GHG Protocol (WRI and WBCSD, 2020). Here, well-to-tank (WTT) emissions per unit figures (BEIS and DEFRA, 2022) are divided by the unit price, reflecting the emissions produced during the extraction, processing, and distribution of fuels and energy. This process, and resulting factors are presented in Table 1.4.

*Table 1.4 - Calculation of WTT emission conversion factors. Units are fuel-specific: petrol, diesel, and oil are measured in litres, while gas and electricity are measured in kilowatt-hours.*

| | | **Scope 1 Related** | | | **Scope 2 Related** |
|---|---|---|---|---|---|
| | Petrol | Diesel | Gas | Oil | Electricity |
| Emissions per Unit (kg CO2e / Unit) | 0.61 | 0.61 | 0.03 | 0.53 | 0.02 |
| Average Unit Price (Unit / £) | 1.31 | 1.35 | 0.03 | 0.43 | 0.15 |
| Expenditure factors (kg CO2e / £) | 0.47 | 0.45 | 1.02 | 1.23 | 0.12 |
| Consumption Split | 68% | 32% | | | |
| **Conversion Factors (kg CO2e / £)** | **0.457** | | **1.13** | | **0.125** |

1.6.4.2 Calculating indirect emissions from purchases

The remaining categories of spend relate to non-energy and fuel purchases. To estimate the upstream emissions associated with this spend, emission conversion factors derived through environmentally extended input-output analysis (IOA) are required. IOA, developed by Wassily Leontief in the 1930s, traces monetary flows between industries and to final demand consumers. Initially a national framework to model how economic shocks propagate through supply chains, IOA has evolved into Multi-Regional Input-Output (MRIO) models, incorporating global economies and then into Environmentally Extended MRIOs (EEMRIOs) (Miller and Blair, 2009). When extended to incorporate greenhouse gases, EEMRIOs are capable of tracing emissions embodied within global supply chains, in a standardized and manageable process (Kitzes, 2013).

Key to IOA is the transaction matrix ($\mathbf{Z}$), shown in Figure 1.3. $\mathbf{Z}$ is the central matrix containing intermediate flows of trade between industries. Reading the transaction matrix on a column basis, $\mathbf{Z}$ indicates the specific inputs required by industries to produce their goods and services, whilst reading $\mathbf{Z}$ on a row basis indicates the various industries a single product is sold to. The elements within $\mathbf{Z}$ ($z_{ij}$) therefore represent monetary flows between industry and product pairs (i and j).

Sales to final demand are represented by the $\mathbf{Y}$ matrix and contain all final sales to end consumers. When combined, the row sums of $\mathbf{Z}$ and $\mathbf{Y}$ can be calculated to find the total output of each industry (vector $\mathbf{x}$). These row sums equal the column sums of $\mathbf{Z}$ plus value added ($\boldsymbol{va}$), which contains the additional value created by an economy (Equation 1.2).

$$\sum_i z_{ij} + y_j = \sum_j z_{ij} + va_i = x_{ij} \tag{1.2}$$

Environmental impacts can then be quantified by extending the core monetary IO system with an environmental impacts vector. Here, $\boldsymbol{e}$ contains each sector's direct emission intensity, expressed as emissions per unit of output. IOA can then be used to track how these impacts are embodied within the global economy (Kitzes, 2013). These impacts can be displayed through embodied emission conversion factors, contained within vector $\boldsymbol{eL}$. These values represent the total amount of upstream emissions that occur anywhere in the economy, in any industry, to ultimately produce one unit of output.

The first step towards calculating $\boldsymbol{eL}$, is to derive the technical coefficient matrix, $\boldsymbol{A}$, through Equation 1.3. $\boldsymbol{A}$ contains elements signifying the unit input from all industries required to produce a unit output.

$$A_{ij} = \frac{Z_{ij}}{x_j} \tag{1.3}$$

With Equations (1.2) and (1.3), (1.4) is produced. Equation (1.4) can then be written in matrix form (1.5) and solved for x (1.6)

$$x_i = A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n + y_i \tag{1.4}$$

$$x = Ax + y \tag{1.5}$$

$$x = (I - A)^{-1}y \tag{1.6}$$

Equation 1.6 is fundamental to the input-output process and is referred to as the Leontief equation. It describes output ($\boldsymbol{x}$) as a function of final demand ($\boldsymbol{y}$). Equation 1.6 introduces the identity matrix $\boldsymbol{I}$, contained within $(I - A)^{-1}$, which is henceforth refer to as $\boldsymbol{L}$.

$\boldsymbol{L}$ is multiplied with the direct intensity vector $\boldsymbol{e}$ to produce $\boldsymbol{eL^{bp}}$, the embodied emission conversion factors expressed in basic prices (bp). Basic prices represent the underlying cost of production and reflect the value received by producers before any taxes, trade, or transport margins. As basic prices exclude these mark-ups, emissions are allocated over a smaller monetary value than what is actually paid (purchasers' prices), which means that embodied intensities expressed in basic prices are typically higher than intensities expressed in purchasers' prices.

Conversion factors in purchasers' prices ($\boldsymbol{eL^{pp}}$) are more suitable when applying IO-based intensities to FTD, as this data includes these mark-ups. To maintain price-level consistency, total industry embodied emissions are first calculated using basic-price conversion factors as ($\boldsymbol{F} = \boldsymbol{eL^{bp}y^{pp}}$). An average purchasers-price intensity can then be obtained by dividing these total emissions by total expenditure, giving an emission intensity per unit of spending at purchasers-prices ($\boldsymbol{eL^{pp}} = \frac{\boldsymbol{F}}{\sum_i \boldsymbol{y_i^{pp}}}$).

Whilst these conversion factors are generally attributed to a unit of final demand, the Leontief inverse assumes sectoral goods are homogenous and thus have the same cradle-to-gate emissions per monetary unit. As such, intermediate inputs have the same unit footprint as products sold to the final consumer (Hertwich and Wood, 2018). The underlying assumption of these conversion factors is that the supply structure of each purchase is approximated by the corresponding economic sector, as represented within the underlying data (Schmidt et al., 2022).

## 1.6.5 Regression Models

### 1.6.5.1 Rationale

Regression modelling provides a framework for interpreting relationships between an outcome (dependent variable), and a set of predictors (independent variables). Once specified, a regression model can be used to predict an outcome given the functions of the models predictors (Gelman and Hill, 2006). In this thesis, both linear and logistic specifications are employed to model continuous and discrete dependent variables respectively. Linear models are used to estimate continuous outcomes in all empirical chapters, while binary logistic models are applied in Chapter 3 to predict discrete outcomes.

Where regression specifications are applied to FTD, a hierarchical approach is adopted. This accounts for the inherent structure of the data, where firms are nested within industry categories. Hierarchical models allow for varying intercepts and slopes across groups, capturing sectoral effects, whist supporting more reliable inference and practical application (Gelman and Hill, 2006).

1.6.5.2 Model specifications

*Ordinary least squares linear specification*

Ordinary least squares (OLS) linear regressions, in their simplest form, models the expected value of an outcome as a linear combination of predictors, and is fitted to ungrouped data (pooled). Equation 1.7 provides a simple OLS regression example.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1.7}$$

Here, $\beta_0$ is the intercept, signifying the expected $y_i$, when all independent variables equal zero. Whilst $\varepsilon_i$ is the error term, capturing variation not explained by the regressors. In a classical OLS linear regression, this error term is assumed to have mean of zero and a normal distribution. The coefficient term $\beta_1$ measures marginal effects, ceteris paribus, which can be used to estimate the change in the dependent variable associated with a unit increase in the independent variable.

The OLS equation is calculated to find the line of best fit, that produces estimates as close as possible to the actual values. Here, 'best' refers to the smallest total of squared residuals between predicted and actual values. The procedure yields coefficient estimates, together with standard errors that quantify sampling uncertainty. Statistical significance of each coefficient is assessed by relating each coefficient to its standard errors, to distinguish genuine associations from random noise (Schroeder et al., 1986).

To account for non-linear relationships, reduce skewness and heteroskedasticity, and adjust for predictors measured on different scales, variables can expressed as the natural logarithm ($ln$) of the raw datapoints (Keene, 1995). This requires strictly positive observations and permits the interpretation of coefficients as elasticities. For example, should $\beta_1$equal 0.6, this can be interpreted as a 1% increase in $x_i$ associates with a 0.6% increase in $y_i$, ceteris paribus.

OLS regressions can also be enhanced through interaction terms. Interactions allow the model to capture situations where two variables may influence each other. In these cases, the effect of one variable can become stronger, weaker, or even change direction depending on the level of the other (Rimpler et al., 2025).

Model performance is summarised by the coefficient of determination (RSQ), which calculates the proportion of variance explained by the specified regression formula (Equation 1.8).

$$RSQ = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \widehat{y_i})^2}{\sum_i (y_i - \overline{y_i})^2} \tag{1.8}$$

Here, $SS_{RES}$ is the residual sum of squares, defined as the sum of squared differences between observed and predicted values, $\sum_i (y_i - \widehat{y_i})^2$. In an OLS regression, the fitted model is chosen to minimise this value. $SS_{TOT}$ is the total sum of squares, defined as the sum of squared differences between observed values and their mean, $\sum_i (y_i - \overline{y_i})^2$. Thus, RSQ measures the proportion of variation in the observed data explained by the model.

Variables are introduced to the specification sequentially, allowing for the evaluation of the required complexity to explain the dependent variables (Cohen et al., 2013). The Akaike Information Criterion (AIC) is useful here (Equation 1.9).

$$AIC = 2k - 2\ln(\hat{L}) \tag{1.9}$$

Where k is the number of parameters in the model and $\hat{L}$ is the maximum log likelihood function for the model. The AIC compares the relative quality of models estimated on the same data, penalising complexity; lower AIC indicates a better balance of fit and parsimony.

*Log-log hierarchical linear specification*

Where the data is hierarchically structured, a pooled linear regression can produce biased coefficients from unobserved group-level effects. In addition, pooled approaches risk noisy estimates in more sparsely sampled groups. Modelling on hierarchical data, like firms grouped within industries, therefore requires an approach that efficiently deals with group-level variations (Equation 1.10)

$$y_i = \left(\beta_1 + \beta_{1,j(i)}\right)x_i + \varepsilon_i \tag{1.10}$$

Equation 1.10 includes the individual and group-level impact of $x_i$, written as $\left(\beta_1 + \beta_{1,j(i)}\right)$. Here, $\beta_1$ is the overall (fixed) effect of $x_i$ with respect to $y_i$, and $\beta_{1,j(i)}$ is the group-specific (j) deviation which individual i belongs, allowing the slope to differ by group. In the context of this research, this heterogeneity is theoretically sensible as industries differ in production technologies and energy intensity, so the elasticity of emissions with respect to variables such as turnover is likely to vary by sector.

This example equation is also estimated without an intercept, which means the line is forced to pass through zero. In practice, this assumes that when the predictor takes the value zero, the outcome is also zero.

When assessing a hierarchical model such as Equation 1.10, both RSQ and AIC remain useful. Under a hierarchical specification, RSQ can be split between marginal and conditional scores. The first refers to the proportion of variance explained by the fixed effects, whilst the latter refers to the proportion explained by the total equation. The larger the gap between the two, the more variance is explained by the group differences.

*Hierarchical Binary Logistic specification*

To estimate discrete outcomes, a hierarchical binomial logistic regression framework can be applied. This specification is used to predict the likelihood of a binary outcome, whilst also accounting for a grouped structure within the data. Equation 1.11 predicts the log odds of $Y_{ij}$ for any data point in group j.

$$logit\left(P(Y_{ij} = 1)\right) = \alpha_j \tag{1.11}$$

Where:

$logit(P(Y_{ij} = 1))$ is the log odds of the probability of the outcome variable Y for data point i in group j. Here, Y = 1 indicates an occurrence, while Y = 0 denotes its absence

$\alpha_j$ represents the variable effect for group j

Model performance is summarised using several complementary metrics. AIC ranks alternative specifications by balancing goodness of fit against model complexity. The median residual describes the typical error; with values near zero suggesting no systematic over- or under-prediction. Both the log-likelihood and the deviance also quantify overall model fit. When comparing models, a higher log-likelihood and a lower deviance indicate better fit, enabling straightforward comparison across competing specifications. Finally, predicted log odds can be converted into probabilities easy with Equation 1.12, enabling interpretation.

$$Probability = \frac{1}{1 + \exp\left(-Y_{ij}\right)} \tag{1.12}$$

1.6.5.3 Prediction and out-of-sample application

After interpretation, regression models can be used to generate predictions for new observations. Once a model has been fitted, its estimated coefficients provide a formula from inputs (the predictors) to produce an expected outcome. Applying this formula to unseen data allows the model to predict outcomes for cases where only the predictors are known.

Assessing predictive accuracy is therefore an important step and is conducted through out-of-sample testing. Out-of-sample testing provides an unbiased measure of predictive accuracy and checks for overfitting, ensuring that the model generalises beyond the training data. In this thesis, a 5-fold cross-validation is used to randomly divide data into five folds, and repeatedly train the model specification on four folds and test on the remaining fold (James et al., 2021). This provides a more stable and reliable estimate of predictive performance than a single split. Stability is assessed by examining whether the mean RSQ across folds is consistent with the base model.

Additional out-of-sample assessments are carried out by comparing the model's predictions with those generated by alternative approaches, or by testing the model against independent datasets, assisting to evaluate validity. In these cases, performance is also assessed using absolute error (AE) and absolute percentage error (APE). These error metrics are summarised by their means and medians.

#### *1.6.6.2 Macroeconomic aggregates from MRIO tables*

Implementing the nowcasting procedure requires an understanding of how the MRIO framework maps to macroeconomic aggregates, specifically regional GDP, exports, and imports. These aggregates are sourced from the base table, projected using exogenous growth rates, with the resulting figures used as a basis for the nowcasted table. The first step is to split the base SMRIO into domestic IO tables (Figure 1.4).

### 1.6.6 Multi-regional Input-Output Projections

Whilst EEMRIO databases provide a robust, standardised framework for estimating the embodied emissions of purchases, their usefulness is limited by significant publication lags. The principal cause of these lags is inherited through the delay in publication of official SUTs. These pairs of tables record how supplies of different kinds of goods and services originate from domestic industries and imports, and how those supplies are allocated between various intermediate and final users. Thus, providing essential detail on the structure of economies through monetary values of transactions between pairs of sectors (Miller and Blair, 2009). It is due to this detail, that SUTs are generally published several years after the period they refer to. The European Transmission programme sets targets for member state to publish SUTs within 36 months after the reference period (European Commission, 2008), however in other parts of the globe, publication is less frequent.

Chapter 4 investigates methodologies for adapting base EEMRIO tables to allow for compatibility with real-time data, such as FTD. For this, a modified version of the Scenario-based Projection of the International Trade Network (SPIN) method (Beaufils and Wenz, 2022) is combined with higher frequency economic datasets to produce nowcasted EEMRIO tables. The SPIN method builds on established input–output principles, utilising the Leontief inverse, RAS-type balancing procedures, and applying the fundamental accounting relationships between regional output, Gross Domestic Product (GDP), international trade, and aggregate demand, making it both a practical and theoretically sound approach to nowcasting.

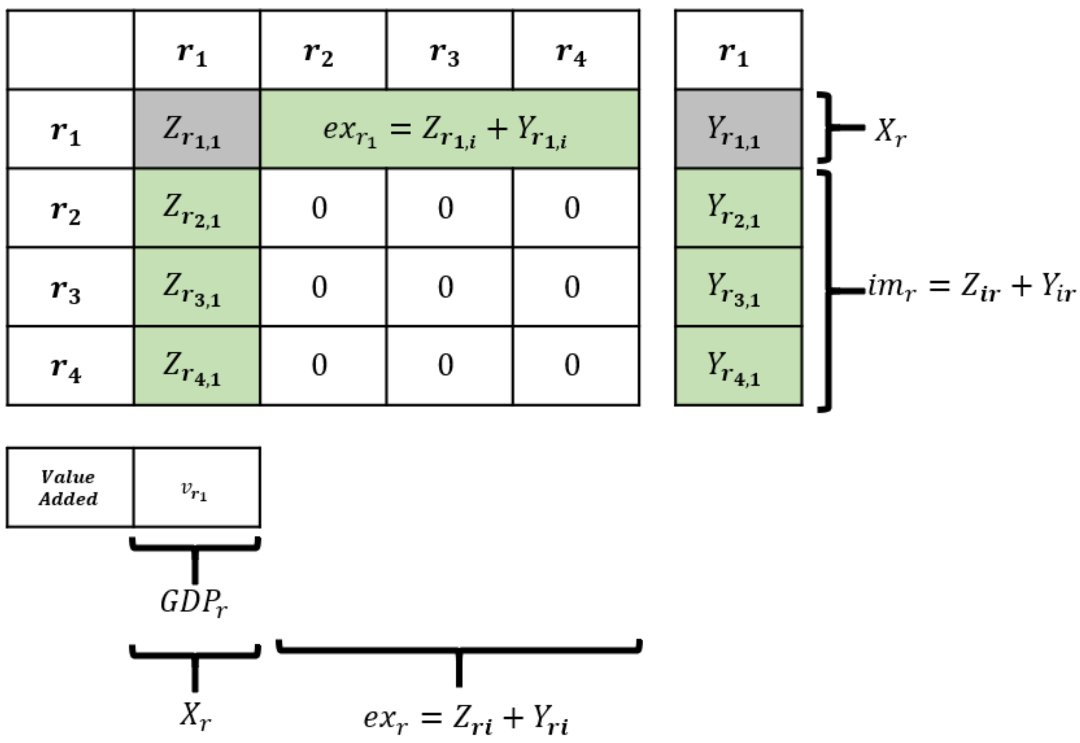
*1.6.6.1 UKMRIO sdjustments*

The UKMRIO is originally in a supply and use table format (Figure 1.3), to keep data as true to the original official data points as possible. Theoretically, this is useful when industries produce more than one product type. The supply table details secondary products as off-diagonal values (products made by industries other than their principal industry). To enable compatibility with the SPIN method the UKMRIO must be converted into a symmetric MRIO table (SMRIO), whilst the single Rest of the World final demand column must be disaggregated into regions.

To transform the UKMRIO into a SMRIO table, the Eurostat manual of supply, use and input-output tables identifies four transformation methods. In this thesis, Model D is used to convert the Supply and use table structured MRIO (SUT-MRIO) into a SMRIO. Model D assumes that each product has its own specific sales structure, independent of the industry in which it is produced (Eurostat, 2008. p296). By adopting Model D, the resulting SMRIO better reflects product-specific patterns of production, use, and trade, providing a more accurate basis for upstream environmental accounting.

To disaggregate the Rest of the World final demand column, supplementary data from FIGARO is used (Eurostat, 2021). FIGARO final demand data is aggregated to match UKMRIO regions (creating a region-by-region matrix) and converted into proportions (excluding UK final demand). This provides the shares for distributing a region's final demand across non-UK destination regions.

*Figure 1.4 - Domestic IO table (region 1 example).*




The common formula for GDP measurement follows the expenditure approach (Equation 1.13). Under this formula GDP is equal to the sum of private consumption (C), investment (I), government spending (G) and net exports (X – M).

$$GDP = C + I + G + (X - M) \tag{1.13}$$

For each region Equation 1.14 is written:

$$gdp_r = c_r + \delta_r \tag{1.14}$$

Where $c_r$ refers to domestic use of final goods and services (C + I + G), whilst $\delta_r$ refers to the balance of payments (net exports). GDP can also be measured though the income measure (Equation 1.15), where GDP straightforwardly links to value added (Wenz et al., 2014).

$$gdp_r = v_r \tag{1.15}$$

Total imports can be found by summing rows indicating non-domestic sectors in the transaction and final demand matrix (Equation 1.16), whilst total exports can be found by summing non-domestic columns in the transaction and final demand matrix (Equation 1.17). Subscript ordering denote source region then destination region ($Z_{\boldsymbol{ir}}$ = intermediate imports for region r from region i).

$$im_r = Z_{\boldsymbol{ir}} + Y_{\boldsymbol{ir}} \tag{1.16}$$

$$ex_r = Z_{\boldsymbol{ri}} + Y_{\boldsymbol{ri}} \tag{1.17}$$

1.6.6.2 Deriving aggregate demand from macroeconomic aggregates

The base table values of GDP, imports and exports are then multiplied by real growth rates, externally sourced from high frequency datasets, to create the projected values that form a basis for constructing the nowcasted table. Projected intermediate exports and imports are easily assigned to relevant the row / column constraint ($ex_r$ and $im_r$). Where values involve domestic production or consumption ($Z_{\boldsymbol{r}_{1.1}}$and $Y_{\boldsymbol{r}_{1.1}}$), gross output is first derived.

To derived gross output, aggregate demand must first be found (Equation 1.18).

$$ad_r = y_{rr} + ex_r \tag{1.18}$$

$ad_r$ is the domestic demand for domestic goods and services ($y_{rr}$) plus foreign demand for domestic goods and services ($ex_r$). Output is then produced to satisfy this demand. To find each of the terms that make up aggregate demand, each component is expressed in terms of GDP, imports, and exports. These terms are known values as they have been projected in the previous section.

$$ex_r = \sum_i \theta^*_{ri} ex_{ri} \tag{1.19}$$

Where $\theta^*_{ri}$ refers to the region shares of each industry's exports, whilst $ex_{ri}$ refers to the export section of the domestic IO table. The second term in Equation 1.18, domestic demand for domestic goods and services, can be expanded to Equation 1.20.

$$y_{rr} = \varphi_r^* \left( gdp_r - \delta_r - \alpha_r^* \sum_i im_{ri} \right) \tag{1.20}$$

Where $\varphi_r^*$ represents the relative market shares of domestic sectors in the domestic final demand of region r, $\alpha_r^*$ refers to the share of final commodities in imports to region r. In Equations 1.19 and 1.20, $\theta_{ri}^*$, $\varphi_r^*$ and $\alpha_r^*$ can either be derived directly from the base table or adapted to reflect economic changes.

1.6.6.3 Deriving output from aggregate demand

The Leontief inverse makes explicit the direct and indirect dependencies interlinking the sectors within the transaction table. Whilst traditionally used to obtain the sectoral production volumes requited to fulfil a given demand for commodities, here the Leontief inverse is applied to find the regional sectoral production ($x_r$) required to fulfil the estimated aggregate demand ($ad_r$). Equation 1.6 can be rearranged to find $x_r$, as the exogenous demand is represented by $ad_r$ (Beaufils and Wenz, 2022)

$$x_r = L_r^* ad_r \tag{1.21}$$

1.6.6.5 Balancing steps and MRIO compilation

Each region's domestic IO tables are balanced using a GRAS algorithm, allowing for the handling of negative coefficients (subsidies). The constraints for the balancing procedure are listed below and provided in Figure 1.5.

The domestic inter industry table's margins should meet the gross production requirement $x_r$, export arrays to each trade zone should equal the volume of exports to the corresponding region as prescribed in input, import arrays plus imported final demand sum up to the volume of imports from the corresponding trade zone, total value added, and domestic final demand are linked to GDP and trade balance.

*Figure 1.5 - Domestic IO table with constraints for GRAS balancing*

| | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_1$ | Row constraints |
|---|---|---|---|---|---|---|
| $r_1$ | $Z_{r_{1,1}}$ | $ex_{r_1} = Z_{r_{1,i}} + Y_{r_{1,i}}$ | | | $Y_{r_{1,1}}$ | $x_1$ |
| $r_2$ | $Z_{r_{2,1}}$ | 0 | 0 | 0 | $Y_{r_{2,1}}$ | $im_{r2}$ |
| $r_3$ | $Z_{r_{3,1}}$ | 0 | 0 | 0 | $Y_{r_{3,1}}$ | $im_{r3}$ |
| $r_4$ | $Z_{r_{4,1}}$ | 0 | 0 | 0 | $Y_{r_{4,1}}$ | $im_{r4}$ |

| | |
|---|---|
| Value Added | $v_{r_1}$ |

| | | | | |
|---|---|---|---|---|
| Column constraints | $x'_1$ | $ex_{r2}$ | $ex_{r3}$ | $ex_{r4}$ |

The second step of balancing adjusts international trade with respect to the exports and imports derived at the regional level. Using regional imports and exports constraints determined from the first balancing step, international trade is optimized using the basic RAS algorithm. Once done, the domestic blocks are inputted to create a new MRIO table. This table is balanced and consistent with the prescribed GDP and trade projections.

## 1.7 Methodological Limitations

### 1.7.1 Limitations associated with spend-based methodologies and input-output models

Spend-based methodologies carry several methodological constraints that should be considered throughout. First, full coverage of the economy is not achievable as certain industries are excluded due to sparse samples or inherent errors in the turnover and emissions estimation. In addition, the absence of receipt-level detail in FTD necessitates assigning emission factors at broad merchant-category level and assuming product mixes (gas/electricity; petrol/diesel). Related to this, Scope 2 estimation is calculated solely on a location-based basis, which may cause differences when compared against firm-level data which may be calculated on a market basis.

Furthermore, interpretability of emissions insights are constrained by coarse category taxonomies (for example “Shopping”), limiting the actionability of findings from emission calculations and often requiring additional information to enhance utility.

Whilst energy and fuel multipliers are narrowly specified, for all other spending categories the embodied emission multipliers are mapped from the 112 sectoral intensities detailed in the UKMRIO to the 620 merchant expenditure codes. This is closely related to the aggregation uncertainty of IO models, where geographical and technological variability of production exists within a particular sector, but uncaptured by the MRIO database which requires the assumption of homogenous products (Kitzes, 2013; Lenzen, 2000).

The key process for this work is the use of monetary values to represent physical flows. This implies that a monetary value can accurately represent a physical volume (Lenzen, 2000), assuming a level of linearity which is not always persistent in the true relationship between monetary and physical values. In reality, constant, fixed proportions of inputs do not create output (Kitzes, 2013), with the existence of features such as economies of scale. Some inputs such as electricity supplied by utility firms have highly volatile prices, which can decrease with high volumes of supply. Such examples violate the proportionality assumption (Lenzen, 2000; Tukker et al., 2018), and require further adjustments.

Validation tests at the firm-level remain challenging, with one-to-one comparisons of self-reported emissions to transaction-derived estimates generally only feasible for Scope 1 and Scope 2, but impracticable for SMEs who do not produce disclosures at scale. Among the larger firms that do meet SECR thresholds and also bank with LBG, sample sizes are very small and anonymity risks arise. Scope 3 validation is particularly challenging due to sparse and incomplete disclosures, whilst several categories of Scope 3 are either unobservable in FTD (for example, use-phase of sold products) or rely on strong assumptions, creating uncertainty.

Finally, this research is conducted for the single year 2021, replicating an annual reporting cycle. As FTD is longitudinal by nature this reduces a key benefit of the data, and year-on-year comparability depends on periodic updates to FTD, and conversion factors.

### 1.7.2 Limitations associated with regression models

The use of linear regression models also carries assumptions which need to be considered before use. First, the mathematical assumptions of these models is that the predicted variable is the product of a linear function of the sperate inputs, and the errors from the prediction line are assumed to be independent, and normally distributed. When these specification are extended to hierarchical models, additional assumptions occur, where each level of the model corresponds to its own set of assumptions of linearity, independence and normality (Gelman and Hill, 2006).

In addition, training these models on estimated rather than directly measured emissions carries forward the limitations of the estimation process, and it is important to be mindful that the fitted regression reproduces proxy-based estimates rather than physical emissions. The subsequent predictions are a snapshot estimate, producing a benchmarking tool rather than one that is capable of tracking individual level changes over time.

This thesis employs parsimonious regressions rather than high-dimensional machine learning, with the aim to produce practical to deploy models and reduce the risk of overfitting. Importantly, more inputs do not guarantee better results when predictors are inconsistently available. Scope 3 emissions illustrates this challenge as improving predictions would require accessible, standardised data on the number and characteristics of suppliers, product use-phase emissions, transportation and logistics metrics such as distance shipped and freight mode, and procurement expenditures disaggregated by category (Wang and Ye, 2025). These inputs are rare for SMEs at scale and attempting to include partial proxies may inadvertently increase error.

## 1.8 Thesis Structure and Alternative Format

The Faculty of Environment at the University of Leeds offers the submission of PhD thesis in the so-called ‘alternative format’. This means that the empirical chapters of this thesis (Chapter 2, Chapter 3, and Chapter 4) consist of first-author, peer-reviewed journal articles at various stages of the submission and publication process. It is required for one of these papers to be accepted for publication, for one paper to be accepted for resubmission, and for one paper to be ready for submission.

To date, Chapters 2 and 3 are published (Phillpotts et al. (2025, 2026)), whilst Chapter 4 is in a format and at a standard ready to be submitted for peer-review. Thus, this thesis meets the conditions for the alternative format.

Following the empirical chapters, Chapter 5 provides the discussion and conclusion to the thesis. For further details on this format please refer to the Faculty of Environment's guide on the alternative style thesis (University of Leeds: Faculty of Environment, 2020).

## 1.9 References

Acquaye, A., Genovese, A., Barrett, J., Lenny Koh, S.C., 2014. Benchmarking carbon emissions performance in supply chains. Supply Chain Management: An International Journal 19, 306–321. https://doi.org/10.1108/SCM-11-2013-0419

Afionis, S., Sakai, M., 2022. Production- and consumption-based accounting: an overview of the debate, in: Handbook on Trade Policy and Climate Change, Chapter 15. Edward Elgar Publishing, pp. 239–255.

Alipour, A., Rahimpour, M., 2020. Sustainability Barriers in SMEs: A study of strength of sustainability barriers and practical solutions in Green product lifecycle at SMEs. Jönköping University. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1462056&dswid=-2267.

Assael, J., Heurtebize, T., Carlier, L., Soupé, F., 2023. Greenhouse Gases Emissions: Estimating Corporate Non-Reported Emissions Using Interpretable Machine Learning. Sustainability 15. https://doi.org/10.3390/su15043391

Bajic, A., Kiesel, R., Hellmich, M., 2023. Handle with care: Challenges in company-level emissions data for assessing financial risks from climate change. Journal of Climate Finance 5, 100017. https://doi.org/10.1016/j.jclimf.2023.100017

Bakos, J., Siu, M., Orengo, A., Kasiri, N., 2020. An analysis of environmental sustainability in small & medium-sized enterprises: Patterns and trends. Business Strategy and the Environment 29, 1285–1296. https://doi.org/10.1002/bse.2433

Banerjee, P., Bose, S., Dhole, S., Truong, C., 2025. Greening up their act: Corporate carbon emissions reduction in response to political risk. Energy Economics 150, 108847. https://doi.org/10.1016/j.eneco.2025.108847

Barendregt, W., Biørn-Hansen, A., Andersson, D., 2020. Users' Experiences with the Use of Transaction Data to Estimate Consumption-Based Emissions in a Carbon Calculator. Sustainability 12, 7777. https://doi.org/10.3390/su12187777

Barrett, J., Glen, P., Wiedmann, T., Scott, K., Lenzen, M., Roelich, K., Le Quere, C., 2013. Consumption-based GHG emission accounting: a UK case study. Climate Policy 13, 451–470. http://dx.doi.org/10.1080/14693062.2013.788858

Beaufils, T., Wenz, L., 2022. A scenario-based method for projecting multi-regional input–output tables. Economic Systems Research 34, 440–468. https://doi.org/10.1080/09535314.2021.1952404

Betts-Davies, S., Owen, A., Barrett, J., Brockway, P., Norman, J., 2025. Quantifying the energy and emissions implications of consumption redistribution in the UK through sustainable consumption corridors. Sci Rep 15, 16499. https://doi.org/10.1038/s41598-025-01495-0

Blanco, C., Caro, F., Corbett, C.J., 2016. The state of supply chain carbon footprinting: analysis of CDP disclosures by US firms. Journal of Cleaner Production 135, 1189–1197. https://doi.org/10.1016/j.jclepro.2016.06.132

Boudt, K., Schoors, K., Van Den Heuvel, M., Weytjens, J., 2022. The consumption response to labour income changes (Working Paper No. 415). NBB Working Paper. https://www.econstor.eu/handle/10419/273119.

British Business Bank (BBB), 2021. Small businesses and net zero. British Business Bank. https://www.british-business-bank.co.uk/research/smaller-businesses-and-the-transition-to-net zero/ (accessed 7.12.23).

Buchenau, N., Oetzel, J., Hechelmann, R.-H., 2025. Category-specific benchmarking of Scope 3 emissions for corporate clusters. Renewable and Sustainable Energy Reviews 208, 115019. https://doi.org/10.1016/j.rser.2024.115019

Busch, T., Johnson, M., Pioch, T., 2022. Corporate carbon performance data: Quo vadis? Journal of Industrial Ecology 26, 350–363. https://doi.org/10.1111/jiec.13008

Caldera, H.T.S., Desha, C., Dawes, L., 2019. Evaluating the enablers and barriers for successful implementation of sustainable business practice in 'lean' SMEs. Journal of Cleaner Production 218, 575–590. https://doi.org/10.1016/j.jclepro.2019.01.239

Cheema-Fox, A., LaPerla, B.R., Serafeim, G., Turkington, D., Wang, H. (Stacie), 2021. Decarbonizing Everything. Financial Analysts Journal 77, 93–108. https://doi.org/10.1080/0015198X.2021.1909943

Climate Change Committee, 2025. Progress in reducing emissions - 2025 report to Parliament. https://www.theccc.org.uk/publication/progress-in-reducing-emissions-2025-report-to-parliament/ (accessed 7.11.25).

Cogo, 2022. Calculate The Carbon Footprint of Your Transactions. https://www.cogo.co/about (accessed 5.23.23).

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd ed. Routledge, New York. https://doi.org/10.4324/9780203774441

Cologna, V., Meiler, S., Kropf, C.M., Lüthi, S., Mede, N.G., Bresch, D.N., Lecuona, O., Berger, S., Besley, J., Brick, C., Joubert, M., Maibach, E.W., Mihelj, S., Oreskes, N., Schäfer, M.S., Linden, S. van der, 2025. Extreme weather event attribution predicts climate policy support across the world. Nat. Clim. Chang. 15, 725–735. https://doi.org/10.1038/s41558-025-02372-4

Competition Markets Authority, 2016. Retail banking market investigation final report. https://assets.publishing.service.gov.uk/media/57ac9667e5274a0f6c00007a/retail-banking-market-investigation-full-final-report.pdf (accessed 1.10.25).

Conway, E., 2015. Engaging small and medium-sized enterprises (SMEs) in the low carbon agenda. Energy, Sustainability and Society 5, 32. https://doi.org/10.1186/s13705-015-0060-x

Davis, S.J., Caldeira, K., 2010. Consumption-based accounting of CO2 emissions. Proceedings of the National Academy of Sciences 107, 5687–5692. https://doi.org/10.1073/pnas.0906974107

Department for Business, Energy & Industrial Strategy (BEIS), 2022. Business population estimates for the UK and regions 2022: statistical release [Dataset]. GOV.UK. https://www.gov.uk/government/statistics/business-population-estimates-2022/business-population-estimates-for-the-uk-and-regions-2022-statistical-release-html (accessed 7.12.23).

Department for Business, Energy & Industrial Strategy (BEIS), 2021. Net Zero Strategy: Build Back Greener (Presented to Parliament pursuant to Section 14 of the Climate Change Act 2008). https://www.gov.uk/government/publications/net zero-strategy (accessed 1.10.24).

Department for Business, Energy & Industrial Strategy (BEIS) and Department for Environmental Food & Rural Affairs (DEFRA), 2022. Greenhouse gas reporting: conversion factors 2021 [Dataset]. GOV.UK. https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2021 (accessed 7.23.24).

Department for Energy Security and Net Zero (DESNZ), 2024. Gas and electricity prices in the non-domestic sector [Dataset]. GOV.UK. https://www.gov.uk/government/statistical-data-sets/gas-and-electricity-prices-in-the-non-domestic-sector (accessed 7.23.24).

Department for Energy Security and Net Zero (DESNZ) and Department for Business, Energy & Industrial Strategy (BEIS), 2022a. UK's 2030 Nationally Determined Contribution (NDC) emissions reduction target under the Paris Agreement. GOV.UK. https://www.gov.uk/government/publications/the-uks-nationally-determined-contribution-communication-to-the-unfccc (accessed 11.12.25).

Department for Energy Security and Net Zero (DESNZ) and Department for Business, Energy & Industrial Strategy (BEIS), 2022b. Digest of UK Energy Statistics (DUKES) 2022 [Dataset]. GOV.UK. https://www.gov.uk/government/statistics/digest-of-uk-energy-statistics-dukes-2022 (accessed 8.23.24).

Department for Environmental Food & Rural Affairs (DEFRA), 2025. Carbon footprint for the UK and England to 2022. GOV.UK. https://www.gov.uk/government/statistics/uks-carbon-footprint/carbon-footprint-for-the-uk-and-england-to-2022 (accessed 12.1.25).

Dragomir, V.D., 2012. The disclosure of industrial greenhouse gas emissions: a critical assessment of corporate sustainability reports. Journal of Cleaner Production 29–30, 222–237. https://doi.org/10.1016/j.jclepro.2012.01.024

Edens, B., Hoekstra, R., Zult, D., Lemmers, O., Wilting, H.C., Wu, R., 2015. A method to create carbon footprint estimates consistent with national accounts. Economic Systems Research 27, 440–457. https://doi.org/10.1080/09535314.2015.1048428

European Commission, 2022. Annual Report on European SMEs 2021/2022. https://data.europa.eu/doi/10.2826/164089 (accessed 1.10.25).

European Commission (Ed.), 2008. Eurostat manual of supply, use and input-output tables, 2008 edition. ed. Publications Office, Luxembourg. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-07-013 (accessed 3.10.25).

Eurostat, 2021. FIGARO methodology. Eurostat, Luxemberg. https://ec.europa.eu/eurostat/documents/51957/12767369/Figaro-methodology.pdf/487255c9-903a-0cb7-f5e2-73f37e35f196?t=1620750704022 (accessed 1.10.25).

Fankhauser, S., Smith, S.M., Allen, M., Axelsson, K., Hale, T., Hepburn, C., Kendall, J.M., Khosla, R., Lezaun, J., Mitchell-Larson, E., Obersteiner, M., Rajamani, L., Rickaby, R., Seddon, N., Wetzer, T., 2022. The meaning of net zero and how to get it right. Nat. Clim. Chang. 12, 15–21. https://doi.org/10.1038/s41558-021-01245-w

Financial Reporting Council, 2021. Thematic Review: Streamlined Energy and Carbon Reporting. https://media.frc.org.uk/documents/Streamlined_Energy_and_Carbon_Reporting_SECR.pdf. (accessed 1.9.24).

Franzen, A., Mader, S., 2018. Consumption-based versus production-based accounting of CO2 emissions: Is there evidence for carbon leakage? Environmental Science & Policy 84, 34–40. https://doi.org/10.1016/j.envsci.2018.02.009

Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Higher Education from Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Goldhammer, B., Busse, C., Busch, T., 2017. Estimating Corporate Carbon Footprints with Externally Available Data. Journal of Industrial Ecology 21, 1165–1179. https://doi.org/10.1111/jiec.12522

Griffin, P.A., Lont, D.H., Sun, E.Y., 2017. The Relevance to Investors of Greenhouse Gas Emission Disclosures. Contemporary Accounting Research 34, 1265–1297. https://doi.org/10.1111/1911-3846.12298

Hertwich, E.G., Peters, G.P., 2009. Carbon Footprint of Nations: A Global, Trade-Linked Analysis. Environ. Sci. Technol. 43, 6414–6420. https://doi.org/10.1021/es803496a

Hertwich, E.G., Wood, R., 2018. The growing importance of scope 3 greenhouse gas emissions from industry. Environ. Res. Lett. 13, 104013. https://doi.org/10.1088/1748-9326/aae19a

Heurtebize, T., Chen, F., Soupé, F., Carvalho, R.L. de, 2022. Corporate Carbon Footprint: A Machine Learning Predictive Model for Unreported Data. https://doi.org/10.2139/ssrn.4038436

HM Government, 2019. Environmental Reporting Guidelines: Including streamlined energy and carbon reporting guidance. https://www.gov.uk/government/publications/environmental-reporting-guidelines-including-mandatory-greenhouse-gas-emissions-reporting-guidance.

Hutton, H., 2020. Small and medium enterprises (SMEs) are vital to UK decarbonisation – how do we meaningfully engage them? https://www.cisl.cam.ac.uk/news/blog/small-and-medium-enterprises-smes-are-vital-to-uk-decarbonisation-how-do-we-meaningfully-engage-them (accessed 11.12.25).

IMF, 2022. World Economic Outlook Database, October 2022 [Dataset]. IMF. https://www.imf.org/en/Publications/WEO/weo-database/2022/October (accessed 6.19.23).

Intergovernmental Panel on Climate Change, 2021. Chapter 3: Human Influence on the Climate System, in: IPCC Sixth Assessment Report: Working Group 1: The Physical Science Basis. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. doi:10.1017/9781009157896.

Intergovernmental Panel on Climate Change (IPCC) (Ed.), 2023. Weather and Climate Extreme Events in a Changing Climate, in: Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp. 1513–1766. https://doi.org/10.1017/9781009157896.013

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning with Applications in R, Second. ed. Springer, New York, United States.

Janda, K.B., Bottrill, C., Layberry, R., 2014. Learning from the “data poor”: energy management in understudied organizations. Journal of Property Investment & Finance 32, 424–442. https://doi.org/10.1108/JPIF-03-2014-0018

Keene, O.N., 1995. The log transformation is special. Statistics in Medicine 14, 811–819. https://doi.org/10.1002/sim.4780140810

Kilian, L., Owen, A., Newing, A., Ivanova, D., 2023. Microdata selection for estimating household consumption-based emissions. Economic Systems Research 35, 325–353. https://doi.org/10.1080/09535314.2022.2034139

Kitzes, J., 2013. An Introduction to Environmentally-Extended Input-Output Analysis. Resources 2, 489–503. https://doi.org/doi:10.3390/resources2040489

Lambert, D.O., Igan, Thomas, 2019. Bank Lobbying: Regulatory Capture and Beyond. IMF. https://www.imf.org/en/Publications/WP/Issues/2019/08/09/Bank-Lobbying-Regulatory-Capture-and-Beyond-45735 (accessed 8.29.25).

Lenzen, M., 2000. Errors in Conventional and Input-Output-based Life-Cycle Inventories. Journal of Industrial Ecology 4, 127–148. https://doi.org/10.1162/10881980052541981

Lenzen, M., Murray, J., 2010. Conceptualising environmental responsibility. Ecological Economics, Special Section: Ecological Distribution Conflicts 70, 261–270. https://doi.org/10.1016/j.ecolecon.2010.04.005

Matsumura, E.M., Prakash, R., Vera-Muñoz, S.C., 2014. Firm-Value Effects of Carbon Emissions and Carbon Disclosures. The Accounting Review 89, 695–724.

Maquet, P., Baron, R., Leré, S., 2024. Imported emissions: From co-dependence to co-operative action. https://www.carbone4.com/en/publication-imported-emissions-en.

May, R., 2019. Start-ups across the UK are going bust - they need more careful management for our economy to boom. The Telegraph.

https://www.telegraph.co.uk/politics/2019/01/24/start-ups-across-uk-going-bust-need-careful-management-economy/ (accessed 20.9.25).

Mazhar, M.U., Domingues, A.R., Bull, R., O'Boyle, S., 2022. Small and medium-sized enterprises: hard to reach, data-poor but rich in creative potential as agents of change for decarbonisation. Presented at the European Council for an Energy Efficient Economy (ECEEE) Summer Study, European Council for an Energy Efficient Economy, Hyères, France, pp. 145–153.

Meath, C., Linnenluecke, M., Griffiths, A., 2016. Barriers and motivators to the adoption of energy savings measures for small- and medium-sized enterprises (SMEs): the case of the ClimateSmart Business Cluster program. Journal of Cleaner Production 112, 3597–3604. https://doi.org/10.1016/j.jclepro.2015.08.085

Miller, R.E., Blair, P.D., 2009. Foundations of Input–Output Analysis. Cambridge University Press, Cambridge.

Newman, R., Noy, I., 2023. The global costs of extreme weather that are attributable to climate change. Nat Commun 14, 6103. https://doi.org/10.1038/s41467-023-41888-1

Nguyen, Q., Diaz-Rainey, I., Kitto, A., McNeil, B.I., Pittman, N.A., Zhang, R., 2023. Scope 3 emissions: Data quality and machine learning prediction accuracy. PLOS Climate 2, e0000208. https://doi.org/10.1371/journal.pclm.0000208

Nguyen, Q., Diaz-Rainey, I., Kuruppuarachchi, D., 2021. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. Energy Economics 95, 105129. https://doi.org/10.1016/j.eneco.2021.105129

Normative, 2024. Survey: small businesses face recurring barriers to carbon reduction. Normative. https://normative.io/insight/smech-survey/ (accessed 8.7.24).

Organisation for Economic Co-operation and Development (OECD), 2023. Assessing greenhouse gas emissions and energy consumption in SMEs. https://www.oecd.org/en/publications/assessing-greenhouse-gas-emissions-and-energy-consumption-in-smes_ac8e6450-en.html.

Office for National Statistics (ONS), 2025. Atmospheric emissions: greenhouse gases by industry and gas. https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/greenhousegasintensityprovisionalestimatesuk/provisionalestimates2023 (accessed 9.4.23).

Office for National Statistics (ONS), 2022. Supply and use tables [Dataset]. https://www.ons.gov.uk/economy/nationalaccounts/supplyandusetables (accessed 8.4.23).

Office for National Statistics (ONS), 2025a. Producer price inflation statistics - Office for National Statistics [Dataset]. https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/producerpriceinflationstatistics (accessed 11.20.25).

Office for National Statistics (ONS), 2025b. Contributions to monthly GDP - Office for National Statistics [Dataset]. https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/contributionstomonthlygdp (accessed 11.20.25).

Office for National Statistics (ONS), 2021. Business energy spending: experimental measures from the Office for National Statistics' business surveys - Office for National Statistics [Dataset]. https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/articles/businessenergyspendingexperimentalmeasuresfromtheofficefornationalstatisticsbusinesssurveys/2022-09-07 (accessed 8.8.24).

Okimoto, T., Takaoka, S., 2024. Credit default swaps and corporate carbon emissions in Japan. Energy Economics 133, 107504. https://doi.org/10.1016/j.eneco.2024.107504

Oluleye, G., Rhodes, A., Hattan, L., Tu, Z., Milan, E., 2025. Interventions to support UK SMEs in deploying low carbon technologies and measures by 2030. London. https://www.imperial-consultants.co.uk/wp-content/uploads/2025/09/Interventions-to-support-UK-SMEs-decarbonise.pdf (accessed 8.1.25).

Owen, A., Büchs, M., 2024. Examining changes in household carbon footprints across generations in the UK using decomposition analysis. Journal of Industrial Ecology 28, 1786–1800. https://doi.org/10.1111/jiec.13567

Owen, A., Kilian, L., 2025. 2025 Data Release of Consumption- based Accounts for the UK: Summary of Methods. https://assets.publishing.service.gov.uk/media/68220ff3db6463b14cd81916/Summary_of_Methods_2025_final.pdf.

Owen, A., Steen-Olsen, K., Barrett, J., Wiedmann, T., Lenzen, M., 2014. A structural decomposition approach to comparing MRIO databases. Economic Systems Research, 26. https://doi.org/10.1080/09535314.2014.935299.

Oyewole, A.T., Adeoye, O.B., Addy, W.A., Okoye, C.C., Ofodile, O.C., 2024. Enhancing Global Competitiveness Of U.S. SMEs Through Sustainable Finance: A Review And Future Directions. International Journal of Management & Entrepreneurship Research 6, 634–647. https://doi.org/10.51594/ijmer.v6i3.876

PCAF, 2022. The Global GHG Accounting and Reporting Standard Part A: Financed Emissions., 2nd ed. https://carbonaccountingfinancials.com/standard (accessed 8.3.24).

Peters, G.P., 2008. From production-based to consumption-based national emission inventories. Ecological Economics 65, 13–23. https://doi.org/10.1016/j.ecolecon.2007.10.014

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2025. Bridging the SME reporting gap: A new model for predicting Scope 1 and 2 emissions. Journal of Industrial Ecology, 29. https://doi.org/10.1111/jiec.70106

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2026. Completing the SME carbon profile: scalable prediction of scope 3 emissions. *Journal of Industrial Ecology*. https://doi.org/10.1007/s44498-026-00003-5

Rimpler, A., Kiers, H.A.L., van Ravenzwaaij, D., 2025. To interact or not to interact: The pros and cons of including interactions in linear regression models. Behav Res Methods 57, 92. https://doi.org/10.3758/s13428-025-02613-6

Schroeder, L.D., Sjoquist, D.L., Stephan, P.E., 1986. Understanding Regression Analysis: An Introductory Guide. SAGE Publications, Inc. https://doi.org/10.4135/9781506361628

Serafeim, G., Velez Caicedo, G., 2022. Machine Learning Models for Prediction of Scope 3 Carbon Emissions. https://doi.org/10.2139/ssrn.4149874

Svalna, 2022. Calculate your climate impact. https://www.svalna.se/web/en (accessed 5.23.23).

Talbot, D., Boiral, O., 2013. Can we trust corporates GHG inventories? An investigation among Canada's large final emitters. Energy Policy 63, 1075–1085. https://doi.org/10.1016/j.enpol.2013.09.054

Trendl, A., Anwyl-Irvine, A., Vomfell, L., Abbey, E., Stewart, N., Atkins, D., Llewellyn, D.J., Gathergood, J., Leake, D., 2025. Early Behavioral Markers of Loss of Financial Capacity. JAMA Netw Open 8, e2515894. https://doi.org/10.1001/jamanetworkopen.2025.15894

Trendl, A., Owen, A., Vomfell, L., Kilian, L., Gathergood, J., Stewart, N., Leake, D., 2023. Estimating carbon footprints from large scale financial transaction data. Journal of Industrial Ecology 27, 56–70. https://doi.org/10.1111/jiec.13351

Tukker, A., Koning, A. de, Owen, A., Lutter, S., Bruckner, M., Giljum, S., Stadler, K., Wood, R., Hoekstra, R., 2018. Towards Robust, Authoritative Assessments of Environmental Impacts Embodied in Trade. Journal of Industrial Ecology 22, 585–598. https://doi.org/10.1111/jiec.12716

UNFCCC, 2015. The Paris Agreement. https://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf (accessed 30.8.25).

UNFCCC, 1997. The Kyoto Protocol to the Framework Convention on Climate Change. http://unfccc.int/kyoto_protocol/items/2830.php (accessed 9.11.25).

Wang, S.-Y., Ye, N.-Z., 2025. Invisible footprints, visible insights: machine learning reveals Scope 3 emissions. Front. Sustain. 6. https://doi.org/10.3389/frsus.2025.1649150

Wells, J., Trendl, A., Owen, A., Barrett, J., Jobst, N., Leake, D., 2025. Targeting carbon reduction in UK households: A new segmentation model using financial transaction data. Journal of Industrial Ecology. https://doi.org/10.1111/jiec.70020

Wenz, L., Willner, S.N., Bierkandt, R., Levermann, A., 2014. Acclimate—a model for economic damage propagation. Part II: a dynamic formulation of the backward effects of disaster-induced production failures in the global supply network. Environ Syst Decis 34, 525–539. https://doi.org/10.1007/s10669-014-9521-6

West, S.E., Owen, A., Axelsson, K., West, C.D., 2016. Evaluating the Use of a Carbon Footprint Calculator: Communicating Impacts of Consumption at Household Level and Exploring Mitigation Options. Journal of Industrial Ecology 20, 396–409. https://doi.org/10.1111/jiec.12372

Wiedmann, T., 2009. A review of recent multi-region input–output models used for consumption-based emission and resource accounting. Ecological Economics 69, 211–222. https://doi.org/10.1016/j.ecolecon.2009.08.026

Wiedmann, T., Minx, J., 2008. A Definition of Carbon Footprint. CC Pertsova, Ecological Economics Research Trends 2, 55–65. https://wiki.epfl.ch/hdstudio/documents/articles/a%20definition%20of%20carbon%20footprint.pdf.

Wiedmann, T., Wood, R., Minx, J.C., Lenzen, M., Guan, D., Harris, R., 2010. A Carbon Footprint Time Series Of The UK – Results From A Multi-Region Input–Output Model. Economic Systems Research 22, 19–42. https://doi.org/DOI: 10.1080/09535311003612591

World Meteorological Organization, 2024. WMO confirms that 2023 smashes global temperature record. https://wmo.int/news/media-centre/wmo-confirms-2023-smashes-global-temperature-record (accessed 30.8.25).

World Business Council for Sustainable Development and World Resources Institute (WRI & WBCSD), 2001. The greenhouse gas protocol- A corporate accounting and reporting standard. https://ghgprotocol.org/corporate-standard.

World Business Council for Sustainable Development and World Resources Institute (WRI and WBCSD), 2020. Corporate Value Chain (Scope 3) Accounting and Reporting Standard. World Resources Institute and World Business Council for Sustainable Development. https://ghgprotocol.org/sites/default/files/standards/Corporate-Value-Chain-Accounting-Reporing-Standard_041613_2.pdf.

Yacob, P., Wong, L.S., Khor, S.C., 2018. An empirical investigation of green initiatives and environmental sustainability for manufacturing SMEs. Journal of Manufacturing Technology Management 30, 2–25. https://doi.org/10.1108/JMTM-08-2017-0153

# Chapter 2. Bridging the SME Reporting Gap: A new model for predicting Scope 1 and 2 emissions.

## 2.1 Abstract

We present a novel statistical model for predicting Scope 1 and 2 emissions for small and medium-sized enterprises (SMEs). Trained on financial transaction data from over 100,000 UK SMEs, the model targets a business segment excluded from formal emissions reporting and often under-engaged in sustainability efforts. By leveraging scalable, objective data, our approach offers an accessible alternative to existing methods that rely on either coarse sectoral averages or detailed, resource-intensive firm-level activity data. In developing the model, we evaluate a range of predictors and find that incorporating industry-level variables beyond basic emission intensity significantly enhances predictive accuracy. We also observe diminishing returns from additional model complexity, reinforcing the value of a parsimonious, low-input design. The final model achieves RSQ values of 0.89 for Scope 1 and 0.72 for Scope 2, improves accuracy by up to 50% compared to sector-level estimates, and performs reliably on out-of-sample data. Our findings provide a simpler approach to emissions estimation for SMEs, supporting broader climate engagement among smaller actors.

## 2.2 Introduction

Although small and medium-sized enterprises (SMEs) constitute 99% of the United Kingdom's business population (Department for Business, Energy & Industrial Strategy (BEIS), 2022), their environmental impact remains insufficiently understood. Unlike larger firms, SMEs are exempt from formal emissions reporting requirements under frameworks such as the UK's Streamlined Energy and Carbon Reporting (SECR) scheme, which mandates the disclosure of emissions as part of broader extra-financial reporting[1] (HM Government, 2019; Meath et al., 2016). Globally, frameworks like SECR have increased the number of companies measuring and reporting their emissions (Assael et al., 2023; Tomar, 2023); however, the focus on large enterprises has led to a situation where only a subset of firms consistently report their emissions (Nguyen et al., 2021).

Extending reporting obligations to SMEs presents a complex challenge. While SMEs have a significant environmental impact collectively, their individual contributions are often minimal. Moreover, SMEs face numerous internal and external barriers to enhancing their sustainability, many of which are unique to firms of their size (Alipour & Rahimpour, 2020; Normative, 2024; Oyewole et al., 2024). Among these, information barriers and the inability to measure and monitor emissions are frequently cited as key challenges (British Business Bank (BBB), 2024). The process of emission measurement involves gathering detailed data on industrial processes and technologies from multiple sources, tracking annual consumption of fuel and energy in physical units, and applying calculations using relevant conversion factors. This is a challenging, laborious task, even for dedicated teams within big corporations (Assael et al., 2023; Goldhammer et al., 2017). These barriers, combined with regulatory exemptions, have allowed SMEs to remain largely under-engaged on the topic of sustainability (BBB, 2024).

To bridge the SME reporting gap, we develop a prediction model using a series of hierarchical regressions to estimate Scope 1 and 2 emissions requiring only minimal, widely available inputs: company turnover, industry classification, and industry-level characteristics drawn from public accounts. To tailor our model to the SME use case, it is fitted to emission estimates derived from anonymized financial transaction data (FTD) for over 100,000 UK SMEs.

[1] SECR disclosure requirements cover companies who meet at least two of the following criteria: (1) turnover of £36m or more; (2) balance sheet assets of £18m or more; (3) 250 employees or more. Companies consuming less than 40,000 kWh of energy per year are exempt.

By relying on routinely collected FTD rather than detailed activity data, this model places fewer requirements on SMEs while still capturing the key drivers of Scope 1 and 2 emissions. FTD-based emission estimates capture a simplified version of emissions, and their validation against actual emissions is challenging due to the scarcity of alternative data. Here, we argue that, for SMEs, precision is not paramount. A degree of absolute accuracy can be reasonably traded for substantial gains in efficiency and scalability. Prior work confirms FTD as a credible alternative to survey-based emissions data (Trendl et al., 2023), noting key advantages such as reduced calculation effort, scalability, and real-time application (Wells et al., 2025). Moreover, using a standardized, transparent methodology across firms helps avoid the inconsistencies associated with errors-prone self-reported data (Financial Reporting Council, 2021).

Through this work, we contribute to the understanding of emissions within this critical segment of the economy and provide a simple methodology to underpin more practical and accessible solutions for SMEs' emission estimation. We argue that developing this model into a publicly accessible tool would enable SMEs and other stakeholders to generate reliable emissions estimates, while encouraging greater engagement with sustainability among currently disengaged smaller actors.

The remainder of this article is organised as follows. First, we provide an overview of emission accounting for firms, the stakeholders of SME emissions data, and the current approaches for estimating non-reported emissions. Following this, we introduce the materials and methods used to calculate Scope 1 and 2 emissions from FTD and the statistical techniques used in developing our prediction model.  Subsequently, we present the results of our model, along with robustness checks and out-of-sample performance evaluations. The article concludes with a discussion of our key findings, their practical implications, limitations, and suggestions for future research directions.

## 2.3 Background and Context

### 2.3.1 Defining Emission Scopes

A firm's carbon footprint refers to the greenhouse gases (GHGs) that are produced as a by-product of its economic activities. The concept is defined by the accounting principles, system boundaries, reporting standards, and framework for calculations laid out in the GHG Protocol Corporate Accounting and Reporting Standards (WRI & WBCSD, 2004). This framework outlines clear, differentiated "Scopes" that make up a firm's footprint:

Scope 1 - Direct GHG emissions from operations that are owned or controlled by the company. This includes stationary combustion, mobile combustion, and process emissions.

Scope 2 - Indirect GHG emissions from the generation of purchased electricity, steam and heating/cooling consumed by the company.

Scope 3 - Other indirect GHG emissions from the upstream and downstream value chain.

These definitions indicate that one firm's Scope 2 and 3 emissions are the allocated Scope 1 emissions of another, emphasising that purchasers should account for emissions driven by their demand and highlighting the need for broader emission reduction strategies beyond their own production (Hertwich & Wood, 2018).

Emission reporting regulations, such as SECR, have historically prioritized the mandatory disclosure of energy-related Scope 1 and 2 emissions, with Scope 3 emissions often considered ambiguous, voluntary and deferred for future consideration (Assael et al., 2023; Hansen et al., 2022; Hettler & Graf-Vlachy, 2024; Nguyen et al., 2021). The focus on Scopes 1 and 2 facilitates manageable measurement, but provides a limited view, as Scope 3 can make up the majority of a firm's emission profile (Harangozo & Szigeti, 2017; Hertwich and Wood, 2018; Matthews et al., 2008; Stein & Khare, 2009). Given the distinct scopes, here we estimate Scope 1 and 2 emissions independently, noting the additional theoretical considerations for Scope 3 as well as the opportunity and novelty involved in leveraging FTD as a record of upstream emissions.

### 2.3.2 Implications of the SME Reporting Gap

Emission measurement is widely recognized as a critical first step for businesses in advancing sustainability, grounded in the principle "what gets measured gets managed" (WRI & WBCSD, 2004, p. 11). Research supports this, showing that firm emission measurement and disclosure is often followed by reductions (Tomar, 2023), through the creation of emission baselines and an improved understanding of emission sources (Harangozo and Szigeti, 2017; Talbot and Boiral, 2013). In the absence of this information, SMEs remain limited in their ability to identify and implement effective decarbonisation strategies. The lack of emissions data also places SMEs at a competitive disadvantage with both downstream consumers (Deloitte, 2023; McKinsey, 2023) and upstream suppliers (Hettler and Graf-Vlachy, 2024; World Economic Forum, 2022) increasingly focusing on environmental factors in decision-making. Firms with clearly defined sustainability strategies and emission measurement approaches may be favoured over those without, leading to important long-term effects on company performance.

Additionally, policymakers and financial institutions (FIs) are impacted by the absence of comprehensive SME emissions data. For the policymaker, increased access to and a greater understanding of SME emissions data is essential for developing effective environmental policies capable of driving sector-wide behavioural and operational change. This presents a significant challenge to realising the UK's legally binding Carbon Budgets, which are designed to provide a structured and measurable pathway to achieving net zero greenhouse gas (GHG) emissions by 2050 (BEIS, 2021). For FIs, these data gaps make it difficult to design, assess, and implement financed emissions strategies, limiting their ability to align lending and investment portfolios with climate targets. Under the Partnership for Carbon Accounting Financials (PCAF, 2022) framework, FIs must measure and disclose their financed emissions. Without accurate SME emission data, FIs rely on estimation techniques, which introduce uncertainty and reduce the precision of portfolio-level climate impact assessments. Improving the quality and accessibility of SME emissions data would enable financial institutions to pursue more targeted green financing, with important knock-on effects for the wider economy (OECD, 2024).

### 2.3.3 Methods for Estimating Non-reported Emissions

In the absence of reported emissions and internal physical usage data, a common approach for predicting firm-level emissions involves combining industry-average emission intensity factors, derived from official statistical data or established environmentally extended input-output tables, with firm-specific variables such as revenue and employment (Hirvonen et al., 2021; OECD, 2023). This methodology is employed in both financed emission calculation (PCAF, 2022), and in broader assessments of the overall contribution of SMEs emissions. Such estimates suggest that SMEs may account for approximately half of all UK business-driven emissions (BBB, 2021, p. 15).

Whilst straightforward and simple to produce, sole reliance on industry emission intensity factors ignore firm-level heterogeneities within an industry and assume a constant relationship between emissions and the specified economic variable (European Commission, 2022). Additionally, industry averages are skewed by larger firms and may not represent smaller businesses. Despite their physical numbers, SMEs contribute far less to total industry turnover, labour, and emissions.

Beyond basic estimates, more sophisticated models have been developed using machine learning and statistical methods, summarised in Table 1 (Assael et al., 2023; Goldhammer et al., 2017; Griffin et al., 2017; Heurtebize et al., 2022b; Nguyen et al., 2021). These models rely on publicly disclosed firm-level emission data, resulting in their development being predominantly based on larger firms. This limits applicability to smaller firms (Assael et al., 2023) and can introduce inconsistencies due to quality and reliability issues of self-reported data (Dragomir, 2012; Talbot & Boiral, 2013). Moreover, the reliance on self-reported data places an important limitation on model sample sizes, with the largest study involving 4,360 firm-year observations (Assael et al., 2023).

*Table 2.1 Summary of existing prediction models for company emissions*

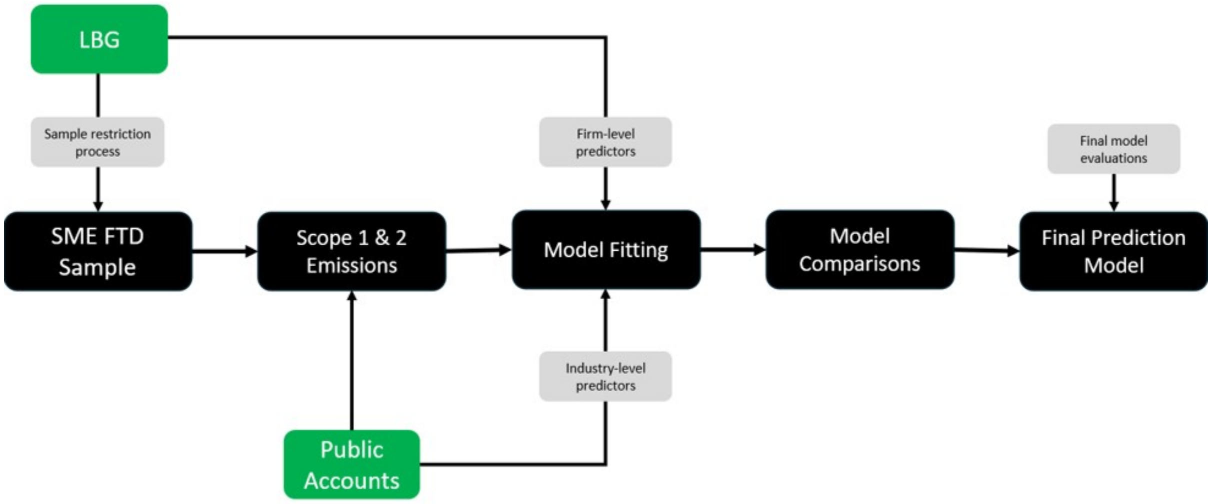
| Model | Observations | Sample Data | Model | Scope 1 & 2 Treatment |
|---|---|---|---|---|
| Goldhammer et al. (2017) | 93 | Self-reported figures | Ordinary Least Squares | 1 and 2 jointly |
| Griffin et al. (2017) | 1,657 | | Ordinary Least Squares | 1 and 2 jointly |
| Nguyen et al. (2021) | 2,289 | | Machine Learning | 1 and 2 individually |
| Heurtebize et al. (2022) | 3,500 | | Machine Learning | 1 and 2 individually |
| Assael et al. (2023) | 4,360 | | Machine Learning | 1 and 2 individually |

## 2.4 Materials and Methods

The prediction model presented here is developed using the anonymised FTD of businesses banking with Lloyds Banking Group (LBG), one of the UK's largest banks, with a business banking client base of over 900,000 firms (Lloyds Banking Group, 2023, p. 12). All businesses included in our analysis are SMEs, henceforth defined as firms with an annual turnover of less than £36 million. The model is based on the single year 2021, given the availability of both internal banking data and official industry data. The intention of this paper is to serve as a case study, under the understanding that additional model years would rely on annual data updates. These model parameters would be expected to change only gradually over time, supporting most practical applications.

The methodological approach to developing our model is described in the following subsections, whilst Figure 2.1 provides a schematic overview of this process. First, SME FTD is obtained from LBG, and an iterative sample restriction process is implemented. The FTD of the resulting sample is then combined with spend-based emission conversion factors, sourced from official UK government datasets, to estimate Scope 1 and 2 emissions. Starting with the most basic specification and increasing in complexity, we then fit a series of hierarchical regression models to the FTD emission estimates. For this, firm-level predictors are sourced from LBG, while industry-level predictors are drawn from official UK sources, including government databases and the Office for National Statistics (ONS). These models are evaluated and compared against each other for model performance. Our preferred model is selected, and further model evaluation testing is conducted.

*Figure 2.1 - Overview of prediction model development. Black rectangles indicate the key stages of the modeling process, green denotes data sources, and gray highlights the specific inputs and processes required at each step. LBG, Lloyds Banking Group.*




## 2.4.1 Sample Selection

To ensure the analysis is based on high-quality and comprehensive financial data histories, a series of emission scope-specific restrictions is required. A full account of this process, including rationale and impact of each step is provided within Supporting Information S2.1. These steps are designed to isolate SMEs who use LBG as their primary banking provider, excluding those with significant financial activities through other institutions.

Additionally, it is necessary to restrict our sample to SMEs for which financial data can reasonably proxy both firm revenue and emission-generating activities. As a result, we exclude agricultural SMEs due to their substantial process-related emissions not directly reflected in FTD. We also omit certain service-sector SMEs, such as those in insurance, real estate, and legal or accounting services, whose frequent handling of non-turnover credit lines (e.g., deposits) leads to systematic overestimation of turnover. Table S2 of Supporting

Information S2.1 provides a full list of explanations of exclusions where relevant, along with sample observations for each industry.

With the sample restrictions steps in place, our final sample for Scope 1 consists of 39,704 companies across 44 industries. Our Scope 2 sample is larger, covering 92,714 companies across 54 industries, as restrictions are based solely on electricity spend, whilst Scope 1 restrictions span multiple spend categories. Collectively, these companies account for an estimated £70 billion in annual revenue and encompass over 95 million financial transactions in 2021.

### 2.4.2 FTD-based Emission Estimates

#### *2.4.2.1 FTD Classification and Adjustments*

A transaction is defined as any spend that occurs on a business account or credit card, including electronic transfers, online transactions, and card payments. Each transaction is classified according to the bank's internal transaction classification system (TCS), which assigns recipient-based categories (Trendl et al., 2023; Wells et al., 2025). Of these categories, two are relevant for the estimation of Scope 1 and 2 emissions: "Energy & Utilities" and "Vehicle Fuelling". The former refers to 4 sub-categories of expenditure, containing energy providers selling electricity, natural gas, and heating oil, whilst the latter refers to diesel and petrol purchased for mobile combustion. With these categories, we capture the Scope 1 and 2 emissions generated from a firm's energy-use, mirroring what is required of larger firms under the SECR framework.

After categorization, we apply two additional adjustments to the spending amounts classified under "Energy & Utilities". First, in instances where UK utility firms are known to provide electricity, gas, and heating oil to consumers, we disaggregate spending into Scope 1 and Scope 2 using industry-level averages on energy spend by energy type (ONS, 2021). Second, due to energy tariff bands, businesses consuming more energy typically pay less on a per-unit basis. We therefore use UK statistics on consumption bands and energy unit prices (DESNZ, 2024) to calculate adjustment factors for different consumption bands. We provide our methodology for these adjustments in Supporting Information S2.2.

*2.4.2.2 Scope 1 and 2 Emission Conversion Factors*

Next, we create a set of conversion factors to capture the emissions associated with a unit purchase of different energy and fuel types. We divide emission per unit figures (CO2e / kwh (or litre); BEIS and DEFRA, 2022) by average price paid by non-domestic consumers per unit (£ / kwh (or litre); DESNZ, 2024) to return emissions produced per pound spent (CO2e / £).

As spending within "Vehicle Fuelling" reflects the consumption of both petrol and diesel, we produce a single conversion factor through a weighted average based on the UK's consumption split between petrol and diesel (DESNZ and BIES, 2022b) and the unit price difference. For gas and oil, we use an average conversion factor given the near zero difference in individual factors. Finally, for electricity, our emission conversion factor is based on the UK grid's emission intensity (location-based method). These conversion factors are multiplied with the annual expenditures of our sample to estimate Scope 1 and 2 emissions based on their energy consumption. Tables containing direct emission conversion factor calculation, and subsequent sample emission summary statistics can be found in Supporting Information S2.3.

## 2.5 Hierarchical Regression Models

Given the nature of the data, where firms are nested within industries, we specify a hierarchical regression. This allows for industry-specific effects to vary whilst including non-independent observations within the same industry (Gelman and Hill, 2006).

To support the statistical reasoning of our model, we ensure a minimum industry sample size of 50 and adequate representation across turnover brackets, requiring at least 5 firms in each: below £250,000, £250,000–£1 million, and above £1 million. Whilst there is no defined minimum number of observation required in each group to perform a hierarchical regression (Gelman and Hill, 2006), this is relatively close to the 50/50 rule, where a minimum of 50 groups and 50 units per group assists the hierarchical method of estimation (Ali et al., 2019).

### 2.5.1 Dependent Variable and Predictor Variables

Following Goldhammer et al. (2017), we predict the natural logarithm (ln) of the firm's Scope 1 and 2 emissions. We apply the same transformation to all independent variables, allowing for coefficients to be interpreted as elasticities (the percentage change in the dependent variable for a one percent change in the independent variable). Table 2 introduces the predictor variables, providing the rationale for their inclusion and the method of procurement.

*Table 2.2 Predictor Variables for the hierarchical regression model*

| Variable | Unit | Description | Calculation / Procurement |
|---|---|---|---|
| Industry-level | | | |
| Industry Intensity | ***kg Co2e / £*** | Representing the emissions produced per unit of output across each industry, this is the only industry variable to change to reflect the scope of emissions in focus. Currently, industry intensity is used exclusively in non-reported emissions estimation methods, as it serves as a key indicator of variation in emissions-generating activity across industries. For this reason, it is an important feature in our model. | This calculation utilizes the 2021 UKMRIO model (Owen and Kilian, 2025), a model of the UK economy designed to study the UK economy and its environmental impact.<br><br>For scope 1 emissions, the calculation involves simply dividing industry direct emissions (F) by industry output in purchasers' prices (X). For scope 2 emissions, we follow an approach outlined by Hertwich and Wood (2018) to return industry scope 2 emissions, which are divided by industry output in purchasers prices (X). |
| Industry Turnover Skew | ***%*** | A variable capturing the share of industry turnover from large firms assesses industry skew. We interact this variable with industry intensity, under the hypothesis that where industry skew is high, industry intensity is heavily influenced by larger firms in the industry, potentially reducing the suitability of industry averages for SME. | ONS industry data is used to generate a percentage figure for the share of turnover produced by firms with an annual turnover exceeding £1 million (ONS, 2024). |
| Firm-level | | | |
| SIC Code | ***SIC*** | The Standard Industrial Classification (SIC) identifies a firm's primary business activities. Including an industry identifier allows for more accurate comparisons of expected emission intensity among firms performing similar activities. | SIC classifications are the standard industry categories in the UK, with official economic and environmental data commonly published on a 2 digit SIC code basis, whilst it is standard practice for firms banking with LBG to be assigned SIC codes at both the 2- and 3-digit level by relationship managers. Of the 88 possible two-digit SIC codes (See Supporting Information S2.1), 44 are included in the Scope 1 sample and 54 in the Scope 2 sample. |
| Turnover | ***£*** | Turnover serves as an indicator of a firm's operational scale and is included in the analysis because, all else being equal, higher turnover typically correlates with greater emissions. However, firms with higher turnover may also benefit from economies of scale or possess greater capacity to invest in emission-reduction technologies, thereby decreasing their emission intensity. | Credit turnover is derived by summing the credit transactions recorded within a firm's bank account in the year. |
| Margin | ***%*** | Firm margin is included as firms operating at a greater financial margin may have capacity to allocate resources towards sustainable practices, potentially lowering their emission intensity. | We calculate profitability by simply subtracting total debit from total credit and make the variable relative by dividing the result by credit turnover. |

| | | | |
|---|---|---|---|
| Capital Spend – Assets | % | Capital expenditure reflects a firm's investment in assets. This variable is included because high capital spending on newer energy-efficient technologies may serve to lower a firm's emission intensity. For ease of calculation, we define capital expenditure as purchases that require financing. | Asset capital spend is calculated by summing the percentage of total spend, spent on asset financing and plant hire for the three-year period ending 2021. We take an average of these percentages to gain a single figure, capturing asset spending across the three-year window. |
| Capital Spend – Vehicles | % | | Vehicle capital spend is calculated by summing the percentage of total spend, spent vehicle financing for the three-year period ending 2021. As above, we calculate a single figure, capturing vehicle spending across the three-year window. |

### 2.5.2 Model Evaluation

We develop four progressively complex models to estimate firm-level emissions for each emissions scope, under the following hierarchical regression structure:

$$ln\, Y_i = \beta_n(fixed\_effects) + \alpha_n(variable\_effects) + \varepsilon_i$$

Where:

$Y_i$ represents the absolute emissions of firm i

$\beta_n$ represents the coefficients for fixed predictors whose effects do not vary across industries

$\alpha_n$ represents the coefficients for the variable predictors whose effects do vary across industries

ε represents the residual error term

We examine predictor influence sequentially, starting with two variables and increasing to eight, evaluating the required complexity to explain the dependent variables (Cohen et al., 2013). The baseline model includes only industry classification and turnover. Building on this, we add industry-average emissions intensity to align with data used under PCAF, followed by an industry skewness variable to test whether the industry-average provides sufficient sectoral detail. Model 4 then introduces additional firm-level variables to capture intra-industry heterogeneity.

For each model, we present coefficients and statistical significance, indicating the strength and direction of predictor relationships to the dependent variables, along with their standard errors, reflecting the estimate uncertainties. We evaluate both model fit (R-squared - RSQ) and model fit with respect to complexity (Akaike information criterion - AIC).

Following model comparison, our preferred model is subjected to further rounds of evaluation. Predictive accuracy is evaluated against the PCAF framework, which assumes uniform firm-level emissions intensity within industries. Comparing predicted intensity (predicted emissions divided by turnover) with PCAF estimates provides a reference point for assessing prediction error, while identifying areas of model strength and weakness. We then test for stability and out-of-sample performance through an 80/20 train-test split with stratified sampling, applying a 5-fold cross-validation (James et al., 2021) to ensure balanced representation and robust performance evaluation. A further round of out-of-sample testing is conducted on a sample of 50 large businesses that report under SECR. As SMEs are not required to disclose emissions, model performance can only be assessed against larger firms for external one-to-one validation. In the SECR sample, annual turnover ranges from £38 million to £211 million. By contrast, our model is fitted exclusively to emissions estimates for SMEs with turnover below £36 million, with 80% falling under £1 million. We also identified several inconsistencies and errors in the self-reported SECR data, introducing further uncertainty into this approach. Finally, we perform a sensitivity analysis to evaluate the robustness of results under varying sample selection criteria.

## 2.6 Results

### 2.6.1 Model Performance Comparison

#### *2.6.1.1 Scope 1*

Table 3 presents summary statistics and coefficient estimates for each model iteration. Across all four model specifications, firm turnover consistently shows a strong, positive relationship with Scope 1 emissions. A 1% increase in turnover corresponds to an average of 0.77% increase in Scope 1 emissions, with the effect strengthening in more complex models.

In Model 2, we find the inclusion of industry average only slightly improves model performance (an RSQ increase of only 0.003), with its coefficient being relatively weak and negative. Accounting for industry skew in Model 3 shifts the industry intensity coefficient to positive, and increases its magnitude. A 1% increase in industry intensity is now related to an 0.84% increase in Scope 1 emissions. The interaction term between industry intensity and industry skew suggests that for industries involving larger, dominant firms, industry intensity has a weaker effect on SME emissions. This supports the hypothesis that industry-wide emissions metrics may not fully capture SME emissions patterns. The 9% increase in model performance here highlights the importance of incorporating additional industry structure variables, with larger firms appearing to emit less per unit of turnover, which may be explained by an economies of scale effect.

Adding firm-specific factors beyond turnover does not significantly enhance model performance. While industry variables gain strength, firm-level predictors yield mixed results: a 1% increase in assets or vehicle spending leads to only a 0.02% and 0.01% rise in Scope 1 emissions, respectively. In contrast, a 1% increase in firm margin results in a notable 0.57% emissions increase, suggesting more profitable firms emit more, ceteris paribus. Despite these additions, model performance slightly declines (RSQ = 0.882).

*2.6.1.2 Scope 2*

In all Scope 2 models displayed in Table 3, turnover remains a strong predictor of emissions. Whilst the initial strength of this variable is smaller for Scope 2, this coefficient increases to 0.79 with the final model.

Notably, we find the initial impact of industry intensity to be 0.16, significantly different to the relationship observed amongst Scope 1 emissions. The inclusion of industry skew in Model 3 again reverses the industry intensity coefficient, this time to negative, suggesting that firms in industries with a higher Scope 2 intensity, have lower individual Scope 2 emissions, holding all else constant, with the effect diminishing as industry skew increases. This result may be driven by the calculation method of industry Scope 2 emissions, which is based on electricity expenditures from Supply and Use Tables. Firms in industries with high electricity consumption generally secure lower unit prices, with larger firms benefiting most. This potential underestimation of total industry emissions may create a discrepancy between industry-wide and firm-level Scope 2 emissions. While the negative, statistically significant interaction term suggests an explanation centred on the influence of industry skew, other contributing factors could include regional tariff variations or industry-specific differences in electricity consumption patterns.

In Model 4, additional firm-level predictors improve RSQ to 0.728 and reduces AIC to 208,299. As in Scope 1, margin has the largest impact, while our capital assets variable has minimal influence (coefficient = 0.02). Interestingly, vehicle spending has a negative coefficient, where a 1% increase is associated with a 0.02% decrease in Scope 2 emissions.

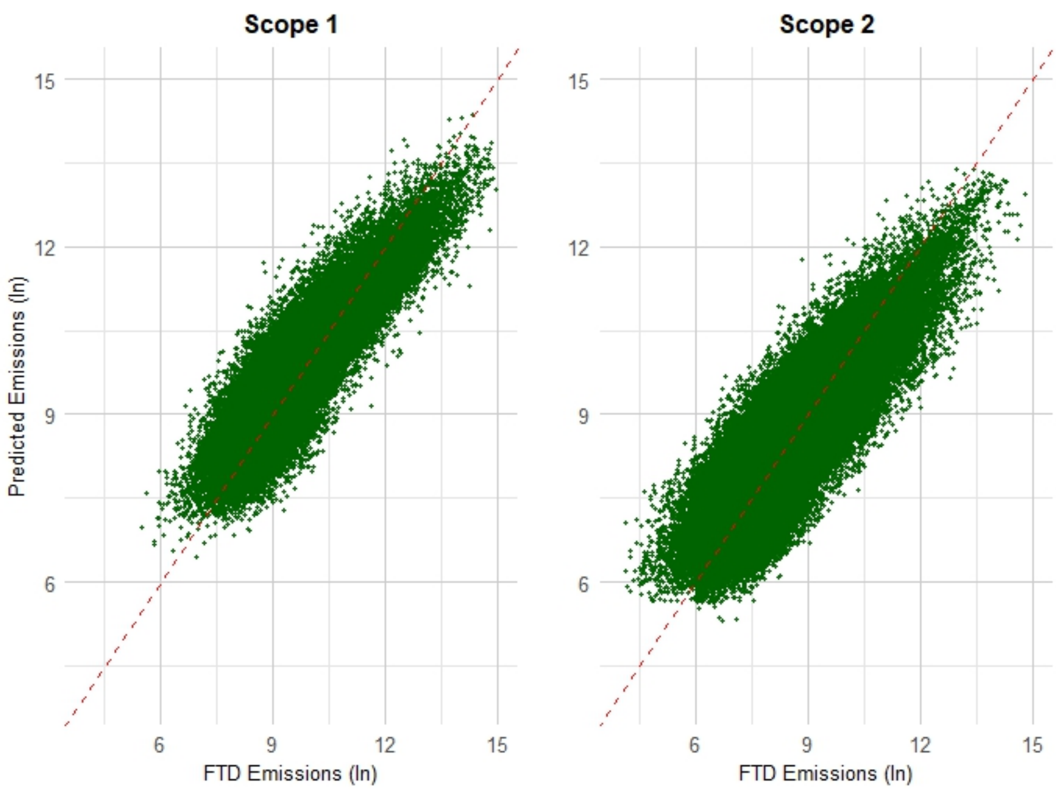
*Table 2.3 Model summary. Asterisks indicate significance level where P values < 0.05 = *, < 0.01 = **, and < 0.001****

| | Scope 1 | | | | | | | | Scope 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
| | **Coef** | **Std** | **Coef** | **Std** | **Coef** | **Std** | **Coef** | **Std** | **Coef** | **Std** | **Coef** | **Std** | **Coef** | **Std** | **Coef** | **Std** |
| ln turnover | 0.79*** | 0.01 | 0.77*** | 0.01 | 0.77*** | 0.02 | 0.79*** | 0.02 | 0.70*** | 0.01 | 0.77*** | 0.01 | 0.78*** | 0.01 | 0.79*** | 0.01 |
| ln scp1_intensity | | | -0.08*** | 0.01 | 0.84*** | 0.21 | 1.26*** | 0.21 | | | | | | | | |
| ln scp2_intensity | | | | | | | | | | | 0.16*** | 0.00 | -1.49*** | 0.10 | -1.28*** | 0.10 |
| ln skew | | | | | -0.47*** | 0.03 | -1.03*** | 0.04 | | | | | 0.01 | 0.03 | -0.51*** | 0.04 |
| intensity*skew | | | | | -0.34*** | 0.05 | -0.42*** | 0.05 | | | | | 0.38*** | 0.03 | 0.34*** | 0.03 |
| ln assets | | | | | | | 0.02*** | 0.00 | | | | | | | 0.02*** | 0.00 |
| ln vehicles | | | | | | | 0.01*** | 0.00 | | | | | | | -0.02*** | 0.00 |
| ln margin | | | | | | | 0.57*** | 0.02 | | | | | | | 0.55*** | 0.01 |
| RSQ (fixed) | 0.655 | | 0.609 | | 0.408 | | 0.425 | | 0.567 | | 0.630 | | 0.586 | | 0.585 | |
| RSQ (total) | 0.798 | | 0.801 | | 0.891 | | 0.882 | | 0.690 | | 0.721 | | 0.717 | | 0.728 | |
| AIC | 79,427 | | 79,341 | | 79,111 | | 78,299 | | 211,955 | | 210,662 | | 210,362 | | 208,299 | |

### *2.6.1.3 Prediction Model Selection*

The preceding results indicate that the inclusion of additional firm-level variables did not significantly enhance model performance, despite the increased complexity and data resources needed to acquire these predictors. This indicates diminishing returns to model complexity; a simple, parsimonious solution can reach useful levels of predictive power whilst being practical to apply. Based on performance and applicability, we select Model 3 as our preferred model for each scope. Figure 2.2 presents the model prediction versus FTD emission estimates.

*Figure 2.2 - Model predictions against FTD emission estimates.*




## 2.6.2 Prediction Model Evaluation

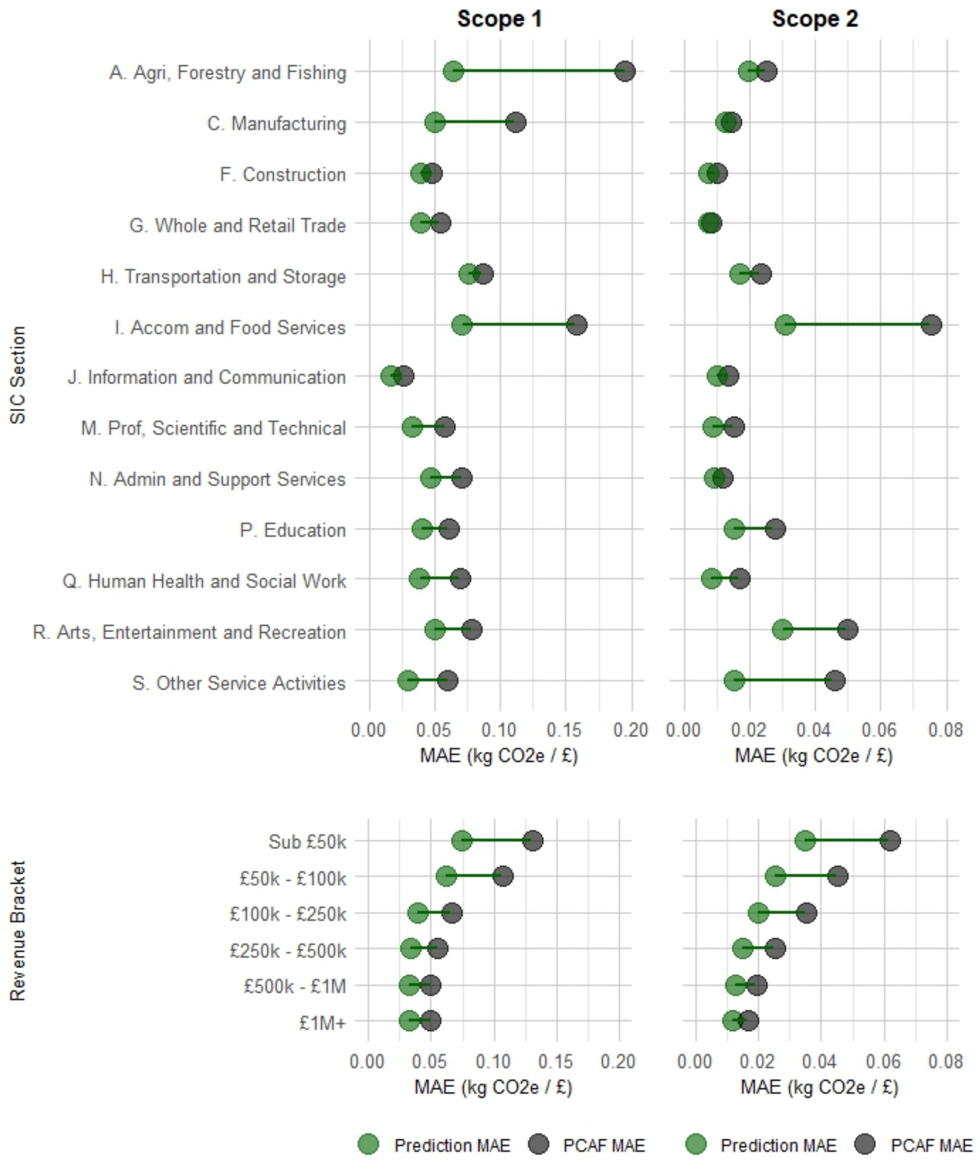
To evaluate accuracy, we calculate the Absolute Error (AE) and Absolute Percentage Error (APE) of Model 3's predictions relative to their FTD counterparts. Given the left-skewed distribution of firm sizes in the sample due to greater physical numbers of smaller firms, we report both the mean and median values of AE and APE.

*2.6.2.1 PCAF Comparison*

Assessing differences in AE and APE between our model and the PCAF framework reveals large error reductions for Scope 1 emissions. Here, the PCAF framework results in a Mean AE of 0.108 kg CO2e/£. Whilst Model 3 improves this significantly, reducing the error to 0.047 kg CO2e/£, a reduction of over 50%. The decrease in Median AE is smaller in absolute terms but falling from 0.092 to 0.033 kg CO2e/£ represents another reduction of over 50%. For Scope 2 emissions, the reduction is more modest, with a decrease in both Mean AE (0.033 kg CO2e/£ to 0.022 kg CO2e/£) and Median AE (0.022 to 0.014 kg CO2e/£).

We plot both PCAF and predicted Mean AE by aggregated SIC section and turnover brackets in Figure 2.3. We observe the greatest reduction in error occurs in industries with the largest starting error. However, error reductions are observed across all industries. Assessing the Mean AE by firm size demonstrates that initial errors decrease with firm size, and that predicted errors diminish consistently across all revenue categories. This highlights the model's effectiveness, significantly reducing the errors for smaller firms, which typically experience the highest error under the average approach.

*Figure 2.3 - Comparison of mean absolute error (MAE) from Model 3 and Partnership for Carbon Accounting Financials (PCAF) across industry sectors and turnover brackets.*




## *2.6.2.2 Out-of-Sample Performance*

Table 4 presents the output from the 5-fold cross-validation for Model 3. For each fold we report Mean and Median AE and APE, as well as the RSQ of the test data. The model consistently produces similar results across the folds, indicating stability. For Scope 1 the Median AE across all folds was 6.8 $tCO_2e$ (std. = 0.03 $tCO_2e$), and average RSQ was 0.78 (s.d. = 0.00) for Scope 1. For Scope 2, the Median AE across all folds was 2.4 $tCO_2e$ (s.d. = 0.01

$tCO_2e$), with an average RSQ of 0.73 (s.d. = 0.00). This indicates that the model performs robustly when used to predict out-of-sample FTD-based emission estimates.

*Table 2.4 Performance Metrics for Model 3 Test and Train Iterations*

| | Scope 1 | | | | | Scope 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | | Median | | | Mean | | Median | |
| **Fold** | **RSQ** | **AE** | **APE** | **AE** | **APE** | **RSQ** | **AE** | **APE** | **AE** | **APE** |
| | | t $CO_2e$ | % | t $CO_2e$ | % | | t $CO_2e$ | % | t $CO_2e$ | % |
| 1 | 0.778 | 29.7 | 63.24 | 6.9 | 41.40 | 0.730 | 9.2 | 73.50 | 2.4 | 48.74 |
| 2 | 0. 778 | 31.0 | 63.35 | 6.8 | 41.45 | 0.730 | 9.6 | 74.65 | 2.4 | 48.69 |
| 3 | 0.782 | 30.5 | 62.57 | 6.8 | 41.35 | 0.730 | 9.4 | 73.96 | 2.4 | 48.21 |
| 4 | 0.782 | 30.9 | 62.13 | 6.9 | 41.12 | 0.729 | 9.2 | 73.95 | 2.4 | 48.53 |
| 5 | 0.776 | 30.4 | 62.75 | 6.9 | 41.26 | 0.728 | 9.7 | 74.22 | 2.4 | 48.79 |
| **Mean** | **0.779** | **30.5** | **62.81** | **6.8** | **41.31** | **0.729** | **9.4** | **74.05** | **2.4** | **48.59** |
| **Std** | **0.003** | **0.5** | **0.50** | **0.03** | **0.13** | **0.000** | **0.2** | **0.42** | **0.01** | **0.23** |

Further out-of-sample testing is presented in Supporting Information S2.4, where we apply our models to predict emissions for 50 large businesses reporting under the SECR framework. We observed a median absolute percentage error (APE) of 25.32% for Scope 1 emissions. Errors are substantially higher for Scope 2 emissions, with a median APE of 65.08%. Although these results are broadly in line with error levels exhibited in Table 4, we observe low RSQ scores and high mean errors, largely driven by a small number of extreme outliers.

### *2.6.2.3 Impact of Sample Selection*

Up to this point, the results presented are based on the restrictions steps laid out in Supporting Information S2.1. For transparency, in Table 5, we now present a sensitivity analysis on Model 3's performance as we alter trimming thresholds of energy and fuel spend intensity (Step 5 in Table 6). In both scopes, we observe increasing RSQ value with increased levels of trimming. Both the Mean and Median AE significantly reduces as the trimming threshold increases, with Median APE reaching 30% and 34% for Scope 1 and 2 respectively. To balance data quality with sample size, we apply a conservative 10% trimming, removing outliers while preserving diversity and representativeness.

*Table 2.5 Performance Metrics for Model 3 Across Different Sample Restrictions*

| Metric | | Unit | 5% Trim | 10% Trim | 15% Trim | 20% Trim | 25% Trim |
|---|---|---|---|---|---|---|---|
| **Scope 1** | | | | | | | |
| Firms | | | 50,290 | 39,704 | 30,217 | 21,962 | 15,140 |
| Industries | | | 46 | 44 | 41 | 36 | 34 |
| RSQ (Fixed) | | | 0.379 | 0.408 | 0.440 | 0.418 | 0.438 |
| RSQ (Total) | | | 0.855 | 0.879 | 0.898 | 0.921 | 0.932 |
| Mean | AE | t $CO_2$e | 35.03 | 30.46 | 26.56 | 23.46 | 20.99 |
| | APE | % | 75.46% | 62.71% | 53.40% | 46.30% | 40.26% |
| Median | AE | t $CO_2$e | 7.72 | 6.80 | 6.02 | 5.33 | 4.71 |
| | APE | % | 45.29% | 41.22% | 37.36% | 33.92% | 30.43% |
| **Scope 2** | | | | | | | |
| Firms | | | 104,272 | 92,714 | 81,094 | 69,477 | 57,903 |
| Industries | | | 54 | 54 | 54 | 52 | 52 |
| RSQ (Fixed) | | | 0.540 | 0.586 | 0.628 | 0.664 | 0.697 |
| RSQ (Total) | | | 0.662 | 0.717 | 0.761 | 0.798 | 0.832 |
| Mean | AE | t $CO_2$e | 10,925 | 9,476 | 8,397 | 7,547 | 6,897 |
| | APE | % | 89.98% | 73.60% | 62.12% | 53.22% | 45.87% |
| Median | AE | t $CO_2$e | 2,663 | 2,364 | 2,098 | 1,848 | 1,604 |
| | APE | % | 53.68% | 48.52% | 43.49% | 38.61% | 33.91% |

## 2.7 Discussion

### 2.7.1 Key Findings

New modelling approaches, such as the one detailed in this paper, offer a practical way forward to addressing emissions data gaps among SMEs. Here, we have provided a use-case for utilising SMEs' financial transaction data (FTD) as proxies for emission-generating activities, estimating the emissions for an otherwise unobserved subset of the business population. We then use this emissions data to develop new models that predict Scope 1 and 2 emissions.

We identify models that can achieve an RSQ of 0.89 for Scope 1 emissions and 0.72 for Scope 2 emissions. This translates to a reduction in error of over 50%, when compared to the simple estimation techniques currently used under frameworks such as PCAF. This error reduction can be attributed to three main findings. First, we observe a consistent, strong and significant relationship between *ln* turnover and our dependent variables (*ln* Scope 1 & 2 emissions) with a coefficient of 0.7 at the lowest. This log-log relationship implies that the relationship between the raw variables is not linear, as is assumed under simple estimation techniques. Second, we find that accounting for industry characteristics beyond basic emission intensity significantly improves model performance, particularly for Scope 1 emissions. We observe a model performance increase of nearly 10% by incorporating industry skew, underscoring the importance of additional factors currently overlooked. Finally, we find that the inclusion of detailed, firm-specific variables to try and explain heterogeneities beyond size leads to diminishing model performance gains, whilst significantly disadvantaging the practical application of the model. A pragmatic approach balances complexity with accuracy to enhance accessibility.

With these findings, this study supports the notion that regression analyses can provide an alternative method to estimating emission data, in the absence of reported data (Goldhammer et al., 2017). By tailoring the model specifically to the SME context, we address a critical limitation in the current literature (Assael et al., 2023), ensuring inclusion of firms with the highest likelihood of emissions data gaps.

### 2.7.2 Practical Implications

Regulatory exclusions and firm size-specific barriers have kept SMEs largely absent from emissions reporting, limiting their engagement in improving sustainability and leaving their environmental impact poorly understood. The emission prediction models described here are made available in Supporting Information S2.6, enabling users to produce SME emission estimates from minimal inputs (namely turnover and industry), while improving on the accuracy of sector-level average estimates.

These models give SMEs a crucial starting point to both understand and reduce their environmental impact - an unmet need in the current landscape (BBB, 2024). When delivered through behaviourally informed design packages tailored to the specific barriers SMEs face, emissions estimates can help simplify the complex landscape of SME net zero support. This approach tackles structural barriers like limited funding and expertise (Caldera et al., 2019; Menon and Ravi, 2021), while also allowing tailored support for common SME constraints, such as occupying rented premises (Mazhar et al., 2024).

The benefits of increased SME engagement through model availability drives the adoption of sustainability practices, helps SMEs maintain competitiveness against larger firms, and removes obstacles to accessing external funding (Appiah-Kubi et al., 2024a; Madrid-Guijarro and Duréndez, 2024). Additional impact may be achieved when emission estimates are distributed by FIs with academic endorsement, leveraging established trusted relationships.

With this paper, we also begin to build the argument for FTD as a viable option for emission estimation at scale, particularly amongst smaller actors. This is an argument developed in the household space by Trendl et al. (2023) and Wells et al. (2025), but one not yet applied in the corporate emission reporting environment. This approach enables automated emission insights for SMEs while allowing external parties to understand emissions drivers at scale. Integrating FTD-based emissions data into the PCAF framework for estimating portfolio financed emissions could incentivise FIs to leverage their extensive data assets to develop methodologies that serve the public good.

### 2.7.3 Limitations

While modelling approaches improves accessibility, key limitations exist. Some industries are excluded due to limited sample sizes or unreliable turnover estimates. Whilst the method is also less suitable for sectors like agriculture, where large proportions of emissions arise from processes not captured in FTD. As a result, full sectoral coverage is not achieved.

A full list of excluded SIC codes and the implication of their removal is provided in Table S7 within Supporting Information S2.1. To assess the impact of exclusions, we compare business population data (BEIS, 2022) against the prediction model coverage to identify SIC codes that have the highest exclusion impact. We find that the model covers over 71% of UK SME population, with scope to expand this further if trade-offs in accuracy are acceptable. Of the excluded SMEs, 25% are excluded due to our initial exclusions, with an additional 4% excluded because of sample size within our restriction process. Of the excluded industries, most represent a minimal share of the UK's SME population. Notable exceptions include agriculture and service sectors such as insurance, real estate, legal and accounting which contain SME populations of 5% and 16% of the SME population respectively.

Additionally, FTD lack the granularity of receipt-level analysis. Emission conversion factors are therefore applied at the merchant category level only (Trendl et al., 2023). This means transaction-based emission estimation relies on assumptions regarding energy mix, since many energy suppliers provide both natural gas and electricity, and petrol stations sell both diesel and petrol. Although these assumptions are informed by official industry surveys, they remain approximations. Related to this, Scope 2 emissions are estimated using location-based emission intensity factors. For SMEs on renewable energy tariffs, emissions are overestimated when looking from a market-based perspective. Our model is primarily designed to predict energy-use emissions rather than precisely reflect specific tariff choices and in a publicly available version of the tool, users would have the option to toggle renewable tariff use on or off. This would both highlight the emission reduction opportunities of green tariffs and present a clear, actionable step for firms not yet on a renewable tariff.

This analysis reflects a one-year snapshot, aligning with practical deployment. Future iterations of the model can be easily updated, with new FTD and emissions conversion factors. In addition, the model is developed with UK data, its applicability is therefore limited to non-cash-based, digital economies with comprehensive FTD. SME activity is often shaped by regional factors, model development for other economies requires changes to conversion factors, currencies, grid intensities, and other differences. This presents an opportunity for collaborations between regional policymakers and financial institutions.

### 2.7.4 Future Research Directions

While this paper focuses on Scope 1 and 2 emissions due to regulatory emphasis, a firm's complete footprint includes Scope 3 emissions. As a growing area of regulatory interest, Scope 3 presents a strong use case for FTD. Future research could explore applying FTD to estimate these emissions, potentially enabling broader reporting of this underrepresented category while reducing the effort and cost associated with measurement

(Hettler and Graf-Vlachy, 2024). This may require integrating new data sources or methods to better capture supply chain emissions and provide a more complete SME profile.

A second area for further research is to compare transaction-based emission estimates for firms with alternative firm-level microdata. While Trendl et al. (2023) verifies FTD as a credible alternative to survey-based household estimates, equivalent validation in the commercial sector remains unexplored. Addressing this gap would help clarify the current uncertainties around the validity of FTD emissions estimation. Potential comparable emissions data include self-disclosed reports such as those submitted under the SECR framework. However, these are constrained by limited sample sizes, business size thresholds, and the need for manual data collection. In this study, a manual compilation of SECR reports conducted for out-of-sample testing revealed several challenges with the dataset, these are discussed in detail in Supporting Information S2.4. Survey data, such as the Annual Business Survey (ONS, 2024), offer broader sectoral and firm coverage, whilst avoiding the inconsistencies within self-reported methodologies. However, such sources present access barriers and do not permit one-to-one firm-level comparisons. Accountancy data also offers a promising avenue, with software providers increasingly integrate emissions profiling into their platforms (Sage Earth, 2025; Xero, 2025).

## 2.8 Conclusion

This paper introduces a new statistical model for predicting the Scope 1 and 2 emissions of SMEs using widely accessible information such as company turnover, industry classification, and sectoral characteristics from public accounts. By fitting the model to the FTD data from over 100,000 UK SMEs, we capture key firm and industry-level relationships that enable it to outperform commonly used benchmarking approaches.

A key finding from this research is that relatively simple model requiring only four variables, two publicly available industry variables and a firm's industry and annual turnover, can generate accurate and robust emission estimates while remaining practical for real-world application. These types of modelling solutions can address important gaps in emission reporting for SMEs, who lack the resources and time to designate to sustainability challenges. By simplifying access to more accurate emission profiles, we can increase engagement among SMEs and provide their stakeholders, including financial lenders and policymakers, with information for targeting sustainability strategies.

## 2.9 Notes

### 2.9.1 Acknowledgements

We would first like to thank Lara Vomfell, Jasmine Wells, Alistair Gilfillan, and the rest of the behavioural science team at Lloyds Banking Group for sharing their invaluable expertise at key points of this research, as well as their continued interest and support. Secondly, we thank Hamza Suleman, whose contributions and technical skills were instrumental in developing the model described here into a working and accessible implementation. Finally, we extend our sincere gratitude to Ranil Boteju, Trystan Davies, Andrew McCulloch, Neil Oliver and Jaideep Mann for making this project possible and supporting its successful publication.

### 2.9.1 Funding Information

Alec Phillpotts reports financial support was provided by Lloyds Banking Group. Jonathan Norman's time was funded by the UK Research and Innovation (UKRI) Council, as part of the Energy Demand Research Centre, grant reference number EP/Y010078/1.

## 2.10 References

Ali, A., Ali, S., Khan, S.A., Khan, D.M., Abbas, K., Khalil, A., Manzoor, S., Khalil, U., 2019. Sample size issues in multilevel logistic regression models. PLoS One 14, e0225427. https://doi.org/10.1371/journal.pone.0225427

Alipour, A., Rahimpour, M., 2020. Sustainability Barriers in SMEs: A study of strength of sustainability barriers and practical solutions in Green product lifecycle at SMEs. Jönköping University. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1462056&dswid=663

Appiah-Kubi, E., Koranteng, F.O., Dura, C.C., Mihăilă, A.A., Drigă, I., Preda, A., 2024. Green financing and sustainability reporting among SMEs: The role of pro-environmental behavior and digitization. Journal of Cleaner Production 478, 143939. https://doi.org/10.1016/j.jclepro.2024.143939

Assael, J., Heurtebize, T., Carlier, L., Soupé, F., 2023. Greenhouse Gases Emissions: Estimating Corporate Non-Reported Emissions Using Interpretable Machine Learning. Sustainability 15. https://doi.org/10.3390/su15043391

British Business Bank (BBB), 2024. Small Business Finance Markets Report 2024. https://www.british-business-bank.co.uk/about/research-and-publications/small-business-finance-markets-report-2024 (accessed 25.06.25).

British Business Bank (BBB), 2021. Small businesses and net zero. British Business Bank. https://www.british-business-bank.co.uk/research/smaller-businesses-and-the-transition-to-net zero/ (accessed 07.12.23).

Caldera, H.T.S., Desha, C., Dawes, L., 2019. Evaluating the enablers and barriers for successful implementation of sustainable business practice in 'lean' SMEs. Journal of Cleaner Production 218, 575–590. https://doi.org/10.1016/j.jclepro.2019.01.239

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd ed. Routledge, New York. https://doi.org/10.4324/9780203774441

Deloitte, 2023. Sustainability & Consumer Behaviour 2023. https://www.deloitte.com/uk/en/Industries/consumer/research/sustainable-consumer.html (accessed 10.03.24).

Department for Business, Energy & Industrial Strategy (BEIS), 2022. Business population estimates for the UK and regions 2022: statistical release [Dataset]. GOV.UK. https://www.gov.uk/government/statistics/business-population-estimates-2022/business-population-estimates-for-the-uk-and-regions-2022-statistical-release-html (accessed 07.12.23).

Department for Business, Energy & Industrial Strategy (BEIS), 2021. Net Zero Strategy: Build Back Greener (Presented to Parliament pursuant to Section 14 of the Climate Change Act 2008). https://www.gov.uk/government/publications/net zero-strategy (accessed 25.06.25).

Department for Business, Energy & Industrial Strategy (BEIS) and Department for Environmental Food & Rural Affairs (DEFRA), 2022. Greenhouse gas reporting: conversion factors 2021 [Dataset]. GOV.UK. https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2021 (accessed 23.07.24).

Department for Energy Security and Net Zero (DESNZ), 2024. Gas and electricity prices in the non-domestic sector [Dataset]. GOV.UK. https://www.gov.uk/government/statistical-data-sets/gas-and-electricity-prices-in-the-non-domestic-sector (accessed 23.07.24).

Department for Energy Security and Net Zero (DESNZ) and Department for Business, Energy & Industrial Strategy (BEIS), 2022. Digest of UK Energy Statistics (DUKES) 2022 [Dataset]. GOV.UK. https://www.gov.uk/government/statistics/digest-of-uk-energy-statistics-dukes-2022 (accessed 23.08.24).

Dragomir, V.D., 2012. The disclosure of industrial greenhouse gas emissions: a critical assessment of corporate sustainability reports. Journal of Cleaner Production 29–30, 222–237. https://doi.org/10.1016/j.jclepro.2012.01.024

European Commission, 2022. Annual Report on European SMEs 2021/2022. https://op.europa.eu/en/publication-detail/-/publication/c45665ad-fd9a-11ec-b94a-01aa75ed71a1/language-en (accessed 25.06.2025).

Financial Reporting Council, 2021. Thematic Review: Streamlined Energy and Carbon Reporting. https://www.frc.org.uk/news-and-events/news/2021/09/frc-publishes-review-findings-on-streamlined-energy-and-carbon-reporting/ (accessed 25.06.2025).

Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Higher Education from Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Goldhammer, B., Busse, C., Busch, T., 2017. Estimating Corporate Carbon Footprints with Externally Available Data. Journal of Industrial Ecology 21, 1165–1179. https://doi.org/10.1111/jiec.12522

Griffin, P.A., Lont, D.H., Sun, E.Y., 2017. The Relevance to Investors of Greenhouse Gas Emission Disclosures. Contemporary Accounting Research 34, 1265–1297. https://doi.org/10.1111/1911-3846.12298

Hansen, A.D., Kuramochi, T., Wicke, B., 2022. The status of corporate greenhouse gas emissions reporting in the food sector: An evaluation of food and beverage manufacturers. Journal of Cleaner Production 361, 132279. https://doi.org/10.1016/j.jclepro.2022.132279

Harangozo, G., Szigeti, C., 2017. Corporate carbon footprint analysis in practice – With a special focus on validity and reliability issues. Journal of Cleaner Production 167, 1177–1183. https://doi.org/10.1016/j.jclepro.2017.07.237

Hertwich, E.G., Wood, R., 2018. The growing importance of scope 3 greenhouse gas emissions from industry. Environ. Res. Lett. 13, 104013. https://doi.org/10.1088/1748-9326/aae19a

Hettler, M., Graf-Vlachy, L., 2024. Corporate scope 3 carbon emission reporting as an enabler of supply chain decarbonization: A systematic review and comprehensive research agenda. Business Strategy and the Environment 33, 263–282. https://doi.org/10.1002/bse.3486

Heurtebize, T., Chen, F., Soupé, F., Carvalho, R.L. de, 2022. Corporate Carbon Footprint: A Machine Learning Predictive Model For Unreported Data. BNP Paribas. https://www.institutlouisbachelier.org/wp-content/uploads/2022/03/corporate-carbon-footprint-en.pdf (accessed 25.06.2025)

Hirvonen, A., Anu, K., Ville, T., 2021. Measuring the carbon footprint of loans to domestic nonfinancial corporations (NFC) by banks. Manuscript. https://www.imf.org/-/media/Files/Conferences/2021/9th-stats-forum/33FinalPaperVilleTolkki.ashx#:~:text=6)%20The%20carbon%20footprint%20of,net%20zero%20by%20the%202035 (accessed 25.06.2025).

HM Government, 2019. Environmental Reporting Guidelines: Including streamlined energy and carbon reporting guidance. https://www.gov.uk/government/publications/environmental-reporting-guidelines-including-mandatory-greenhouse-gas-emissions-reporting-guidance accessed 25.06.2025)

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning with Applications in R, Second. ed. Springer, New York, United States.

Lloyds Banking Group (LBG), 2023. 2023 Lloyds Banking Group Annual Review. https://www.lloydsbankinggroup.com/assets/pdfs/investors/annual-report/2023/2023-lbg-annual-review.pdf (accessed 25.06.2025).

Madrid-Guijarro, A., Duréndez, A., 2024. Sustainable development barriers and pressures in SMEs: The mediating effect of management commitment to environmental practices. Business Strategy and the Environment 33, 949–967. https://doi.org/10.1002/bse.3537

Matthews, H.S., Hendrickson, C.T., Weber, C.L., 2008. The Importance of Carbon Footprint Estimation Boundaries. Environ. Sci. Technol. 42, 5839–5842. https://doi.org/10.1021/es703112w

Mazhar, M.U., Domingues, A.R., Yakar-Pritchard, G., Bull, R., Ling, K., 2024. Reaching for net zero: The impact of an innovative university-led business support programme on carbon management strategy and practices of small and medium-sized enterprises. Business Strategy and the Environment 33, 6940–6960. https://doi.org/10.1002/bse.3844

McKinsey, 2023. Do consumers care about sustainability & ESG claims?. https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/consumers-care-about-sustainability-and-back-it-up-with-their-wallets (accessed 08.07.24).

Meath, C., Linnenluecke, M., Griffiths, A., 2016. Barriers and motivators to the adoption of energy savings measures for small- and medium-sized enterprises (SMEs): the case of the ClimateSmart Business Cluster program. Journal of Cleaner Production 112, 3597–3604. https://doi.org/10.1016/j.jclepro.2015.08.085

Menon, R.R., Ravi, V., 2021. Analysis of barriers of sustainable supply chain management in electronics industry: An interpretive structural modelling approach. Cleaner and Responsible Consumption 3, 100026. https://doi.org/10.1016/j.clrc.2021.100026

Nguyen, Q., Diaz-Rainey, I., Kuruppuarachchi, D., 2021. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. Energy Economics 95, 105129. https://doi.org/10.1016/j.eneco.2021.105129

Normative, 2024. Survey: small businesses face recurring barriers to carbon reduction. Normative. https://normative.io/insight/smech-survey/ (accessed 08.07.24).

Organisation for Economic Co-operation and Development (OECD), 2024. Financing SMEs and Entrepreneurs 2024: An OECD Scoreboard, Financing SMEs and Entrepreneurs. OECD. https://doi.org/10.1787/fa521246-en

Organisation for Economic Co-operation and Development (OECD), 2023. Assessing greenhouse gas emissions and energy consumption in SMEs. https://www.oecd.org/en/publications/assessing-greenhouse-gas-emissions-and-energy-consumption-in-smes_ac8e6450-en.html (accessed on 25.06.2025).

Office for National Statistics (ONS), 2024. Annual Business Survey 2024. https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/annualbusinesssurvey (accessed 02.06.25).

Office for National Statistics (ONS), 2024. Turnover by division and size, 2019 [Dataset].

https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/adhocs/1986turnoverbydivisionandsize2019 (accessed 26.07.24).

Office for National Statistics (ONS), 2021. Business energy spending: experimental measures from the Office for National Statistics' business surveys - Office for National Statistics [Dataset]. https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/articles/businessenergyspendingexperimentalmeasuresfromtheofficefornationalstatisticsbusinesssurveys/2022-09-07 (accessed 08.08.24).

Owen, A., Kilian, L., 2025. 2025 Data Release of Consumption- based Accounts for the UK: Summary of Methods. https://assets.publishing.service.gov.uk/media/68220ff3db6463b14cd81916/Summary_of_Methods_2025_final.pdf.

Oyewole, A.T., Adeoye, O.B., Addy, W.A., Okoye, C.C., Ofodile, O.C., 2024. Enhancing Global Competitiveness Of U.S. Smes Through Sustainable Finance: A Review And Future Directions. International Journal of Management & Entrepreneurship Research 6, 634–647. https://doi.org/10.51594/ijmer.v6i3.876

Partnership for Carbon Accounting Financials (PCAF), 2022. The Global GHG Accounting and Reporting Standard Part A: Financed Emissions., 2nd ed. https://carbonaccountingfinancials.com/standard (accessed 05.02.25).

Sage Earth, 2025. Carbon Accounting Software. https://www.sage.com/en-gb/sage-business-cloud/sage-earth/ (accessed 05.02.25).

Stein, M., Khare, A., 2009. Calculating the carbon footprint of a chemical plant: a case study of akzonobel. J. Env. Assmt. Pol. Mgmt. 11, 291–310. https://doi.org/10.1142/S1464333209003373

Talbot, D., Boiral, O., 2013. Can we trust corporates GHG inventories? An investigation among Canada's large final emitters. Energy Policy 63, 1075–1085. https://doi.org/10.1016/j.enpol.2013.09.054

Tomar, S., 2023. Greenhouse Gas Disclosure and Emissions Benchmarking. Journal of Accounting Research 61, 451–492. https://doi.org/10.1111/1475-679X.12473

Trendl, A., Owen, A., Vomfell, L., Kilian, L., Gathergood, J., Stewart, N., Leake, D., 2023. Estimating carbon footprints from large scale financial transaction data. Journal of Industrial Ecology 27, 56–70. https://doi.org/10.1111/jiec.13351

Wells, J., Trendl, A., Owen, A., Barrett, J., Jobst, N., Leake, D., 2025. Targeting carbon reduction in UK households: A new segmentation model using financial transaction data. Journal of Industrial Ecology. https://doi.org/10.1111/jiec.70020

World Economic Forum, 2022. How 4 leading companies are tackling supply chain emissions. World Economic Forum. https://www.weforum.org/agenda/2022/11/scope3-supply-chain-emissions-cop27-ikea-philips-zf-unilever/ (accessed 10.03.24).

World Business Council for Sustainable Development and World Resources Institute (WRI & WBCSD), 2004. The greenhouse gas protocol- A corporate accounting and reporting standard- Revised edition. https://ghgprotocol.org/corporate-standard (accessed 05.02.25).

Xero, 2025. Carbon Accounting and Management Apps - Carbon Footprint Tracking and Management Apps for Small Business that Integrates with Xero. Xero. https://apps.xero.com/uk/function/carbon-accounting (accessed 05.02.25).

# Chapter 3. Completing the SME Carbon Profile: Scalable Prediction of Scope 3 Emissions

## 3.1 Abstract

Despite growing recognition of significance, a business's Scope 3 emissions remain rarely measured and as a result are poorly understood. This situation is particularly common amongst small and medium-sized enterprises (SMEs), which face additional obstacles to emission measurement. With this paper, we present a transaction-based approach to facilitate SME Scope 3 engagement. Using financial transaction data for 150,000+ UK SMEs, we produce spend-based Scope 3 estimates across key Greenhouse Gas Protocol categories. We then fit a series of hierarchical regression models to both quantify and identify firm-level Scope 3 emissions, with minimal user inputs. We find that this approach is effective in predicting the upstream emissions of both purchased goods & services (RSQ = 0.87) and fuel and energy-related activities (RSQ = 0.89 and 0.72), whilst weaker for more targeted categories such as business travel. We also find a small number of recurrent industry hotspots tend to account for 75% of a firm's upstream emissions. By leveraging objective, standardised data to estimate emissions, this method provides a low-input alternative to costly micro-studies for generating Scope 3 insights, extending the visibility of emissions beyond a firm's direct operations, revealing emission hotspots and supporting the development of value chain decarbonisation strategies.

## 3.2 Introduction

Business emission reporting has traditionally focused on the emissions produced directly by operations (Scope 1) and energy-use (Scope 2). Whilst these emissions are both relatively easy to quantify and directly under a business's organisational control, this neglects the substantial share of emissions that occur across the broader value chain (Scope 3) (Serafeim and Velez Caicedo, 2022). Typically these indirect emissions far exceed a firms combined Scope 1 and 2 totals (Huang et al., 2009). Overlooking Scope 3 emissions both limits business decarbonization efforts (Hettler and Graf-Vlachy, 2024) and risks inefficient policy interventions (Hertwich and Wood, 2018).

Recognition of the importance of Scope 3 emissions is gaining momentum. Policy developments to integrate Scope 3 into disclosure frameworks are underway in Europe (DESNZ, 2023a; EU, 2022), while Australia is phasing in the introduction of mandatory Scope 3 emissions reporting for larger businesses. However, extending these frameworks to small and medium-sized enterprises (SMEs) requires careful consideration.

Scope 3 emissions are complex relative to Scope 1 and 2 emissions (Cheema-Fox et al., 2021). Their calculation necessitates subjective decisions on system boundaries, whilst requiring extensive external data on supplier locations, production techniques, and knowledge on various other factors within global supply chains (Acquaye et al., 2014). Under the Greenhouse Gas (GHG) Protocol, Scope 3 emissions are divided into 15 distinct and mutually exclusive categories, capturing embodied upstream emissions from purchased goods and services, as well as downstream emissions from sold goods and services, including distribution, use, and end-of-life treatment (WRI and WBCSD, 2020).

For larger firms with complex supply networks, significant regulatory exposure, and notable individual environmental impact, firm-specific Life Cycle Assessments (LCAs) are the preferred approach. LCAs involve bottom-up techniques to account for the specific processes, suppliers, and regions involved in each stage of the life cycle. The resulting data are primary, and therefore considered more reliable (Busch et al., 2022). Whilst guided by international standards, and adaptable to the GHG Protocol Scope 3 categories, LCAs are inherently resource-intensive and methodologically complex, even for dedicated teams within large corporations (Goldhammer et al., 2017). Calculating these emissions requires technical knowledge, and involves significant operational and transactional costs (Isil and Sebastianelli, 2020). Moreover, differences in methodological approaches undermine data comparability (Robinson et al., 2015), while frequent errors further compromise data quality (Patchell, 2018).

To mitigate the substantial operational and transactional costs of primary data collection, firms often turn to secondary data sources. These can include pre-existing LCAs, or industry-average emission conversion factors derived from environmentally extended multi-regional input-output (EEMRIO) models (Nguyen et al., 2023). Whilst simpler, these approaches still require a level of technical expertise, and for some, the reliance on secondary factors introduces inaccuracies (Busch et al., 2022).

Despite this, different levels of accuracy and completeness are necessary to address different needs (Huang et al., 2009), with process-based data alone insufficient for producing robust and comparable estimates at scale (Minx et al., 2008). For SMEs, sole reliance on expensive micro-analysis presents a substantial barrier to the widespread adoption of Scope 3 emission measurement.

In our earlier work (Phillpotts et al., 2025), we build the argument for applying simple statistical models to proxy Scope 1 and 2 emissions amongst SMEs. Our approach uses large-scale financial transaction data (FTD) to fit models that are simple to apply, requiring two inputs: turnover and industry. This efficient, and highly scalable approach addresses the data gap created by the exclusion of SMEs from current reporting requirements. FTD-based emission estimates capture a simplified version of emissions, and its validation against actual emissions is challenging due to the scarcity of alternative data. However, we argue that transparent, standardised, and objective measurement represents a justified trade-off, at the expense of some absolute accuracy.

This paper builds on these foundations by extending the approach to SME Scope 3 emissions. Existing approaches to model unreported Scope 3 emissions have used regression and machine learning techniques (Carbon Disclosure Project, 2020; Nguyen et al., 2023; Serafeim and Velez Caicedo, 2022; Shakdwipee and Lee, 2016). These models rely on self-reported data for their training, presenting inherent limitations; inconsistent methodologies across reporting firms undermine comparability, while the training data underrepresents SMEs. With this paper, we utilise the FTD of over 150,000 UK SMEs to estimate Scope 3 emission categories, creating a large dataset for a previously unobserved segment of the business population.

Drawing on our novel, micro-level dataset, we make two contributions: first we explore how different transaction categories contribute to a firm's Scope 3 emissions, and second, we develop a series of statistical models to quantify and identify Scope 3 emissions. Through this work, we demonstrate how FTD can support the mainstreaming of Scope 3 emission measurement and introduce a simple benchmarking tool to generate Scope 3 emissions insights. Thereby transforming extensive, granular, otherwise private data into accessible macro-level insights to facilitate a wider adoption of Scope 3 measurement amongst smaller, often under-engaged actors (British Business Bank (BBB), 2024).

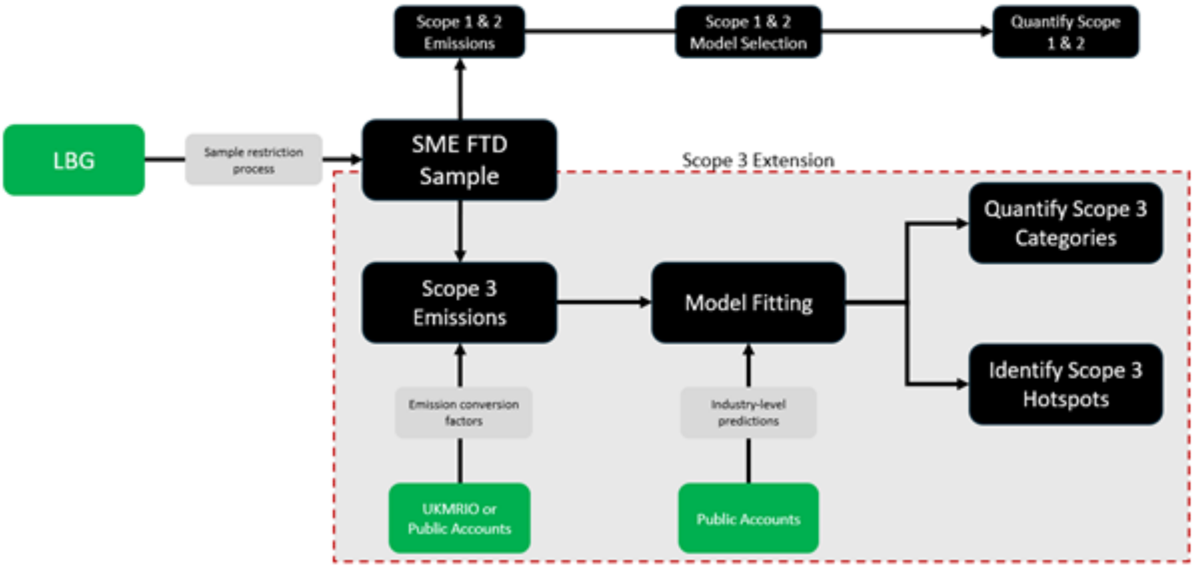
The remainder of this article is structured as follows. First, we describe the data and method used to calculate Scope 3 categories, identify hotspot thresholds, and outline the statistical techniques applied in developing our models. Next, we present model results, interpretations, and stability tests. Finally, we conclude with a discussion of the key findings, their practical implications, limitations, and recommendations for future research.

## 3.3 Materials and Methods

In partnership with Lloyds Banking Group (LBG), one of the UK's largest retail banks (LBG, 2023), this study accessed a substantial dataset of transactions made by UK SMEs in the calendar year 2021. Here, we define SMEs as firms with an annual turnover of less than £36 million (Companies Act, 2006).

Our methodological approach is outlined in the following subsections, with Figure 3.1 providing a schematic overview of the Scope 3 extension to the model. First, FTD is obtained from LBG, and an iterative sample restriction process is implemented. The FTD from this sample is then used to estimate Scope 3 emission categories. We then produce prediction models by fitting a series of hierarchical regression models to the FTD-based emission estimates, using firm-level predictors from LBG, and industry predictors from public accounts when required. In some instances, the model fitting process requires additional sample restrictions to ensure model convergence. For each category, we evaluate model performance and select a preferred specification, before conducting further validation tests.

Figure 3.1 - Overview of prediction model development. Black rectangles indicate key stages of the modelling process, green denotes data sources, and grey highlights the specific inputs and processes required at each step. Adapted from Phillpotts et al. (2025).




### 3.3.1 Sample Selection

A series of sample restrictions are required to ensure our analysis is based on comprehensive financial data histories, isolating SMEs who use LBG as their primary banking provider, and excluding those with significant financial activities through other institutions (Phillpotts et al., 2025). A full account of this process, including rationale and impact of each step is provided within Supporting Information S3.1.

We also restrict our sample to only SMEs for which FTD can reasonably serve as a proxy for both firm revenue and Scope 3 emission-generating activities. We therefore exclude SMEs in sectors including agriculture, energy, and water, where substantial industry-specific Scope 3 emissions require more tailored modelling approaches. Certain service-sector SMEs are also omitted, such as those in insurance, real estate, and legal or accounting services, due to frequent handling of non-turnover credit lines causing a systematic overestimation of turnover (Phillpotts et al., 2025). Table S2 in Supporting Information S3.1 provides a full list of explanations of exclusions where relevant, along with the sample and population figures for each industry.

With these restriction steps in place, our sample consists of 167,169 companies across 57 industries. Collectively, these companies account for an estimated £103 billion in annual revenue, with over 101 million financial transactions collectively in 2021.

## 3.3.2 FTD-based Emissions Estimates

### 3.3.2.1 Mapping to greenhouse Gas (GHG) protocol categories

The GHG Protocol (2011) defines 15 categories of Scope 3 emissions. These categories are intended to encompass all indirect emissions from business activities. Consequently, a firm's FTD does not directly translate to all Scope 3 categories, and assumptions may be required to enable compatibility. Table 3.1 introduces each Scope 3 category, noting its coverage by FTD and the need for any assumptions in its estimation.

*Table 3.1 GHG Protocol Scope 3 categories and their alignment with FTD.*

| **Scope 3 Category** | **Captured by FTD?** | | **Included in Model?** | |
|---|---|---|---|---|
| 1. Purchased goods & services | Yes | FTD reflects the goods and services a business purchases when paid via the business account. These transactions can be matched to emissions using appropriate emission factors. | Yes | |
| 2. Capital goods | Partially | Capital asset purchases are captured in FTD; however, without additional context, FTD alone cannot reliably distinguish capital goods from regular purchases, nor identify the specific type of capital asset involved. | No | |
| 3. Fuel- and energy-related | Yes | Fuel and energy purchases are visible in FTD and can be used to calculate Scope 1, 2 and associated Scope 3 emissions. | Yes | |
| 4. Upstream transport & distribution | Partially | Payments to logistics providers are visible when made directly. However, transport and delivery fees are often bundled within purchases and may not be separately identifiable. | Yes | We assume upstream transport and distribution costs are included within the purchases of goods. These emissions are therefore included within Category 1. |
| 5, Waste from operations | Partially | Payments to waste service providers are visible. However, FTD does not capture waste volume, type, or treatment method. Additional data or assumptions are beneficial for measurement. | No | |
| 6. Business travel | Partially | Travel-related purchases such as flights, hotels, and car hire are identifiable in company FTD. However, expenses that are paid by employees and later invoiced to the company are not captured and thus remain unmeasured. | Yes | We assume all business travel spend in our sample is made through the business account. |
| 7. Employee commuting | No | Commute-related spending is made by employees and does not typically appear in business transactions. | No | |

| | | | | |
|---|---|---|---|---|
| 8. Upstream leased assets | Partially | Lease payments are visible, but emissions depend on asset type and use, which cannot be determined by FTD alone. | No | |
| 9. Downstream transportation and distribution | Partially | Courier and distribution payments are visible when made directly. However, as with upstream transport, bundled costs and lack of directionality (up vs. downstream) can make attribution difficult. | Yes | We assume all payments to courier, distribution and logistics firms are related to the downstream transportation of goods and services. |
| 10. Processing of sold products | No | No direct purchase associated | No | |
| 11. Use of sold products | No | No direct purchase associated | No | |
| 12. End-of-life treatment of sold products | No | No direct purchase associated | No | |
| 13. Downstream leased assets | No | Lease income may appear in FTD, but usage-based emissions are not measurable with FTD alone. | No | |
| 14. Franchises | No | Payments such as franchise fees or royalties may be visible, but operational emissions would be challenging to estimate. Moreover, SMEs are unlikely to have associated franchises, due to their size. | No | |
| 15. Investments | No | Investment income or flows may appear, but emissions from investments require detailed portfolio data. In addition, SMEs, are unlikely to be holding large sums of financial investments. | No | |

### 3.3.2 Scope 3 Emission Conversion Factors

The primary source for Scope 3 emission conversion factors is the UKMRIO (Owen and Kilian, 2024). The UKMRIO is a Single-country National Accounts Consistent (SNAC) EEMRIO model, used to produces the official annual consumption-based emissions for the UK, making it ideal for a UK focused methodology (Tukker et al., 2018). EEMRIO conversion factors reflect the embodied environmental load of a unit of final demand (e.g. kg $CO_2e$ / £) (Wiedmann et al., 2011), a full derivation of which, can be found in Kitzes (2013, pp. 497-498). Whilst these conversion factors are generally attributed to a unit of final demand, the Leontief inverse assumes sectoral goods are homogenous and thus have the same cradle-to-gate emissions per monetary unit. As such, intermediate inputs have the same unit footprint as products sold to the final consumer (Hertwich and Wood, 2018). The underlying assumption of EEMRIO conversion factor use is that the supply structure of each purchase can be approximated by the corresponding economic sector, as represented in the UK Supply and Use Tables (ONS, 2022; Schmidt et al., 2022).

The UKMRIO produces consumption-based conversion factors for 112 industries, denoted by their Standard Industrial Classification of economic activities (SIC code). To make these conversion factors compatible with FTD, SIC industries are mapped to all but two headings of spend: “Energy & Utilities” and “Vehicle Fuelling”. The mapping table for this process can be found in Supporting Information S3.2.

Spend categorised under “Energy & Utilities” and “Vehicle Fuelling” can instead be combined with direct conversion factors to estimate Scope 1 and Scope 2 emissions (Phillpotts et al., 2025) or as in this case, well-to-tank (WTT) conversion factors. The resulting estimates reflect the emissions produced during the extraction, processing, and distribution of fuels and energy. Table 3.2 displays the resulting interquartile range and median emissions per estimated Scope 3 category, along with observation numbers, whilst further details of the conversion factor calculations and their mapping to transactions are provided in Supporting Information S3.3.

*Table 3.2 Number of observations, interquartile and median emissions by aggregated industry and Scope 3 category. Here, SIC Section A – Agriculture, Forestry and Fishing, excludes agriculture SMEs.*

| SIC Section | Sample | Category 1. Purchased Goods and Services | | | Category 3.1 Fuel-related activities | | | Category 3.2 Energy-related activities | | | Category 6. Business Travel | | | Category 9. Downstream transportation and distribution | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Q1** | **Median** | **Q2** | **Q1** | **Median** | **Q2** | **Q1** | **Median** | **Q2** | **Q1** | **Median** | **Q2** | **Q1** | **Median** | **Q2** |
| | *#* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* |
| A - Agriculture | 696 | 2,139 | 57,261 | 160,016 | 1,872 | 3,650 | 10,601 | 207 | 429 | 973 | 577 | 1,286 | 2,689 | 45 | 308 | 1,821 |
| C- Mining | 78 | 61,734 | 142,245 | 460,682 | 2,637 | 6,951 | 23,407 | 232 | 647 | 2,391 | 771 | 1,511 | 3,557 | 423 | 3,815 | 18,208 |
| F - Manufacturing | 15,497 | 20,591 | 60,863 | 197,533 | 1,296 | 2,733 | 8,350 | 65 | 203 | 597 | 416 | 974 | 2,115 | 89 | 827 | 5,701 |
| G - Construction | 44,901 | 14,158 | 31,430 | 88,309 | 2,153 | 4,358 | 9,775 | 199 | 427 | 1,012 | 500 | 1,031 | 1,990 | 14 | 83 | 805 |
| H - Wholesale & Retail | 34,107 | 21,483 | 63,890 | 196,064 | 1,736 | 5,187 | 16,430 | 293 | 690 | 1,810 | 228 | 602 | 1,419 | 70 | 489 | 3,831 |
| I - Transportation | 2,244 | 2,883 | 8,100 | 30,409 | 3,261 | 5,991 | 10,280 | 693 | 1,376 | 2,417 | 495 | 1,260 | 2,942 | 6,244 | 22,753 | 83,916 |
| J - Hospitality | 16,897 | 17,642 | 43,888 | 92,906 | 403 | 810 | 1,748 | 163 | 405 | 991 | 202 | 683 | 3,320 | 10 | 54 | 369 |
| M - Prof Activities | 3,317 | 9,094 | 20,636 | 67,617 | 909 | 1,733 | 3,790 | 164 | 381 | 891 | 274 | 656 | 1,545 | 17 | 79 | 593 |
| N - Admin Services | 7,127 | 8,682 | 19,281 | 60,308 | 1,104 | 2,253 | 4,847 | 98 | 229 | 502 | 340 | 787 | 1,644 | 15 | 54 | 377 |
| P - Public | 9,171 | 4,792 | 11,540 | 34,089 | 2,510 | 18,210 | 78,787 | 567 | 2,712 | 9,364 | 350 | 794 | 1,620 | 12 | 56 | 461 |
| Q - Education | 5,419 | 6,604 | 17,150 | 53,743 | 2,024 | 4,120 | 9,067 | 324 | 716 | 1,706 | 774 | 1,368 | 2,666 | 14 | 81 | 905 |
| R - Health | 11,953 | 8,996 | 38,942 | 136,825 | 1,176 | 2,529 | 7,312 | 222 | 501 | 1,580 | 191 | 501 | 1,212 | 16 | 52 | 194 |
| S - Arts | 6,388 | 11,397 | 41,184 | 131,622 | 700 | 1,086 | 1,721 | 181 | 280 | 448 | 142 | 369 | 871 | 45 | 56 | 249 |

### 3.3.2.3 Defining Scope 3 emission hotspots

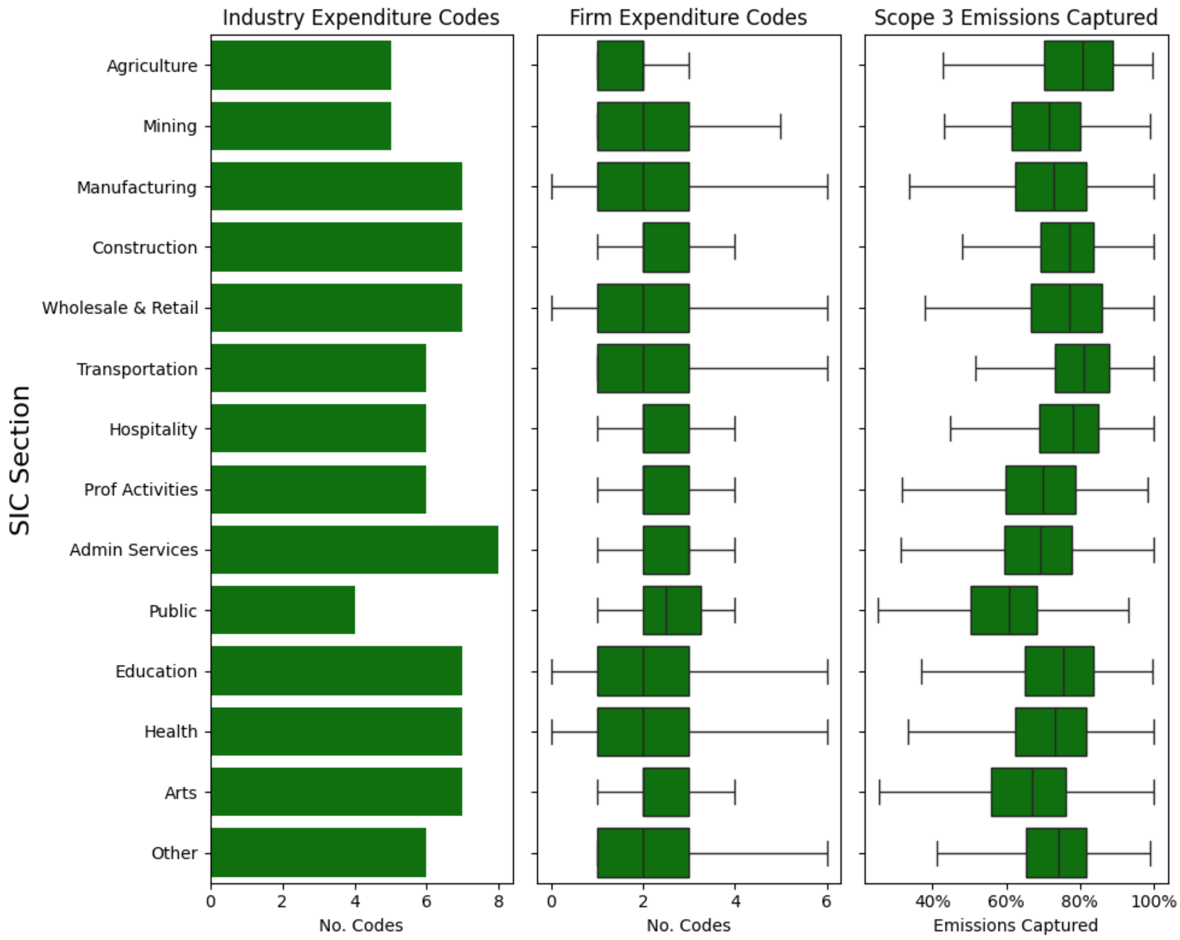
While some Scope 3 categories are narrow, Category 1. Purchased Goods and Services span diverse, sector-specific emission sources. To compare patterns across firms and sectors, we express each emission source as a proportion of total calculated Scope 3 emissions and use these proportions to identify recurrent hotspots. Table 3.3 displays different minimum-contribution thresholds, showing the number of expenditure codes used at the industry level, the number used per firm, and the share of emissions captured. For each metric we report the mean, median, and standard deviation.

Using a 10% contribution threshold, we capture, on average, 74.5% of emissions while reducing the median number categories from 24 to 2 per firm. We define hotspots as categories contributing above this 10% threshold and visualise industry mean values in Figure 3.2.

*Table 3.3 The impact of minimum contribution thresholds on Scope 3 emission sources.*

| **Min Scope 3 Contribution** | **Industry Expenditure Codes** | | | **Firm Expenditure Codes** | | | **Emission Captured** | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***Mean*** | ***Median*** | ***Std*** | ***Mean*** | ***Median*** | ***Std*** | ***Mean*** | ***Median*** | ***Std*** |
| None | 72.3 | 74 | 15.6 | 26.8 | 24 | 13.0 | 100% | 100% | 0 |
| 1% | 21 | 21 | 3.2 | 8.8 | 8 | 3.5 | 95.7% | 96.1% | 0 |
| 5% | 8.8 | 9 | 1.1 | 3.6 | 3 | 1.5 | 83.7% | 84.9% | 0.1 |
| **10%** | **5.6** | **6** | **0.8** | **2.3** | **2** | **0.9** | **74.5%** | **76.2%** | **0.1** |
| 20% | 3.9 | 4 | 0.5 | 1.4 | 1 | 0.6 | 61.7% | 64.6% | 0.2 |

*Figure 3.2 - Industry breakdown of hotspot spend categories and the level of emissions captured by them. Here, SIC Section A – Agriculture, Forestry and Fishing, excludes agriculture SMEs.*




## 3.3.3 Hierarchical Regression Models

To develop models for predicting Scope 3 emissions, we fit a series of hierarchical regression models to the FTD-based emissions estimates. To maintain the statistical reasoning outlined in (Phillpotts et al., 2025), turnover representativeness and adequate group sizes are ensured (Ali et al., 2019).

### 3.3.3.1 Quantifying Scope 3 emission categories

To quantify estimates for Scope 3 emission categories, we utilise a hierarchical linear regression specification, allowing for industry-specific effects to vary (Gelman and Hill, 2006). Following Goldhammer et al. (2017), we specify a log-log relationship, allowing for coefficients to be interpreted as elasticities.

3.3.3.1.1 Variable selection

To predict Categories 1, 6 and 9, we rely on a selection of firm-level variables, outlined in Table 3.4. Additional variables to capture Scope 3 emissions would need to reflect more granular operational details, such as supplier characteristics or transportation metrics on distance or freight mode (Wang and Ye, 2025). Standardised versions of these inputs are rarely available for SMEs at scale, and the inclusion of weak proxies risk introducing measurement error and biasing coefficient estimates (Hausman, 2001).

For Category 1, we retain the full sample, whilst Categories 6 and 9 require further targeted sample refinement to support model convergence. Here, we exclude SMEs with zero expenditure in the relevant category (43% of the sample for downstream transportation, and 17% for business travel) and remove extreme values by trimming the top and bottom 10% of category spend, relative to total expenditure. The resulting samples comprise 111,098 SMEs for Category 6 and 75,622 for Category 9, with industry coverage maintained at 57 industries.

*Table 3.4 Predictor variables for the hierarchical regression model. Adapted from Phillpotts et al. (2025).*

| **Variable** | **Unit** | **Description** | **Calculation** |
|---|---|---|---|
| Turnover | ***£*** | Turnover serves as an indicator of a firm's operational scale and is included in the analysis because, all else being equal, higher turnover typically correlates with greater emissions. | Credit turnover is derived by summing the credit transactions recorded within a firm's bank account in the year (excluding identified non-turnover credit lines). |
| 2 Digit SIC Code | ***SIC*** | The Standard Industrial Classification (SIC) identifies a firm's primary business activities. Including this industry identifier allows for more accurate comparisons of expected emission intensity among firms performing similar activities. | SIC classifications are the standard industry categories in the UK with official economic and environmental data commonly published on a 2 digit SIC code basis, whilst it is standard practice for firms banking with LBG to be assigned SIC codes at both the 2- and 3-digit level by relationship managers. For a fuller understanding of these classifications see (Companies House, 2018). |
| 3 Digit SIC Code | ***SIC*** | | |
| Capital Spend – Assets | ***%*** | Capital expenditure reflects a firm's investment in assets. This variable is included because high capital spending on newer energy-efficient technologies may serve to lower a firm's emission intensity. For ease of calculation, we define capital expenditure as purchases that require financing. | Asset capital spend is calculated by summing the percentage of total spend, spent on asset financing and plant hire for the three-year period ending 2021. We take an average of these percentages to gain a single figure, capturing asset spending across the three-year window. |
| Capital Spend – Vehicles | ***%*** | | Vehicle capital spend is calculated by summing the percentage of total spend, spent vehicle financing for the three-year period ending 2021. As above, we calculate a single figure, capturing vehicle spending across the three-year window. |

As Category 3. Fuel- and energy-related activities emissions should directly align with a firm's Scope 1 and 2 emissions, the model specifications developed and validated in Phillpotts et al. (2025) for Scope 1 and 2 emissions are applied, adapting each to estimate WTT emissions of fuels and electricity respectively. Here, models rely on turnover and 2-digit SIC code, as well as variables on industry-level Scope 1 and 2 emission intensity, as well as the industry distribution of turnover.

3.3.3.1.2 Model fitting and evaluation

Models for Categories 1, 6 and 9 are built sequentially, starting with two and increasing to five variables, evaluating the required complexity to explain the dependent variables (Cohen et al., 2013). Under a transaction-based methodology, higher Scope 3 emissions are associated with both greater absolute levels of expenditure and a greater allocation of spend to high intensity categories. We first reflect these dynamics with just turnover and industry classification in Model 1, allowing the effect of turnover to vary by industry. We then extend the model by nesting additional three-digit SIC codes within the industry groupings (Model 2) and subsequently incorporate capital variables to capture further firm-level heterogeneity (Model 3). These variables are specified as fixed effects to mitigate convergence issues. The hierarchical regression structure of these models are as follows:

Model 1: $ln\, y_i = \left(\beta_t + \beta_{t,j(i)}\right) ln\, turnover_i + \varepsilon_i$

Model 2: $ln\, y_i = \left(\beta_t + \beta_{t,j(i)} + \beta_{t,k(i)}\right) ln\, turnover_i + \varepsilon_i$

Model 3: $ln\, y_i = \beta_t\, ln\, turnover_i + \beta_2\, ln\, assets_i + \beta_3\, ln\, vehicles_i +$
$\beta_{t,j(i)}\, ln\, turnover_i + \beta_{t,k(i)}\, ln\, turnover_i + \varepsilon_i$

Where:

$y_i$ represents the absolute emissions each category of firm i's Scope 3 emissions

$ln$ represents the natural logarithm of the variable

$\beta_t$ represents the fixed coefficients for firm turnover

$\beta_{t,j(i)}$represents the variable coefficient for industry j of firm i

$\beta_{t,k(i)}$represents the variable coefficient for the additional industry grouping (k) of firm i

$\beta_2$ represents the fixed coefficients for the firms' assets variable

$\beta_3$ represents the fixed coefficients for the firms 'vehicles variable

$\varepsilon_i$ represents the residual error term

Model specifications for estimating Category 3 are taken directly from Phillpotts et al. (2025), this emissions category is therefore split between Category 3.1 (fuels) and 3.2 (electricity), whilst only results for the final model is presented.

Fuels (WTT emissions related to Scope 1):

$$ln\, y_i = \beta_1\, ln\, intensity_i + \beta_2\, ln\, skew_i + \beta_3\left(ln\, intensity_i \cdot ln\, skew_i\right) + \beta_{t,j(i)}\, ln\, turnover_i + \varepsilon_i$$

Electricity (WTT emissions related to Scope 2):

$$\ln y_i = \beta_1 \ln intensity_i + \beta_2 \ln skew_i + \beta_3(\ln intensity_i \cdot \ln skew_i) + \beta_{t,j(i)} \ln turnover_i + \varepsilon_i$$

Where:

$y_i$ represents the absolute emissions of firm i's fuel or energy Scope 3 emissions

$ln$ represents the natural logarithm of the variable

$\beta_{1,2,3}$ represents the fixed coefficients for industry variables

$\beta_{t,j(i)}$ represents the variable coefficient of turnover for industry j of firm i

$\varepsilon_i$ represents the residual error term

For each model, we present coefficients and statistical significance, indicating the strength and direction of predictor relationships to the dependent variables, along with their standard errors, reflecting the estimate uncertainties. We evaluate both model fit (RSQ) and model fit with respect to complexity (AIC). We then select a preferred model and test out-of-sample performance using an 80/20 stratified train-test split with 5-fold cross-validation (James et al., 2021).

3.3.3.2 Identifying Scope 3 emission hotspots

Next, we fit a hierarchical binomial logistic regression framework on the proportional emissions data of Table and Figure 3.2. This approach enables the estimation of an outcome probability (Gelman and Hill, 2006), thereby allowing us to identify and present hotspot categories based on firm-level predictors. Here, our dependent variable is defined as the log odds of an expenditure category contributing over 10% to calculated emissions. We run an iteration of this regression for each of the 624 expenditure categories.

The analysis was computationally demanding in R as separate hierarchical binomial logistic regressions were required to estimate each transaction code, with different coefficients computed across the 57 industry groupings, resulting in 624 outcome iterations. The base sample included approximately 170,000 firms, each associated with up to six hotspot outcomes, creating substantial processing demands in terms of runtime. Even the simpler specifications exhibited 72 hour plus runtimes, while more complex models did not complete after more than 168 hours, with convergence becoming more problematic as complexity increased. To keep the analysis tractable, two model formulations were implemented for each spend category, providing a balance between computational feasibility and operational results. First, Model 4 predicts the log odds of hotspot occurrence given firm industry. Second, Model 5 predicts the log odds of hotspot occurrence given both firm industry and turnover, allowing for changes in hotspot likelihood with firm size.

As we run 624 iterations of each model, we report the average performance of each model specification. In addition to the AIC, we report the median residual as a measure of central tendency, where values close to zero indicate accurate predictions. We also present the model's log-likelihood, reflecting explanatory power, with higher values indicating a better fit. Finally, we convert the predicted log odds into probabilities, to enable visualisation and interpretation. Full model specifications, results and probability calculations can be found in Supporting Information S3.4.

## 3.4 Results

### 3.4.1 Hierarchical Linear Regression Results

Summaries for the three models used to predict Categories 1, 6 and 9 are presented in Table 3.5. In Model 1, turnover returns a strong, positive coefficient. The elasticity is highest for Category 1 (0.87), whilst lower for Category 6 (0.55) and Category 9 (0.47). Explanatory power varies substantially, with RSQ values ranging from 0.87 for Category 1 to 0.55 and 0.41 for Categories 6 and 9, reflecting differences in how well turnover and industry alone explain each emissions source.

The introduction of more granular industry detail in Model 2 expands groupings from 58 to 212 SIC codes. These groupings slightly improve model fit across all categories, as seen in lower AIC values and small gains in RSQ.

The inclusion of asset and vehicle variables produces mixed effects across categories. Turnover coefficients remain largely unchanged across each category's models, with additional disaggregation and predictors unable to weaken its strong relationship with emissions. For Category 1, both additional variables have small negative coefficients, slightly reducing emissions after controlling for turnover. For Category 6, the vehicles variable shows a small positive relationship, while assets have no significant effect. For Category 9, the effect of the two additional variables is strongest, with elasticities of -0.07 and -0.12, suggesting that firms with greater in-house capacity may rely less on outsourced distribution services.

*Table 3.5 Model summary. Asterisks indicate significance level where P values < 0.05 = *, < 0.01 = **, and < 0.001****

| | **Model 1** | | **Model 2** | | **Model 3** | |
|---|---|---|---|---|---|---|
| | **Coef** | **Std** | **Coef** | **Std** | **Coef** | **Std** |
| **Category 1. Purchased goods & services** | | | | | | |
| ln turnover | 0.87*** | 0.01 | 0.87*** | 0.01 | 0.86*** | 0.01 |
| ln assets | | | | | -0.02*** | 0.00 |
| ln vehicles | | | | | -0.01*** | 0.00 |
| RSQ (fixed) | 0.74 | | 0.75 | | 0.75 | |
| RSQ (total) | 0.87 | | 0.88 | | 0.88 | |
| AIC | 269,421 | | 248,957 | | 247,866 | |
| **Category 3.1. Fuel-related activities** | | | | | | |
| ln turnover | | | | | 0.77*** | 0.02 |
| ln scp1_intensity | | | | | 1.35*** | 0.21 |
| ln skew | | | | | -0.82*** | 0.02 |
| ln (intensity * skew) | | | | | -0.46*** | 0.05 |
| RSQ (fixed) | | | | | 0.41 | |
| RSQ (total) | | | | | 0.89 | |
| AIC | | | | | 80,304 | |
| **Category 3.2. Electricity-related activities** | | | | | | |
| ln turnover | | | | | 0.78*** | 0.01 |
| ln scp2_intensity | | | | | -1.07*** | 0.10 |
| ln skew | | | | | -0.54*** | 0.03 |
| ln (intensity * skew) | | | | | 0.28*** | 0.03 |
| RSQ (fixed) | | | | | 0.59 | |
| RSQ (total) | | | | | 0.72 | |
| AIC | | | | | 210,345 | |
| **Category 6. Business Travel** | | | | | | |
| ln turnover | 0.55*** | 0.01 | 0.54*** | 0.01 | 0.55*** | 0.01 |
| ln assets | | | | | -0.00 | 0.00 |
| ln vehicles | | | | | 0.04*** | 0.00 |
| RSQ (fixed) | 0.27 | | 0.28 | | 0.27 | |
| RSQ (total) | 0.55 | | 0.56 | | 0.55 | |
| AIC | 321,419 | | 318,082 | | 317,793 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Category 9. Downstream transportation and distribution** | | | | | | |
| ln turnover | 0.47*** | 0.02 | 0.47*** | 0.02 | 0.46*** | 0.02 |
| ln assets | | | | | -0.07*** | 0.00 |
| ln vehicles | | | | | -0.12*** | 0.01 |
| | | | | | | |
| RSQ (fixed) | 0.07 | | 0.07 | | 0.09 | |
| RSQ (total) | 0.41 | | 0.43 | | 0.43 | |
| AIC | 322,529 | | 321,169 | | 320,393 | |

Models for both Categories 3.1 and 3.2 return identical RSQ values to what is observed in Phillpotts et al. (2025), indicating that changes in emission factors did not introduce unexpected variation in model performance. Across both models, all coefficients retain consistent correlations, with only slight changes in magnitude, reflecting the absolute differences between Scope 1 and 2 emissions and their WTT counterparts. The foremost difference lies in the skew term where the previously insignificant coefficient for electricity now returns as negative and highly significant (-0.54***).

The summaries in Table 3.5 indicate that the inclusion of additional firm-level variables did not significantly improve model performance, despite increasing model complexity and data requirements. Model 3 delivers the best or near-best fit across all categories but gains in explanatory power over earlier models are modest, just one to two percentage points higher than the baseline Model 1.

For Category 1, a simple and parsimonious model provides strong predictive power while remaining practical to implement, these emissions are well explained by a firm's turnover, reflecting scale, and its industry, capturing spending patterns. Model performance is further supported by the standardised calculation of emissions and broad emission category. In contrast, for Categories 6 and 9, explanatory power is weaker, with the most complex model failing to significantly improve performance. This reflects the higher uncertainties introduced by the necessary assumptions underlying emissions estimation.

Alongside the models for Categories 3.1 and 3.2, Model 1 is selected as the preferred specification for the remaining categories, and five-fold cross-validation is conducted to evaluate model robustness. The results indicate consistent performance across folds, with minimal variation in both RSQ and median errors. Category 1 shows the strongest performance, with a median absolute error of 8.5 kg $CO_2e$ (28%), while Categories 3.1 and

3.2 produce smaller absolute errors (1.0 kg CO2e and 0.2 kg CO2e, respectively) reflecting the low magnitude of WTT emissions but higher relative percentage errors (both around 49%). Categories 6 and 9 exhibit weaker fits, with median errors of 0.4 kg CO2e (59%) and 0.18 kg CO2e (93%), respectively. Overall, the stable RSQ values across folds indicate that, regardless of absolute fit strength, the models capture relationships that are reproducible and not dependent on specific training samples. Test and train results for each iteration can be found in Supporting Information S3.4.

### 3.4.2 Hierarchical Binomial Logistic Results

Table S7 in Supporting Information S3.4 presents average summaries for the binomial regression iterations. Under Model 4, we observe an average median residual of -0.0018, while average log-likelihood is -2,562. The inclusion of turnover in Model 5 improves all performance metrics, reducing the median residual slightly to -0.0015 respectively, and increasing the average log-likelihood to -2,424. Comparing the average AIC between models, we observe a slight decrease from 5,127 to 4,856.

While metric improvements are observed, the initial near-zero median residuals suggest that the basic model does not systematically mis-predict hotspots. Turnover's inclusion in the logistic regression does not drastically enhance model fit but adds complexity to computation and deployment. Model 4 produces a single log-odds value for each industry-expenditure code combination. These log-odds are easily converted to probabilities, a full matrix of which is provided in Supporting Information S3.5. Figure 3.3 displays an aggregated version of this matrix, plotting the most likely hotspot heading occurring by SIC Section.

*Figure 3.3 - Hotspot matrix showing the likeliest hotspot spend category, by SIC section. Here, SIC Section A – Agriculture, Forestry and Fishing, excludes agriculture SMEs.*


Even at this aggregate level, we observe a complex web of hotspots. We observe frequent, industry-specific hotspots, where firm SIC section aligns closely with the spend category. For example, Section Q. Human Health and Social Work Activities exhibits a hotspot in Health-related spending. Supporting Information S3.5 shows that this is largely driven by SIC codes 86. Human Health Activities and 87. Residential Care Activities), with corresponding hotspots in categories such as 296 (Medical Facilities & Services) and 306 (Assisted Living and Long-Term Care). Similar patterns emerge across other industries, suggesting that a substantial share of firms' Scope 3 emissions arises from intra-industry transactions. This aligns with supply and use table patterns, where the largest values occur along the leading diagonal, reflecting industries' specialisation in their primary products (Australian Bureau of Statistics, 2021).

In addition, several expenditure codes consistently emerge as hotspots across multiple SIC sections. The General Business & Industry heading is particularly prominent, containing hotspots for 11 of the 14 modelled sections, while Food & Drink and Business Travel contain hotspots in five and four sections, respectively. Supporting Information S3.5 shows that General Business & Industry encompasses a diverse set of spend categories that frequently return high hotspot probabilities, especially among industrial SIC codes such as manufacturing and construction. Within this heading, Category 132 (Office Supplies) is the most common hotspot, appearing in 27 of the 57 SIC divisions modelled. Other recurring hotspots include Shopping (27 divisions) and Food & Drink (19 divisions).

## 3.5 Discussion

### 3.5.1 Key Findings

The FTD-based approach of this paper offers new insights into the drivers of firms' Scope 3 emissions and facilitates analysis of the extent to which these patterns differ across industries. We identify that across all industries, a small number of spend categories typically account for an average of 75% of a firms calculated Scope 3 emissions. These categories are largely industry-specific, with a median of six per industry, while individual firms within those industries typically exhibit more concentrated spending patterns, with a median of two per firm.

We then construct models to predict these emissions, a practice often used by emission data providers (Serafeim and Velez Caicedo, 2022). We find that with just industry and turnover inputs; a hierarchical log-log regression specification achieves an RSQ of 0.87 for Category 1. Purchased Goods and Services. In this model, we observe a strong and significant relationship between turnover and Category 1 emissions with an elasticity of 0.87. The strong explanatory power of this simple model reflects both the standardized process of emissions calculation and the relatively coarse view it provides of emissions, whilst remaining consistent with previous studies showing that firm size and industry are primary determinants of Scope 3 emissions (Buchenau et al., 2025).

For Category 6. Business Travel and Category 9. Downstream Transportation and Distribution, model performance is weaker, with the simplest specifications returning RSQ values of 0.55 and 0.41, respectively. Whilst more complex iterations offering little improvement, these categories rely on stronger assumptions, which may contribute to the unexplained variance. For Business Travel, we assume all employee travel is paid directly via the business account, excluding cases where employees pay and later invoice the firm. For Downstream Transportation, many firms (43% of the sample) record zero spending in the relevant categories, further limiting model applicability. With these categories, emissions may reflect firm behaviours that extend beyond size and industry alone.

In the case of Category 3. Fuel- and energy-related activities, we find that the models developed in Phillpotts et al. (2025), can be redeployed as well-to-tank alternatives. The results are consistent with those reported in the paper, with the same model specifications yielding RSQ values of 0.87 and 0.72 respectively and a median average percentage error of 49%.

Our hotspot predictions complement the emission estimates by identifying spend categories that contribute disproportionately to calculated Scope 3 emissions. We observe both industry-specific hotspots, closely tied to sectoral activities, and shared categories that recur across multiple industries. These findings may indicate that many Scope 3 emissions originate within close supplier relationships, while also showing that certain categories of spending consistently drive emissions across sectors.

### 3.5.2 Practical Implications

Mainstreaming the measurement of Scope 3 emissions remains highly challenging due to the complexity of measurement and the absence of harmonised reporting requirements. As a result, current Scope 3 data is often incomplete, and existing measurements are regularly inconsistent (Nguyen et al., 2023). This paper examines the potential of FTD as a transparent and standardised basis for calculating Scope 3 emissions, minimising subjective decision-making and reducing common challenges such as truncation error (Ward et al., 2018). For SMEs, which face financial, resource, and influence obstacles to

Scope 3 measurement, the advantages of a low-cost, time-efficient solution are pronounced (Serafeim and Velez Caicedo, 2022).

Recent studies on household emissions (Trendl et al., 2023; Wells et al., 2025) highlight the application of FTD approaches, and our findings extend these insights to business emission accounting. For Scope 3 emissions the advantages of FTD are particularly salient given current regulatory gaps and limited awareness. Drawing on intrinsic business data that directly reflects operational behaviours, FTD offers a simplified pathway for SMEs to begin understanding and managing their supply chain impacts, which typically account for much of their environmental footprint.

To demonstrate the practical benefits of FTD, we use detailed, micro-level data to construct macro-level pathways that substitute for transaction records where these are unavailable. With only two user inputs, we generate robust estimates for Category 1. Purchased Goods & Services and Category 3. Fuel and Energy-related activities. These estimates can be refined by identifying recurring emission hotspots that arise from typical industry spending patterns.

To approximate total Scope 3 emissions, one accurately measured category can be scaled using benchmarking proportions to estimate the remaining categories (Buchenau et al., 2025). This approach constructs a complete picture of Scope 3 emissions, while avoiding reliance on the weaker predictions associated with Categories 6 and 9. An illustration of this approach is provided in Supporting Information S3.6.

When combined with Scope 1 and 2 estimates (Phillpotts et al., 2025), a simple model requiring only two user inputs produces a complete SME emissions profile. Guidance on accessing the model is provided in Supporting Information S3.6. Providing such estimates address the structural barriers faced by SMEs (Caldera et al., 2019; Menon and Ravi, 2021) and fosters broader engagement with emissions profile understanding among firms not yet subject to mandatory disclosure requirements (BBB, 2024). Understanding Scope 3 emissions enables businesses to identify their most material value-chain hotspots. Upstream reduction levers may include supplier engagement and procurement reform, with evidence indicating greater supply-chain engagement encourages more companies to disclose their emissions through voluntary frameworks (Science Based Targets initiative, 2018).

For policymakers, the findings highlight the value of leveraging FTD held by institutions to generate automated, standardised emissions insights. While financial institutions are increasingly aware of this potential (e.g., Bankers for Net Zero, 2022), policy guidance and incentives remain necessary to facilitate widespread, standardised deployment. Policymakers therefore have a role in creating the conditions that allow transaction-based approaches to complement existing reporting systems. Furthermore, our analysis also shows that upstream emissions are highly concentrated and overlapping across industries, revealing shared hotspots that can act as policy leverage points (Meadows, 1999), where modest but well-coordinated interventions could materially reduce overall emissions (Science Based Targets initiative, 2018).

### 3.5.3 Limitations

While these modelling solutions offer advantages in terms of accessibility and efficiency, it is important to acknowledge some key limitations. Our study includes 57 of the 88 SIC codes. To assess the impact of exclusions, Table S3.2 in Supporting Information S1 compares business population data (BEIS, 2022) with model coverage, showing that the model captures over 75% of the UK SME population. Among excluded SMEs, agriculture accounts for 5% and service sectors for 16%, both omitted due to inherent limitations in the FTD process.

Additionally, FTD lack the level of detail available in receipt-level data (Trendl et al., 2023). Emission conversion factors are therefore applied at the merchant category level only, making it difficult to maximise interpretation from hotspot estimates. For example, the general “Shopping” spend classification, would require examples of relevant merchants to assist transferability.

Related to this point, the validity of FTD-based Scope 3 emission estimates is difficult to assess, owing to accessibility and inconsistency limitations in alternative data. This challenge is compounded by the mapping of EEMRIO conversion factors to spend categories, which are derived from broad industry averages and applied to non–product-specific expenditures, exact matches are often unattainable. Instead, they serve only to differentiate between higher- and lower-embodied emission categories.

In practice, analysts of Scope 3 emissions should treat both reported and estimated values with caution and account for potential data error in their analyses (Nguyen et al., 2023). With an approach based on FTD, transparent and standardised methodologies come at the expense of absolute accuracy.

Here we demonstrate that some Scope 3 categories are either entirely unrepresented by FTD or require assumptions for compatibility (Table 1). This means that potentially significant emissions like Use of Sold Products are unmeasured and thus require their own measurement process, whilst those that require assumptions can be measured with an unknown degree of uncertainty. To address this, we propose the use of benchmarking approaches; however, these techniques rely on self-reported emissions data, which are often limited, incomplete, and inconsistent.

Finally, in addition to the limitations of a transaction approach, this methodology shares the limitations experienced by any IOA method. These include the time lags of EEMRIO databases, sectoral aggregations, price; temporal and spatial variation, and proportionality assumptions (Lenzen, 2000).

### 3.5.4 Future Research Directions

Among the limitations discussed, two key directions for future research emerge. The first refers to the limitations of EEMRIO emission factors. While these multipliers capture upstream emissions associated with each unit of expenditure, their publication is subject to significant time lags. This delay introduces inaccuracies, as economic structures and prices may change during the lag period. In the UK and across much of Europe, this issue is especially relevant due to recent experiences with prolonged periods of high inflation. A significant benefit to FTD is its real-time nature, and lagged multipliers depreciate this. To mitigate this challenge, techniques such as nowcasting and inflation adjustments have been proposed (OECD, 2017; Stadler et al., 2018). However, there remains a notable gap in the literature evaluating the effectiveness of these approaches in reducing inaccuracies associated with time lags.

A second avenue for future research is to identify the most suitable financial data for SME emission accounting. While this study focuses on FTD, accountancy-level data offers an alternative view of business operations. Its established role in financial reporting and subsequent completeness could reduce reliance on assumptions whilst improving coverage and accuracy. Similarly, the validity of FTD-based estimates should be tested against survey or physical data, to strengthen its use case in policy. Trendl et al. (2023) demonstrates the strong alignment of FTD with survey data across UK households; a comparable assessment for Scope 3 would be valuable, though currently constrained by data availability.

## 3.6 Conclusion

This paper demonstrates the potential of FTD as a scalable and transparent methodology for calculating key categories of Scope 3 emissions thereby addressing an important gap in current measurement practices. Using FTD-based estimates, we find emission sources are typically concentrated in a small set of hotspot spend categories, with both persistent industry-specific hotspots and common hotspots identified across multiple sectors. The models produced here robustly estimate Category 1 and Category 3 Scope 3 emissions, whilst Categories 6 and 9 are less well captured, reflecting their reliance on stronger assumptions.

FTD-based approaches substantially reduce reporting burdens and dependence on subjective measurement decisions, providing an accessible entry point for SMEs. Using the approach outlined here, SMEs can begin to explore decarbonisation strategies that extend beyond their direct operations, whilst the close relationships observed between firms Scope 3 emission sources highlight opportunities for coordinated, high-impact policy interventions.

## 3.7 Notes

### 3.7.1 Acknowledgements

We would first like to thank Lara Vomfell, Jasmine Wells, Alistair Gilfillan, and the rest of the behavioural science team at Lloyds Banking Group for sharing their invaluable expertise at key points of this research, as well as their continued interest and support. Secondly, we thank Hamza Suleman, whose contributions and technical skills were instrumental in developing the model described here into a working and accessible implementation. Finally, we extend our sincere gratitude to Ranil Boteju, Trystan Davies, Andrew McCulloch, Neil Oliver and Jaideep Mann for making this project possible and supporting its successful publication.

### 3.7.2 Funding Information:

Alec Phillpotts reports financial support was provided by Lloyds Banking Group. Jonathan Norman's time was funded by the UK Research and Innovation (UKRI) Council, as part of the Energy Demand Research Centre, grant reference number EP/Y010078/1.

## 3.8 References

Acquaye, A., Genovese, A., Barrett, J., Lenny Koh, S.C., 2014. Benchmarking carbon emissions performance in supply chains. Supply Chain Management: An International Journal 19, 306–321. https://doi.org/10.1108/SCM-11-2013-0419

Ali, A., Ali, S., Khan, S.A., Khan, D.M., Abbas, K., Khalil, A., Manzoor, S., Khalil, U., 2019. Sample size issues in multilevel logistic regression models. PLoS One 14, e0225427. https://doi.org/10.1371/journal.pone.0225427

Australian Bureau of Statistics, 2021. Australian System of National Accounts: Concepts, Sources and Methods. p. Chapter 22. https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/australian-system-national-accounts-concepts-sources-and-methods/2020-21/chapter-22-input-output-tables (accessed 25.6.25).

Bankers for Net Zero, 2022. Scoping Exercise: The role of banks in reducing GHG emissions of UK SMEs. https://www.bankersfornetzero.co.uk/wp-content/uploads/2024/05/Smart-Data-Foundry-B4NZ-Scoping-Exercise-Full-report-September-2022-Final-version.pdf (accessed 20.8.25).

British Business Bank, 2024. Small Business Finance Markets Report 2024. https://www.british-business-bank.co.uk/about/research-and-publications/small-business-finance-markets-report-2024 (accessed 25.6.25).

Buchenau, N., Oetzel, J., Hechelmann, R.-H., 2025. Category-specific benchmarking of Scope 3 emissions for corporate clusters. Renewable and Sustainable Energy Reviews 208, 115019. https://doi.org/10.1016/j.rser.2024.115019

Busch, T., Johnson, M., Pioch, T., 2022. Corporate carbon performance data: Quo vadis? Journal of Industrial Ecology 26, 350–363. https://doi.org/10.1111/jiec.13008

Caldera, H.T.S., Desha, C., Dawes, L., 2019. Evaluating the enablers and barriers for successful implementation of sustainable business practice in 'lean' SMEs. Journal of Cleaner Production 218, 575–590. https://doi.org/10.1016/j.jclepro.2019.01.239

CDP, 2020. CDP Full GHG Emissions Dataset: Technical Annex IV: Scope 3 Overview and Modelling. https://assets.ctfassets.net/v7uy4j80khf8/4wN17Ev2pHtHqmqkDNvcA4/0ec997081f246af9a1d6d1bf69afbcd0/Technical_Annex_III_Scope_3_Overview_and_Modelling_2024.pdf (accessed 12.7.24).

Cheema-Fox, A., LaPerla, B.R., Serafeim, G., Turkington, D., Wang, H. (Stacie), 2021. Decarbonizing Everything. Financial Analysts Journal 77, 93–108. https://doi.org/10.1080/0015198X.2021.1909943

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd ed. Routledge, New York. https://doi.org/10.4324/9780203774441

Companies Act, 2006. Legislation.gov. https://www.legislation.gov.uk/ukpga/2006/46/contents (accessed 12.12.24).

Department for Business, Energy & Industrial Strategy, 2022. Business population estimates for the UK and regions 2022: statistical release [Dataset]. GOV.UK. https://www.gov.uk/government/statistics/business-population-estimates-2022/business-population-estimates-for-the-uk-and-regions-2022-statistical-release-html (accessed 7.12.23).

Department for Energy Security and Net Zero, 2023. UK greenhouse gas emissions reporting: Scope 3 emissions [Dataset]. GOV.UK. https://www.gov.uk/government/calls-for-evidence/uk-greenhouse-gas-emissions-reporting-scope-3-emissions (accessed 1.7.25).

EU, 2022. Directive (EU) 2022/2464 of the European Parliament and of the Council, OJ L. https://eur-lex.europa.eu/eli/dir/2022/2464/oj/eng (accessed 11.8.25).

Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Higher Education from Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Goldhammer, B., Busse, C., Busch, T., 2017. Estimating Corporate Carbon Footprints with Externally Available Data. Journal of Industrial Ecology 21, 1165–1179. https://doi.org/10.1111/jiec.12522

Hausman, J., 2001. Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. Journal of Economic Perspectives 15, 57–67. https://doi.org/10.1257/jep.15.4.57

Hertwich, E.G., Wood, R., 2018. The growing importance of scope 3 greenhouse gas emissions from industry. Environ. Res. Lett. 13, 104013. https://doi.org/10.1088/1748-9326/aae19a

Hettler, M., Graf-Vlachy, L., 2024. Corporate scope 3 carbon emission reporting as an enabler of supply chain decarbonization: A systematic review and comprehensive research agenda. Business Strategy and the Environment 33, 263–282. https://doi.org/10.1002/bse.3486

Huang, Y., Lenzen, M., Weber, C., Murray, J., Matthews, H., 2009. The role of input-output analysis for the screening of corporate carbon footprints. Economic Systems Research 21, 217–242. https://doi.org/10.1080/09535310903541348

Isil, O., Sebastianelli, R., 2020. Arcs of carbon awareness in the value chain and their antecedents. Business Strategy and the Environment 29, 503–518. https://doi.org/10.1002/bse.2383

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning with Applications in R, Second. ed. Springer, New York, United States.

Kitzes, J., 2013. An Introduction to Environmentally-Extended Input-Output Analysis. Resources 2, 489–503. https://doi.org/doi:10.3390/resources2040489

Lenzen, M., 2000. Errors in Conventional and Input-Output-based Life-Cycle Inventories. Journal of Industrial Ecology 4, 127–148. https://doi.org/10.1162/10881980052541981

Lloyds Banking Group, 2023. 2023 Lloyds Banking Group Annual Review. https://www.lloydsbankinggroup.com/assets/pdfs/investors/annual-report/2023/2023-lbg-annual-review.pdf (accessed 25.6.25).

Meadows, D.H., 1999. Leverage Points: Places to Intervene in a System. The Sustainability Institute. https://donellameadows.org/archives/leverage-points-places-to-intervene-in-a-system/ (accessed 23.8.25).

Menon, R.R., Ravi, V., 2021. Analysis of barriers of sustainable supply chain management in electronics industry: An interpretive structural modelling approach. Cleaner and Responsible Consumption 3, 100026. https://doi.org/10.1016/j.clrc.2021.100026

Minx, J., Wiedmann, T., Barrett, J., Suh, S., 2008. Methods review to support the PAS for the calculation of the embodied greenhouse gas emissions of goods and services.

https://www.academia.edu/13468787/Methods_review_to_support_the_PAS_process_for_the_calculation_of_the_greenhouse_gas_emissions_embodied_in_goods_and_services (accessed 23.5. 25).

Nguyen, Q., Diaz-Rainey, I., Kitto, A., McNeil, B.I., Pittman, N.A., Zhang, R., 2023. Scope 3 emissions: Data quality and machine learning prediction accuracy. PLOS Climate 2, e0000208. https://doi.org/10.1371/journal.pclm.0000208

Organisation for Economic Co-operation and Development (OECD), 2017. Nowcasting techniques to improve timeliness. Presented at the Committee for the Coordination of Statistical Activities, Muscat.

Office for National Statistics (ONS), 2022. Supply and use tables. [Dataset] https://www.ons.gov.uk/economy/nationalaccounts/supplyandusetables (accessed 8.4.23).

Owen, A., Kilian, L., 2024. 2024 Data Release of Consumption- based Accounts for the UK: Summary of Methods. https://assets.publishing.service.gov.uk/media/68220ff3db6463b14cd81916/Summary_of_Methods_2025_final.pdf (accessed 1.9.25)

Patchell, J., 2018. Can the implications of the GHG Protocol's scope 3 standard be realized? Journal of Cleaner Production 185, 941–958. https://doi.org/10.1016/j.jclepro.2018.03.003

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2025. Bridging the SME reporting gap: A new model for predicting Scope 1 and 2 emissions. Journal of Industrial Ecology. https://doi.org/10.1111/jiec.70106

Robinson, O., Kemp, S., Williams, I., 2015. Carbon management at universities: a reality check. Journal of Cleaner Production, Bridges for a more sustainable future: Joining Environmental Management for Sustainable Universities (EMSU) and the European Roundtable for Sustainable Consumption and Production (ERSCP) conferences 106, 109–118. https://doi.org/10.1016/j.jclepro.2014.06.095

Schmidt, M., Nill, M., Scholz, J., 2022. Determining the Scope 3 Emissions of Companies. Chemical Engineering & Technology 45, 1218–1230. https://doi.org/10.1002/ceat.202200181

Science Based Targets initiative, 2018. Value Change in the Value Chain: Best Practices in Scope 3 Greenhouse Gas Management. https://files.sciencebasedtargets.org/production/files/SBT_Value_Chain_Report-1.pdf (accessed 7.10.25).

Serafeim, G., Velez Caicedo, G., 2022. Machine Learning Models for Prediction of Scope 3 Carbon Emissions. https://doi.org/10.2139/ssrn.4149874

Shakdwipee, M., Lee, L.-E., 2016. Filling the Blanks: Comparing Carbon Estimates Against Disclosures (MSCI ESG Research Issue Brief). MSCI. https://www.msci.com/documents/10199/139b2ab7-c95f-4f09-9d33-fdc491c5316e (accessed 7.10.24).

Stadler, K., Wood, R., Bulavskaya, T., Sodersten, C.-J., Simas, M., Schmidt, S., Usubiaga, A., Acosta-Fernandez, J., Kuenen, J., Bruckner, M., Giljum, S., Lutter, S., Merciai, S., Theurl, M.C., Plutzar, C., Kastner, T., Eisenmenger, N., Erb, K.-H., Koning, A. de, Tukker, A., 2018. EXIOBASE 3 Developing a Time Series of Detailed Environmentally Extended Multi-Regional Input-Output Tables. Journal of Industrial Ecology 22, 502–515. https://doi.org/10.1111/jiec.12715

Trendl, A., Owen, A., Vomfell, L., Kilian, L., Gathergood, J., Stewart, N., Leake, D., 2023. Estimating carbon footprints from large scale financial transaction data. Journal of Industrial Ecology 27, 56–70. https://doi.org/10.1111/jiec.13351

Tukker, A., Koning, A. de, Owen, A., Lutter, S., Bruckner, M., Giljum, S., Stadler, K., Wood, R., Hoekstra, R., 2018. Towards Robust, Authoritative Assessments of Environmental Impacts Embodied in Trade. Journal of Industrial Ecology 22, 585–598. https://doi.org/10.1111/jiec.12716

Ward, H., Wenz, L., Steckel, J.C., Minx, J.C., 2018. Truncation Error Estimates in Process Life Cycle Assessment Using Input-Output Analysis. Journal of Industrial Ecology 22, 1080–1091. https://doi.org/10.1111/jiec.12655

Wells, J., Trendl, A., Owen, A., Barrett, J., Jobst, N., Leake, D., 2025. Targeting carbon reduction in UK households: A new segmentation model using financial transaction data. Journal of Industrial Ecology. https://doi.org/10.1111/jiec.70020

Wiedmann, T., Wilting, H.C., Lenzen, M., Lutter, S., Palm, V., 2011. Quo Vadis MRIO? Methodological, data and institutional requirements for multi-region input–output analysis. Ecological Economics 70, 1937–1945. https://doi.org/10.1016/j.ecolecon.2011.06.014

World Business Council for Sustainable Development and World Resources Institute (WRI and WBCSD), 2020. Corporate Value Chain (Scope 3) Accounting and Reporting Standard. World Resources Institute and World Business Council for Sustainable Development. https://ghgprotocol.org/sites/default/files/standards/Corporate-Value-Chain-Accounting-Reporing-Standard_041613_2.pdf (accessed 06.06.2025)

# Chapter 4. Towards Timelier Embodied Emission Factors: Investigating methods to enable real-time EEMRIO applications

## 4.1 Abstract

The time lags associated with Environmentally Extended Multi-Regional Input–Output (EEMRIO) databases are a significant limitation to their practical utility. Policymakers, researchers, and businesses are increasingly requiring near real-time assessments to monitor progress toward climate targets yet are constrained by data that are typically published with a multi-year delay. To address this challenge, this paper develops three methods, each requiring different levels of assumptions and data requirement, for nowcasting EEMRIO databases to enable more timely analysis. We test these nowcasting methods with the UKMRIO for a set of historic data (2018–2022) and evaluate nowcasted performance against their known counterparts.

We find that average error levels between nowcasted embodied emission factors and their known counterparts are similar across methods, with mean absolute percentage errors ranging between 16.5% and 17.5%. The most effective balance between complexity and accuracy is achieved by an intermediate approach, which applies regional-level adjustments without requiring extensive industry-level modifications. Performance is highly dependent on the economic stability of the nowcasted period and remains sensitive to methodological choices in constructing yearly data points. These results demonstrate both the potential and limitations of MRIO nowcasting and point to the need for a single-country national accounts consistent approach that maximises the use of domestic data to improve both timeliness and relevance of application.

## 4.2 Introduction

Understanding the relationship between environmental degradation and economic activities has become increasingly important in the context of today's global sustainability challenge (Kasman and Duman, 2015). Achieving this requires robust methodologies for attributing emissions to their respective economic sources. At present, international mitigation targets are predominantly territorial-based, reflecting the jurisdictional limitations of national regulatory agencies (Afionis and Sakai, 2022; Tukker et al., 2020). However, reducing the emissions associated with national consumption is also important to reduce global emissions. Consumption-based approaches attribute the responsibility of greenhouse gas (GHG) emissions to the consumer, ensuring that those whose demand is driving production processes, bear some responsibility for the associated emissions (Davis and Caldeira, 2010; Shue, 1999). At the national level, this entails accounting for carbon leakages and the outsourcing of greenhouse gas-intensive production. For individual actors, such as households and businesses, it involves recognising responsibility for the emissions generated through supply chains to satisfy their consumption demands.

As a system capable of calculating consumption-based emissions, Environmentally Extended Multi-Regional Input-Output (EEMRIO) databases offer a promising route to delivering a globally consistent environmental accounting system, given their ability to efficiently track embodied emissions across global supply chains in a robust and consistent (Tukker et al., 2018; Wood et al., 2019). However, harnessing the power of input-output systems to capture indirect emissions is not without limitation. The multi-year publication delay of EEMRIO databases is one of the most frequently cited barriers to the uptake of input-output techniques (Lenzen et al., 2012). This lag creates a gap between the underlying data and the application of policy, research questions and other real-time use, weakening the argument for their adoption (Wiedmann et al., 2011). In the UK, the 30-month delay in publishing EEMRIO derived consumption-based emissions has been highlighted by the Climate Change Committee as a significant risk. Here, the Committee stressed that greater investment in the collection and reporting of consumption-based emissions is needed to ensure statistics are produced on a timely, annual basis. Without such improvements, the UK's consumption-based accounts will continue to rely on lagged data and may fail to capture genuine, real-time reductions in the nation's carbon footprint (CCC, 2022. p494).

Over the lagged period, economies depicted within an EEMRIO database may undergo substantial shifts in trade patterns, purchasing powers, and industrial emission intensity. The extent of such change varies throughout regional economic cycles, with temporal lags less consequential in periods of stability, but increasingly problematic amid economic volatility (Maluck and Donner, 2015). In the wake of the 2008 financial crisis, the need for the timely data publication has heightened across all domains (OECD, 2017) and statistical offices are taking steps towards the prompt production of national emission figures (Brown et al., 2021; OECD, 2017; Office for National Statistics, 2022a). Given recent major economic shocks such as Brexit, COVID-19, the Russia-Ukraine war, and subsequent high inflation across much of Europe, continuing to allow EEMRIO databases to be constrained by significant time lags presents a serious limitation to their practical utility.

In this paper, we explore methods to address the inherent time lag of EEMRIO databases, with the aim of quantifying the uncertainties arising from using nowcasted databases as well as providing practitioners with an easy approach to enable real-time data applications. Specifically, we combine straightforward Multi-Regional Input-output (MRIO) projection techniques (Beaufils and Wenz, 2022), with high frequency, exogenously sourced economic datasets, to produce three nowcasting procedures for a period of known years (2018 to 2022). This provides insights into the expected accuracy of results based on the information known at that time (Brown et al. 2021). With each approach, the underlying assumptions and data requirements increase: the first only adjusts for inflation, the second additionally applies regional changes in production and trade, and the most complex allows for industry-specific adjustments.

The rest of this paper is organised as follows; first we provide a background and context section, which introduces EEMRIO databases, their use in environmental accounting, associated time lags and current nowcasting approaches. Next, in the methods and materials section we introduce the EEMRIO used here, the projection techniques and the data requirements underpinning each nowcasting method. Subsequently, we present the results of our comparison. These results are presented by key EEMRIO component. The paper then concludes with a discussion of our key findings, their practical implications, limitations, and suggestions for future research directions.

## 4.3 Background and Context

### 4.3.1 Time lags associated with EEMRIO models

Input-Output Analysis (IOA), developed by Wassily Leontief in the 1930s, traces monetary flows between industries and to final demand consumers. Initially a national framework to model how economic shocks propagate through supply chains, IOA evolved into MRIO models to incorporate global economies and then into EEMRIOs (Miller and Blair, 2009). When extended to incorporate GHG emissions, the EEMRIO framework traces emissions embodied across global supply chains, linking production activities to final consumption, revealing the environmental impacts driven by demand (Kitzes, 2013). A particularly versatile product of the EEMRIO framework are the industry specific emission factors that can be derived from an EEMRIO database, reflecting the total environmental load of a unit of final demand (e.g. kg $CO_2e$ / £) (Wiedmann et al., 2011). For the full derivation of these multipliers, see Kitzes (2013, pp. 497-498). In the literature, these multipliers are widely employed to analyse upstream emissions in various contexts (Owen and Büchs, 2024; Schmidt et al., 2022; Trendl et al., 2023; Wells et al., 2025b)

The construction of EEMRIO databases is a challenging, laborious and complicated process (Tukker et al., 2018; Wood et al., 2019). They are intrinsically data intensive, particularly when covering a number of regions, where data can often be hard to both source and manage (Afionis and Sakai, 2022; Mangır and Ülkü Alver, 2022). Whilst the challenges associated with the construction of MRIOs do lead to delays in database publication, the principal cause of database time lag is inherited through the delay in publication of official Supply and Use Tables (SUTs). These pairs of tables record how supplies of different kinds of goods and services originate from domestic industries and imports, and how those supplies are allocated between various intermediate and final users. Thus, providing essential detail on the structure of economies through monetary values of transactions between pairs of sectors (Miller and Blair, 2009). It is due to this detail, that

SUTs are generally published several years after the period they refer to. The European Transmission programme sets targets for member state to publish SUTs within 36 months after the reference period (Eurostat, 2025), however in other parts of the globe, publication is less frequent still.

Whilst automating the MRIO construction process can shorten the time to publish an MRIO, this reduction is limited to only a few months (Wiedmann et al., 2011; Wilting et al., 2009). It is for this reason that it is beneficial to explore methodologies capable of utilizing additional, more frequent, datasets to generate timelier embodied emission multipliers.

### 4.3.2 Nowcasting EEMRIO Models

Many studies have focused on projecting the raw data underlying MRIOs, IOTs (Temursho et al., 2020), and SUTs (Temurshoev et al., 2011; Valderas-Jaramillo et al., 2019), whilst scenario-based modelling of MRIOs is also applied in practice (Beaufils and Wenz, 2022; Wiebe et al., 2018). Nowcasting is a specific form of MRIO projection, defined as 'the prediction of the present, the very near future and the very recent past', through the use of higher frequency data to obtain an early estimate before the target period is available (Banbura et al., 2013. p2). This is a process used in other contexts such as weather (Mostajabi et al., 2019), economics and financial markets (Bank of England, 2017; Huber et al., 2021; Xie, 2023). Compensating for missing data is also common practice in IOA, where projections are frequently needed during the compilation, balancing, and revision of SUTs (Department of Economic and Social Affairs, 2018).

Efforts to nowcast MRIOs exist (Wiebe et al., 2018), with key databases employing procedures to nowcast inter-country MRIO makeups (Miao and Fortanier, 2017). However, methodologies and assumptions vary, whilst levels of error and uncertainties are typically not disclosed. We consolidate and summarise existing nowcasting approaches and assumptions in Table 4.1

*Table 4.1 Summary of key MRIO databases existing approaches to nowcasting*

| Database | Nowcasting Documentation | Key Assumptions | Reference |
|---|---|---|---|
| Eora | Yes | Growth rates: GDP, import and export growth rates from the IMF WEO are used to scale national totals<br><br>Structure: Sectoral composition is assumed fixed; the entire industrial structure of each country is inherited from the last observed SUT (no sector-specific variation)<br><br>Balancing: Not explicitly detailed; nowcasted updates are applied proportionally to existing SUT structures without modelling bilateral trade partner shifts | (Casella et al., 2019) |
| FIGARO | Yes | Growth rates: GDP, output, final demand, and trade growth are used to scale the previous year's SUT<br><br>Structure: Technical coefficients and sector relationships are assumed stable<br><br>Balancing: RAS-style balancing ensures the updated tables remain internally consistent | (Eurostat, 2021) |
| EXIOBASE | Yes | Growth rates: Regional macroeconomic indicators (GDP, etc.) from the IMF WEO are used to scale MRIO tables forward<br><br>Structure: Technical coefficients incorporate historical structural change, with sector-specific short-term growth rates extrapolated conservatively<br><br>Balancing: MRIO balancing is maintained, assuming structural evolution follows historical trends | (Stadler et al., 2014; Stadler et al., 2015) |
| WIOD | No | | |
| GTAP | No | | |

Following the 2013 World Investment Report, the UNCTAD-Eora Global Value Chain (GVC) database emerged as a popular tool for analysts studying the globally fragmented nature of production processes. Based on the Eora MRIO dataset and national supply-use tables (SUTs), its usefulness was limited by the two- to three-year lag between the latest available data and the time of analysis. To address this, analysts sought to reduce the multi-year lag between the latest available data and the time of analysis. As of 2019, the UNCTAD-Eora database contained actual data from 1990 to 2015, with nowcasted estimates for the years 2016, 2017 and 2018 (Casella et al., 2019). The nowcasting procedure employed by the UNCTAD-Eora database incorporates data on annual change in the GDP, imports, and exports of each country from the International Monetary Fund's (IMF) World Economic Outlook Database (WEO). These statistics are used to direct figures within the database to reflect changes observed within the global economy. Given the absence of sectoral datasets with the granularity of SUTs, the composition of the economy in nowcasted years is fully inherited from the last year that the data is available (Casella et al., 2019). As a result, output increases are constant across regional industries, reflecting only changes in relative regional performance. The incorporation of global trends is limited by the lack of detail on trade partners, with the WEO database reporting only percentage changes in import and export volumes by country, without specifying source or destination regions.

FIGARO also adopts an approach that assumes that short-term economic developments can be captured by applying National Accounts growth rates to update SUTs and trade flows, while maintaining broadly stable technical coefficients. Price and volume effects are treated separately, and environmental extensions are scaled in line with economic activity when new environmental data are unavailable. Together, these assumptions are used to generate provisional estimates for years where full data are not yet published (Eurostat, 2021).

EXIOBASE is another MRIO database, offering comprehensive regional coverage, high sectoral detail, and the incorporation of multiple environmental satellite accounts (Stadler et al., 2018). To address the publication lag, EXIOBASE applies a nowcasting procedure that uses regional development indicators from the WEO to generate initial estimates of the most recent MRIO tables (Stadler et al., 2014). EXIOBASE 3.9 is constructed from supply-and-use tables available up to 2020, with later years produced through nowcasting. Unlike the UNCTAD-Eora approach, EXIOBASE explicitly accounts for the disproportionate and sector-specific growth patterns observed in real economies. This is achieved by calculating technical coefficients that reflect historically observed structural change within the EXIOBASE time series. These coefficients adopt a conservative approach by extrapolating short-term, industry-level growth rates forward in time (Stadler et al., 2015).

## 4.4 Materials and Methods

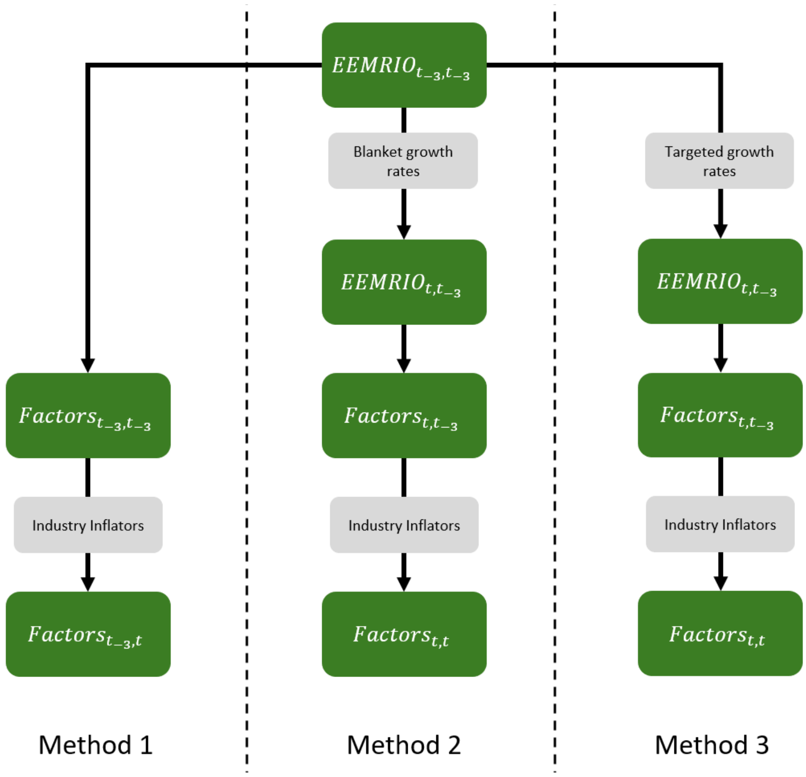
Put simply, the requirement to nowcast an EEMRIO database is a base table and exogenous data for the uncaptured period. The amount of exogenous data used depends on the assumptions and complexity of the nowcasting approach. In practice, three distinct adjustments must be made, each relating to a core component of the system: updating the economic structure captured in the MRIO matrices, projecting the environmental extension vector, and aligning monetary values to the correct price level.

The data used, and projection methods are described in the following subsections, whilst Figure 4.1 provides a schematic overview of the three nowcasting processes. Method 1 provides the simplest adjustment to lagged multipliers by applying inflators to embodied emissions factors, so they are expressed at real-time price levels. This practical step reduces emission overestimation from inflation and keeps estimates comparable over time. The method is targeted and simple to implement, but it does not alter the base EEMRIO. It therefore assumes no structural change during the lag period: production recipes and trade flows are unchanged, and real sectoral emission intensities remain at their base-year levels.

Method 2 then addresses changing production and global trade levels by applying regional real growth rates. Under assumptions used by the UNCTAD-Eora (Casella et al., 2019), output increases are constant across regional industries, reflecting only changes in relative regional performance. Once a balanced, nowcasted table is achieved, the inflation procedure is applied.

Finally, Method 3 accounts for disproportional industry growth by further refining the projection procedure for the core MRIO matrices. In the standard projection framework, sectoral allocations are derived from the base table proportions prior to balancing. This assumption can be relaxed to incorporate exogenous information on sectoral dynamics. This approach preserves the overall growth rate while allowing sectors to expand or contract at different rates. Once a balanced, nowcasted table is achieved, the inflation procedure is applied.

*Figure 4.1 Overview of nowcasting methods. EEMRIOs and factors are denoted with two-time subscripts: the first referring to the structure, the second referring to price level; here t = nowcasted year whilst t-3 = base year*




### 4.4.1 The UKMRIO

In this study, we use the UKMRIO (Owen and Kilian, 2025), an EEMRIO database developed at the University of Leeds with the primary purpose of producing and publishing the UK's consumption-based emissions accounts (Barrett et al., 2013). To maintain an official status, the database is constructed with an emphasis on preserving the original tables released by the Office for National Statistics (ONS), making it a Single-country National Accounts Consistent (SNAC) MRIO, a framework specifically designed for credible, robust country-specific analysis (Edens et al., 2015; Tukker et al., 2018)

In the UK, SUTs are released with a lag of 22 months following the reference period, typically in October each year (ONS, 2022). Following their release, the UKMRIO is constructed for the calculation of the UK's consumption-based emissions accounts. These figures are subsequently published in June, approximately 30 months after the reference year.

Here, we use the 2025 publication of the UKMRIO database (Owen and Kilian, 2025), which provides a time series of data for the years 1990–2022, to recreate data for the most recent five years available. These years capture: periods of relative stability (2018-2019), disruption (2020-2021), and subsequent high inflation (2022). We use key matrices and vectors as benchmarks to compare against nowcasted alternatives. To simulate the typical three-year publication lag, we use base tables from 2015 to 2019 to generate nowcasted target years.

### 4.4.2 Exogenous data

A selection of higher frequency, external datasets are used to nowcast the UKMRIO across the lagged period, ensuring that the resulting nowcasted table is grounded in insights from official datasets. Table 4.2 summarises the datasets used, their publication frequency, and the methods they are used in. Each dataset and specific data points used are then described in more detail in the subsections that follow. In addition, Supporting Information S4.1 contains each dataset used for nowcasting.

*Table 4.2 Data inputs to nowcasting methods, with sources and publication frequency.*

| Dataset | Ref | Publication Frequency | Used in |
|---|---|---|---|
| ONS Price Indexes | (ONS, 2025a) | Monthly/Quarterly | Methods 1,2,3 |
| World Economic Outlook Database | (IMF, 2022a) | Biannually | Methods 2, 3 |
| Contributions to monthly GDP | (ONS, 2025b) | Monthly | Method 3 |

#### 4.4.2.1 ONS Price Indexes

The ONS publishes a comprehensive range of price indices on a monthly, quarterly, and annual basis, typically with a short lag of one to two months. These indices allow data to be inflated or deflated to account for temporal price changes. In all methods, the embodied emission factors produced by the nowcasted UKMRIO are expressed in constant prices and therefore require price-level adjustments to the target year. To accurately capture fluctuations in the prices of intermediate goods and services, industry-specific inflation indices are applied across all 112 industries in the UKMRIO. Relevant data sources include the Agricultural Price Index, Fuel Price Indices, the Producer Price Index, and the Services Producer Price Index (ONS, 2025a). The classification of industries in these indices aligns with those in the UKMRIO because both follow the product and industry structures prescribed by the System of National Accounts and the UK national accounting framework.

4.4.2.2 World Economic Outlook Database

The World Economic Outlook (WEO) database contains a range of macroeconomic datasets which are presented in the statistical appendix of the WEO report. The WEO is published biannually, in April and October each year, and is noted for its high country resolution (Stadler et al., 2015).

The WEO publishes GDP statistics under the follow conditions; Constant Prices (National Currency), Constant Prices (Percent Change), Current Prices (USD). As real GDP is not published in a single currency, a procedure is outlined in the IMFs Country Data Documentation (IMF, 2022b) to calculate real growth rates which can be used in the UKMRIO.

Data on changes to imports and exports are also available within the WEO database, however only in percentage terms for each nation (each nation has a +/- annual percentage figure for imports and exports). This is limited by the lack of detail on trade partners, with the WEO database only reporting volume changes by country, without specifying source or destination. These figures refer exclusively to the aggregate change in the quantities of total imports meaning that the goods and services and their prices are held constant (IMF, 2022a).

With real GDP and trade growth rates, the base table is projected in constant prices, meaning that without the relevant inflation adjustment, nowcasted tables remain in the price of the base year.

4.4.2.3 ONS Monthly Contributions to GDP

To incorporate sectoral heterogeneity in economic growth, monthly ONS data on contributions to GDP growth are sourced (ONS, 2025b). This data, released with a two-month publication lag, report the proportionate contributions of broad industry groupings to GDP growth. The series is transformed to produce year-on-year contribution measures, and the high-level industry groups are disaggregated to align with the 112 industries represented in the UKMRIO. This procedure enables national GDP growth to be allocated more accurately across industries while preserving the aggregate growth rate implied by the WEO dataset.

The non-UK components of the UKMRIO is projected at the regional level, due to the absence of comparable high-frequency datasets for all the regions represented in the model. An alternative is to extrapolate trends observed from historical tables (Stadler et al., 2014), however this approach is not chosen for two reasons. First, the UKMRIO is not produced as a methodologically consistent time series (Owen and Kilian, 2025), which limits its suitability for year-to-year analysis. Second, undertaking an additional round of inflation and deflation to observe real industry growth rates would introduce avoidable uncertainty into the nowcasting process.

In practical terms, this approach strengthens the UK-focused analysis by allowing domestic embodied emissions to reflect recent variation in UK industry growth. The international component is projected at a more aggregated level, so emissions embodied in imports capture broad regional trends rather than detailed sectoral shifts. Overall, the method prioritises coherence and robustness in the nowcasted UKMRIO, while accepting reduced granularity in the non-UK elements.

### 4.4.3 Nowcasting the core UKMRIO matrices

To project the UKMRIO we apply a modified version of the Scenario-based Projection of the International Trade Network (SPIN) method (Beaufils and Wenz, 2022). The SPIN method builds on established input–output principles, utilising the Leontief inverse, RAS-type balancing procedures, whilst ensuring consistency with the fundamental relationships between regional output, GDP, international trade, and aggregate demand, making it both a practical and theoretically sound approach to nowcasting. In its simplest form, the SPIN method requires only a base, symmetric MRIO table and exogenously specified targets for regional GDP, imports, and exports levels. The method then produces a balanced EEMRIO database for the target year aligning with these exogenous values, while allocating domestic production, exports, and imports to intermediate and final demand, based on the sectoral shares derived from the base table.

With minor modifications, the SPIN method can be utilised for nowcasting. Instead of relying on exogenously prescribed GDP and trade values, base-year figures can be calculated, and projected to the target year using exogenously sourced real growth rates. This projection may take the form of a blanket adjustment, applying a single growth rate across all sectors within a region (Method 2), or, where data permit, a more granular approach in which individual sectors are allowed to grow at different rates while still maintaining consistency with aggregate growth (Method 3).

To enable compatibility with the SPIN method, the UKMRIO must be converted into a symmetric MRIO table (SMRIO), whilst the single Rest of the World final demand column must be disaggregated into regions. To transform the UKMRIO into a symmetric MRIO (SMRIO) table, the Eurostat manual of supply, use and input-output tables identifies four transformation methods. Here, Model D is used to convert the SUT-MRIO into a SMRIO. Model D assumes that each product has its own specific sales structure, independent of the industry in which it is produced (Eurostat, 2008. p296). By adopting this approach, the resulting SMRIO better reflects product-specific patterns of production, use, and trade, providing a more accurate basis for upstream environmental accounting.

To disaggregate the Rest of the World final demand column, supplementary data from FIGARO is used (Eurostat, 2021). FIGARO's final demand data is aggregated to match UKMRIO regions (creating a region-by-region matrix) and converted into proportions (excluding UK final demand). This provides the shares for final demand across non-UK destination regions. Full descriptions of these procedures are provided in Supporting Information S4.2, noting immaterial differences between the consumption-based emission estimates of the UK produced by the original model, against those produced by the symmetric version.

*Figure 4.2 Domestic IO table, extracted for each region from the UKMRIO.*

| | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_1$ | Row constraints |
|---|---|---|---|---|---|---|
| $r_1$ | $Z_{r_{1,1}}$ | $ex_{r_1} = Z_{r_{1,i}} + Y_{r_{1,i}}$ | | | $Y_{r_{1,1}}$ | $x_1$ |
| $r_2$ | $Z_{r_{2,1}}$ | 0 | 0 | 0 | $Y_{r_{2,1}}$ | $im_{r2}$ |
| $r_3$ | $Z_{r_{3,1}}$ | 0 | 0 | 0 | $Y_{r_{3,1}}$ | $im_{r3}$ |
| $r_4$ | $Z_{r_{4,1}}$ | 0 | 0 | 0 | $Y_{r_{4,1}}$ | $im_{r4}$ |

| | |
|---|---|
| Value Added | $v_{r_1}$ |

| | | | | |
|---|---|---|---|---|
| Column constraints | $x'_1$ | $ex_{r2}$ | $ex_{r3}$ | $ex_{r4}$ |

To run the SPIN method, we create domestic IO tables for each region using the base symmetric UKMRIO. Each table contains a domestic transaction table, with columns for intermediate imports and imports for final demand, rows for domestic industry exports (both to other industries and final demand), and domestic industry value added. A four-region example of a domestic IO table is presented in Figure 4.2.

Base-year GDP, import totals, and export totals are then derived from the domestic IO tables (Equations 4.1-4.3). Here, for each region, GDP equals value added, imports equal the sum of imports for intermediate and final demand, and exports equal the sum of exports for intermediate and final demand. Subscript ordering denote source region then destination region ($Z_{\boldsymbol{ir}}$ = intermediate imports from region i to region r).

$$gdp_r = v_r \tag{4.1}$$

$$im_r = Z_{\boldsymbol{ir}} + Y_{\boldsymbol{ir}} \tag{4.2}$$

$$ex_r = Z_{\boldsymbol{ri}} + Y_{\boldsymbol{ri}} \tag{4.3}$$

Base-year totals are then projected using WEO real growth rates, to produce the key inputs for the construction of the nowcasted IO tables. Projected exports and imports are easily assigned to relevant the row / column constraint, whilst values involving domestic production or consumption ($Z_{r_{1.1}}$ and $Y_{r_{1.1}}$ in Figure 4.2) require further steps to derive output ($x_i$). For these rows, we calculate projected aggregate demand for each region (the domestic final demand for commodities produced by domestic sectors plus the total exports to trade partners - $ad_r$), as expressed in terms of GDP, Imports and Exports.

$$ad_r = \varphi_r^* \cdot \left(gdp_r - \delta_r - \alpha_r^* \sum_i im_{ri}\right) + \sum_i \theta_{ri}^* ex_{ri} \tag{4.4}$$

Where; $\varphi_r^*$ refers to the sectoral shares of domestic final demand

$gdp_r - \delta_r$ refers to final demand for locally produced commodities (gdp less balance of payments),

$\alpha_r^*$ refers to the share of final commodities in imports,

$\sum_i im_{ri}$ refers to the sum of imports across regions i,

$\theta_{ri}^*$ refers to the region shares of each sectors exports,

$ex_{ri}$ refers to the export section of the domestic IO table,

Here, $\theta_{ri}^*$, $\varphi_r^*$ and $\alpha_r^*$ can either be derived directly from the base table, or in the case of the UK section of Method 3, adapted to reflect variation in industry growth. With the derived aggregate demand ($ad_r$), the Leontief inverse can be used to find the regional sectoral production required to fulfil the estimated aggregate demand. We substitute $ad_r$ into the traditional Leontief equation (Equation 4.5), to produce Equation 4.6, with $x_r$ equalling the nowcasted estimate for output.

$$x = Ly \tag{4.5}$$

$$x_r = L_r^* ad_r \tag{4.6}$$

After deriving all constraints for the nowcasted domestic IO table, a GRAS algorithm is applied to balance the domestic block. The implied regional import and export constraints are subsequently harmonised across regions using a basic RAS algorithm. This produces a fully balanced, nowcasted MRIO table that is consistent with the prescribed GDP and trade projections. For a comprehensive description of the original SPIN method please see, Beaufils and Wenz (2022), whilst a step-by-step derivation of the procedure used here is also provided in Supporting Information S4.3.

#### 4.4.3.1 Nowcasting the emission extension vector

The emission extension vector contains the annual $CO_2$e emissions generated by each industrial sector. In the UK, the ONS disaggregates UK production emissions into the 112-sectors modelled in the UKMRIO and reports on these emissions as Environmental Accounts, published at a lag of 2 years (ONS, 2025). Nowcasting for this data, and the corresponding international element is challenging, due to the scarcity in alternative, compatible emissions data sources. Examples include the Emissions Database for Global Atmospheric Research (EDGAR), which has a lag of less than one year (typically in the autumn following the reference year) but is not provided at the same level of sectoral disaggregation. Alternative sources include the UNFCCC National Inventory Submissions, published annually by individual countries, or the Eurostat Air Emission Accounts (AEA), which are generally available with a two-year lag.

Here, we instead adopt a statistical approach to predicting changes in regional emissions based on historically observed relationships with real macroeconomic growth. This method captures underlying emission trends while controlling for shocks that affect regional output. It also removes the need to use the UKMRIO as a time series, as well as avoiding extensive emissions data procurement and complicated data alignment steps.

For each nowcasted target year, we fit a simple ordinary least squares linear regression (Equation 4.7), specifying regional year-on-year emissions growth as the target variable ($\Delta F_t$) with the exogenously sourced, real GDP ($\Delta Y_t$), and import growth rates ($\Delta M_t$) as covariates. Here, the subscript t denotes growth rate year, with historical observations included in each regression to predict changes in future periods. We exclude exports to avoid collinearity arising from the near inevitable positive correlation between GDP and exports, a problem noted in trade-growth literature (Michaely, 1977).

$$\Delta F_t = \alpha + \beta_1 \Delta Y_t + \beta_2 \Delta M_t + \varepsilon_t \tag{4.7}$$

To replicate the three-year lag, the model for each target year is only fitted on data available three years earlier. This approach ensures out-of-sample predictions consistent with real implementation. As a result, a distinct regression is estimated for each region and every forecast year. These models are used to predict emission growth rates which can be applied to the observed levels in the base table to estimate for the target year. Regional totals are then allocated to industries on base-year shares. We provide a full account of the rationale and processes of these regressions in Supporting Information S4.4.

#### 4.4.3.2 Adjusting the price level

The aforementioned projection techniques apply real growth rates for GDP, imports, and exports. Consequently, without further adjustment, each nowcasted table is expressed in the price levels of the base year. Benchmark tables on the other hand, are presented in the current prices of each respective year. Inflationary adjustments are required to enable a comparison and in practice ensure compatibility with the nowcasted EEMRIO database and real-time data. Implementing real adjustments to a global database poses challenges, due to the varying inflationary pressures experienced by different nations during the same period. This, along with the need to maintain a balanced database, makes it difficult to adopt a nation-specific approach to inflating the database.

Given the SNAC construction of the UKMRIO, one option is to apply a single inflation index such as the Consumer Price Index (CPI), since purchases in the UK are made in UK prices (Wiedmann et al., 2010). However, this general adjustment does not capture industry-specific drivers of inflation. Instead, it is preferable to use industry-specific indices, which can be practically applied after multiplier calculation, rather than inflating the entire nowcasted MRIO table. Ultimately, the choice of inflation index depends on the intended use of the model.

In this paper, we use annual, industry-specific Producer Price Indices (PPIs) to align price levels between the UKMRIO emission multipliers and firm-level transaction data, enabling the estimation of Scope 3 emissions. This approach is feasible due to the alignment between UKMRIO sector definitions and published ONS inflation series (Owen and Kilian, 2025). Operationally, practitioners have the option to adjust multipliers (annually) or the transaction data (annually, quarterly, or monthly). Full details of the indices used are provided in Supporting Information S4.1.

### 4.4.4 Nowcasting Evaluation

To evaluate a nowcasting procedure's ability to reproduce actual UKMRIO years we rely on a selection of matrix difference statistics. There is no single statistical test for assessing the accuracy with which one matrix or vector corresponds with another, thus it is common practice to use different metrics simultaneously (Butterfield and Mules, 1980). The metrics selected are intended to complement each other, ensuring that weaknesses in one test are compensated and accounted for in other tests. Whilst these statistics often have differing assumptions, they all involve a quantitative description of some aspects of the difference between a predicted and benchmark value (Knudsen and Fotheringham, 1986). Because the inflation adjustment is applied only to the emissions multipliers, our comparison includes two price-invariant components (the emissions vector and the direct requirements matrix) and the resulting inflation-adjusted emissions multipliers. The subsections below introduce the metrics used and describes each matrix and vector included in the comparison.

#### 4.4.4.1 Matrix difference statistics

R-Squared (RSQ) or coefficient of determination: One of the most commonly employed goodness-of-fit statistics is RSQ (Knudsen and Fotheringham, 1986). This measure is the square of the correlation coefficient between observed and predicted values ($y_i$ and $\hat{y}_i$). As RSQ approaches 1, the model is seen as a good fit for the actual data. However, RSQ in isolation is considered an imperfect statistic and must be treated with caution (Butterfield and Mules, 1980; Knudsen and Fotheringham, 1986; Smith and Hutchinson, 1981).

$$RSQ = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \tag{4.8}$$

Here, $SS_{RES}$ is the residual sum of squares, defined as the sum of squared differences between observed and predicted values, $\sum_i (y_i - \hat{y}_i)^2$. In an OLS regression, the fitted model is chosen to minimise this value. $SS_{TOT}$ is the total sum of squares, defined as the sum of squared differences between observed values and their mean, $\sum_i (y_i - \bar{y}_i)^2$. Thus, RSQ measures the proportion of variation in the observed data explained by the model.

To complement RSQ, we calculate the mean absolute error (MAE) and mean absolute percentage error (MAPE). These metrics provide the average that each nowcasted element is larger or smaller than its target value (in both absolute and proportional terms).

$$MAE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| y_{ij} - y_{ij}^{true} \right| \tag{4.9}$$

$$MAPE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\left| y_{ij} - y_{ij}^{true} \right|}{\left| y_{ij}^{true} \right|} \times 100 \tag{4.10}$$

4.4.5.2 Matrices and vectors evaluated

We evaluate the nowcasting methods against a benchmark using three complementary diagnostics. First, we examine the industry-level emissions vector ($\boldsymbol{f}$) to assess accuracy of the emissions levels of the nowcasted table. We then compare the direct-requirements matrix ($\boldsymbol{A}$) to gauge the accuracy of the supply-chain structure used to distribute those emissions. Finally, we assess the product-level emission-intensity vector ($\boldsymbol{eL^{pp}}$ - embodied emission factors) to establish accuracy of the multipliers used in real-time.

Vector of emissions by industry ($\boldsymbol{f}$)

The first vector assessed is the emissions extension vector. This contains the emissions produced by every industry and region combination represented in the UKMRIO. We first compare annual, regional emissions totals before evaluating on an industry basis. We evaluate the whole vector (1232 industries) before focusing on the UK (112 industries).

As the same method for nowcasting these emissions are used for each method, only one assessment is required.

Matrix of direct requirements matrix ($\boldsymbol{A}$)

We then assess the matrix of direct requirements (technical coefficients). This matrix details the inputs from each industry required to produce the outputs of each industry. The matrix is normalised, with each element in $\boldsymbol{A}$ showing the proportional requirement of each input ($a_{ij}$ = $\frac{z_{ij}}{x_{ij}}$) for a single unit of output. The matrix contains every industry and region represented in the UKMRIO, with dimensions of 1232 x 1232. We assess the entire matrix, before focusing on UK domestic requirement. Three versions of A are produced, the first is matrix A from 3 years prior (no adjustment is made to the base table in Method 1), whilst Method 2 and 3 produce new versions of $\boldsymbol{A}$.

Vector of emission intensity by product ($\boldsymbol{eL^{pp}}$)

Finally, we assess the emission intensity vector of products consumed in the UK, in purchasers' prices. This vector is calculated in purchasers' prices, to replicate multipliers that could be combined with the transactions of businesses to calculate the embodied emissions of their purchases. Given our UK focus, we only look at the accuracy of nowcasted multipliers of each method for UK industries.

# 4.5 Results

## 4.5.1 Vector of emissions by industry ($\boldsymbol{f}$)

Table 4.3 presents the error metrics associated with different sections of the emissions vector. A total of 75 regression equations are estimated (15 regions across 5 years), as such, a summary table of regional model performance along with further interpretation, is provided in Supporting Information S4.4. These results are based on models trained using the full dataset, whereas the emissions estimates reported in the main analysis are derived from individual out-of-sample forecasts thus replicating real-world implementation. We begin by assessing the predicted regional production-based emission totals, which exhibit low percentage error levels. This suggests that the regression method effectively estimates total GHG emissions observed within the UKMRIO. The consistently high RSQ values further indicate that the model reliably captures variation in emissions across regions, given known real GDP and import rates. On average, the model yields an error of 4.77% across the study period, with the lowest errors observed during more stable economic years (2018 and 2019).

Figure 4.3 overlays the out-of-sample predictions onto actual emissions for the comparison period, as well as for the subsequent years (2023–2025), demonstrating the model's nowcasting capability. Regional trends are generally well replicated. However, more notable deviations are observed for some regions such as South Africa, where the model produces larger errors. These regions are noted to have more irregular patterns of emissions and emission intensity, creating challenges for prediction (See Supporting Information S4.4). In contrast, predictions for the UK and its major trading partners (EU, USA, China) closely align.

*Table 4.3 Comparison of emissions nowcasted, against benchmark. Green highlighting reflects best performing years for each method-metric combination (highest RSQ, lowest MAE and MAPE).*

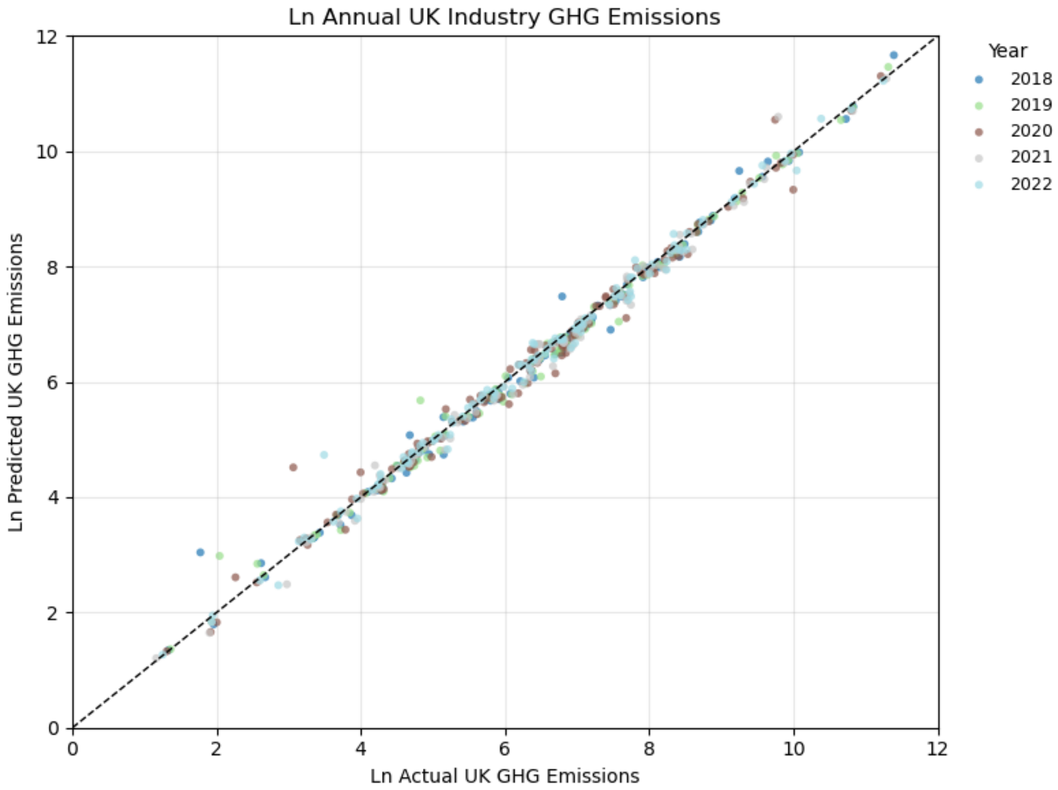
| | | 2018 | 2019 | 2020 | 2021 | 2022 | Mean |
|---|---|---|---|---|---|---|---|
| **Emissions by Regions** | | | | | | | |
| R2 | | 0.999 | 0.999 | 0.994 | 0.995 | 0.996 | 0.997 |
| MAE | tCo2e | 76,040 | 88,295 | 217,654 | 214,957 | 178,794 | 155,148 |
| MAPE | % | 2.25 | 3.57 | 6.9 | 5.78 | 5.37 | 4.77 |
| **Emissions by Industries** | | | | | | | |
| R2 | | 0.996 | 0.995 | 0.992 | 0.992 | 0.992 | 0.993 |
| MAE | tCo2e | 2,661 | 2,458 | 3,518 | 3,357 | 3,258 | 3,050 |
| MAPE | % | 9.6 | 9.3 | 16.3 | 12.4 | 9.4 | 11.4 |
| **Emissions by UK Industries** | | | | | | | |
| R2 | | 0.933 | 0.981 | 0.931 | 0.945 | 0.988 | 0.956 |
| MAE | tCo2e | 578 | 372 | 548 | 460 | 317 | 455 |
| MAPE | % | 14.87 | 12.93 | 16.54 | 12.07 | 11.35 | 13.552 |

*Figure 4.3 Nowcasted total regional production emissions.*


At the industry level, RSQ values remain strong but show a slight decline compared to the regional assessments. Meanwhile, MAPE more than doubles across all years. This indicates that while the model effectively predicts total emissions, it is less precise in distributing emissions across individual industries. In some cases, industries with high emissions in the base year unexpectedly exhibit much lower values in the benchmark data. In other cases, predicted industry emissions are consistently below observed levels, suggesting the allocation procedure fails to track rapidly expanding industries. The most prominent example is China's electric power generation sector, which produces the largest absolute error in all five years (average annual error of 462ktCO2e or 8.28%). These discrepancies contribute to higher element-level errors, despite the overall regional accuracy.

Focusing on UK industries, predictions maintain consistently high RSQ values, although slightly lower; the average MAPE across the study period also reflects this, increasing to 13.6%. Figure 4.4 visualises the relationship between logged emissions of UK industries and their nowcasted counterparts across all years. Among the ten industries with the highest annual percentage errors, 'Support Activities for Mining' appears in three separate years, while 'Manufacture of Tobacco Products'. This is due to their small absolute values (making absolute error very low). ‘Air Transport’ appears twice among the industries with the highest percentage errors (in 2020 and 2021), and among the top ten in terms of absolute error. This pattern is likely linked to the sharp reduction in air transport activity during the COVID-19 pandemic, a disruption that the lagged, base-year allocation method is unable to reflect. Table 4.3 also reveals a lower level of accuracy for UK industries in 2018, with performance closest to that observed for 2020 - a year marked by unprecedented disruption due to COVID-19. This reduced accuracy is likely linked to differences in the construction of the underlying base tables. The UKMRIO is designed to align as closely as possible with official national statistics published by the ONS. As a result, the 2015 base table (used to nowcast 2018) is built using detailed analytical tables in which the combined use table is fully disaggregated into a domestic use matrix and separate components for imports, product taxes, and value added. By contrast, from 2016 onwards, only less detailed analytical tables are available, and these provide only the Domestic Use table. This discrepancy in data availability and table construction means that comparisons between the nowcasted 2018 table (based on 2015) and the published 2018 table may reflect underlying structural differences rather than genuine projection error (Owen and Kilian, 2025).

*Figure 4.4 Nowcasted emissions of UK industries against benchmark counterparts.*




## 4.5.2 Matrix of direct requirements ($\boldsymbol{A}$)

Table 4.4 presents the error metrics for each method's direct requirements matrix ($\boldsymbol{A}$). At the full-matrix level, all methods yield consistently high RSQ values, indicating strong overall alignment with the benchmark structure. Because the elements of $\boldsymbol{A}$ are proportions and therefore small in magnitude, MAE values remain low, while MAPE values become inflated due to the prevalence of near-zero coefficients. Methods 2 and 3 show a pronounced increase in MAPE during 2021–2022, consistent with the economic volatility following COVID-19 period, which affects both the input data used for projection and the benchmark outcomes. Notably, the performance metrics in 2020 are consistent with those observed in surrounding years. The 2018 results, again, stand out as a clear outlier, with substantially lower accuracy for both the global matrix and, more prominently, the UK submatrix. As discussed earlier, this is likely driven by differences in the construction of the underlying base tables. These inconsistencies in table construction introduce structural differences that reduce nowcasting feasibility.

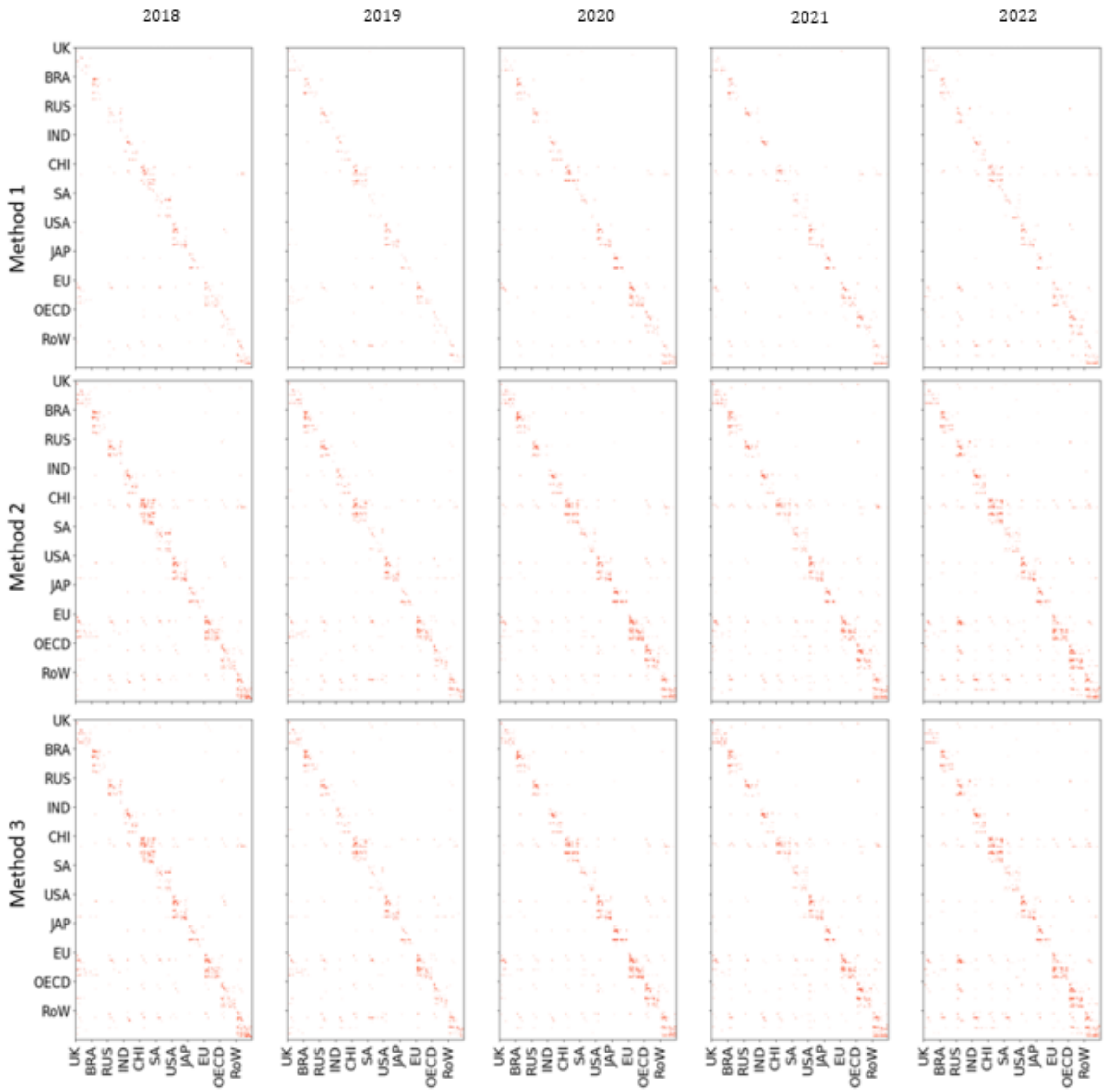
*Table 4.4 Comparison of* $\boldsymbol{A}$ *nowcasted matrices against benchmark. Green highlighting reflects best performing years for each method-metric combination (highest RSQ, lowest MAE and MAPE)*

| | | 2018 | 2019 | 2020 | 2021 | 2022 | Mean |
|---|---|---|---|---|---|---|---|
| **Entire A Matrix** | | | | | | | |
| Method 1 | RSQ | 0.92 | 0.95 | 0.96 | 0.96 | 0.95 | 0.95 |
| | MAE | 8.47E-05 | 7.23E-05 | 7.53E-05 | 7.25E-05 | 8.21E-05 | 7.74E-05 |
| | MAPE | 2.74E+09 | 1.24E+08 | 3.99E+05 | 9.20E+09 | 8.99E+10 | 2.04E+10 |
| Method 2 | RSQ | 0.92 | 0.95 | 0.96 | 0.96 | 0.95 | 0.95 |
| | MAE | 8.55E-05 | 7.22E-05 | 7.52E-05 | 7.26E-05 | 8.23E-05 | 7.76E-05 |
| | MAPE | 4.46E+09 | 1.14E+09 | 4.10E+08 | 2.90E+18 | 2.18E+19 | 4.94E+18 |
| Method 3 | RSQ | 0.92 | 0.95 | 0.96 | 0.96 | 0.95 | 0.95 |
| | MAE | 8.56E-05 | 7.25E-05 | 7.55E-05 | 7.25E-05 | 8.24E-05 | 7.77E-05 |
| | MAPE | 4.60E+09 | 1.17E+09 | 4.06E+08 | 2.88E+18 | 2.19E+19 | 4.96E+18 |
| **UK Section Only** | | | | | | | |
| Method 1 | RSQ | 0.59 | 0.93 | 0.89 | 0.89 | 0.92 | 0.85 |
| | MAE | 1.04E-03 | 8.11E-04 | 9.54E-04 | 9.45E-04 | 8.54E-04 | 9.22E-04 |
| | MAPE | 3.86E+09 | 1.73E+03 | 8.95E+02 | 1.73E+02 | 2.18E+03 | 7.72E+08 |
| Method 2 | RSQ | 0.60 | 0.94 | 0.89 | 0.89 | 0.92 | 0.85 |
| | MAE | 1.04E-03 | 8.09E-04 | 9.58E-04 | 9.50E-04 | 8.55E-04 | 9.23E-04 |
| | MAPE | 3.72E+09 | 1.29E+05 | 3.64E+03 | 1.08E+04 | 1.59E+04 | 7.44E+08 |
| Method 3 | RSQ | 0.59 | 0.93 | 0.87 | 0.87 | 0.92 | 0.84 |
| | MAE | 1.05E-03 | 8.07E-04 | 1.03E-03 | 1.01E-03 | 8.51E-04 | 9.5E-04 |
| | MAPE | 3.72E+09 | 1.29E+05 | 3.74E+03 | 1.11E+04 | 1.56E+04 | 7.4E+08 |

Within the UK submatrix more generally, RSQ values are systematically lower than in the global matrix, with all methods returning approximately 0.6 in 2018. Unlike the global results, the UK section does not exhibit the late-period MAPE spike, suggesting relatively greater stability on the UK level. Across all metrics, there is no clear performance advantage associated with the additional complexity of Methods 2 and 3, indicating that their expanded data requirements do not improve the accuracy of the projected UK A matrix.

Figure 4.5 presents heatmaps of annual MAE, illustrating where the projection errors occur across methods and years. Errors in Method 1 are highly localised within domestic transaction blocks (along the diagonal), with China and the Rest of the World showing the largest concentrations. In contrast, Method 2 produces more widespread deviations, with additional errors across multiple international flows, indicating that this approach introduces structural changes that are not reflected in the target-year data. Method 3 shows a similar performance to Method 2 and aligns with the metrics reported in Table 4.4, where industry-specific adjustments fail to meaningfully reduce errors within the UK portion of the database. Overall, the heatmaps indicate that Method 1 yields the most structurally consistent projections, despite not incorporating nowcasted GDP or trade adjustments.

*Figure 4.5 Top 15% largest sources of MAE within nowcasted A matrices.*




### 4.5.3 Vector of product emission intensity ($\boldsymbol{eL^{pp}}$)

Table 4 summarises the error metrics for the embodied emission multipliers for the UK, and a time-series for each metric is also shown in Figure 4.6. Across the period, Methods 1 and 2 show broadly similar performance, whereas Method 3 displays greater volatility.

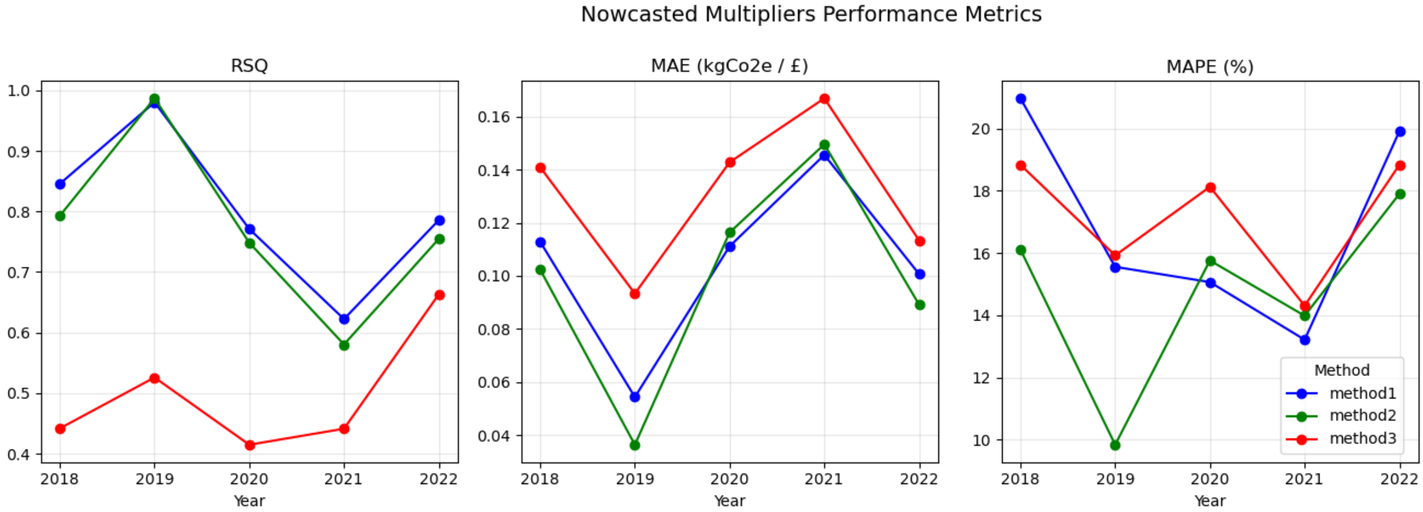
Methods 1 and 2 achieve comparable RSQ values, averaging 0.80 and 0.77 respectively, with increases and declines occurring in the same years. Their error magnitudes are also similar: both record average MAE values of 0.1 kgCo2e / £, and MAPE values in the

15-17% range, with Method 2 performing slightly better on average, due to larger errors observed in 2018 by Method 1.

*Table 4.5 - Comparison of $\boldsymbol{eL^{pp}}$ nowcasted vectors against benchmark. Green highlighting reflects best performing years for each method-metric combination (highest RSQ, lowest MAE and MAPE)*

| | 2018 | 2019 | 2020 | 2021 | 2022 | Mean |
|---|---|---|---|---|---|---|
| **Method 1** | | | | | | |
| R2 | 0.85 | 0.98 | 0.77 | 0.62 | 0.79 | 0.802 |
| MAE (kgCo2e / £) | 0.11 | 0.05 | 0.11 | 0.15 | 0.10 | 0.104 |
| MAPE (%) | 20.97 | 15.55 | 15.06 | 13.21 | 19.91 | 16.94 |
| **Method 2** | | | | | | |
| R2 | 0.79 | 0.99 | 0.75 | 0.58 | 0.75 | 0.772 |
| MAE (kgCo2e / £) | 0.10 | 0.04 | 0.12 | 0.15 | 0.09 | 0.10 |
| MAPE (%) | 16.11 | 9.84 | 15.76 | 13.98 | 17.90 | 14.718 |
| **Method 3** | | | | | | |
| R2 | 0.44 | 0.53 | 0.42 | 0.44 | 0.66 | 0.498 |
| MAE (kgCo2e / £) | 0.14 | 0.09 | 0.14 | 0.17 | 0.11 | 0.13 |
| MAPE (%) | 15.71 | 18.83 | 15.92 | 18.13 | 14.30 | 18.83 |

*Figure 4.8 Annual error metrics of nowcasted emission factors by method.*




Method 3's performance heavily depends on the metric used. Its RSQ values are substantially lower (averaging around 0.5), although they show some improvement in the 2022. In absolute terms, MAE generally follows the same broad pattern as for Methods 1 and 2, at a slightly higher level. The comparatively low RSQ alongside similar MAE and MAPE values indicates that, while the model predicts the overall magnitude of multipliers reasonably well, it fails to capture their underlying variation across industries.

Overall, these patterns suggest that Method 3, whilst attempting to capture UK industry growth variation, fails to increase accuracy relative to Method 1 and 2. By contrast, Methods 1 and 2 track the period more smoothly and with greater consistency. The main distinction between the two models is Method 2's noticeably lower MAPE in 2018 and 2019.

## 4.6. Discussion

### 4.6.1 Key Findings

The nowcasting approaches developed in this paper build on existing MRIO projection techniques and are designed to be practical to implement while capturing the key economic changes occurring during the publication lag period. Method 1 applies only an inflation adjustment, retaining the structural relationships of the base year. Method 2 extends this by incorporating real GDP and trade growth to update emissions estimates and generate a nowcasted MRIO structure. Method 3 adds industry-specific adjustments using additional UK-focused data.

Across the period studied, differences in average error across methods are relatively modest. The mean MAPE for the product multipliers remains between 14.7% and 17.2% for all three methods. On average, the nowcasted multipliers slightly underestimate their benchmark values. Method 2 achieves the lowest average MAPE, indicating that incorporating macroeconomic indicators offers a marginal improvement, though the gains over a simple inflation adjustment remain small.

Error levels are lowest in 2019, while the years 2020–2022 exhibit higher errors, consistent with the sharp disruptions brought about by the COVID-19 pandemic, supply-chain instability, and rapid inflation. Comparisons of $\boldsymbol{A}$ and $\boldsymbol{eL^{pp}}$ also show distinctly poor performance in 2018, indicating strong sensitivity to differences in the construction of the base tables used for nowcasting.

Higher-than-average errors in the $\boldsymbol{A}$ matrix for Methods 2 and 3 highlight the difficulty of projecting technical coefficients, each of which changes over time even when underlying structural patterns appear persistent (Torres-González and Yang, 2019). These challenges also intensify during periods of economic disruption. The UK block of $\boldsymbol{A}$ avoids the very large global errors, but Method 3 introduces additional volatility in its attempts to incorporate industry-specific adjustments.

The approach used to project regional emissions totals performs well for many regions and provides a practical solution requiring no additional data. Other comparisons of MRIO databases have found the emissions vector to be large explainers of different database consumption-based estimates (Owen et al., 2014), whilst Tukker et al., (2018) note that the environmental figures used are the most significant factor contributing uncertainty of estimates. Projections are best for regions economically or geographically close to the UK, where emissions trajectories appear relatively stable. Larger errors arise for regions such as South Africa and Japan, where emissions intensities exhibit substantial volatility in the base data. Mining Support Services is a notable example, returning MAPE values above 100% due to much higher base-year levels than in the target year. Importantly, these errors occur on small absolute totals (21, 35, and 124 MtCO2e). This reflects a broader trend: smaller values tend to be associated with disproportionately higher relative errors, as shown in Figure 4.4, which shows overestimation is concentrated at the lower end of the distribution.

Across the assessment period, Methods 1 and 2 return broadly comparable and stable error levels for the UK embodied emission multipliers ($\boldsymbol{eL^{pp}}$), with similar MAE and MAPE values and RSQs around 0.8. By contrast, Method 3 shows much greater volatility and substantially lower RSQ values, indicating that it struggles to reproduce industry-level variation despite generating multiplier magnitudes similar to the other methods. Overall, the added complexity of Method 3 does not translate into improved accuracy, while Methods 1 and 2 provide more reliable and consistent estimates for practical use.

### 4.6.2 Practical Implications

The time lag associated with EEMRIOs poses a significant barrier to their practical application (Lenzen, 2000; Wiedmann et al., 2011). The results presented here do not show sufficient error reduction to justify the additional complexity of Methods 2 and 3. While the regional emissions totals can be projected with reasonable accuracy, distributing these emissions across industries and projecting the $\boldsymbol{A}$ matrix remains challenging.

The emission factors produced by EEMRIO databases are highly versatile and as a result, a growing body of research integrates EEMRIO-derived emission factors with alternative data sources to estimate the embodied emissions of various actors. For households, Trendl et al (2023) demonstrate that FTD-based emission estimates align closely with those derived from household spend surveys, whilst Wells et al (2025) uses FTD-based emissions to segment individuals with similar sociodemographic, consumption behaviours and carbon footprints. Similarly, Owen and Büchs (2024) utilise household spend surveys with an EEMRIO database to investigate generational differences in consumption-based emissions. When combined with the FTD of businesses, these emission factors can also be used estimate the embodied supply chain emissions of purchases (Scope 3 emissions) (Schmidt et al., 2022; Phillpotts et al., 2026). These bodies of work, alongside others (Andersson, 2020; Cogo, 2022; Schmidt et al., 2022), have demonstrated the distinct advantages of leveraging real-time FTD to generate scalable, standardized emission estimates.

The UKMRIO's alignment with ONS inflation indices means that external inflation adjustments offer a simple and effective means of updating emission factors. To support wider adoption of EEMRIO tools in real-time applications, including business value-chain emissions estimation, this paper provides evidence that inflation adjustments remain an appropriate and low-burden strategy at present. When combined with real-time data like financial transactions, users have the option to apply deflation series directly to the transaction, this can be done on an annual, quarterly, or even monthly basis.

#### 4.6.3 Limitations

Several limitations contribute to the error levels observed. First, no additional data were incorporated to proxy changes in the environmental extension vector. While supplementary datasets exist, they lack the appropriate level of sectoral detail or are released too infrequently to be useful for annual nowcasting. Incorporating the most recent year of environmental accounts would still leave two unavailable years, limiting its usefulness. Second, industry-level emissions are allocated using fixed base-year shares. Although pragmatic, this approach weakens performance when sectoral emissions change from their historical distribution. For example, the sharp decline in UK aviation emissions during 2020 and 2021 was not captured through this approach. More sophisticated alternatives, such as using historical emissions trends or incorporating aggregated energy-use data, would require far greater model complexity.

#### 4.6.4 Future Research Directions

The results highlight the need for a Single-country National Accounts Consistent (SNAC) approach to nowcasting national MRIOs such as the UKMRIO. The Climate Change Committee has repeatedly identified the MRIO publication lag as a major obstacle to effective climate-policy monitoring (CCC, 2022). The methods developed here represent progress, but they remain grounded in techniques originally designed for global MRIO systems, which prioritise balanced accuracy across regions rather than national specificity.

Future work should nowcast the UKMRIO directly in its SUT form, avoiding the additional assumptions introduced when converting to symmetric IO tables. A SNAC approach would anchor projections more firmly in UK-specific data, and produce results more aligned with domestic policy requirements. Such a framework would provide timelier and more reliable inputs for assessing national progress and enable a wider set of real-time applications, including consumption-based emissions monitoring and Scope 3 analysis.

In the near term, this paper demonstrates that simple inflation adjustments remain an effective and scalable method for updating EEMRIO-derived emission factors. Establishing a standardised inflation-adjustment methodology, or publishing regularly updated multipliers, would support their broader adoption in real-time analytical settings.

## 4.7. Conclusion

This paper has developed and tested three practical approaches for nowcasting the UKMRIO, aimed at addressing the long-standing challenge of time lags in EEMRIO availability. The results demonstrate that while average error levels are relatively similar across methods, moderate adjustments to regional sizes and trade flows (Method 2) provide the best balance between complexity and predictive performance. Performance is affected in years of economic stability and weakest during periods of economic disruption, reflecting the inherent difficulty of replicating the complex data structures represented by an EEMRIO. Nowcasting procedures are also highly sensitive to changes in the construction process of EEMRIOS. For practitioners, the findings underline that additional methodological complexity does not necessarily yield proportional improvements in accuracy, and that straightforward adjustments to account for changing price levels may often be more or equally effective.

# 4.8 Notes

### 4.8.1 Acknowledgements

We would first like to thank Lara Vomfell, Jasmine Wells, Alistair Gilfillan, and the rest of the behavioural science team at Lloyds Banking Group for sharing their invaluable expertise at key points of this research, as well as their continued interest and support.

Finally, we extend our sincere gratitude to Ranil Boteju, Trystan Davies, Andrew McCulloch, Neil Oliver and Jaideep Mann for making this project possible and supporting its successful publication.

### 4.8.2 Funding Information:

Alec Phillpotts reports financial support was provided by Lloyds Banking Group. Jonathan Norman's time was funded by the UK Research and Innovation (UKRI) Council, as part of the Energy Demand Research Centre, grant reference number EP/Y010078/1.

## 4.9 References

Afionis, S., Sakai, M., 2022. Production- and consumption-based accounting: an overview of the debate, in: Handbook on Trade Policy and Climate Change, Chapter 15. Edward Elgar Publishing, pp. 239–255.

Andersson, D., 2020. A novel approach to calculate individuals' carbon footprints using financial transaction data – App development and design. J. Clean. Prod. 256, 120396. https://doi.org/10.1016/j.jclepro.2020.120396

Banbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow (No. 1564). European Central Bank, Brussels. https://doi.org/10.1016/B978-0-444-53683-9.00004-9

Bank of England, 2017. Peering into the present: the Bank's approach to GDP nowcasting. Q. Bull. 2017 Q2 Artic. https://www.bankofengland.co.uk/quarterly-bulletin/2017/q2/peering-into-the-present-the-banks-approach-to-gdp-nowcasting (accessed 8.7.23).

Barrett, J., Glen, P., Wiedmann, T., Scott, K., Lenzen, M., Roelich, K., Le Quere, C., 2013. Consumption-based GHG emission accounting: a UK case study. Clim. Policy 13, 451–470. http://dx.doi.org/10.1080/14693062.2013.788858

Beaufils, T., Wenz, L., 2022. A scenario-based method for projecting multi-regional input–output tables. Econ. Syst. Res. 34, 440–468. https://doi.org/10.1080/09535314.2021.1952404

Brown, N., Berglund, M., Roth, S., Statistics Sweden, 2021. Improving timeliness of statistics on air emissions with a consumption perspective. https://seea.un.org/sites/seea.un.org/files/brown_improving-timeliness-of-statistics-on-air-emissions-with-a-consumption-perspective_paper.pdf (accessed 8.4.23).

Butterfield, M., Mules, T., 1980. A Testing Routine For Evaluating Cell By Cell Accuracy In Short-Cut Regional Input-Output Tables. J. Reg. Sci. 20, 293–310. https://doi.org/10.1111/j.1467-9787.1980.tb00648.x

Casella, B., Bolwijn, R., Moran, D.D., Kanemoto, K., 2019. Improving the Analysis of Global Value Chains: The UNCTAD-Eora Database. Transnatl. Corp. J. 26. https://unctad.org/system/files/official-document/diaeia2019d3a5_en.pdf (accessed 8.4.23).

Cogo, 2022. Calculate The Carbon Footprint of Your Transactions. https://www.cogo.co/about (accessed 23.5.23).

Committee on Climate Change (CCC), 2022. Progress in reducing emissions: 2022 Report to Parliament. London. https://www.theccc.org.uk/wp-content/uploads/2022/06/Progress-in-reducing-emissions-2022-Report-to-Parliament.pdf (accessed 4.5.23).

Davis, S.J., Caldeira, K., 2010. Consumption-based accounting of CO2 emissions. Proc. Natl. Acad. Sci. 107, 5687–5692. https://doi.org/10.1073/pnas.0906974107

Department of Economic and Social Affairs, 2018. Chapter 18. Projecting supply, use and input-output tables, in: Handbook on Supply and Use Tables and Input-Output Tables with Extensions and Applications. Department of Economic and Social Affairs. https://unstats.un.org/unsd/nationalaccount/docs/SUT_IOT_HB_Final_Cover.pdf (accessed 2.5.23).

Edens, B., Hoekstra, R., Zult, D., Lemmers, O., Wilting, H.C., Wu, R., 2015. A method to create carbon footprint estimates consistent with national accounts. Econ. Syst. Res. 27, 440–457. https://doi.org/10.1080/09535314.2015.1048428

Eurostat, 2025. Supply and use tables for individual countries. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Supply_and_use_tables_for_individual_countries (accessed 11.17.25).

Eurostat, 2021. FIGARO methodology. Eurostat, Luxemberg. https://ec.europa.eu/eurostat/documents/51957/12767369/Figaro-methodology.pdf/487255c9-903a-0cb7-f5e2-73f37e35f196?t=1620750704022 (accessed 1.10.25).

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., Schreiner, J., 2021. Nowcasting in a pandemic using non-parametric mixed frequency VARs. J. Econom. 232, 52–69. https://doi.org/10.1016/j.jeconom.2020.11.006

International Monetary Fund (IMF), 2022. World Economic Outlook Database, October 2022 [Dataset]. IMF. https://www.imf.org/en/Publications/WEO/weo-database/2022/October (accessed 6.19.23).

International Monetary Fund (IMF), 2022. October 2022 WEO Database - Country Data Documentation. https://www.imf.org/external/pubs/ft/weo/2022/02/weodata/co.pdf (accessed 6.19.23).

Kasman, A., Duman, Y.S., 2015. CO2 emissions, economic growth, energy consumption, trade and urbanization in new EU member and candidate countries: A panel data analysis. Econ. Model. 44, 97–103.

Kitzes, J., 2013. An Introduction to Environmentally-Extended Input-Output Analysis. Resources 2, 489–503. https://doi.org/doi:10.3390/resources2040489

Knudsen, D.C., Fotheringham, A.S., 1986. Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modelling. Int. Reg. Sci. Rev. 127–147. http://dx.doi.org/10.1177/016001768601000203

Lenzen, M., 2000. Errors in Conventional and Input-Output-based Life-Cycle Inventories. J. Ind. Ecol. 4, 127–148. https://doi.org/10.1162/10881980052541981

Lenzen, M., Kanemoto, K., Moran, D., Geschke, A., 2012. Mapping the Structure of the World Economy. Environ. Sci. Technol. 46, 8374–8381. https://doi.org/10.1021/es300171x

Maluck, J., Donner, R.V., 2015. A Network of Networks Perspective on Global Trade. PLOS ONE 10, e0133310. https://doi.org/10.1371/journal.pone.0133310

Mangır, N., Ülkü Alver, Ş., 2022. An environmentally extended global multi-regional input–output analysis of consumption-based and embodied import-based carbon emissions of Turkey. Environ. Sci. Pollut. Res. 29, 54813–54826. https://doi.org/10.1007/s11356-022-19290-z

Miao, G., Fortanier, F., 2017. Nowcast TiVA estimates: methodology. https://one.oecd.org/document/STD/CSSP/WPTGS(2018)4/en/pdf (accessed 9.4.23).

Michaely, M., 1977. Exports and growth: An empirical investigation. J. Dev. Econ. 4, 49–53. https://doi.org/10.1016/0304-3878(77)90006-2

Miller, R.E., Blair, P.D., 2009. Foundations of Input–Output Analysis. Cambridge University Press, Cambridge.

Mostajabi, A., Finney, D., Rubinstein, M., Rachidi, F., 2019. Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. Npj Clim. Atmospheric Sci. 2, 41. https://doi.org/10.1038/s41612-019-0098-0

Organisation for Economic Co-operation and Development (OECD), 2017. Nowcasting techniques to improve timeliness. Presented at the Committee for the Coordination of

Statistical Activities, Muscat. https://unstats.un.org/unsd/statcom/51st-session/documents/2020-5-UNSystem-EE.pdf (accessed 9.4.23).

Office for National Statistics (ONS), 2025. Atmospheric emissions: greenhouse gases by industry and gas [Dataset]. https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/greenhousegasintensityprovisionalestimatesuk/provisionalestimates2023 (accessed 9.4.23).

Office for National Statistics (ONS), 2022a. Developing quarterly greenhouse gas emissions accounts, UK - Office for National Statistics. https://www.ons.gov.uk/economy/environmentalaccounts/articles/developingquarterlygreenhousegasemissionsaccountsuk/may2022 (accessed 8.22.23).

Office for National Statistics (ONS), 2022b. Supply and use tables [Dataset]. https://www.ons.gov.uk/economy/nationalaccounts/supplyandusetables (accessed 8.4.23).

Office for National Statistics (ONS), 2025a. Producer price inflation statistics - Office for National Statistics [Dataset]. https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/producerpriceinflationstatistics (accessed 11.20.25).

Office for National Statistics (ONS), 2025b. Contributions to monthly GDP - Office for National Statistics [Dataset]. https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/contributionstomonthlygdp (accessed 11.20.25).

Owen, A., Büchs, M., 2024. Examining changes in household carbon footprints across generations in the UK using decomposition analysis. J. Ind. Ecol. 28, 1786–1800. https://doi.org/10.1111/jiec.13567

Owen, A., Kilian, L., 2025. 2025 Data Release of Consumption- based Accounts for the UK: Summary of Methods. https://assets.publishing.service.gov.uk/media/68220ff3db6463b14cd81916/Summary_of_Methods_2025_final.pdf.

Owen, A., Steen-Olsen, K., Barrett, J., Wiedmann, T., Lenzen, M., 2014. A structural decomposition approach to comparing MRIO databases. Economic Systems Research, 26. https://doi.org/10.1080/09535314.2014.935299.

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2026. Completing the SME carbon profile: scalable prediction of scope 3 emissions. Journal of Industrial Ecology. https://doi.org/10.1007/s44498-026-00003-5

Schmidt, M., Nill, M., Scholz, J., 2022. Determining the Scope 3 Emissions of Companies. Chem. Eng. Technol. 45, 1218–1230. https://doi.org/10.1002/ceat.202200181

Shue, H., 1999. Global Environment and International Inequality. Int. Aff. 75, 531–545. https://doi.org/10.1111/1468-2346.00092

Smith, D.P., Hutchinson, B.G., 1981. Goodness of fit statistics for trip distribution models. Transp. Res. Part Gen. 15, 295–303. https://doi.org/doi.org/10.1016/0191-2607(81)90011-X

Stadler, K., Wood, R., Bulavskaya, T., Kuenen, J., Eisenmenger, N., Theurl, M.C., Koning, A. de, Giljum, S., Lutter, S., Wieland, H., Usubiaga, A., Acosta-Fernandez, J., Merciai, S., Schmidt, J., Tukker, A., 2014. DESIRE: Interim Report on Data Processing Creating EE IO Time Series and Now-Casted Data, D5.2 interim Report. https://2-0-lca.com/publications/show/d5-2-interim-report-on-data-processing-creating-ee-io-time-series-and-now-casted-data/ (accessed 8.4.23).

Stadler, K., Wood, R., Bulavskaya, T., Sodersten, C.-J., Simas, M., Schmidt, S., Usubiaga, A., Acosta-Fernandez, J., Kuenen, J., Bruckner, M., Giljum, S., Lutter, S., Merciai, S., Theurl, M.C., Plutzar, C., Kastner, T., Eisenmenger, N., Erb, K.-H., Koning, A. de, Tukker, A., 2018. EXIOBASE 3 Developing a Time Series of Detailed Environmentally Extended Multi-Regional Input-Output Tables. J. Ind. Ecol. 22, 502–515. https://doi.org/10.1111/jiec.12715

Stadler, K., Wood, R., Simas, M., Bulavskaya, T., Kuenen, J., Acosta-Fernandez, J., Usubiaga, A., Merciai, S., Schmidt, J., Theurl, M.C., Kastner, T., Eisenmenger, N., Giljum, S., Lutter, S., Lieber, M., Wieland, H., Tukker, A., 2015. Desire: Integrated Report on EE IO related

Macro Resource Indicator Time Series, D5.3 Integreted Report. https://lca-net.com/files/DESIRE_D5_3_main-plus-annexes-v2.pdf (accessed 8.4.23).

Temursho, U., Cardenete, M.A., Wojtowicz, K., Rey Los Santos, L., Weitzel, M., Vandyck, T., Saveyn, B., 2020. Projecting input-output tables for model baselines. Joint Research Centre. https://publications.jrc.ec.europa.eu/repository/handle/JRC120513 (accessed 8.4.23).

Temurshoev, U., Webb, C., Yamano, N., 2011. Projection of Supply and Use Tables: Methods and their Empirical Assessment. Econ. Syst. Res. 23, 91–123. https://doi.org/10.1080/09535314.2010.534978

Torres-González, L.D., Yang, J., 2019. The persistent statistical structure of the US input–output coefficient matrices: 1963–2007. Econ. Syst. Res. 31, 481–504. https://doi.org/10.1080/09535314.2018.1561425

Trendl, A., Owen, A., Vomfell, L., Kilian, L., Gathergood, J., Stewart, N., Leake, D., 2023. Estimating carbon footprints from large scale financial transaction data. J. Ind. Ecol. 27, 56–70. https://doi.org/10.1111/jiec.13351

Tukker, A., Koning, A. de, Owen, A., Lutter, S., Bruckner, M., Giljum, S., Stadler, K., Wood, R., Hoekstra, R., 2018. Towards Robust, Authoritative Assessments of Environmental Impacts Embodied in Trade. J. Ind. Ecol. 22, 585–598. https://doi.org/10.1111/jiec.12716

Tukker, A., Pollitt ,Hector, and Henkemans, M., 2020. Consumption-based carbon accounting: sense and sensibility. Clim. Policy 20, S1–S13. https://doi.org/10.1080/14693062.2020.1728208

Valderas-Jaramillo, J.M., Rueda-Cantuche, J.M., Olmendo, E., Beutel, J., 2019. Projecting supply and use tables: new variants and fair comparisons. Econ. Syst. Res. 31, 423–444. https://doi.org/10.1080/09535314.2018.1545221

Wells, J., Trendl, A., Owen, A., Barrett, J., Jobst, N., Leake, D., 2025. Targeting carbon reduction in UK households: A new segmentation model using financial transaction data. J. Ind. Ecol. https://doi.org/10.1111/jiec.70020

Wiebe, K.S., Bjelle, E.L., Tobben, J., Wood, R., 2018. Implementing exogenous scenarios in a global MRIO model for the estimation of future environmental footprints. Econ. Struct. 7. https://doi.org/10.1186/s40008-018-0118-y

Wiedmann, T., 2009. A review of recent multi-region input–output models used for consumption-based emission and resource accounting. Ecol. Econ. 69, 211–222. https://doi.org/10.1016/j.ecolecon.2009.08.026

Wiedmann, T., Wilting, H.C., Lenzen, M., Lutter, S., Palm, V., 2011. Quo Vadis MRIO? Methodological, data and institutional requirements for multi-region input–output analysis. Ecol. Econ. 70, 1937–1945. https://doi.org/10.1016/j.ecolecon.2011.06.014

Wiedmann, T., Wood, R., Minx, J.C., Lenzen, M., Guan, D., Harris, R., 2010. A Carbon Footprint Time Series Of The UK – Results From A Multi-Region Input–Output Model. Econ. Syst. Res. 22, 19–42. https://doi.org/DOI: 10.1080/09535311003612591

Wilting, H., Wiedmann, T., Lutter, F.S., Palm, V., Giljum, S., Wadeskog, A., Nijdam, D., 2009. Development of a methodology for the assessment of global environmental impacts of traded goods and services.

Wood, R., Moran, D.D., Rodrigues, J.F.D., Stadler, K., 2019. Variation in trends of consumption-based carbon accounts. Sci. Data 6. https://doi.org/10.1038/s41597-019-0102-x

Xie, J., 2023. Identifying Optimal Indicators and Lag Terms for Nowcasting Models. IMF. https://doi.org/10.5089/9798400235177.001

# Chapter 5. Discussion and Conclusion

This chapter begins with a summary of the key findings from the empirical chapters in light of this thesis' research questions, followed by a discussion of the impacts, methodological contributions, practical implementation, and policy implications of this work. In addition, this chapter discusses the key limitations and future directions highlighted by this research, before finishing on some concluding remarks. This chapter, therefore, summarises the novel contributions of this thesis and where it places within the wider context of small- and medium-sized enterprise (SME) emission measurement, disclosure, and estimation as well as within the broader sustainability topic.

## 5.1 Summary of Findings

In Chapter 1 of this thesis, 3 research questions (RQs) are set out. These RQs are then addressed through the empirical Chapters 2, 3, and 4. This section briefly summarises the results of these chapters, and how they answer the RQs they intended to address.

### 5.1.1 RQ1: Can FTD be used to produce models that capture variation in SME emissions across SMEs and industries?

Chapters 2 and 3 describe the processes required to transform large-scale Financial Transaction Data (FTD) into emissions estimates for SMEs, and the subsequent models that can be produced from this dataset. These cover direct emissions from fuel use (Scope 1), indirect emissions from purchased energy (Scope 2), and other upstream emissions arising from their purchases (Scope 3). Owing to methodological differences in estimation, as well as the distinct external stakeholders and reporting purposes associated with each, Chapter 2 focuses on Scopes 1 and 2, while Chapter 3 extends the framework to Scope 3. Within these chapters, a novel process is developed for calculating SME emissions from their bank accounts, using publicly available UK datasets and the UKMRIO. Using country-specific data enhances the robustness and relevance of these estimates (Edens et al., 2015; Tukker et al., 2018). This is especially important for Scopes 1 and 2, where aligning spend-based estimates with physical-activity-based estimates requires applying the same underlying physical intensity factors, adjusted to make them compatible with expenditure data.

Hierarchical regression formulas are then specified, using a selection of potential industry and firm-level predictors, which are sequentially applied to the resulting emissions datasets and their performances are assessed. The practical implementation of these models enables FTD-based emission estimates to be predicted in instances where extensive FTD are not available.

In Chapter 2, models used to estimate Scope 1 and 2 emissions achieve an RSQ of up to 0.89 for Scope 1, and 0.73 for Scope 2. Potential predictors include firm characteristics on size (annual turnover), industry, and insights on business attributes that may influence emissions, such as annual margin (revenues less expenses), and capital spending (on vehicles and other assets such as plant hire). Within these models, a consistent, strong, and significant relationship between ln turnover and ln Scope 1 and 2 emissions is observed, with coefficients of at least 0.7. This log-log form shows that the underlying relationship between turnover and emissions is non-linear, which contrasts with the assumptions of simple estimation methods using industry averages. In addition, accounting for industry characteristics beyond average emission intensity markedly improves model performance, particularly for Scope 1. Incorporating industry skew increases explanatory power by almost 10%, highlighting the important factors that are often overlooked when using industry averages.

In Chapter 3, the approach is applied to the Scope 3 emission categories that can be captured by a firm's transactions. Here, the same firm-level variables are included, with the addition of more granular industry groupings, to further capture variation of patterns of spending. Model performance is dependent on the Scope 3 emission category under investigation; Category 1 (Purchased Goods and Services) produces RSQ values of up to 0.88 whilst Category 3 (Energy- and Fuel-Related Emissions) is disaggregated into elements corresponding to the well-to-tank (WTT) emissions associated with Scope 1 and Scope 2, which return equivalent RSQ values (up to 0.89 and 0.73). Model performances, however, are nearly 20% lower for Categories 6 (Business Travel) and 9 (Downstream transportation and distribution), where the highest RSQ values only achieve 0.55 and 0.43 respectively. Weaker performance for these categories reflects their reliance on stronger assumptions, which likely contributes to the higher unexplained variance. For business travel, the assumption that all employee travel is paid directly through the business account is required, excluding cases where employees pay upfront and are later reimbursed. For downstream transportation, a substantial share of firms (43% of the sample) report zero spending in the relevant spend categories, further limiting the model's applicability. Emissions in these categories may therefore reflect firm behaviours not fully explained by size and industry characteristics, with bank account data providing only a partial view of total exposure. In absolute terms, these emissions are also much smaller than those in Category 1.

Findings in these chapters not only exemplify the usability of FTD in producing emission estimates for an overwise uncaptured segment of businesses but also introduce an approach to producing models that are able to capture the insights gained from granular FTD-based approaches. These models are designed to serve as an accessible resource, allowing both SMEs and their external stakeholders to obtain benchmark emissions profiles.

When accessed by SMEs, the estimates can inform decision-making and target setting. When used by external stakeholders, they offer a mechanism to approximate unreported emissions, supporting the evaluation of financed climate risks, Scope 3 calculations or policy interventions. The use of hierarchical regression specifications enhances practicality and transparency, utilising only a small number of input variables compared to more complex alternatives, such as machine learning approaches.

### 5.1.2 RQ2: What level of accuracy can be retained whilst minimising user input requirement to produce SME emission estimates?

Following the development of models capable of estimating SME emissions, a selection process is required to determine the specification that both achieves robust, accurate results, whilst remaining practical to deploy. Model results indicate that the inclusion of additional firm-level variables fails to significantly enhance model performance, despite the required increases in complexity and data resources. These diminishing returns to model complexity suggest that a simple, parsimonious solution can reach useful levels of predictive power whilst being practical to apply. In all cases, increasing model complexity yields only marginal gains in explanatory power, with improvements in RSQ of no more than 10% relative to simpler models. A model which incorporates just two user inputs (turnover and industry) is therefore able to predict SME emissions well.

To assess the accuracy of these models it is important to contextualise model performance through out-of-sample testing, such as comparisons against alternative prediction techniques, withheld sample data or, where possible, self-reported data. Given the current regulatory focus on Scope 1 and 2 emissions, there is more opportunity to do so for these emissions than Scope 3. First, the Scope 1 and 2 model prediction accuracy is compared to the Partnership for Carbon Accounting Financials (PCAF) approach which assumes uniform firm-level emission intensities within industries. The findings indicate that the model integrating firm turnover, industry classification, industry emission intensity, and industry skew halves error when compared to the coarse industry-average method for Scope 1 and 2 emissions. The simple application of firm turnover to an industry average therefore overlooks important variation within the data, which can be captured through the modelling framework outlined in Chapter 2.

Further model evaluations include test and train validation, which allows for an unbiased account of model performance and stability. For this, models for all scopes are assessed through an 80/20 train-test split with stratified sampling, applying a 5-fold cross-validation (James et al., 2021). When applied to out-of-sample FTD-based emissions estimates, models return stable RSQ values for all iterations (STD values less than 0.00 for all selected models), whilst median absolute percentage error ranges from 28% to 49% depending on emission assessed.

An additional layer of out-of-sample testing was applied to Scope 1 and 2 estimates, given self-reported data can be accessed for larger firms, within SECR disclosures. A sample of 50 large businesses is collected and used to compare self-reported emissions values to model estimated counterparts. In this sample, annual turnover ranges from £38 million to £211 million, whilst the models specified in Chapter 2 are exclusively trained on SMEs with turnover below £36 million, with 80% of the sample falling under £1 million. Despite this, median absolute percentage error of 25% are returned Scope 1 emissions, whilst error is higher for Scope 2 emissions, with a median absolute percentage error of 65%. Although these results are broadly in line with error levels exhibited in the test and train evaluation, low RSQ scores and high mean errors occur and are largely driven by a small number of extreme outliers.

These findings reveal the high levels of accuracy that can be retained while substantially minimising user input requirements, with a simple modelling approach. A parsimonious model relying primarily on turnover and industry classification consistently delivers robust emissions estimates, demonstrates stable out-of-sample performance, and withstands multiple forms of validation, including benchmarking against industry-average methods and comparisons with self-reported data. While additional complexity yields only marginal gains, the simple model achieves a favourable balance between accuracy and practicality, making it well suited for scalable deployment in real-world SME emissions estimation, both for internal use by firms and for external analysis by third parties.

### 5.1.3 RQ3: What processes are required to ensure FTD is applied to timely emission conversion factors, thus enabling real-time estimates?

Chapter 4 examines processes that produce timely and credible emissions conversion factors for real-time applications, through the development and evaluation of three distinct nowcasting methods. The results show that updating conversion factors through simple inflation adjustment provides a surprisingly strong foundation for real-time use. Method 1, which applies only an inflation adjustment and leaves the MRIO structure unchanged, performs to a similar level of accuracy as the more complex approaches. Across the assessed period, all three methods produce mean average absolute percentage (MAPE) error values for product emission multipliers, ranging from 14.7% percent to 17.2%. Method 2, which incorporates real GDP and trade growth to update the MRIO structure, obtains the lowest MAPE but only marginally. Method 3, which applies additional industry-specific adjustments based on UK-focused data, does not reduce error and in some cases introduces greater volatility. This is especially evident in the technical coefficients, where adjustments to the $\boldsymbol{A}$ matrix results in higher error levels during periods of economic instability (2020-2022).

These findings indicate that inflation-based updates are a reliable, and practical, process for keeping emissions factors aligned with changing price levels and economic conditions.

The performance across years highlights that economic stability has a strong influence on model accuracy. Predicting 2019 data produces notably lower errors, while the disrupted period of 2020 to 2022 produces significantly higher errors. This reflects the effects of COVID-19 on supply chain disturbances and the sharp energy price increases following the Russia-Ukraine war. Attempts to adjust the $\boldsymbol{A}$ matrix are particularly sensitive to such disruption reinforcing the conclusion that simple inflation adjustments are both more practical and more robust for real-time estimation. Differences in the construction of the 2015 base table, which was used to nowcast 2018, also introduced significant performance challenges and resulted in lower accuracy relative to other years.

Drawing these findings together, the chapter concludes that an inflation adjustment represents the most effective and operationally feasible process for updating EEMRIO-based conversion factors for use with FTD. Since the UKMRIO already aligns with ONS inflation statistics, this strengthens the reliability of this approach allowing users to combine inflation-adjusted multipliers or deflated transactional data to generate real-time estimates at annual, quarterly, or monthly frequencies. It also identifies the longer-term need for a fully developed Single-country National Accounts Consistent (SNAC) updating framework to support more accurate and timely projections.

## 5.2 Project Impacts

This PhD was conducted as a joint research collaboration between the Sustainability Research Institute (SRI) at the University of Leeds and Lloyds Banking Group (LBG). The partnership combines academic expertise with industry interests and data assets, demonstrating the value of collaborative research in addressing key sustainability challenges such as SME emission measurement. As a project delivered in direct partnership with a large retail bank, the findings have been widely disseminated within the organisation throughout the research period. The work was overseen by the Managing Director for Portfolio Analytics, Client Products in Lloyds Banking Group's Commercial Banking division, and required monthly presentations to update senior leadership on research progress and emerging results. Ongoing engagement with senior members of the Environmental Sustainability and Business & Commercial Banking teams helped ensure that the research direction remained commercially relevant while also meeting academic standards of rigour, novelty and contribution to existing gaps in the literature. At the start of this research in September 2022, Lloyds' emission estimation programme (hereafter referred to as Carbon Insights) was at a prototype stage. It had been deployed only to a small group of opt-in commercial banking clients, and its approach was not yet aligned with GHG Protocol standards. At this point, Carbon Insights relied on lagged embodied emissions multipliers with no inflation adjustment, and the potential role of FTD in the wider context of SME decarbonisation challenges had not yet been defined.

The immediate impacts of this work included establishing the GHG Protocol as a core criterion for producing business-level emission estimates. This involved mapping internal transaction categories to the appropriate emission Scopes and calculating GHG Protocol-compatible spend-based multipliers capable of estimating Scope 1, 2 and 3 emissions without overlap. These procedures are detailed in the Methods section of Chapters 2 and 3.

A further contribution was the incorporation of an inflation adjustment within the estimation process. Together with Lara Vomfell and Anna Trendl, a process was developed to deflate FTD values on a monthly basis to match the price level of the multiplier year, ensuring consistency between real-time FTD and lagged EEMRIO factors. This represented a practical and necessary step for the operational use of EEMRIO data alongside FTD, justified by the results described in Chapter 4.

To date, Carbon Insights has not yet officially rolled out, as the bank continues to assess the most appropriate implementation pathway. One area under consideration is the potential shift to accountancy data to address the challenge posed by multi-banked clients, whose financial activity is not fully observable through a single bank account. The models developed in Chapters 2 and 3, and operationalised through the CRAN package, were presented to senior leadership as a tool for inclusion on the Lloyds Sustainability webpage. The proposal positioned the tool as a non-intrusive, easy-to-use interface that could provide

SMEs with indicative emissions estimates and link them to relevant green finance offerings. At this stage, this has not been taken up.

Through engagement with external organisations such as the British Business Bank (BBB), the Organisation for Economic Co-operation and Development (OECD), and the Office for National Statistics (ONS), the approach developed in this thesis provided a basis for exchanging ideas on how firm-level FTD can support more effective measurement of SME emissions. With BBB and the OECD, this centred on the shared challenge of better isolating, measuring, and understanding the emissions contribution of SMEs, where firm-level FTD offers potential to strengthen existing approaches. In particular, the work highlighted limitations in methods that rely on gross industry-average emissions intensities, which are often shaped disproportionately by the largest firms within each industry. For the ONS, the research demonstrated how Lloyds' FTD can be combined with official economic and environmental indices to support more detailed and policy-relevant analysis of business emissions.

Work from this thesis was also accepted for presentation at various UK and international conferences, showcasing the interest amongst varying research contexts. Preliminary results for Chapter 4 were presented at the 29th International Input-Output Association Conference in Alghero, Sardinia (2023). As this conference took place early in my PhD, it provided a valuable opportunity to meet leading figures in the input-output community and to reflect on how their work related to my research. Combining transaction data with input-output analysis attracted significant interest, with the data asset seen as a real novelty. Although my own results were still preliminary and I had not yet engaged fully with the FTD, discussions with researchers were highly productive. Speaking with Konstantin Stadler, one of the developers of EXIOBASE, allowed me to explore different approaches to nowcasting. I also met with Timothé Beaufils, who had recently published the SPIN method (Beaufils and Wenz, 2022), and our conversations directly influenced my decision to apply this method in Chapter 4. This paper stood out to me as a highly practical approach to nowcasting the UKMRIO, making effective use of exogenous data while remaining firmly grounded in input-output principles. It was unsurprising that the original SPIN paper went on to receive the 2023 Sir Richard Stone Award for the best publication in Economic Systems Research. I also had the chance to connect with Sanjiv Mahajan, who is Head of Methods and Research Engagement at the ONS and was President of the International Input-Output Association at the time.

Chapters 2 and 3 were then presented at conferences of various themes. First at the Corporate Responsibility Research Conference (CRRC) in Leeds, UK (2024), where the conference's qualitative emphasis offered valuable new perspectives, while also providing space for my empirical work to add a complementary dimension. Here my presentation was held in the 'Corporate Social Responsibility and Sustainability for SMEs' session, chaired by Richard Bull and Ana Rita Domingues, who have conducted SME focused research which helped guide the use cases developed in the thesis (Mazhar et al., 2024). This conference also provided an opportunity to discuss my work with Alice Owen, who presented on differentiated emissions policies for SMEs. These conversations helped reinforce the relevance of developing practical, data-driven approaches for SME emissions estimation, and highlighted how the methods presented in this thesis can support more targeted and equitable policy design.

Second, Chapter 2 was presented at the Sustainable Consumption Research and Action Initiative (SCORAI) conference on Mainstreaming Sustainable Consumption in Lund, Sweden (2025). This aligned with the themes of this chapter, which demonstrates how intrinsically collected data can be used to generate low-input emissions insights for SMEs. In Lund I was able to connect with the David Andersson, the Chief Executive Officer of Svalna, who's research papers formed a strong basis for the beginning research period of my PhD (Andersson, 2020; Barendregt et al., 2020; Svalna, 2022). I was also questioned on the implications of this work for Scope 3 emissions and was able to explain that this is the focus of my current research (now Chapter 3. Phillpotts et al. 2026). The keynote talks, particularly those by Manisha Anantharaman and Tim Jackson, were not directly related to my empirical work, but were influential in helping me view my research through a broader sustainability lens and reflect on how it fits within the complex landscape of 'Sustainability'. On a personal level, it was both nerve-racking and exciting to learn that my presentation would be delivered on the main stage, immediately before the conference's closing remarks.

Finally, but by no means least, I presented Chapter 3 in the 12th International Conference on Industrial Ecology, in Singapore (2025). This conference felt like a natural point of reflection, as I was finalising my empirical results and beginning to assemble this thesis with the aim of submitting in late 2025. Its theme of interconnectivity resonated strongly with my work, which sits at the intersection of academia and industry, by focusing on supporting SMEs through collaboration between a university and a major bank. I was able to reconnect with familiar faces and meet new researchers at various stages of their careers. A career workshop with Edgar Hertwich provided an opportunity to ask questions and receive guidance on navigating the field. My work also attracted interest from conference sponsor FootprintLab, whose Co-Founder, Tim Bayes, highlighted its relevance to the Australian context, where compulsory Scope 3 disclosures are being introduced into the regulatory framework.

To date, Chapter 2. Bridging the SME Reporting Gap: A new model for predicting Scope 1 and 2 emissions (Phillpotts et al., 2025) and Chapter 3. Completing the SME Carbon Profile: Scalable Prediction of Scope 3 Emissions (Phillpotts et al., 2026) are open-access papers published in the Journal of Industrial Ecology. The Journal of Industrial Ecology has an acceptance rate of 31% and an impact factor of 5.4. In addition to these research paper contributions, the models developed in Chapter's 2 and 3 have been combined and published on CRAN, the primary package repository in the R community, to enhance accessibility and reproducibility. To date this package has been downloaded 250 times, whilst the code for the nowcasting procedures has also been made public via Zendo (Phillpotts, A., 2026).

Chapter 4 has been prepared as an academic paper ready for submission to *Economic Systems Research* (acceptance rate 9%; impact factor 1.6). The chapter tackles a widely cited limitation in the EEMRIO literature, which presents as a significant barrier to the implementation of EEMRIOs in real-time applications (CCC, 2022).

## 5.3 Contribution to Knowledge Base

This research draws on FTD from a large sample of UK SMEs, providing a unique opportunity for large-scale analysis of an otherwise difficult-to-observe segment of the economy. Owing to this novelty, the methods, results, and interpretations presented in the first two empirical chapters offer original contributions to different areas within the existing literature. Chapter 4, while not directly using FTD, addresses one of the central limitations of applying EEMRIO models to real-time data. This is a widely recognised challenge that remains underexplored in the current literature.

### 5.3.1 Methodological Contributions

Methodologically, this thesis sits within three areas of literature. First, it advances methods for estimating SME emissions using financial transaction data (FTD); second, it contributes to modelling approaches for SME emissions at scale; and third, it assesses nowcasting frameworks for EEMRIO databases.

#### 5.3.1.1 Estimating SME emissions with FTD

This thesis makes a methodological and practical contribution to supporting the measurement of SME emissions using FTD. While the GHG Protocol recognises expenditure-based estimation (WRI & WBCSD, 2004), no standardised procedures currently exists for applying it at the firm level, nor are expenditure-based emission multipliers officially published for the UK that are compatible with the Protocol. This research therefore develops a transparent, standardised, and replicable process for producing expenditure-based emission estimates consistent with the Protocol's Scope definitions. The approach derives unit spend emissions multipliers from official UK datasets, making use of government conversion factors, energy prices, and fuel-use statistics. This is a replicable process that can be adapted for other regions given equivalent data.

Beyond the estimation process, the thesis also builds the argument for FTD as a viable basis for emission estimation among smaller actors - an idea previously explored for households (Trendl et al., 2023) but not yet applied to SMEs. By enabling automated, high-resolution insights into SME emissions, the approach has the potential to support SME disclosure practices and enhance financial-sector applications such as financed-emission accounting and climate-risk assessment, while also providing external stakeholders with more accurate data to enable their Scope 3 emissions measurement.

This approaches contribution to enabling the mainstreaming Scope 3 emissions provides an additional layer of benefit, where methodology complexity and costs surrounding these emissions presents a current substantial barrier to measurement (Acquaye et al., 2014; Cheema-Fox et al., 2021; Hettler and Graf-Vlachy, 2024; Isil and Sebastianelli, 2020). Further to this, differences in methodological approaches undermine data comparability (Robinson et al., 2015). For SMEs, the sole reliance on costly, data-intensive micro-analysis poses a major barrier to the widespread adoption of Scope 3 emissions measurement. In contrast, FTD-based calculations offer a scalable alternative that overcomes these structural barriers while ensuring standardised and transparent data. This approach enables further analysis of firm-level spending patterns to identify Scope 3 emissions hotspots across industries, providing valuable insights for the SME and policymakers.

#### 5.3.1.2 Modelling SME emissions

The models developed in this thesis differ fundamentally from existing estimation approaches in the literature due to the nature of the underlying data. Whereas most previous models rely on self-reported emissions (Assael et al., 2023; CDP, 2020; Goldhammer et al., 2017; Griffin et al., 2017; Heurtebize et al., 2022a; Nguyen et al., 2023, 2021; Serafeim and Velez Caicedo, 2022; Shakdwipee and Lee, 2016; Wang and Ye, 2025), this research uses FTD-based emission estimates. This addresses several key limitations of self-reported data, which are prone to reporting errors (Financial Reporting Council, 2021), and vary in methodology across firms, ultimately undermining comparability (Dragomir, 2012; Talbot and Boiral, 2013). By contrast, FTD-based estimation ensures that emissions for all firms in the sample are calculated using a standardised and transparent process, as detailed in Chapters 2 and 3. This consistency enhances the comparability of data, assisting for the inherent limitation of using self-reported data, thus allowing more robust benchmarking across firms and sectors.

A further contribution is the impact of the removal of the dependency on self-disclosures on sample inclusion. The FTD approach enables the inclusion of firms that traditionally fall outside existing emission datasets, expanding both the size and representativeness of the sample: the only prerequisite for inclusion is that a firm holds and uses a transaction account with the participating financial institution (LBG). Previous models largely exclude SMEs due to their limited disclosure activity, whilst the models developed in this thesis are trained entirely on SME data, uniquely tailoring them to the SME context and significantly improving the coverage of emission insights for this crucial yet underrepresented segment of the economy. SMEs stand to benefit most from accessible and efficient emissions estimation methods (Goldhammer et al., 2017), given their financial, time and knowledge limitations (Assael et al., 2023), whilst external stakeholders benefit from accessing these emission estimates given the lack of alternative data on SME emissions (Deloitte, 2024).

By deploying an OLS regression framework, the approach taken to produce the models in this thesis are comparable to previous models. Goldhammer et al. (2017) use the joint Scope 1 and 2 emissions of 93 European companies sourced from self-reported datasets to produce a regression-based procedure, requiring predictors sourced from their financials. Here the authors show model performance gains when treating different industries individually, with individual regressions or a single regression with industry dummy variables performing greater than a pooled alternative. The authors also make attempts to further describe intra industry firm heterogeneity aside from size, with metrics such as vertical integration and capital intensity. Whilst benefitting from being standardised metrics that can be procured from company financials, these variables regularly returned non-significant coefficients.

The models described in this thesis use a hierarchical regression framework that incorporates industry-level variables as an essential part of the modelling process. This approach allows firms within the same industry to be modelled together, while still recognising differences between sectors. The hierarchical framework shares information across industries, meaning that sectors with fewer firms can still benefit from patterns observed elsewhere, ensuring stability for smaller sectors. With a dataset of over 30,000 SMEs, this partial pooling produces more accurate and generalisable results (Gelman and Hill, 2006), balancing firm-level variation with the stability of industry-level trends. In contrast to Goldhammer et al. (2017), the models developed in this thesis found capital intensity to be statistically significant, yet coefficients were small, and the improvement in model fit was modest.

The machine learning approaches developed in the literature are intended to train models on as much information as possible to maximise the accuracy of predictions. Whilst this is beneficial for the use case of extrapolating emissions estimates in emission datasets (Nguyen et al., 2021), simpler models achieve both a low user burden and accurate predictions when compared to more complex alternatives, matching the use case of the thesis.

#### 5.3.1.3 Nowcasting EEMRIO databases

This thesis then addresses one of the most persistent challenges facing EEMRIO users: the multi-year publication lag. This delay limits the operational use of EEMRIOs for real-time applications. To explore methods for producing timelier emission factors, Chapter 4 systematically evaluates a set of methods, ranging from a minimal inflation-only adjustment through to region- and industry-specific updates. By testing and comparing these methods, the chapter shows how differences in data requirements and methodological complexity influence nowcasting accuracy.

To project base-year tables, Chapter 4 employs an adapted version of the SPIN method (Beaufils and Wenz, 2022). In the original paper, the authors use the SPIN method to explore alternative, hypothetical scenarios following the 2008 financial crisis. To the best of my knowledge, this thesis presents the first attempt to reproduce the SPIN method specifically to compare projections against known MRIO tables. Using this technique in a nowcasting context is valuable because it emphasises operational simplicity while remaining faithful to core input-output principles. This makes it suitable for use by non-technical practitioners and is particularly relevant for mainstreaming EEMRIOs in contexts such as SME Scope 3 emission estimation.

Chapter 4 also establishes an error-quantification framework that evaluates the performance of each projection approach over a known historical window (2018 to 2022). This illustrates how methodological choices, and changing annual contexts, affect the stability of key EEMRIO components ($\boldsymbol{f}, \boldsymbol{A}$ and $\boldsymbol{eL^{pp}}$). At present, the literature contains no published assessment of the error and uncertainty involved in using nowcasted MRIO tables to estimate environmental impacts or generate output multipliers. For nowcasted EEMRIOs to be adopted in official or policy settings, this uncertainty must be understood, communicated, and eventually reduced.

Finally, the chapter provides practical guidance for the operational use of EEMRIOs, particularly in the context of SME Scope 3 emissions. The finding that simple inflation-based adjustments perform comparably to more complex projection methods has important implications. It suggests that for many real-time applications, robust price adjustments may be more effective than attempting to replicate detailed structural change. This insight provides a foundation for integrating MRIO-based emission factors with high-frequency datasets such as FTD, which can support the wider adoption of EEMRIOs in real-time emissions analysis.

### 5.3.2 Practical Implications

The outputs of this thesis are primarily designed to address the persistent under-engagement of SMEs in emission measurement (BBB, 2024), which represents both a regulatory and practical omission within current emission accounting frameworks. SMEs make up the majority of the business population and collectively contribute a significant amount to total business emissions, making their engagement and action crucial to achieving any decarbonisation targets. However, individually, SME contributions are often minimal, whilst their exposure to additional regulatory pressures are higher than larger firms (Oluleye et al., 2025). Extending reporting obligations to SMEs therefore presents a complex challenge.

The models outlined in this thesis, and operationalised into an SME emission prediction tool, addresses the structural barriers faced by SMEs such as limited funding, resources, and expertise (Caldera et al., 2019; Menon and Ravi, 2021), providing a starting point for SMEs to engage with strategies to reduce their associated emissions, whilst removing the burden of emission measurement from the business. The tool, currently available on CRAN, would gain greatest credibility if distributed through financial institutions with academic validation, allowing SMEs to use trusted channels while avoiding both the burden of emissions measurement and the confusion created by the wide and inconsistent range of private carbon calculators. This would also create opportunities for FIs to promote targeted green financing products that align with each firm's emissions profile make-up, offering the most relevant and impactful improvements. Whilst affordability isn't impacted, connecting relevant products to SMEs overcomes some barriers to accessing external funding (Appiah-Kubi et al., 2024b). When adopted, green financing has been shown to generate energy efficiency gains, cost savings, and enhanced environmental disclosure, as firms seek both to demonstrate efforts and to track improvements over time (Bouchmel et al., 2024; Liu et al., 2022).

Beyond the direct use to SMEs, external stakeholders also have a growing demand for this type of emissions data. Large corporations require reliable information on their suppliers' Scope 1 and 2 emissions to calculate their own Scope 3 footprints. Likewise, financial institutions depend on emissions data to calculate financed emissions and climate-related financial risks across their portfolios. At present SME data is missing. Providing consistent, high-quality emissions data at the SME level therefore supports wider transparency and accountability within both supply-chain and financial-sector disclosure frameworks.

Finally, this thesis has highlighted the missed opportunity created by the absence of a centralised repository for SECR disclosures. Consolidating these reports into a single structured dataset would provide a valuable resource for analysts, researchers, and policymakers, enabling more robust understanding of firm-level emissions and sectoral trends. It is therefore recommended that, going forward, SECR disclosures be provided both within annual accounts and through submission to a central, accessible database, in order to maximise their analytical and policy value.

### 5.3.3. Policy Implications

#### 5.3.3.1 Spend-based methods

FTD provides a new basis for data-driven insights that were previously impossible to obtain at scale. Spending patterns can reveal valuable proxies for energy use and fuel consumption, the identification of spending patterns, and upstream emissions drivers. In this thesis, by capturing real financial activity across thousands of firms, FTD has enabled the identification of supply-chain emissions drivers with far greater efficiency than traditional methods. These insights can directly inform net zero strategies by highlighting where emissions are generated and where interventions are likely to have the greatest impact. Our analysis shows that upstream emissions are highly concentrated and overlapping across industries, revealing shared hotspots that can act as policy leverage points (Meadows, 1999), where modest but well-coordinated interventions could materially reduce overall emissions (Science Based Targets initiative, 2018).

This thesis also demonstrates that FTD can provide a credible and scalable means of filling this SME data gap. By leveraging routinely collected financial data already held by financial institutions, FTD-based estimation enables automated and standardised emissions calculations without imposing new reporting burdens on firms. The untapped potential of FTD as a low burden method of generating automated emissions data is recognised by initiatives such as the UK focused Bankers for Net Zero (2022).

However, the final year of this research, 2025, saw the demise of the Net Zero Banking Alliance, a global coalition of financial institutions, following a wave of departures that exposed growing divergence in global attitudes toward coordinated climate action. This began in response to the shifting political environment linked to the 2024 US election with six major US banks leaving. These exits triggered subsequent UK departures; HSBC stepped back to pursue its own transition strategy, while Barclays argued that the alliance “no longer has the membership to support our transition” after most global peers had already withdrawn (Messenger, 2025). In this context, there is both an opportunity and a clear need for UK policy to provide clearer guidance on how banks can best use their data assets to support the net zero transition in a way that aligns with national strategy.

In the UK, Scope 1 and 2 emission factors are published only in activity-based form. Activity data is, and remains, the preferred basis for Scope 1 and 2 reporting under SECR for larger firms, but this approach is far less practical for SMEs, who lack the resources to collect detailed operational data and conduct accurate emissions calculations. Enabling SMEs to calculate emissions using spend data removes a significant burden and increases the likelihood of meaningful engagement. It would be a positive action to include the GHG Protocol aligned spend-based multipliers, derived in this thesis, in the annual emission multiplier dataset (BEIS and DEFRA, 2022)

UKMRIO emission factors are also not currently provided in a form that aligns with the GHG Protocol or incorporates transparent inflation adjustments to address the time lag in MRIO publication. To mainstream the use of input–output analysis for Scope 3 calculations, UK-focused EEIO multipliers derived from the UKMRIO should be made available through an accessible interface designed specifically to support Scope 3 estimation. Establishing a nationally endorsed, regularly updated set of UK spend-based multipliers, supported by clear methodological guidance, would enable firms to produce emissions estimates that are consistent, credible, and aligned with national policy expectations. Recent tools such as Climatiq, which offers EXIOBASE-derived multipliers via API, demonstrate both the strong demand for simple spend-based factors and the feasibility of delivering them in a user-friendly, well-documented format (Climatiq, 2025).

#### 5.3.3.2 External Prediction Models

The models developed in this thesis offer a non-intrusive way to harness the emissions insights that can be derived from FTD. By aggregating micro-level transaction data into sector-level insights, the approach avoids breaching privacy boundaries, whilst converting what would otherwise remain private data into publicly usable insights.

These models provide a practical mechanism for larger companies to improve Scope 3 emissions reduction schemes. Although these models do not replace the use of EEIO multipliers for Scope 3 estimation, they offer a useful starting point for supplier engagement between larger firms and SMEs. By providing indicative emissions estimates based on simple information such as industry classification and turnover, the models allow reporting firms to identify where engagement may be most effective without requiring SMEs to generate detailed emissions data. This reduces the reporting burden placed on SMEs and lowers the risk that smaller suppliers are disadvantaged or excluded from value chains due to limited disclosure capacity.

The models in this thesis also have direct application for financial institutions reporting financed emissions under the PCAF. Banks often struggle to estimate emissions for SME borrowers, since most do not disclose primary emissions data, resulting in reliance on low-quality estimates. The models presented here offer a practical alternative, that improves on the accuracy of estimates by up to 50% (Phillpotts et al., 2025).

Recent developments within PCAF reinforce the value of such approaches. In 2025, PCAF integrated the Comprehensive Environmental Data Archive (CEDA) into its database to expand access to standardised proxy emission factors. This shift highlights the growing need for credible and transparent estimation methods when primary emissions data are unavailable. The SME models developed in this thesis complement this direction by offering an additional data source tailored to smaller firms, a group that remains largely unrepresented in existing emissions datasets (PCAF, 2025).

#### 5.3.3.3 Nowcasted EEMRIO

The nowcasting procedures developed in this thesis begin to address the multi-year publication lag associated with EEMRIO databases, a limitation that has been explicitly highlighted by the UK Climate Change Committee (CCC). Here, the Committee stressed that greater investment in the collection and reporting of consumption-based emissions is needed to ensure statistics are produced on a timely, annual basis. Without such improvements, the UK's consumption-based accounts will continue to rely on lagged data and may fail to capture genuine, real-time reductions in the nation's carbon footprint (CCC, 2022. p494). By demonstrating that relatively simple adjustments, such as inflation-based updates to MRIO-derived emission factors, can achieve accuracy levels close to more complex projection techniques, this thesis offers a practical route for producing interim updates to the UKMRIO.

The relevance of this work becomes especially clear in the context of the UK's forthcoming Carbon Border Adjustment Mechanism (CBAM), scheduled to begin phasing in from 2027 (HM Treasury and HM Revenue & Customs, 2024). An effective CBAM requires accurate and up-to-date estimates of the embodied emissions in imported products, yet current MRIO data lags limit the UK's capacity to assess these impacts. Timelier MRIO insights would allow policymakers to better understand how domestic decarbonisation interacts with international trade patterns at a time when more jurisdictions are introducing similar border-carbon measures.

## 5.4 Limitations and Challenges

### 5.4.1 Validity of Emission Estimates

FTD-based emission estimates provide a simplified representation of firms' emissions, and their validation is difficult due to the limited availability of alternative data. Whilst the methodology used in this thesis is transparent and theoretically grounded, the lack of comparable datasets introduces uncertainty around the accuracy of FTD-based estimates. For Scope 1 and 2 emissions, the validation challenge is more manageable. These emissions relate to direct operational activities, such as fuel use and electricity consumption, which are typically captured through metered data, billing records, or other quantifiable sources. While self-disclosed Scope 1 and 2 data, including SECR submissions, have restricted coverage, may contain errors, and often require manual compilation, they nonetheless provide an identifiable reference point for indicative comparison.

Scope 3 presents a different limitation. The issue is not only that data coverage is limited, but that comparable alternative data is largely absent. Scope 3 emissions encompass wide-ranging and complex value-chain activities, and variations in category boundaries, supplier assumptions, and emission factors mean that firms apply varied methodological approaches. This introduces substantial subjectivity into existing Scope 3 assessments. Consequently, disclosed Scope 3 figures cannot be treated as empirical truths or definitive benchmarks, as they reflect methodological choices as much as underlying activity.

## 5.4.2 Coverage of the Model

### 5.4.2.1 Industry Coverage

While the modelling approaches presented within this thesis improves accessibility, some industries are excluded due to limited sample sizes or unreliable turnover estimates. The method is also less suitable for sectors such as agriculture, where a large share of emissions arises from process-based activities rather than fuel combustion and is therefore not captured in FTD. A full list of excluded SIC codes and the implication of their removal is provided in Table S2 of Supporting Information S2.1. To assess the impact of exclusions, business population data (BEIS, 2022) is compared to the prediction model coverage, identifying the SIC codes that have the highest exclusion impact. The model produces complete emissions profiles for over 71% of UK SME population, with scope to expand this further if trade-offs in accuracy are acceptable. Of the excluded SMEs, 25% are excluded due to our initial exclusions, with an additional 4% excluded because of sample size within our restriction process. Of the SIC codes excluded from the analysis, most account for only a small proportion of the UK's SME population. The main exceptions are agriculture and service sectors such as insurance, real estate, legal, and accounting services, which together contain approximately 5% and 16% of UK SMEs respectively. Industries with distinctive emissions profiles, such as agriculture, require specialised estimation tools to account for their unique production processes and resource use. In contrast, the service-sector SMEs, were excluded due to the frequent occurrence of non-turnover credit flows, which cause a systematic overestimation of turnover when calculated from FTD.

### 5.4.2.2 Emission Coverage

As mentioned previously, FTD-based emission estimates capture a simplified version of emissions. For Scope 1 and 2 emissions, switching away from physical activity data means monetary units are used to infer consumption quantities, and their subsequent emissions. This is particularly problematic in electricity purchases due to energy tariff bands as businesses consuming more energy typically pay less on a per-unit basis. Adjustments are therefore required to account for this.

Additionally, FTD lack the granularity of receipt-level analysis. Emission conversion factors are therefore applied at the merchant category level only (Trendl et al., 2023). This means transaction-based emission estimation relies on assumptions regarding energy mix, since many energy suppliers provide both natural gas and electricity, and petrol stations sell both diesel and petrol. Consequently, a single conversion factor produces emissions from petrol station purchases, whilst all electricity purchases are converted on a location-basis (national grid intensity factor), regardless of tariff. In addition, models only capture the Scope 1 and 2 emissions generated from a firm's energy-use, mirroring what is required of larger firms under the SECR framework. This means emissions that occur from a firm's process are not included. It is for this reason that industries like agriculture, water and waste are not included.

For the estimation of Scope 3 emissions, the UKMRIO is used to calculate selected Scope 3 categories. This approach assumes that the supply structure of each purchase can be approximated by the corresponding economic sector as represented in the UK Supply and Use Tables (ONS, 2022; Schmidt et al., 2022). In this study, emissions are modelled for five of the fifteen Scope 3 categories, excluding several major sources such as the use of sold products and capital goods.

#### 5.4.2.3 Regional and Longitudinal Coverage

This analysis reflects a one-year snapshot, aligning with practical deployment. Future iterations of the model can be easily updated, with new FTD and emissions conversion factors. In addition, the model is developed with UK data, its applicability is therefore limited to non-cash-based, digital economies with comprehensive FTD. SME activity is often shaped by regional factors, model development for other economies requires changes to conversion factors, currencies, grid intensities, and other differences. This presents an opportunity for collaborations between policymakers and financial institutions in other regions.

### 5.4.3 Nowcasting the UKMRIO

Several limitations help explain the error levels observed across the nowcasting methods of Chapter 4. First, the analysis does not incorporate additional GHG emissions data, instead modelling emissions changes through relationships with macroeconomic growth rates. Although supplementary datasets exist, they are not available at the level of aggregation required for the UKMRIO and are published far less frequently than macroeconomic indicators such as GDP. In principle, the most recent UK environmental accounts could be added, but doing so would still leave two missing years to estimate, limiting the practical benefit. A related limitation is that industry-level emissions are allocated using base-year shares. This is a pragmatic choice in the absence of consistent, up-to-date emissions or energy data, but it weakens accuracy in sectors with volatile emissions. For example, aviation emissions fell sharply in 2020 and 2021, yet the method retains the base-year distribution, leaving this reduction unrepresented.

A further limitation stems from the need to convert the UKMRIO from its Supply and Use Table structure into a symmetric input–output table before nowcasting. This transformation inevitably introduces its own assumptions and potential distortions, which may take away from the SNAC focus of the UKMRIO.

## 5.5 Future Research Directions

### 5.5.1 Addressing the Limitations

There are some clear paths to overcoming some of the key limitations mentioned above. First, comparisons of FTD-based emission estimates against alternative firm-level microdata is entirely possible. For Scope 1 and 2 emissions, potential comparable emissions data include self-disclosed reports such as those submitted under the SECR framework. However, these are constrained by limited sample sizes, business size thresholds, and the need for manual data collection.

Survey data, such as the Annual Business Survey (ONS, 2024), offers a path to comparing Scope 3 estimates. Trendl et al. (2023) conducts a comparison between household FTD and household survey-based estimates, finding FTD serves as a credible alternative, with a broad agreement between the two data sources across key emission areas and demographic characteristics. An equivalent validation in the commercial sector remains unexplored and would address this gap, helping to clarify the current uncertainties around the validity of FTD Scope 3 emission estimation.

The most informative validation exercises would compare emissions estimates derived from FTD, accountancy records, physical activity data, and survey-based approaches in parallel. This would allow the differences between methods to be identified and quantified. Such a study, however, would require active participation from SMEs to supply their accountancy and activity data, cooperation from the financial institution providing FTD, and access to the relevant business survey.

Related to this point, the inclusion of accountancy data in this could allow for the inclusion of some service-sector firms, such as those in insurance, real estate, and legal or accounting services, which were excluded due to uncertainty in their FTD-based turnover estimate, from frequent handling of non-turnover credit lines (e.g., deposits) leading to systematic overestimation of turnover. In Chapters 2 and 3, these exclusions accounted for 16% the SME population uncaptured by the model. In principle, these firms could be incorporated if turnover were measured using accountancy-based figures rather than FTD credit lines. This was not feasible within the scope of the present study as accountancy data are not routinely collected by financial institutions.

Other industries, such as agriculture, energy and water, were excluded because they involve substantial sector-specific emission sources that would require tailored modelling approaches. Although many of these sectors have relatively small SME populations, some contain a significant number of SMEs. Agriculture is the most notable example, with 5% of all SMEs operating in this sector (BEIS, 2022). Wells et al. (2025) demonstrate how a sector-specific approach can be developed for this industry, using a hierarchical framework similar to the methods used in this thesis. Their model, which incorporates variables such as pigs, sheep, dairy cows, fuel spend, cattle and farm size (expressed relative to turnover), achieves an RSQ of 0.91 in explaining variation in emission intensity across farm types.

Moreover, to address the categories of Scope 3 emissions uncaptured in Chapter 4, a benchmarking approach has been adopted using data from the CDP corporate disclosure database (Buchenau et al., 2025). However, this adjustment relies on self-reported figures, which are subject to well-documented issues of completeness, representativeness, and methodological inconsistency. Modelling frameworks for quantifying downstream Scope 3 impacts would make a valuable contribution.

Finally, future nowcasting focused research should focus on developing a SNAC framework for EEMRIOs (like the UKMRIO). The approaches used in this thesis replicated assumptions currently employed, which operate for global MRIOs, which optimise accuracy across regions rather than prioritising national relevance. A dedicated SNAC method would instead draw directly on UK-specific data sources, including national accounts, sectoral statistics, trade data and environmental accounts, ensuring projections are firmly aligned with the domestic economy and more useful for policy analysis. Importantly, future efforts should aim to nowcast the UKMRIO in its original SUT form rather than converting to a symmetric IO table, since this conversion introduces additional assumptions and potential distortions. A SNAC-based SUT nowcasting approach would better preserve accounting consistency and provide more robust foundations for national-level consumption-based emissions monitoring. In the UK, Supply and Use Tables are constructed from 349 individual data sources, reflecting the depth and detail of the underlying system (ONS, 2025c). A future nowcasted iteration of the UKMRIO could examine how many of these sources are reported more frequently and whether they can be used to construct preliminary, nowcasted SUTs during the lagged period.

### 5.5.2 Wider Research Application

There are also several wider research opportunities that follow directly from this thesis. The potential applications of FTD are extensive, given its unprecedented coverage and the limited academic literature that currently exists on its use for emissions estimation.

One valuable direction is the development of mixed-methods studies that combine the vast, quantitative insights (exemplified in this thesis), with qualitative engagement from firms. This would require direct engagement between the financial institution and participating firms, which may increase the cost and time required to conduct the study. However, this would enable the creation of a voluntary sample of businesses who can provide contextual insights, beyond the FTD estimations, through interviews or surveys. Researchers can then assess both the empirical accuracy of FTD-derived estimates and the perspectives of firms regarding their credibility and usefulness. Such work could also identify how businesses interpret the emission information derived from FTD and where they perceive uncertainty or misalignment with operational realities. Related to this, future work could examine how the FTD-based emissions tool developed in this thesis should be designed to maximise SME engagement. A trial phase involving a sample of firms could explore preferred formats, levels of detail and types of guidance that transform emissions estimates into actionable next steps. This would also provide an opportunity to understand behavioural and organisational barriers to decarbonisation and to test whether the provision of automated estimates influences decision making in practice.

A further research avenue is the integration of FTD-based estimates with existing Scope 3 accounting practices. Although FTD does not replace EEMRIO methods for larger corporate emissions reporting, it can complement these frameworks by offering supplier-specific indicators that refine hotspot identification, supplier screening and uncertainty assessment. Future studies could investigate how FTD-derived insights align with or diverge from spend-based EEIO estimates and whether hybrid approaches could improve the reliability or granularity of value-chain emissions assessments. There is also significant potential to use FTD to disaggregate industry-level emissions between SMEs and larger firms. Current environmental accounts typically report emissions at broad industry levels without distinguishing contributions from firms of different sizes. Aggregated FTD-based emissions profiles, combined with structural business statistics or environmental accounts, could enable researchers to estimate size-specific emission intensities and to understand how emissions are distributed across firm types within industries. This would refine assumptions used in EEMRIO modelling.

Finally, the sectoral detail available in FTD creates opportunities for industry-specific research. Future studies could focus on a single sector, such as manufacturing, construction or accommodation and food services, to examine how spending behaviours, supply-chain structures and emissions profiles vary within industries. Such sector-focused analyses could reveal heterogeneity that is obscured in aggregate models and would provide insights directly relevant to industry-level policy design.

## 5.6 Smaller Actors within a Fragmenting Net Zero Context

To sustain momentum, environmental policies should be grounded in realistic, evidence-driven outcomes that consider impact, operational feasibility, and the wider economic context. Globally, and within the UK, the long-standing consensus on net zero and climate policy has begun to fracture. This period of political agreement may have fostered a degree of complacency, leaving certain policies vulnerable to criticism. Legitimate concerns have been raised about the offshoring of carbon-intensive industries, causing significant job losses, and impacted both the economic stability of local communities and vulnerability to external global shocks. The failure to anticipate and manage these transitions effectively is concerning (CCC, 2025). As global emissions continue to increase, these impacts undermine the integrity of the UK's net zero commitments and risks eroding public confidence, support, and engagement with the transition.

The findings of this thesis demonstrate that emission measurement and management can be practically achievable through FTD, and subsequent models. By applying transaction-based methods, emissions can be transparently and consistently estimated without imposing excessive reporting burdens on SMEs. This creates a pragmatic middle ground which supports engagement, without overwhelming SMEs, helping to maintain participation in the transition.

The SME population, targeted in this thesis, will play a vital role in the collective transition toward a more sustainable future. A future where the global economy stops pushing the plant to its environmental limits. The individual choices of smaller actors aggregate into wider outcomes, meaning that small-scale action replicated across millions of firms, can have a substantial cumulative impact. The provision of emissions estimates through the models in this thesis offer SMEs a clearer starting point for taking practical decarbonisation actions. For SMEs, Scope 1 and 2 emissions can be reduced through practical steps such as the electrification of machinery and vehicle fleets, the adoption of more efficient logistics and operations, and investment in modern, lower-emission technologies. For Scope 3 emissions, the insights provided by the model can help SMEs reflect on procurement choices, pursue low-carbon or locally produced alternatives, strengthen waste management practices, and encourage lower-emission employee behaviours such as public transport use, cycling or car-sharing. In this way, modelled estimates can support SMEs in understanding where emissions arise within their activities and where targeted interventions may be most effective.

It is, however, important to acknowledge the constraints many SMEs face. In addition to heightened economic vulnerabilities, many have limited capacity or incentive to invest in long-term decarbonisation measures. SMEs operating out of rented or shared premises lack the autonomy to install solar panels, retrofit buildings, or replace heating systems with heat pumps (Mazhar et al., 2024). Similarly, SMEs face barriers to fleet electrification due to high upfront costs. These constraints mean that while SMEs can make incremental improvements, their potential to decarbonise independently remains limited without support.

The effectiveness of SME decarbonisation therefore depends on the systemic enablers that determine whether sustainable options are accessible, affordable, and reliable (Pinkse et al., 2024). Smaller actors operate within economic and institutional systems that strongly shape the boundaries of their choices. Among the most critical enabler is access to low-cost low-carbon energy. The economy can be understood as a system that transforms energy into goods, services, and waste, a view grounded in ecological economics (Daly, 1996; Georgescu-Roegen, 1971). In the case of SMEs, livelihoods are closely tied to the continuity and viability of these firms, making energy accessibility and competitiveness important. Efforts to accelerate grid decarbonisation must be designed with care, given SMEs face higher economic and regulatory exposure than larger businesses (Oluleye et al., 2025).

The CCC identifies lowering the cost of clean electricity as the single most important action for decarbonising the UK economy (CCC, 2025). Yet, paradoxically, high electricity prices are often cited as one of the main arguments against renewable grid expansion (Holton and Twidale, 2025). Addressing this tension requires coordinated policy action to both manage energy prices and improve transparency around the true costs and benefits of the energy transition. Doing so is essential to sustain public and business confidence in renewable energy investment.

Recent evidence underscores the broader economic benefit of renewables: O'Shea et al. (2025) estimate a net benefit of £104.3 billion to UK consumers from wind power investment between 2010 and 2023. Such findings should be widely communicated and used to expose and correct systemic inequities in the energy system, where, as the authors note, "natural gas users, who pay nothing towards wind investment, have enjoyed 82% of the benefits". In addition, grid capacity and infrastructure, and access to finance are equally important enablers. Without sufficient grid capacity investment, the electrification of heat, transport, and industrial processes will stall (Butler et al., 2023). Similarly, affordable financing mechanisms and tailored green loan products are essential for helping SMEs overcome upfront cost barriers and invest in energy efficiency or renewable technologies (Oyewole et al., 2024).

The role of SMEs in the UK must also be understood within a global context. Although the UK now contributes only a small fraction of global territorial emissions, it carries a disproportionate historical responsibility as one of the earliest industrialised economies and remains among the world's top five historical emitters (Carbon Brief, 2023). Furthermore, emissions measured on a consumption basis and adjusted for population reveal the UK's per capita footprint is approximately 65% higher than the global average (Ritchie et al., 2023). The UK's territorial emission reductions should therefore not be seen as the sole focus of its climate strategy. As domestic emissions fall, a growing share of the UK's carbon footprint is embedded in imported goods and services, highlighting the need to address patterns of overconsumption alongside production-based measures (Barrett et al., 2013; DEFRA, 2025a). It is also essential to address not only domestic emissions, but the wider environmental impacts embedded in global supply chains, ensuring that the UK's progress toward net zero is not achieved at the expense of increasing emissions abroad.

The UK's role in this context involves developing and deploying technological and policy innovations that can be replicated globally, from low-carbon industrial processes to financial mechanisms that incentivise greener supply chains. Although the UK's current direct impact on global emissions is limited, its influence lies in demonstrating how advanced economies can decouple prosperity from carbon intensity whilst taking responsibility for the full impact of its consumption.

Crucially, the framing of the transition to net zero requires careful consideration. Beyond its environmental benefits, decarbonising energy systems represents a strategy for long-term economic stability and national security. Reliance on volatile fossil fuel markets exposes economies to price shocks, supply disruptions, and geopolitical tensions (Energy Crisis Commission, 2024). In recent years the Russia-Ukraine war has exemplified the vulnerabilities of the UK's current energy systems, with £39.3 billion of government support required between October 2022 and March 2023 (DESNZ, 2023b), and a total of £140bn from the start of the war (2021) to the end of 2024 (Energy & Climate Intelligence Unit, 2025), all the while energy prices remain almost 50% higher in late 2025 than in the winter of 2021/22 (Bolton, 2025). In contrast, sustained investment in renewable energy protects firms and households from future volatility. Moreover, not only do the economic costs of inaction far outweigh the costs of a planned transition (The Economist Intelligence Unit, 2015; World Bank, 2024), but recent evidence suggests that the energy transition can yield significant benefits in pure financial terms (O'Shea et al., 2025).

Climate policy framed as a robust economic strategy transforms net zero from an environmental obligation into a long-term investment in national resilience. Such an approach helps policy to transcend election cycles, which importantly ensures continuity and confidence along the transition process. For smaller actors, this perspective rephrases participation in decarbonisation as an investment in their own resilience and competitiveness.

## 5.7 Conclusion

This chapter has demonstrated how the preceding sections addressed each of the RQs set out in Chapter 1. Following this a discussion on the impacts, contributions, limitations, and potential future directions of this work is laid out. It then places this research in the wider context of progressing the decarbonisation agenda. The remaining section demonstrates how this thesis fulfils its overarching aim, before ending on some concluding remarks.

### 5.7.1 Overarching Aim

The objective of this thesis, laid out in Chapter 1, was to explore FTD as a means to deliver real-time, standardised estimates of SME emissions, thereby lowering knowledge barriers and supporting practical pathways to sustainable actions. This aim is achieved throughout the empirical chapters. Chapter 2 presents a process of converting FTD into Scope 1 and 2 emissions, before developing and accessing a model capable of predicting these emission insights, in the absence of this granular, private data. Chapter 3 then extends this work to Scope 3 emissions, where compatible categories of Scope 3 are calculated, and subsequently predicted using the same modelling approach. These chapters found that a simple model, requiring two user inputs of firm industry and annual turnover, was suitable in capturing a large proportion of the variation in the data. This provides a parsimonious approach to emission modelling, which has the potential to benefit both the SME, and external stakeholders such as policy makers and financial institutions. This thesis therefore addresses the two research questions; “Can FTD be used to produce models that capture variation in SME emissions across SMEs and industries?” and “What level of accuracy can be retained whilst minimising user input requirement to produce SME emission estimates?”

Chapter 5 then investigates methods to overcome the multi-year time lag associated with EEMRIOs, a limitation that restricts their mainstream use for applications such as Scope 3 estimation. The results show that, although average error levels are similar across approaches, moderate adjustments to regional size and trade flows (Method 2) offer a balance between complexity and performance. Accuracy is strongest in stable years and weakest during periods of economic disruption, highlighting the difficulty of replicating the detailed structure of an EEMRIO, further reinforced by uncovered sensitivities to changes in how EEMRIOs are constructed. For practitioners, these findings indicate that at this stage, additional methodological complexity does not consistently improve accuracy, and that simple adjustments, such as accounting for changing price levels, may often be as effective. Taken together, these results address the research question: “What processes are required to ensure FTD is applied to timely emission conversion factors, thus enabling real-time estimates?".

### 5.7.2 Concluding Remarks

Achieving any regional net zero ambition requires coordinated action across all parts of the economy, including the many SMEs whose emissions remain difficult to quantify and manage. Addressing the current disconnect between the economic activity and environmental harm of SMEs is therefore essential. This thesis has demonstrated how FTD, a resource already collected at scale, can help bridge this gap by generating credible estimates of firm-level emissions and revealing patterns that would otherwise remain unobserved. It also investigates the processes required to maximise the high frequency of FTD through nowcasting and inflationary adjustment techniques. Although these approaches cannot substitute for the structural and infrastructural advancements required for complete decarbonisation, they demonstrate the value of data-driven methods in supporting a more transparent and inclusive net zero transition. Through these methodological and empirical advances, the thesis contributes to ongoing efforts to expand participation in climate action and ensure that progress toward net zero is both timely and equitable.

## 5.8 References

Acquaye, A., Genovese, A., Barrett, J., Lenny Koh, S.C., 2014. Benchmarking carbon emissions performance in supply chains. Supply Chain Manag. Int. J. 19, 306–321. https://doi.org/10.1108/SCM-11-2013-0419

Andersson, D., 2020. A novel approach to calculate individuals' carbon footprints using financial transaction data – App development and design. J. Clean. Prod. 256, 120396. https://doi.org/10.1016/j.jclepro.2020.120396

Appiah-Kubi, E., Koranteng, F.O., Dura, C.C., Mihăilă, A.A., Drigă, I., Preda, A., 2024. Green financing and sustainability reporting among SMEs: The role of pro-environmental behavior and digitization. J. Clean. Prod. 478, 143939. https://doi.org/10.1016/j.jclepro.2024.143939

Assael, J., Heurtebize, T., Carlier, L., Soupé, F., 2023. Greenhouse Gases Emissions: Estimating Corporate Non-Reported Emissions Using Interpretable Machine Learning. Sustainability 15. https://doi.org/10.3390/su15043391

Bankers for Net Zero, 2022. Scoping Exercise: The role of banks in reducing GHG emissions of UK SMEs. https://www.bankersfornetzero.co.uk/wp-content/uploads/2024/05/Smart-Data-Foundry-B4NZ-Scoping-Exercise-Full-report-September-2022-Final-version.pdf (accessed 20.8.25).

Barendregt, W., Biørn-Hansen, A., Andersson, D., 2020. Users' Experiences with the Use of Transaction Data to Estimate Consumption-Based Emissions in a Carbon Calculator. Sustainability 12, 7777. https://doi.org/10.3390/su12187777

Barrett, J., Glen, P., Wiedmann, T., Scott, K., Lenzen, M., Roelich, K., Le Quere, C., 2013. Consumption-based GHG emission accounting: a UK case study. Clim. Policy 13, 451–470. http://dx.doi.org/10.1080/14693062.2013.788858

Beaufils, T., and Wenz, L., 2022. A scenario-based method for projecting multi-regional input–output tables. Econ. Syst. Res. 34, 440–468. https://doi.org/10.1080/09535314.2021.1952404

Bolton, P., 2025. Gas and electricity prices during the 'energy crisis' and beyond.

Bouchmel, I., Ftiti, Z., Louhich, W., Omri, A., 2024. Financing sources, green investment, and environmental performance: Cross-country evidence. J. Environ. Manage. 353, 120230. https://doi.org/10.1016/j.jenvman.2024.120230

British Business Bank, 2024. Small Business Finance Markets Report 2024. https://www.british-business-bank.co.uk/about/research-and-publications/small-business-finance-markets-report-2024 (accessed 25.6.25).

Buchenau, N., Oetzel, J., Hechelmann, R.-H., 2025. Category-specific benchmarking of Scope 3 emissions for corporate clusters. Renew. Sustain. Energy Rev. 208, 115019. https://doi.org/10.1016/j.rser.2024.115019

Butler, T., Azih, M., Mitchell, A., Baltac, S., 2023. Future opportunities for electrification to decarbonise UK industry, A report for Department for Energy Security and Net Zero. https://assets.publishing.service.gov.uk/media/67295df0541e1dfbf71e8b50/future-opportunities-for-electrification-to-decarbonise-uk-industry.pdf (accessed 1.12.25).

Caldera, H.T.S., Desha, C., Dawes, L., 2019. Evaluating the enablers and barriers for successful implementation of sustainable business practice in ‘lean’ SMEs. J. Clean. Prod. 218, 575–590. https://doi.org/10.1016/j.jclepro.2019.01.239

Carbon Brief, 2023. Revealed: Colonial rule nearly doubles UK’s historical contribution to climate change. Carbon Brief. https://www.carbonbrief.org/revealed-colonial-rule-nearly-doubles-uks-historical-contribution-to-climate-change/ (accessed 11.8.25).

CDP, 2020. CDP Full GHG Emissions Dataset: Technical Annex IV: Scope 3 Overview and Modelling. https://assets.ctfassets.net/v7uy4j80khf8/4wN17Ev2pHtHqmqkDNvcA4/0ec997081f246af9a1d6d1bf69afbcd0/Technical_Annex_III_Scope_3_Overview_and_Modelling_2024.pdf (accessed 12.7.24).

Cheema-Fox, A., LaPerla, B.R., Serafeim, G., Turkington, D., Wang, H. (Stacie), 2021. Decarbonizing Everything. Financ. Anal. J. 77, 93–108. https://doi.org/10.1080/0015198X.2021.1909943

Climate Change Committee (CCC), 2025. Progress in reducing emissions - 2025 report to Parliament. https://www.theccc.org.uk/publication/progress-in-reducing-emissions-2025-report-to-parliament/ (accessed 7.11.25).

Climatiq, 2025. Climatiq. https://www.climatiq.io/methodology (accessed 12.3.25).

Committee on Climate Change (CCC), 2022. Progress in reducing emissions: 2022 Report to Parliament. London. https://www.theccc.org.uk/wp-content/uploads/2022/06/Progress-in-reducing-emissions-2022-Report-to-Parliament.pdf (accessed 4.5.23).

Daly, H.E., 1996. Beyond growth: the economics of sustainable development. Boston: Beacon Press.

Department for Environmental Food & Rural Affairs (DEFRA), 2025. Carbon footprint for the UK and England to 2022. GOV.UK. https://www.gov.uk/government/statistics/uks-carbon-footprint/carbon-footprint-for-the-uk-and-england-to-2022 (accessed 11.10.25).

Deloitte, 2024. Challenges and solutions in measuring and reporting Scope 3 emissions. https://www.deloitte.com/nl/en/issues/climate/challenges-and-solutions-scope-3-emissions (accessed 1.7.25).

Department for Business, Energy & Industrial Strategy (BEIS), 2022. Business population estimates for the UK and regions 2022: statistical release [Dataset]. GOV.UK. https://www.gov.uk/government/statistics/business-population-estimates-2022/business-population-estimates-for-the-uk-and-regions-2022-statistical-release-html (accessed 7.12.23).

Department for Business, Energy & Industrial Strategy (BEIS) and Department for Environmental Food & Rural Affairs (DEFRA), 2022. Greenhouse gas reporting: conversion factors 2021 [Dataset]. GOV.UK. https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2021 (accessed 7.23.24).

Department for Energy Security and Net Zero (DESNZ), 2023. £40 billion spent protecting families and businesses from energy costs. GOV.UK. https://www.gov.uk/government/news/40bn-spent-protecting-families-and-businesses-from-energy-costs (accessed 11.11.25).

Dragomir, V.D., 2012. The disclosure of industrial greenhouse gas emissions: a critical assessment of corporate sustainability reports. J. Clean. Prod. 29–30, 222–237. https://doi.org/10.1016/j.jclepro.2012.01.024

Edens, B., Hoekstra, R., Zult, D., Lemmers, O., Wilting, H.C., Wu, R., 2015. A method to create carbon footprint estimates consistent with national accounts. Econ. Syst. Res. 27, 440–457. https://doi.org/10.1080/09535314.2015.1048428

Energy & Climate Intelligence Unit, 2025. Russian invasion anniversary: £140bn gas bill for UK since crisis…. Energy Clim. Intell. Unit. https://eciu.net/media/press-releases/2025/russian-invasion-anniversary-140bn-gas-bill-for-uk-since-crisis-began (accessed 11.11.25).

Energy Crisis Commission, 2024. The report: protecting the UK from a future energy crisis. https://energycrisiscommission.uk/the-report-2/ (accessed 11.11.25).

Financial Reporting Council, 2021. Thematic Review: Streamlined Energy and Carbon Reporting. https://media.frc.org.uk/documents/Streamlined_Energy_and_Carbon_Reporting_SECR.pdf. (accessed 1.9.24).

Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. High. Educ. Camb. Univ. Press. https://doi.org/10.1017/CBO9780511790942

Georgescu-Roegen, N., 1971. The Entropy Law and the Economic Process: Harvard University Press. https://doi.org/10.4159/harvard.9780674281653

Goldhammer, B., Busse, C., Busch, T., 2017. Estimating Corporate Carbon Footprints with Externally Available Data. J. Ind. Ecol. 21, 1165–1179. https://doi.org/10.1111/jiec.12522

Griffin, P.A., Lont, D.H., Sun, E.Y., 2017. The Relevance to Investors of Greenhouse Gas Emission Disclosures. Contemp. Account. Res. 34, 1265–1297. https://doi.org/10.1111/1911-3846.12298

Hettler, M., Graf-Vlachy, L., 2024. Corporate scope 3 carbon emission reporting as an enabler of supply chain decarbonization: A systematic review and comprehensive research agenda. Bus. Strategy Environ. 33, 263–282. https://doi.org/10.1002/bse.3486

Heurtebize, T., Chen, F., Soupé, F., Carvalho, R.L. de, 2022. Corporate Carbon Footprint: A Machine Learning Predictive Model for Unreported Data. https://doi.org/10.2139/ssrn.4038436

HM Treasury, HM Revenue & Customs, 2024. Consultation on the introduction of a UK carbon border adjustment mechanism - GOV.UK. https://www.gov.uk/government/consultations/consultation-on-the-introduction-of-a-uk-carbon-border-adjustment-mechanism (accessed 12.6.25).

Holton, K., Twidale, S., 2025. Sky-high electricity costs hinder Britain's net zero mission. Reuters. https://www.reuters.com/sustainability/boards-policy-regulation/sky-high-electricity-costs-hinder-britains-net zero-mission-2025-08-28/ (accessed 13.11.25).

Isil, O., Sebastianelli, R., 2020. Arcs of carbon awareness in the value chain and their antecedents. Bus. Strategy Environ. 29, 503–518. https://doi.org/10.1002/bse.2383

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning with Applications in R, Second. ed. Springer, New York, United States.

Liu, H., Yao, P., Latif, S., Aslam, S., Iqbal, N., 2022. Impact of Green financing, FinTech, and financial inclusion on energy efficiency. Environ. Sci. Pollut. Res. 29, 18955–18966. https://doi.org/10.1007/s11356-021-16949-x

Mazhar, M.U., Domingues, A.R., Yakar-Pritchard, G., Bull, R., Ling, K., 2024. Reaching for net zero: The impact of an innovative university-led business support programme on carbon management strategy and practices of small and medium-sized enterprises. Bus. Strategy Environ. 33, 6940–6960. https://doi.org/10.1002/bse.3844

Meadows, D.H., 1999. Leverage Points: Places to Intervene in a System. The Sustainability Institute. https://donellameadows.org/archives/leverage-points-places-to-intervene-in-a-system/ (accessed 23.8.25).

Menon, R.R., Ravi, V., 2021. Analysis of barriers of sustainable supply chain management in electronics industry: An interpretive structural modelling approach. Clean. Responsible Consum. 3, 100026. https://doi.org/10.1016/j.clrc.2021.100026

Messenger, J., 2025. Net Zero Banking Alliance closure shows hurdle of setting single standards, industry body says. Glob. Trade Rev. GTR. https://www.gtreview.com/news/sustainability/net zero-banking-alliance-closure-shows-hurdle-of-setting-single-standards-industry-body-says/ (accessed 12.2.25).

Nguyen, Q., Diaz-Rainey, I., Kitto, A., McNeil, B.I., Pittman, N.A., Zhang, R., 2023. Scope 3 emissions: Data quality and machine learning prediction accuracy. PLOS Clim. 2, e0000208. https://doi.org/10.1371/journal.pclm.0000208

Nguyen, Q., Diaz-Rainey, I., Kuruppuarachchi, D., 2021. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. Energy Econ. 95, 105129. https://doi.org/10.1016/j.eneco.2021.105129

Office for National Statistics (ONS), 2024. Annual Business Survey 2024. https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/annualbusinesssurvey (accessed 2.6.25).

Office for National Statistics (ONS), 2022. Supply and use tables [Dataset]. https://www.ons.gov.uk/economy/nationalaccounts/supplyandusetables (accessed 8.4.23).

Oluleye, G., Rhodes, A., Hattan, L., Tu, Z., Milan, E., 2025. Interventions to support UK SMEs in deploying low carbon technologies and measures by 2030. London. https://www.imperial-consultants.co.uk/wp-content/uploads/2025/09/Interventions-to-support-UK-SMEs-decarbonise.pdf (accessed 8.1.25).

ONS, 2025. Supply and use tables data sources catalogue. https://www.ons.gov.uk/economy/nationalaccounts/supplyandusetables/datasets/supplyandusetablesdatasourcescatalogue/current (accessed 12.6.25).

O'Shea, C., Horne, P., Maslin, M., 2025. Modelling the long-term financial benefits of UK investment in wind energy generation. UCL Open Environ. https://journals.uclpress.co.uk/ucloe/article/id/3584/ (accessed 12.11.25).

Oyewole, A.T., Adeoye, O.B., Addy, W.A., Okoye, C.C., Ofodile, O.C., 2024. Enhancing Global Competitiveness Of U.S. SMEs Through Sustainable Finance: A Review And Future

Directions. International Journal of Management & Entrepreneurship Research 6, 634–647. https://doi.org/10.51594/ijmer.v6i3.876

Partnership for Carbon Accounting Financials (PCAF), 2025. PCAF Integrates CEDA into Database to Improve Emissions Measurement. https://carbonaccountingfinancials.com/newsitem/pcaf-integrates-ceda-into-database-to-improve-emissions-measurement (accessed 12.6.25).

Phillpotts, A: ee21ap (2026) 'ee21ap/Code-and-Supporting-Files---AP-Thesis: Release 1'. Zenodo. doi:10.5281/zenodo.19034950

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2025. Bridging the SME reporting gap: A new model for predicting Scope 1 and 2 emissions. J. Ind. Ecol. n/a. https://doi.org/10.1111/jiec.70106

Phillpotts, A., Owen, A., Norman, J., Trendl, A., Gathergood, J., Jobst, N., Leake, D., 2026. Completing the SME carbon profile: scalable prediction of scope 3 emissions. *Journal of Industrial Ecology*. https://doi.org/10.1007/s44498-026-00003-5

Pinkse, J., Demirel, P., Marino, A., 2024. Unlocking innovation for net zero: constraints, enablers, and firm-level transition strategies. Ind. Innov. 31, 16–41. https://doi.org/10.1080/13662716.2023.2269112

Ritchie, H., Rosado, P., Roser, M., 2023. Are consumption-based $CO_2$ per capita emissions above or below the global average? https://ourworldindata.org/co2-and-greenhouse-gas-emissions (accessed on 8.10.25)

Robinson, O., Kemp, S., Williams, I., 2015. Carbon management at universities: a reality check. J. Clean. Prod., Bridges for a more sustainable future: Joining Environmental Management for Sustainable Universities (EMSU) and the European Roundtable for Sustainable Consumption and Production (ERSCP) conferences 106, 109–118. https://doi.org/10.1016/j.jclepro.2014.06.095

Schmidt, M., Nill, M., Scholz, J., 2022. Determining the Scope 3 Emissions of Companies. Chem. Eng. Technol. 45, 1218–1230. https://doi.org/10.1002/ceat.202200181

Science Based Targets initiative, 2018. Value Change in the Value Chain: Best Practices in Scope 3 Greenhouse Gas Management. https://files.sciencebasedtargets.org/production/files/SBT_Value_Chain_Report-1.pdf (accessed 7.10.25).

Serafeim, G., Velez Caicedo, G., 2022. Machine Learning Models for Prediction of Scope 3 Carbon Emissions. https://doi.org/10.2139/ssrn.4149874

Shakdwipee, M., Lee, L.-E., 2016. Filling the Blanks: Comparing Carbon Estimates Against Disclosures (MSCI ESG Research Issue Brief). MSCI. https://www.msci.com/documents/10199/139b2ab7-c95f-4f09-9d33-fdc491c5316e (accessed 7.10.24).

Svalna, 2022. Calculate your climate impact. https://www.svalna.se/web/en (accessed 5.23.23).

Talbot, D., Boiral, O., 2013. Can we trust corporates GHG inventories? An investigation among Canada's large final emitters. Energy Policy 63, 1075–1085. https://doi.org/10.1016/j.enpol.2013.09.054

The Economist Intelligence Unit, 2015. The cost of inaction. https://impact.economist.com/sustainability/net zero-and-energy/the-cost-of-inaction (accessed 11.11.25).

Trendl, A., Owen, A., Vomfell, L., Kilian, L., Gathergood, J., Stewart, N., Leake, D., 2023. Estimating carbon footprints from large scale financial transaction data. J. Ind. Ecol. 27, 56–70. https://doi.org/10.1111/jiec.13351

Tukker, A., Koning, A. de, Owen, A., Lutter, S., Bruckner, M., Giljum, S., Stadler, K., Wood, R., Hoekstra, R., 2018. Towards Robust, Authoritative Assessments of Environmental Impacts Embodied in Trade. J. Ind. Ecol. 22, 585–598. https://doi.org/10.1111/jiec.12716

Wang, S.-Y., Ye, N.-Z., 2025. Invisible footprints, visible insights: machine learning reveals Scope 3 emissions. Front. Sustain. 6. https://doi.org/10.3389/frsus.2025.1649150

Wells, J., Trendl, A., Owen, A., Barrett, J., Gridley, J., Jobst, N., Leake, D., 2025. A scalable tool for farm-level carbon accounting: evidence from UK agriculture. Environ. Res. Lett. 20, 124046. https://doi.org/10.1088/1748-9326/ae20ab

World Bank, 2024. The Cost of Inaction: Quantifying the Impact of Climate Change on Health in Low- and Middle-Income Countries. https://openknowledge.worldbank.org/entities/publication/be56eef5-60bc-4a14-b403-698d43ebef19 (accessed 11.11.25).

World Business Council for Sustainable Development and World Resources Institute (WRI & WBCSD), 2004. The greenhouse gas protocol- A corporate accounting and reporting standard- Revised edition. https://ghgprotocol.org/corporate-standard (accessed 05.02.25).

# Appendices

## Supporting Information Chapter 2

### Supporting Information S2.1 – Sample Selection Process

This document provides a comprehensive description of the sample selection process. It includes a table detailing the rationale and impact of each step involved in generating the final samples. Additionally, pre- and post-restriction distributions are presented to illustrate the correction of irregularities. A complete list of SIC codes is also provided, with excluded codes clearly identified along with the corresponding justification.

Table S1 outlines the restriction process required to procure our sample for Scope 1 and 2 emission estimation. Each step is provided with corresponding rationale and sample impact (number of firms and industries remaining in the sample). For steps 2, 3, and 4 which involve the creation of metrics used to identify representative data, we also provide pre- and post-distribution plots. This can be used to understand the challenges we intend to correct for, and output distributions.

Finally, in Table S2 we include a full list of SIC codes, with excluded industries identified and the rationale alongside sample and population sizes of each SIC, presenting the impact of industry exclusions. This relates to the rationale of step 1a, and is discussed further in Section 2.7.3.

Table S1. Sample restriction steps and impact

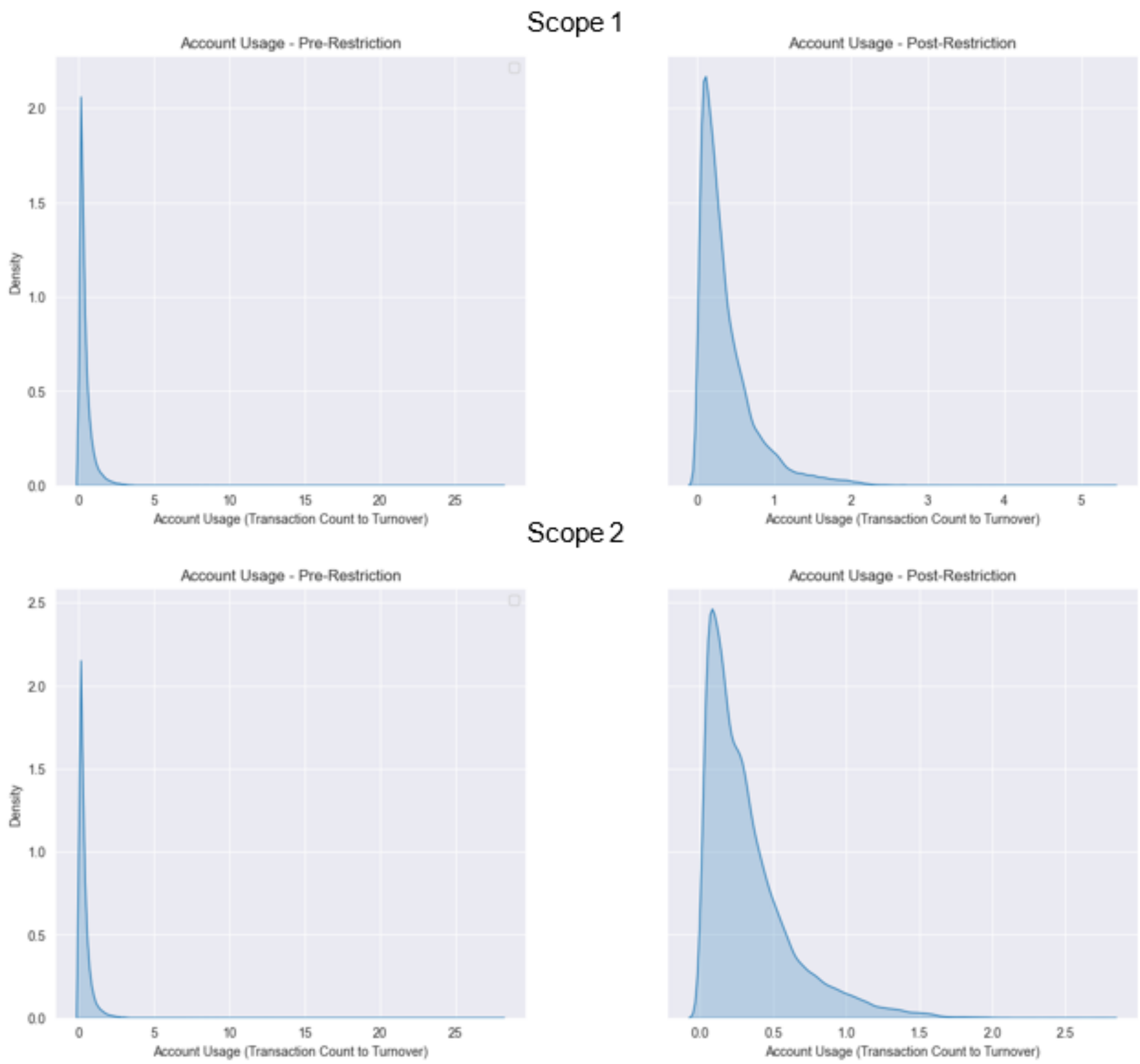
| | Restriction Step | Rationale | Impact | | | |
|---|---|---|---|---|---|---|
| | | | Scope 1 | | Scope 2 | |
| | | | Firms | Industries | Firms | Industries |
| 1a. | Eligible sample | Exclude firms that operate within flagged industries. | 424,764 | 60 | 424,764 | 60 |
| 1b. | | Exclude firms outside defined SME revenue brackets. | 352,811 | 60 | 352,811 | 60 |
| 1c. | | Exclude firms with zero scope related spend. | 97,957 | 54 | 166,169 | 55 |
| 2. | Minimum account activity | Exclude firms with dormant accounts, we set a minimum threshold for total transactions to 60 in 2021 (averaging 5 per month). | 97,608 | 54 | 161,845 | 55 |
| 3. | Account use metric | To identify and exclude firms with unusual bank account activity we divide the number of annual transactions made by firm turnover. We group firms by size and industry and remove the remove the upper and lower 10% of each distribution. | 77,771 | 52 | 129,171 | 55 |
| 4. | Energy and fuel spend to turnover | To remove outliers, caused by artificially low or high turnover levels, we divide annual energy and fuel spend with turnover and exclude the top and bottom 5%. This step ensures turnover figures are representative, consider the firm's energy and fuel spend. | 62,828 | 52 | 116,211 | 55 |
| 5. | Energy and fuel spend to total spend | Finally, we calculate an energy intensity metric, dividing energy and fuel spend by total spend. In this process, we identify firms with artificially high or low energy and fuel spend as a share of their total spend. We group firms by size and industry and remove the upper and lower 15% of each distribution. We conduct a sensitivity analysis to assess the impact of varying the sample restriction thresholds and present these results in Section 3.3. | 39,702 | 44 | 92,714 | 54 |

## Pre- and Post- Distribution Plots

Step 3 – Account Use Metric

After calculating the account usage metric (annual transaction count divided by turnover), we identify issues at both ends of the distribution. Firstly, we notice a substantial number of near-zero values, indicating that account usage is exceedingly low relative to firm size. Secondly, we observe a long right tail, which signifies the presence of firms with exceptionally high account usage relative to their size. To address these disparities, we group firms by industry and size, subsequently trimming the lower and upper 10%. This results in a refined distribution post-restriction.

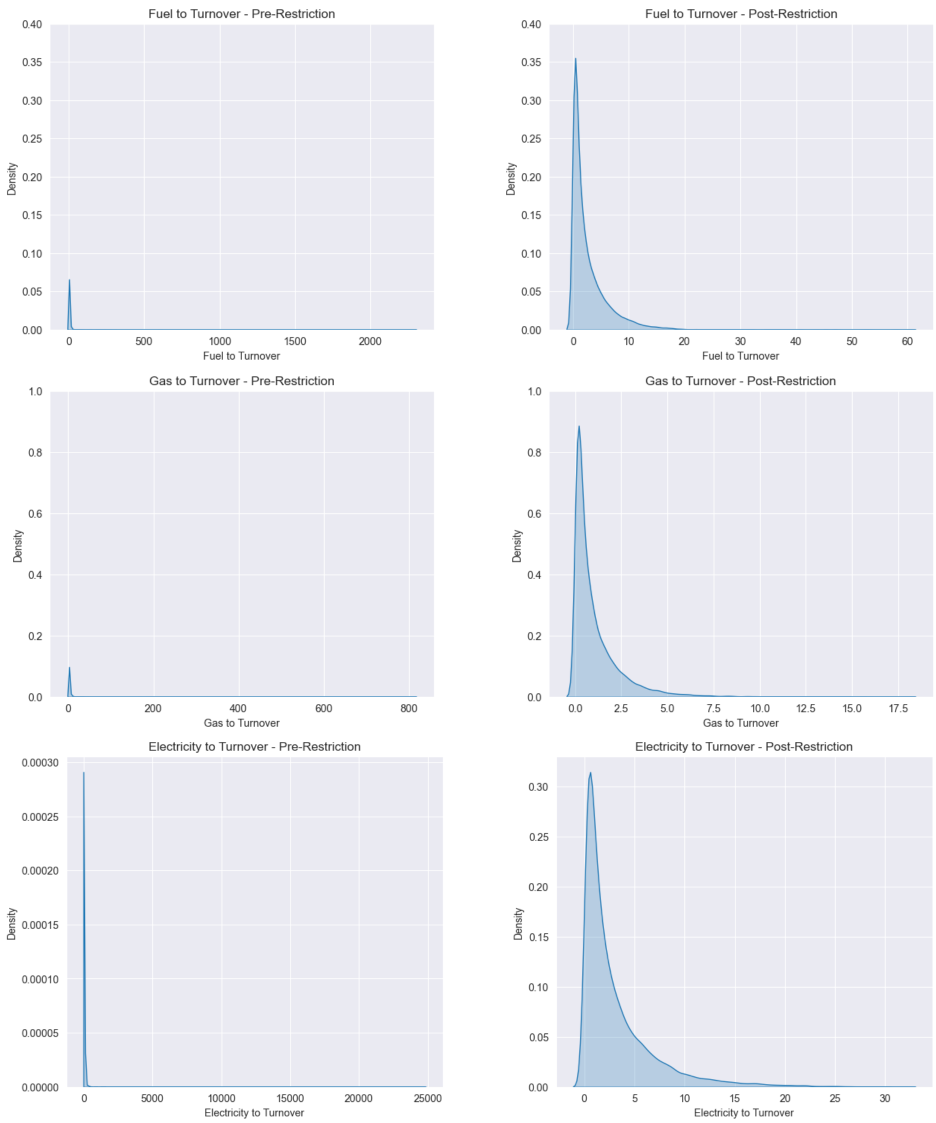
Figure S1. Pre- and post- restriction plots for account usage



Step 4 – Energy and fuel spend to turnover

After calculating energy and fuel expenditures relative to turnover, we again identify issues at both ends of the distribution for each fuel type. As previously noted, a significant number of near-zero values indicate low spending in these categories relative to firm size. Additionally, we observe extremely long right tails, suggesting that some accounts exhibit exceptionally high spending in these areas relative to their size. Given that this step aims primarily to eliminate outliers resulting from artificially high or low observed turnover levels, we only group firms by industry before removing the lower and upper 5% of the distribution. Notably, we find that right tails persist to an extent, particularly in the fuel and gas categories. Upon further investigation, we determine that the firms within these extremes belong to a limited number of industries, leading us to conclude that fuel and gas consumption may indeed be higher in these specific sectors.

Figure S2. Pre- and post- restriction plots for energy and fuel spend to turnover



## Step 5 – Energy and fuel spend to total spend

After resolving issues related to turnover and general account usage, we introduce a restriction to control for the proportion of total expenditure that firms allocate to energy and fuel. Prior to this restriction, we observe that many firms spend close to 0% of their total budget on energy and fuel, while others allocate more than 50%. We hypothesize that this variation stems from some firms maintaining dedicated accounts for energy and fuel, which may not always be captured in our data. To refine the sample, we group firms by industry and size, removing the lower and upper 15% of the distribution. As this step is more arbitrary, we report model performance for different levels of trimming, in Section 2.6.2.3 Impact of Sample Selection.

Figure S3. Pre- and post- restriction plots for energy and fuel spend to total spend

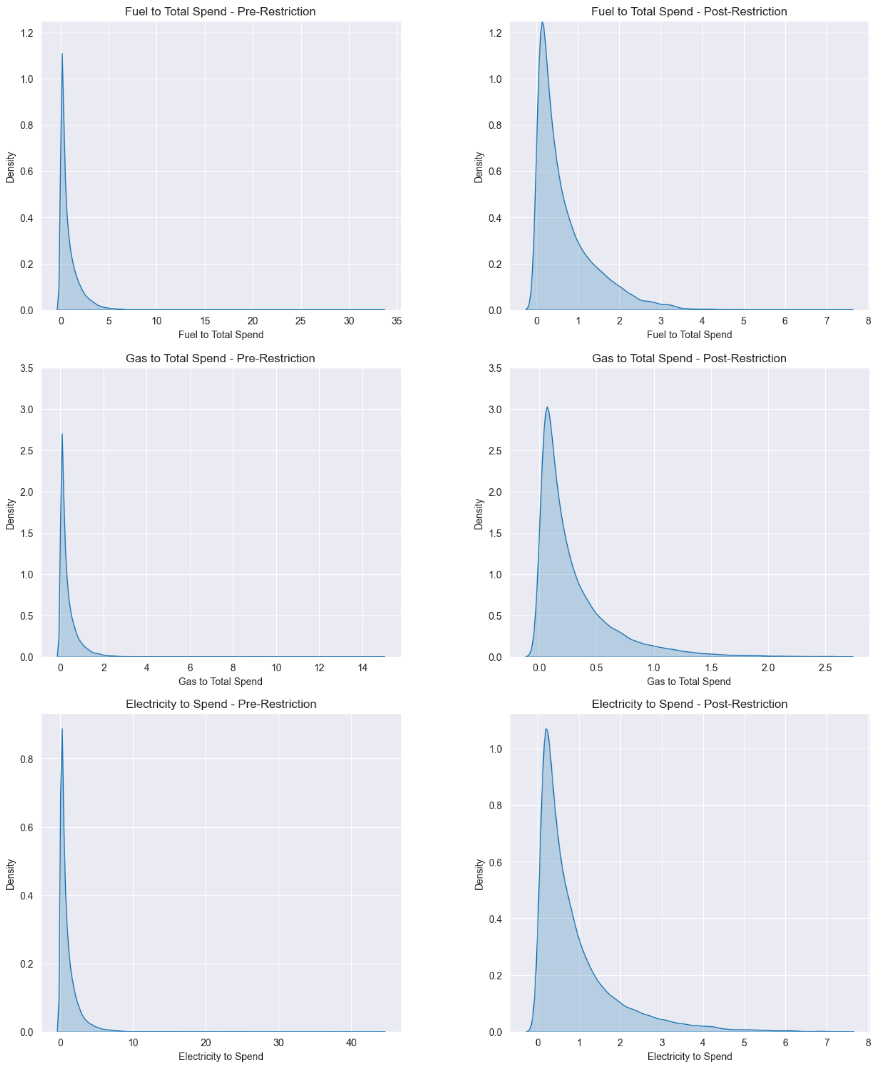


Table S2. SIC code sample and population sizes, exclusion stage and rationale

| | | | | | | Final Sample | | UK Business Population | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Section** | **Division** | **Description** | **Excluded before sample selection?** | **Excluded during sample selection?** | **Starting Sample Size** | **Scp 1** | **Scp 2** | **Total Firms** | **SMEs*** | **SMEs*** | **Notes** |
| | | | | | # | # | # | # | # | % | |
| A | 1 | Crop and animal production, hunting and related service activities | Y | | - | - | - | 132,420 | 132,335 | *99.94* | Industry identified as high error due to limitations of transaction methodology |
| A | 2 | Forestry and logging | | | 726 | 79 | 129 | 4,440 | 4,440 | *100.00* | |
| A | 3 | Fishing and aquaculture | | | 955 | 79 | 153 | 4,160 | 4,155 | *99.88* | |
| B | 5 | Mining of coal and lignite | Y | | - | - | - | 10 | 10 | *100.00* | Missing required industry data for regression input variables |
| B | 6 | Extraction of crude petroleum and natural gas | | Y | - | - | - | 150 | 130 | *86.67* | |
| B | 7 | Mining of metal ores | Y | | - | - | - | - | - | *0.00* | Missing required industry data for regression input variables |
| B | 8 | Other mining and quarrying | | Y | 177 | - | - | 705 | 695 | *98.58* | |
| B | 9 | Mining support service activities | | Y | 310 | - | - | 385 | 375 | *97.40* | |
| C | 10 | Manufacture of food products | | | 1,551 | 227 | 470 | 8,945 | 8,715 | *97.43* | |
| C | 11 | Manufacture of beverages | | | 449 | 66 | 139 | 2,730 | 2,700 | *98.90* | |
| C | 12 | Manufacture of tobacco products | Y | | - | - | - | 5 | 5 | *100.00* | Missing required industry data for regression input variables |
| C | 13 | Manufacture of textiles | | | 1,314 | 134 | 306 | 4,290 | 4,270 | *99.53* | |
| C | 14 | Manufacture of wearing apparel | | Y | 606 | - | 136 | 4,245 | 4,240 | *99.88* | |
| C | 15 | Manufacture of leather and related products | | Y | 157 | - | - | 630 | 625 | *99.21* | |
| C | 16 | Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials | | | 2,533 | 365 | 653 | 9,940 | 9,920 | *99.80* | |
| C | 17 | Manufacture of paper and paper products | | Y | 272 | - | 82 | 1,430 | 1,395 | *97.55* | |
| C | 18 | Printing and reproduction of recorded media | | | 2,463 | 257 | 642 | 10,940 | 10,910 | *99.73* | |
| C | 19 | Manufacture of coke and refined petroleum products | | Y | - | - | - | 90 | 80 | *88.89* | |
| C | 20 | Manufacture of chemicals and chemical products | | | 716 | 88 | 208 | 3,230 | 3,155 | *97.68* | |
| C | 21 | Manufacture of basic pharmaceutical products and pharmaceutical preparations | | Y | - | - | - | 695 | 645 | *92.81* | |
| C | 22 | Manufacture of rubber and plastic products | | | 1,202 | 230 | 419 | 5,550 | 5,460 | *98.38* | |
| C | 23 | Manufacture of other non-metallic mineral products | | | 1,239 | 180 | 338 | 3,810 | 3,765 | *98.82* | |
| C | 24 | Manufacture of basic metals | | | 484 | 100 | 172 | 1,825 | 1,790 | *98.08* | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C | 25 | Manufacture of fabricated metal products, except machinery and equipment | | | 8,011 | 1,382 | 2,345 | 27,805 | 27,730 | *99.73* | |
| C | 26 | Manufacture of computer, electronic and optical products | | | 1,039 | 100 | 248 | 5,755 | 5,695 | *98.96* | |
| C | 27 | Manufacture of electrical equipment | | | 2,349 | 299 | 564 | 3,090 | 3,040 | *98.38* | |
| C | 28 | Manufacture of machinery and equipment nec | | | 2,195 | 411 | 712 | 7,460 | 7,370 | *98.79* | |
| C | 29 | Manufacture of motor vehicles, trailers and semi-trailers | | | 1,015 | 141 | 275 | 3,495 | 3,405 | *97.42* | |
| C | 30 | Manufacture of other transport equipment | | | 893 | 111 | 200 | 2,265 | 2,200 | *97.13* | |
| C | 31 | Manufacture of furniture | | | 1,577 | 259 | 476 | 6,580 | 6,550 | *99.54* | |
| C | 32 | Other manufacturing | | | 2,637 | 218 | 543 | 10,210 | 10,175 | *99.66* | |
| C | 33 | Repair and installation of machinery and equipment | | | 342 | 50 | 100 | 15,085 | 15,030 | *99.64* | |
| D | 35 | Electricity, gas, steam and air conditioning supply | Y | | - | - | - | 5,840 | 5,795 | *99.23* | Industry identified as high error due to limitations of transaction methodology |
| E | 36 | Water collection, treatment and supply | Y | | - | - | - | 105 | 85 | *80.95* | Industry identified as high error due to limitations of transaction methodology |
| E | 37 | Sewerage | Y | | - | - | - | 1,105 | 1,105 | *100.00* | Industry identified as high error due to limitations of transaction methodology |
| E | 38 | Waste collection, treatment and disposal activities; materials recovery | Y | | - | - | - | 5,915 | 5,865 | *99.15* | Industry identified as high error due to limitations of transaction methodology |
| E | 39 | Remediation activities and other waste management services | Y | | - | - | - | 1,150 | 1,145 | *99.57* | Industry identified as high error due to limitations of transaction methodology |
| F | 41 | Construction of buildings | | | 9,365 | 845 | 2,080 | 111,955 | 111,845 | *99.90* | |
| F | 42 | Civil engineering | | | 17,195 | 2,566 | 4,091 | 25,105 | 25,025 | *99.68* | |
| F | 43 | Specialised construction activities | | | 62,387 | 8,597 | 13,901 | 222,640 | 222,530 | *99.95* | |
| G | 45 | Wholesale and retail trade and repair of motor vehicles and motorcycles | | | 19,402 | 3,077 | 5,784 | 78,995 | 78,790 | *99.74* | |
| G | 46 | Wholesale trade, except of motor vehicles and motorcycles | | | 10,937 | 1,553 | 3,079 | 106,735 | 106,250 | *99.55* | |
| G | 47 | Retail trade, except of motor vehicles and motorcycles | | | 43,590 | 4,595 | 12,490 | 220,685 | 220,205 | *99.78* | |
| H | 49 | Land transport and transport via pipelines | Y | | - | - | - | 79,450 | 79,270 | *99.77* | Industry identified as high error due to limitations of transaction methodology |
| H | 50 | Water transport | | Y | 248 | - | - | 1,360 | 1,350 | *99.26* | |
| H | 51 | Air transport | | Y | - | - | - | 1,045 | 1,025 | *98.09* | |
| H | 52 | Warehousing and support activities for transportation | | | 2,539 | 252 | 526 | 17,380 | 17,210 | *99.02* | |
| H | 53 | Postal and courier activities | | Y | - | - | - | 39,170 | 39,140 | *99.92* | |
| I | 55 | Accommodation | | | 6,815 | 1,117 | 2,616 | 18,665 | 18,485 | *99.04* | |
| I | 56 | Food and beverage service activities | | | 24,292 | 3,426 | 9,800 | 148,350 | 147,885 | *99.69* | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| J | 58 | Publishing activities | | Y | 2,291 | - | 202 | 12,075 | 12,015 | *99.50* | |
| J | 59 | Motion picture, video and television programme production, sound recording and music publishing activities | | | 2,216 | 61 | 178 | 29,320 | 29,280 | *99.86* | |
| J | 60 | Programming and broadcasting activities | | Y | 879 | - | 65 | 2,035 | 2,020 | *99.26* | |
| J | 61 | Telecommunications | | Y | 907 | - | 141 | 8,440 | 8,385 | *99.35* | |
| J | 62 | Computer programming, consultancy and related activities | | | 12,766 | 306 | 1,031 | 151,945 | 151,720 | *99.85* | |
| J | 63 | Information service activities | | Y | 707 | - | 62 | 9,150 | 9,115 | *99.62* | |
| K | 64 | Financial service activities, except insurance and pension funding | Y | | - | - | - | 19,645 | 19,505 | *99.29* | Industry identified as high error due to limitations of transaction methodology |
| K | 65 | Insurance, reinsurance and pension funding, except compulsory social security | Y | | - | - | - | 7,355 | 7,300 | *99.25* | Industry identified as high error due to limitations of transaction methodology |
| K | 66 | Activities auxiliary to financial services and insurance activities | Y | | - | - | - | 34,315 | 34,115 | *99.42* | Industry identified as high error due to limitations of transaction methodology |
| L | 68 | Real estate activities | Y | | - | - | - | 105,370 | 105,145 | *99.79* | Industry identified as high error due to limitations of transaction methodology |
| M | 69 | Legal and accounting activities | Y | | - | - | - | 75,620 | 75,355 | *99.65* | Industry identified as high error due to limitations of transaction methodology |
| M | 70 | Activities of head offices; management consultancy activities | Y | | - | - | - | 173,650 | 173,520 | *99.93* | Industry identified as high error due to limitations of transaction methodology |
| M | 71 | Architectural and engineering activities; technical testing and analysis | | | 11,466 | 683 | 1,714 | 93,470 | 93,310 | *99.83* | |
| M | 72 | Scientific research and development | | Y | 1,192 | - | 132 | 5,785 | 5,695 | *98.44* | |
| M | 73 | Advertising and market research | | | 3,718 | 138 | 413 | 23,275 | 23,180 | *99.59* | |
| M | 74 | Other professional, scientific and technical activities | | | 6,939 | 335 | 809 | 77,290 | 77,270 | *99.97* | |
| M | 75 | Veterinary activities | | Y | 942 | - | 261 | 3,890 | 3,880 | *99.74* | |
| N | 77 | Rental and leasing activities | Y | | - | - | - | 18,155 | 18,090 | *99.64* | Industry identified as high error due to limitations of transaction methodology |
| N | 78 | Employment activities | | | 4,251 | 195 | 582 | 30,615 | 30,155 | *98.50* | |
| N | 79 | Travel agency, tour operator and other reservation service and related activities | | Y | - | - | - | 8,670 | 8,630 | *99.54* | |
| N | 80 | Security and investigation activities | | Y | 154 | - | - | 9,910 | 9,835 | *99.24* | |
| N | 81 | Services to buildings and landscape activities | | | 12,575 | 1,012 | 1,672 | 46,525 | 46,225 | *99.36* | |
| N | 82 | Office administrative, office support and other business support activities | | | 10,662 | 522 | 1,310 | 116,340 | 116,160 | *99.85* | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| O | 84 | Public administration and defence; compulsory social security | | Y | 209 | - | - | 7,695 | 7,295 | *94.80* | |
| P | 85 | Education | | | 18,276 | 913 | 3,061 | 45,490 | 43,935 | *96.58* | |
| Q | 86 | Human health activities | | | 21,558 | 1,019 | 4,648 | 57,620 | 57,170 | *99.22* | |
| Q | 87 | Residential care activities | | | 1,384 | 187 | 483 | 10,825 | 10,420 | *96.26* | |
| Q | 88 | Social work activities without accommodation | | | 25,189 | 605 | 2,655 | 36,110 | 35,790 | *99.11* | |
| R | 90 | Creative, arts and entertainment activities | | | 6,933 | 246 | 686 | 29,955 | 29,925 | *99.90* | |
| R | 91 | Libraries, archives, museums and other cultural activities | | Y | 670 | - | 153 | 1,830 | 1,780 | *97.27* | |
| R | 92 | Gambling and betting activities | | Y | - | - | - | 975 | 940 | *96.41* | |
| R | 93 | Sports activities and amusement and recreation activities | | | 21,910 | 951 | 2,618 | 35,495 | 35,265 | *99.35* | |
| S | 94 | Activities of membership organisations | | Y | 4,391 | - | 343 | 22,340 | 22,265 | *99.66* | |
| S | 95 | Repair of computers and personal and household goods | | Y | - | - | - | 9,665 | 9,650 | *99.84* | |
| S | 96 | Other personal service activities | | | 20,597 | 1,727 | 5,818 | 76,275 | 76,245 | *99.96* | |
| T | 97 | Activities of households as employers of domestic personnel | Y | | - | - | - | - | - | *0.00* | Missing required industry data for regression input variables |
| T | 98 | Undifferentiated goods- and services-producing activities of private households for own use | Y | | - | - | - | 5 | 5 | *100.00* | Missing required industry data for regression input variables |
| U | 99 | Activities of extraterritorial organisations and bodies | Y | | - | - | - | - | - | *0.00* | Missing required industry data for regression input variables |
| **Included Firms** | | | | | | | | | | | |
| | | | **2,095,955** | **2,703,915** | **424,764** | **39,704** | **92,714** | **2,765,145** | **2,754,605** | | |
| ***Number of SIC Codes*** | | | ***21*** | ***23*** | ***60*** | ***44*** | ***54*** | ***85*** | ***85*** | | |
| ***Excluded*** | | | ***25.3%*** | ***4.1%*** | | | | | | | |
| ***Captured*** | | | ***74.7%*** | ***95.9%*** | ***74%*** | ***71%*** | ***73%*** | | | | |

## Supporting Information S2.2 – Utility adjustment factors

This supporting information provides the calculation for the splitting of spend on "Energy and Utilities" into Scope 1 and Scope 2 elements, as well as provides the utility adjustment factors for differing consumption levels of energy.

**Splitting "Energy and Utilities" spend**

As many UK utility firms are known to provide electricity, gas, and heating oil to consumers, it is a necessary step to disaggregate spending into Scope 1 and Scope 2. To do this, we source external data which signals the expected split between spend on electricity and other fuels in the UK.

We find that national fuel consumption data is unsuitable, as it is published in physical units, and we require monetary values. Suitable monetary sources include supply and use tables or other survey-based indicators.

While the monetary data from supply and use tables is valid, it exhibits the limitation of requiring the same split percentage across all industries. A more detailed option comes from an ONS dataset on business energy spending intensity, which draws on two otherwise confidential sources: the Annual Business Survey (2019) and the Annual Purchases Survey (2018). This dataset includes the table "Energy Intensity by Industry Section and Energy Type," which provides percentage breakdowns of energy spending Despite being based on 2018 data, we consider this the better option, as it allows for industry-specific variation. Table 8 below reproduces the original data and includes the subsequent calculation of industry-level splits.

To estimate the split between Scope 1 and Scope 2, we first sum the electricity and gas columns, then calculate the contribution of each fuel type. We exclude the "other" category, as supporting source documentation indicates it primarily consists of petroleum products, which are unlikely to be sold by utility companies.

For SIC sections G and I, the gas and other categories are reported as a combined figure due to confidentiality restrictions. To estimate gas intensity for these industries, we calculate the average gas and other intensities across all industries (excluding section D). We then determine the typical gas-to-other ratio based on these averages and apply that ratio to the combined figure to impute the gas intensity

Table S3. Energy & Utilities split ratio calculation (Data highlighted grey represents raw data from source, whilst unhighlighted cells show our calculations).

| | Energy intensity by energy type | | | | | | Suppressed Value Assumption | | Utility Spend Split | |
|---|---|---|---|---|---|---|---|---|---|---|
| SIC Section | Electricity | Petrol/diesel | Natural gas | Other | Suppressed | Total | Gas Share | Utility Spend | Elec Share | Gas Share |
| | % | % | % | % | % | % | % | % | % | % |
| A | 1.2 | 10.0 | 0.1 | 0.6 | 0.0 | 11.9 | | 1.3 | 94 | 6 |
| B | 2.1 | 3.3 | 1.3 | 1.2 | 0.0 | 7.9 | | 3.4 | 61 | 39 |
| C | 1.8 | 0.7 | 0.8 | 9.4 | 0.0 | 12.7 | | 2.6 | 69 | 31 |
| D | 33.5 | 0.3 | 20.8 | 1.7 | 0.0 | 56.3 | | 54.3 | 62 | 38 |
| E | 6.3 | 5.3 | 0.4 | 0.1 | 0.0 | 12.2 | | 6.8 | 94 | 6 |
| F | 0.4 | 2.2 | 0.1 | 0.1 | 0.0 | 2.8 | | 0.5 | 78 | 22 |
| G | 2.5 | 3.5 | | | 3.7 | 9.7 | 1.6 | 4.0 | 62 | 38 |
| H | 1.5 | 14.0 | 0.2 | 8.8 | 0.0 | 24.5 | | 1.7 | 91 | 9 |
| I | 4.6 | 4.9 | | | 4.7 | 14.2 | 1.9 | 6.5 | 71 | 29 |
| J | 2.1 | 0.4 | 0.1 | 0.0 | 0.0 | 2.6 | | 2.2 | 96 | 4 |
| K | 0.8 | 0.2 | 0.1 | 0.0 | 0.0 | 1.0 | | 0.9 | 92 | 8 |
| L | 4.1 | 1.2 | 0.8 | 0.1 | 0.0 | 6.2 | | 4.9 | 83 | 17 |
| M | 0.9 | 1.1 | 0.2 | 0.1 | 0.0 | 2.3 | | 1.1 | 84 | 16 |
| N | 1.4 | 3.5 | 0.4 | 0.4 | 0.0 | 5.7 | | 1.9 | 77 | 23 |
| P | 2.5 | 1.7 | 0.9 | 0.2 | 0.0 | 5.3 | | 3.4 | 73 | 27 |
| Q | 3.6 | 1.3 | 2.0 | 0.1 | 0.0 | 7.0 | | 5.6 | 65 | 35 |
| R | 3.8 | 0.9 | 0.9 | 0.1 | 0.0 | 5.8 | | 4.8 | 81 | 19 |
| S | 2.9 | 4.0 | 0.7 | 0.1 | 0.0 | 7.7 | | 3.6 | 80 | 20 |
| *Mean (Excl D, G. I)* | *2.27* | *2.37* | *0.59* | *0.84* | | *6* | | | 78 | 22 |
| | | | *41* | *59* | | | | | | |

**Utility Adjustment factors**

Transactions made with energy companies require additional adjustment steps, to account for discounts given to consumers of high energy spend. Businesses receive a cheaper unit price for energy as energy spend increases. As a result, we adjust spend upwards for higher consumption levels to reflect actual consumption levels (and thus emissions) – accounting for the lower unit price. We also adjust spend downwards for lower consumption levels to reflect the above average unit price charged to smaller consumers.

Table S4. Annual utility spend floors and the corresponding adjustment factors to apply.

| Annual Utility Spend Floor | Utility Adjustment Factor |
|---|---|
| £0 | 0.77 |
| £4,777 | 0.83 |
| £26,213 | 0.96 |
| £115,690 | 0.98 |
| £264,997 | 0.97 |
| £420,410 | 1.04 |
| £2,161,288 | 1.11 |
| £4,138,796 | 1.12 |
| £14,239,548 | 1.14 |
| £22,533,284 | 1.12 |
| £31,264,609 | 1.10 |

## Supporting Information S2.3 – Direct emission factors and subsequent sample summary statistics

This supporting information provides details on the calculation of key data required to estimate Scope 1 and 2 emissions from financial transactions. This includes the calculation of downstream emission factors, and subsequent sample summary statistics.

Table S5. Calculation of direct emission conversion factors

| | Scope 1 | | | | Scope 2 |
|---|---|---|---|---|---|
| | **Petrol** | **Diesel** | **Gas** | **Oil** | **Electricity** |
| Emissions per Unit (kg CO2e / Unit) | 2.19 | 2.51 | 0.18 | 2.54 | 0.21 |
| Average Unit Price (Unit / £) | 1.31 | 1.35 | 0.03 | 0.43 | 0.15 |
| Expenditure factors (kg CO2e / £) | 1.67 | 1.86 | 5.96 | 5.94 | 1.41 |
| Consumption Split | 68% | 32% | | | |
| **Conversion Factors (kg CO2e / £)** | **1.802** | | **5.949** | | **1.408** |

Table S6. Summary statistics by revenue bracket of Scope 1 and 2 emission estimates (kg $CO_2e$)

| | Annual Revenue Brackets | | | | | |
|---|---|---|---|---|---|---|
| | **< 50k** | **50 - 100k** | **100 - 250k** | **250 - 500k** | **500k - 1M** | **> 1M** |
| **Scope 1** | | | | | | |
| Mean | 5,193 | 9,973 | 15,226 | 28,382 | 49,200 | 221,143 |
| 25th Percentile | 2,494 | 4,921 | 6,970 | 13,079 | 21,112 | 59,473 |
| Median | 4,166 | 8,081 | 11,882 | 22,484 | 39,631 | 122,732 |
| 75th Percentile | 6,659 | 12,790 | 19,726 | 38,163 | 66,507 | 256,564 |
| Min | 235 | 576 | 542 | 707 | 869 | 763 |
| Max | 36,037 | 69,117 | 231,335 | 244,979 | 422,173 | 9,051,760 |
| Count | 5,159 | 6,622 | 8,355 | 6,302 | 5,383 | 7,881 |
| **Scope 2** | | | | | | |
| Mean | 2,148 | 3,748 | 6,909 | 11,423 | 17,840 | 62,589 |
| 25th Percentile | 741 | 1,128 | 1,905 | 3,488 | 5,591 | 13,106 |
| Median | 1,555 | 2,528 | 4,154 | 7,150 | 11,138 | 27,451 |
| 75th Percentile | 2,771 | 4,635 | 8,811 | 15,335 | 22,181 | 58,661 |
| Min | 38 | 101 | 130 | 340 | 588 | 885 |
| Max | 18,472 | 35,958 | 62,804 | 129,489 | 242,181 | 2,650,815 |
| Count | 14,238 | 15,680 | 20,114 | 14,487 | 11,239 | 16,956 |

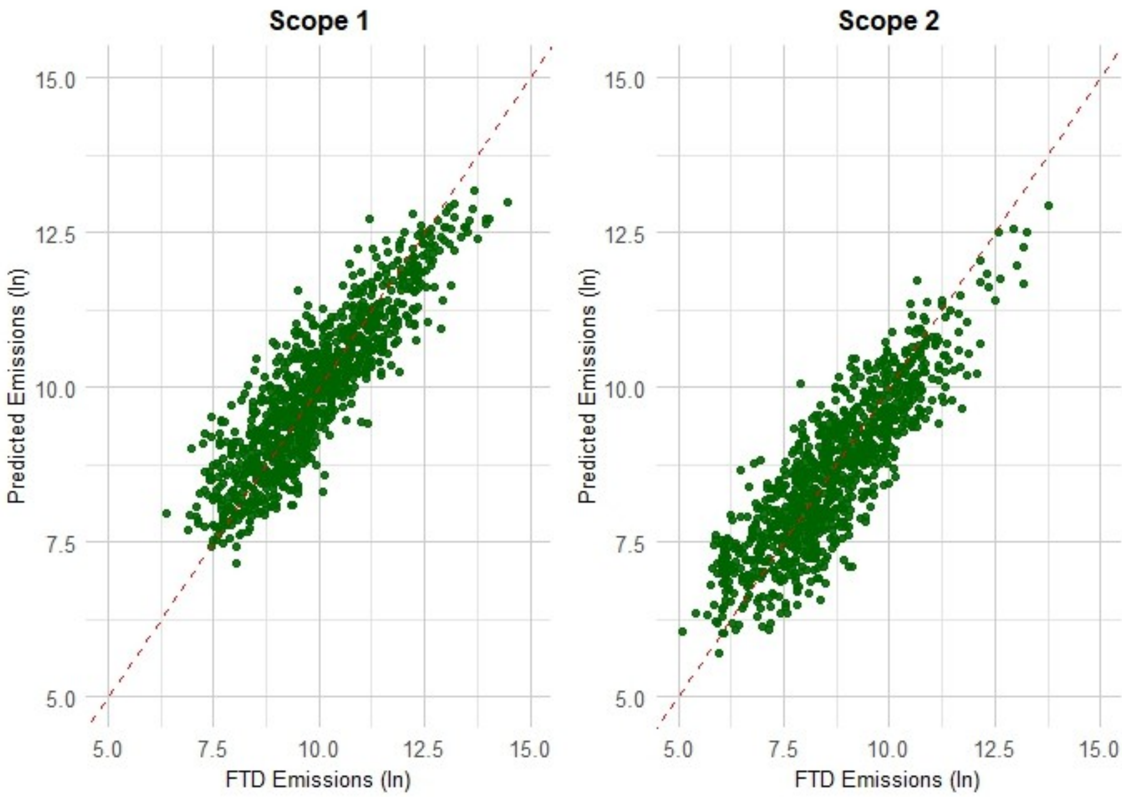
## Supporting Information S2.4 – Out of Sample Performance

This supporting information provides corresponding plots for the out of sample performance demonstrated in Section 4.2.2. We then provide out-of-sample performance metrics and plots for a small sample of self-reported emissions data (SECR).

**5-Fold Test and Train Plot**

We visualise our out-of-sample test and train results presented in Table 2.4 with Figure S4. Here we take a sample of 1,000 firms from the last test sample dataset, and plot predicted values against FTD emissions.

Figure S4 - Model predictions against out-of-sample FTD emission estimates. Underlying data for this figure are available in Table 6 within Supporting Information S5.



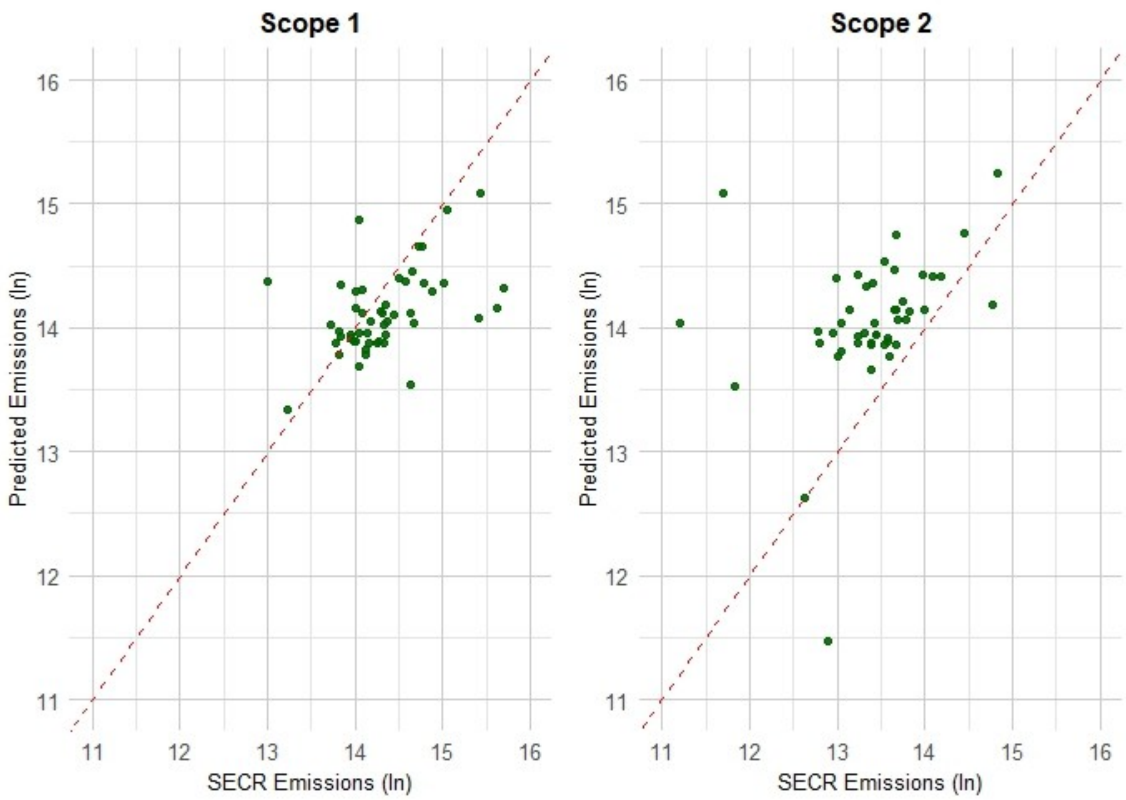

**SECR Out-of-sample Testing**

Further out-of-sample testing is conducted by comparing emissions predicted by our model against emissions self-reported by firms. To do this, we manually sourced and tabulated SECR reports for a sample of 50 firms via the Companies House website. For each firm, we input the reported turnover and SIC code into our model to generate estimates for Scope 1 and 2 emissions. These estimates are then compared to the corresponding self-reported values. Table S7 presents the summary results, with individual firm-level comparisons visualised in Figure S5. Following this, we provide our interpretation of these metrics, along with a detailed discussion of the limitations and key findings arising from the compilation of SECR data.

Table S7. Performance Metrics for Model 3 Test and Train Iterations

| | | Unit | Scope 1 | Scope 2 |
|---|---|---|---|---|
| **RSQ** | | | 0.218 | 0.099 |
| **Mean** | **AE** | *t CO2e* | 714.48 | 704.61 |
| | **APE** | *%* | 33.75 | 193.89 |
| **Median** | **AE** | *t CO2e* | 355.46 | 551.23 |
| | **APE** | *%* | 25.32 | 65.08 |

Figure S5 - Model predictions against out of sample FTD emission estimates. Underlying data for this figure are available in Table 7 within Supporting Information S5




The Table S7 and Figure S5 reveal mixed results when applying our prediction model to SECR-reporting firms. While the majority of predicted values fall close to reported figures, there are notable outliers in both Scope 1 and Scope 2 emissions. These instances appear to be more extreme amongst Scope 2 predictions. When assessing the median AE and APE of Scope 1 predictions, we observe error levels consistent with error observed on out-of-sample FTD emissions, with a median APE of 25.32%. The median APE for Scope 2 is higher, at 65.08%. However, when assessed using mean AE and APE, the model exhibits higher error levels. This is also reflected in the low RSQ scores of 0.2 and 0.1 for Scope 1 and Scope 2, respectively.

Figure S5 also highlights a systematic overestimation of Scope 2 emissions for larger firms. This is consistent with the expectation that energy-use intensity tends to decline with firm size, driven by lower unit energy prices, economies of scale, and disproportionate increases in turnover beyond certain thresholds. Additionally, our model estimates Scope 2 emissions using location-based intensity factors, whereas individual firms often report using a market-based approach that accounts for renewable energy procurement, further contributing to estimate differences.

As previously mentioned, there are important limitations associated with these comparisons, which contribute to uncertainty regarding results. We identify three primary sources of uncertainty in this comparison:

First, and most critically, the SECR firms are substantially larger than those in our model sample. In the SECR sample, annual turnover ranges from £38 million to £211 million. By contrast, our model is trained exclusively on SMEs with turnover below £36 million, 80% of which fall under £1 million. As previously noted, larger firms may differ from SMEs within the same industry in terms of operational complexity, organisational behaviour, and emissions intensity. These differences highlight the need for an SME-specific approach to emissions estimation and contribute to the observed discrepancies in model performance.

Second, the reporting periods covered by SECR submissions do not always align precisely with our model's reference year. Although we restrict our analysis to reports with a start or end date within 2021, company-specific reporting cycles vary and exact temporal alignment cannot be ensured.

Finally, we encountered numerous errors, inconsistencies, and omissions in the SECR reports themselves, making the dataset both difficult to compile and challenging to analyse. A key concern is the lack of consistency in the emissions conversion factors used: firms apply a wide range of factors, some from the relevant year, others from several years prior. Additional issues include unit conversion errors, such as confusion between MWh and kWh or between kg $CO_2$e and t $CO_2$e. This variability introduces further uncertainty into the comparison.

## Supporting Information S2.5 – Accessing the model

Accessing the Model:

To access the model described in this paper, the interested reader is directed to the following public repo:

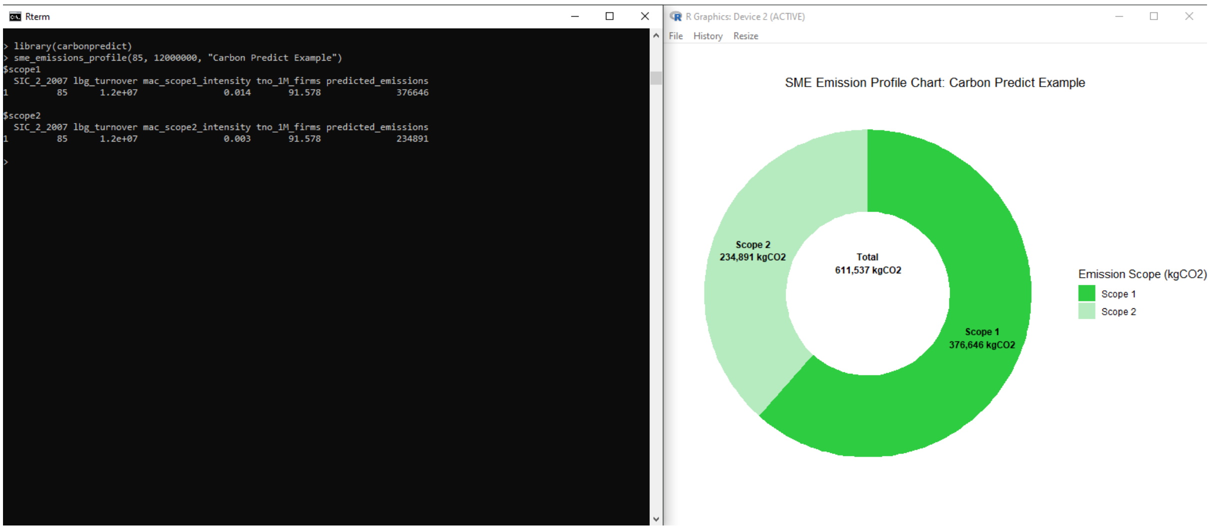
https://david-leake.github.io/carbonpredict/

As well as the following CRAN package:

https://cran.r-project.org/package=carbonpredict

Example function use:

Figure S6 – Prediction model function example.



# Supporting Information: Chapter 3

## Supporting Information S3.1 – Sample Selection and Industry Coverage

This document provides a comprehensive description of the sample selection process. It includes a table detailing the rationale and impact of each step involved in generating the final samples. Additionally, pre- and post-restriction distributions are presented to illustrate the correction of irregularities. A complete list of SIC codes is also provided, with excluded codes clearly identified along with the corresponding justification.

Table S1 outlines the restriction process required to procure our base sample of financial transaction data. Each step is provided with corresponding rationale and sample impact (number of firms and industries remaining in the sample). For steps 2, 3, and 4 which involve the creation of metrics used to identify representative data, we also provide pre- and post-distribution plots. This can be used to understand the challenges we intend to correct for, and output distributions.

Finally, in Table S2 we include a full list of SIC codes, with excluded industries identified and the rationale alongside sample and population sizes of each SIC, presenting the impact of industry exclusions. This provides a sense of the significance of each exclusion. Overall, the model achieves coverage of approximately 75% of UK SMEs.

Two main factors explain the exclusions. Industries in high-intensity sectors such as Agriculture, Energy, Water, Sewage, Waste, and Land Transport were removed because their frequent intra-industry transactions would artificially inflate estimated emissions. Moreover, their Scope 3 emissions are highly sector-specific, requiring a more tailored approach than the general framework applied here. In addition, some service industries (including Financial, Insurance, Real Estate, and Legal) were excluded because FTD-based turnover is grossly overstated in these cases due to non-turnover financial flows and frequent fund transfers. As turnover is a key predictor in our model, such distortions would lead to substantial inaccuracies.

Table S1. Sample restriction steps and impact

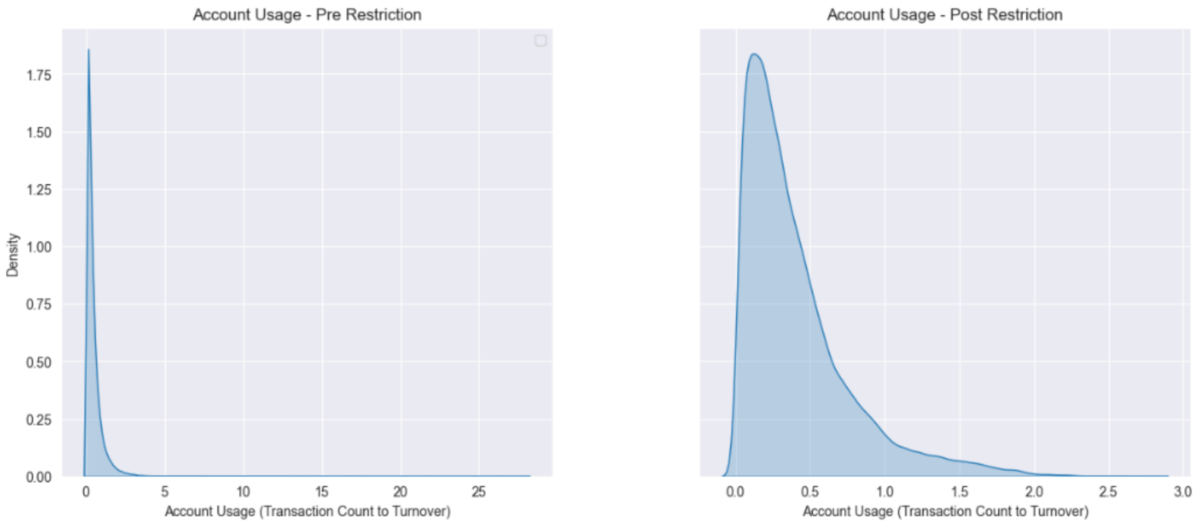
| Restriction Step | | Rationale | Impact | |
|---|---|---|---|---|
| | | | Firms | Industry Types |
| 1a. | Eligible Sample | Exclude firms that operate within flagged industries. | 261,888 | 66 |
| 1b. | | Exclude firms outside defined SME revenue brackets. | 247,010 | 66 |
| 1c. | | Exclude firms with zero fuel or energy spend. | 246,991 | 66 |
| 2. | Minimum Account Activity | To exclude firms with dormant accounts, we set a minimum threshold for total transactions to 60 in 2021 (averaging 5 per month). | 240,534 | 62 |
| 3. | Account use metric | To identify and exclude firms with unusual bank account activity we divide the number of annual transactions made by firm turnover. We group firms by size and industry and remove the remove the upper and lower 10% of each distribution. | 192,055 | 60 |
| 4. | Energy and Fuel spend to turnover | To remove outliers, caused by artificially low or high turnover levels, we divide annual utility and fuel spend with turnover and exclude the top and bottom 5%. This step ensures turnover figures are representative, anchoring it to spend on utilities and fuel. | 172,917 | 59 |
| 5. | Observed Margin | To eliminate clear outliers, we also set a threshold, ensuring that observed spend does not exceed twice the turnover. | 167,169 | 57 |

## Pre- and Post- Distribution Plots

Step 3 – Account Use Metric

After calculating the account usage metric (annual transaction count divided by turnover), we identify issues at both ends of the distribution. Firstly, we notice a substantial number of near-zero values, indicating that account usage is exceedingly low relative to firm size. Secondly, we observe a long right tail, which signifies the presence of firms with exceptionally high account usage relative to their size. To address these disparities, we group firms by industry and size, subsequently trimming the lower and upper 10%. This results in a refined distribution post-restriction.

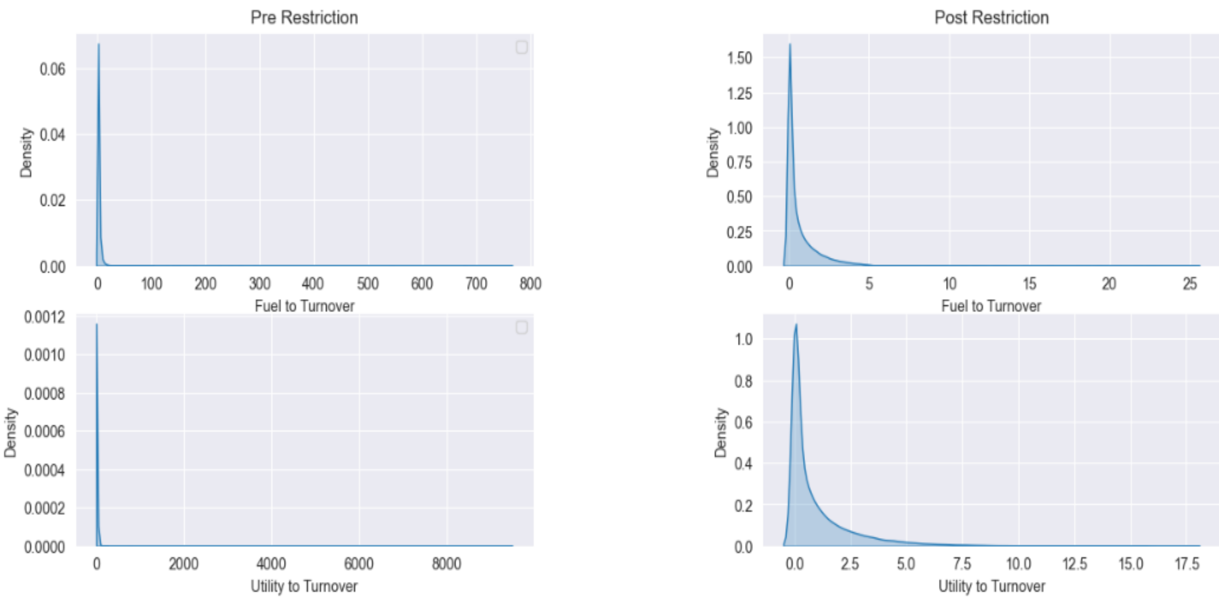
Figure S1. Pre- and post- restriction plots for account usage




Step 4 – Energy and fuel spend to turnover

After calculating energy and fuel expenditures relative to turnover, we again identify issues at both ends of the distribution for each fuel type. As previously noted, a significant number of near-zero values indicate low spending in these categories relative to firm size. Additionally, we observe extremely long right tails, suggesting that some accounts exhibit exceptionally high spending in these areas relative to their size. Given that this step aims primarily to eliminate outliers resulting from artificially high or low observed turnover levels, we only group firms by industry before removing the lower and upper 5% of the distribution. Notably, we find that right tails persist to an extent, particularly in the fuel and gas categories. Upon further investigation, we determine that the firms within these extremes belong to a limited number of industries, leading us to conclude that fuel and gas consumption may indeed be higher in these specific sectors.

Figure S2. Pre- and post- restriction plots for energy and fuel spend to turnover




Step 5 – Observed Margin

As Scope 3 emissions are made up of total spend, rather than spend on few unique categories, the final sample restriction step focuses on how total spend compares to total turnover. As with prior steps, we observe a long right tail of high values. Business profit / loss margins are variable, and it is difficult to make a subjective decision to the extent a business's profit or loss shows irregular bank account activity. As a result, we take a conservative approach, only removing firms with double spend when compared to turnover. This results in a normal distribution, centring over a break-even profit margin.

Figure S3. Pre- and post- restriction plots for firm margin

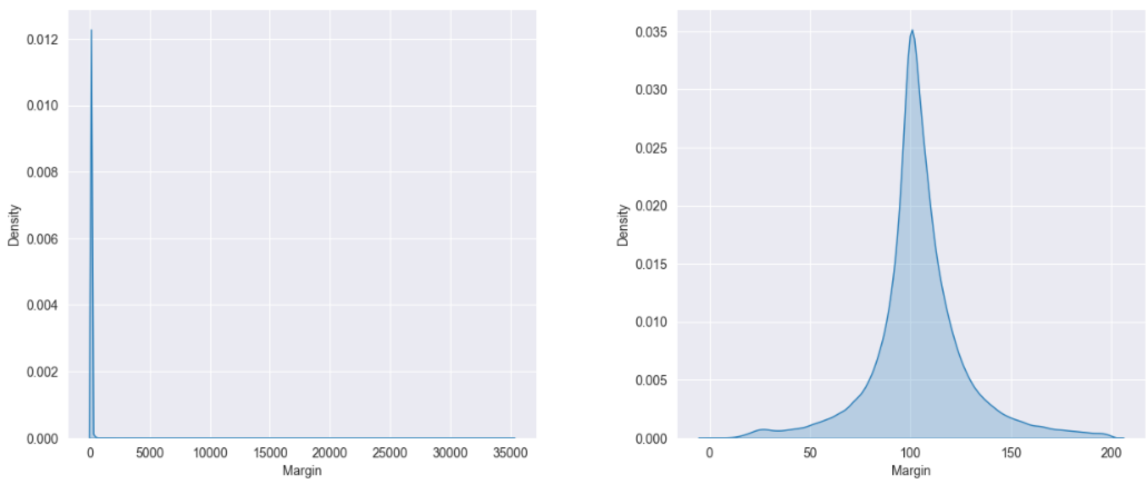


Table S2. SIC code sample and population sizes, exclusion stage and rationale

| Section | Division | Description | Excluded before sample selection? | Excluded during sample selection? | Starting Sample Size | Sample | UK Business Population Total Firms | UK Business Population SMEs* | UK Business Population SMEs* | Notes |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | # | # | # | # | % | |
| A | 1 | Crop and animal production, hunting and related service activities | Y | | - | - | 132,420 | 132,335 | | Industry identified as high error due to limitations of transaction methodology |
| A | 2 | Forestry and logging | | | 520 | 347 | 4,440 | 4,440 | *100.00* | |
| A | 3 | Fishing and aquaculture | | | 532 | 349 | 4,160 | 4,155 | *99.88* | |
| B | 5 | Mining of coal and lignite | Y | | - | - | 10 | 10 | *100.00* | Missing required industry data for regression input variables |
| B | 6 | Extraction of crude petroleum and natural gas | | | 9 | - | 150 | 130 | *86.67* | |
| B | 7 | Mining of metal ores | Y | | - | - | - | - | *0.00* | Missing required industry data for regression input variables |
| B | 8 | Other mining and quarrying | | | 123 | 78 | 705 | 695 | *98.58* | |
| B | 9 | Mining support service activities | | Y | 135 | - | 385 | 375 | *97.40* | |
| C | 10 | Manufacture of food products | | | 1,047 | 727 | 8,945 | 8,715 | *97.43* | |
| C | 11 | Manufacture of beverages | | | 326 | 217 | 2,730 | 2,700 | *98.90* | |
| C | 12 | Manufacture of tobacco products | Y | | - | - | 5 | 5 | *100.00* | Missing required industry data for regression input variables |
| C | 13 | Manufacture of textiles | | | 828 | 567 | 4,290 | 4,270 | *99.53* | |
| C | 14 | Manufacture of wearing apparel | | | 329 | 211 | 4,245 | 4,240 | *99.88* | |
| C | 15 | Manufacture of leather and related products | | | 87 | - | 630 | 625 | *99.21* | |
| C | 16 | Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials | | | 1,959 | 1,381 | 9,940 | 9,920 | *99.80* | |
| C | 17 | Manufacture of paper and paper products | | | 187 | 124 | 1,430 | 1,395 | *97.55* | |
| C | 18 | Printing and reproduction of recorded media | | | 1,454 | 988 | 10,940 | 10,910 | *99.73* | |
| C | 19 | Manufacture of coke and refined petroleum products | | Y | 25 | - | 90 | 80 | *88.89* | |
| C | 20 | Manufacture of chemicals and chemical products | | | 464 | 291 | 3,230 | 3,155 | *97.68* | |
| C | 21 | Manufacture of basic pharmaceutical products and pharmaceutical preparations | | Y | 35 | - | 695 | 645 | *92.81* | |
| C | 22 | Manufacture of rubber and plastic products | | | 941 | 642 | 5,550 | 5,460 | *98.38* | |
| C | 23 | Manufacture of other non-metallic mineral products | | | 917 | 636 | 3,810 | 3,765 | *98.82* | |
| C | 24 | Manufacture of basic metals | | | 388 | 260 | 1,825 | 1,790 | *98.08* | |
| C | 25 | Manufacture of fabricated metal products, except machinery and equipment | | | 5,894 | 4,071 | 27,805 | 27,730 | *99.73* | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| C | 26 | Manufacture of computer, electronic and optical products | | | 603 | 371 | 5,755 | 5,695 | *98.96* | |
| C | 27 | Manufacture of electrical equipment | | | 1,641 | 1,112 | 3,090 | 3,040 | *98.38* | |
| C | 28 | Manufacture of machinery and equipment nec | | | 1,647 | 1,097 | 7,460 | 7,370 | *98.79* | |
| C | 29 | Manufacture of motor vehicles, trailers and semi-trailers | | | 735 | 495 | 3,495 | 3,405 | *97.42* | |
| C | 30 | Manufacture of other transport equipment | | | 590 | 396 | 2,265 | 2,200 | *97.13* | |
| C | 31 | Manufacture of furniture | | | 1,197 | 843 | 6,580 | 6,550 | *99.54* | |
| C | 32 | Other manufacturing | | | 1,309 | 882 | 10,210 | 10,175 | *99.66* | |
| C | 33 | Repair and installation of machinery and equipment | | | 278 | 186 | 15,085 | 15,030 | *99.64* | |
| D | 35 | Electricity, gas, steam and air conditioning supply | Y | | - | - | 5,840 | 5,795 | *99.23* | Industry identified as high error due to limitations of transaction methodology |
| E | 36 | Water collection, treatment and supply | Y | | - | - | 105 | 85 | *80.95* | Industry identified as high error due to limitations of transaction methodology |
| E | 37 | Sewerage | Y | | - | - | 1,105 | 1,105 | *100.00* | Industry identified as high error due to limitations of transaction methodology |
| E | 38 | Waste collection, treatment and disposal activities; materials recovery | Y | | - | - | 5,915 | 5,865 | *99.15* | Industry identified as high error due to limitations of transaction methodology |
| E | 39 | Remediation activities and other waste management services | Y | | - | - | 1,150 | 1,145 | *99.57* | Industry identified as high error due to limitations of transaction methodology |
| F | 41 | Construction of buildings | | | 5,356 | 2,649 | 111,955 | 111,845 | *99.90* | |
| F | 42 | Civil engineering | | | 13,059 | 8,975 | 25,105 | 25,025 | *99.68* | |
| F | 43 | Specialised construction activities | | | 47,221 | 33,277 | 222,640 | 222,530 | *99.95* | |
| G | 45 | Wholesale and retail trade and repair of motor vehicles and motorcycles | | | 14,578 | 10,102 | 78,995 | 78,790 | *99.74* | |
| G | 46 | Wholesale trade, except of motor vehicles and motorcycles | | | 7,273 | 4,886 | 106,735 | 106,250 | *99.55* | |
| G | 47 | Retail trade, except of motor vehicles and motorcycles | | | 27,609 | 19,119 | 220,685 | 220,205 | *99.78* | |
| H | 49 | Land transport and transport via pipelines | Y | | - | - | 79,450 | 79,270 | *99.77* | Industry identified as high error due to limitations of transaction methodology |
| H | 50 | Water transport | | | 154 | 100 | 1,360 | 1,350 | *99.26* | |
| H | 51 | Air transport | | Y | 29 | - | 1,045 | 1,025 | *98.09* | |
| H | 52 | Warehousing and support activities for transportation | | | 1,473 | 944 | 17,380 | 17,210 | *99.02* | |
| H | 53 | Postal and courier activities | | | 1,805 | 1,200 | 39,170 | 39,140 | *99.92* | |
| I | 55 | Accommodation | | | 5,233 | 3,484 | 18,665 | 18,485 | *99.04* | |
| I | 56 | Food and beverage service activities | | | 19,502 | 13,413 | 148,350 | 147,885 | *99.69* | |
| J | 58 | Publishing activities | | | 585 | 367 | 12,075 | 12,015 | *99.50* | |

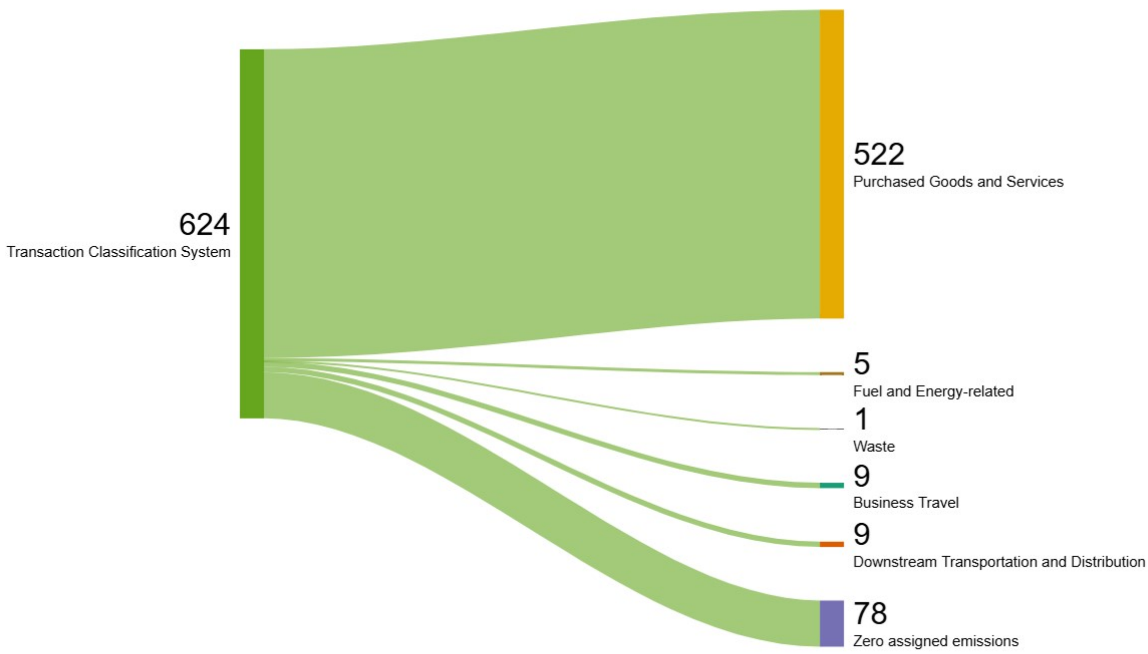
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| J | 59 | Motion picture, video and television programme production, sound recording and music publishing activities | | | 780 | 519 | 29,320 | 29,280 | *99.86* | |
| J | 60 | Programming and broadcasting activities | | Y | 310 | - | 2,035 | 2,020 | *99.26* | |
| J | 61 | Telecommunications | | | 453 | 281 | 8,440 | 8,385 | *99.35* | |
| J | 62 | Computer programming, consultancy and related activities | | | 3,181 | 2,016 | 151,945 | 151,720 | *99.85* | |
| J | 63 | Information service activities | | | 216 | 134 | 9,150 | 9,115 | *99.62* | |
| K | 64 | Financial service activities, except insurance and pension funding | Y | | - | - | 19,645 | 19,505 | *99.29* | Industry identified as high error due to limitations of transaction methodology |
| K | 65 | Insurance, reinsurance and pension funding, except compulsory social security | Y | | - | - | 7,355 | 7,300 | *99.25* | Industry identified as high error due to limitations of transaction methodology |
| K | 66 | Activities auxiliary to financial services and insurance activities | Y | | - | - | 34,315 | 34,115 | *99.42* | Industry identified as high error due to limitations of transaction methodology |
| L | 68 | Real estate activities | Y | | - | - | 105,370 | 105,145 | *99.79* | Industry identified as high error due to limitations of transaction methodology |
| M | 69 | Legal and accounting activities | Y | | - | - | 75,620 | 75,355 | *99.65* | Industry identified as high error due to limitations of transaction methodology |
| M | 70 | Activities of head offices; management consultancy activities | Y | | - | - | 173,650 | 173,520 | *99.93* | Industry identified as high error due to limitations of transaction methodology |
| M | 71 | Architectural and engineering activities; technical testing and analysis | | | 5,275 | 3,546 | 93,470 | 93,310 | *99.83* | |
| M | 72 | Scientific research and development | | | 317 | 175 | 5,785 | 5,695 | *98.44* | |
| M | 73 | Advertising and market research | | | 1,295 | 839 | 23,275 | 23,180 | *99.59* | |
| M | 74 | Other professional, scientific and technical activities | | | 3,176 | 2,163 | 77,290 | 77,270 | *99.97* | |
| M | 75 | Veterinary activities | | | 585 | 404 | 3,890 | 3,880 | *99.74* | |
| N | 77 | Rental and leasing activities | Y | | - | - | 18,155 | 18,090 | *99.64* | Industry identified as high error due to limitations of transaction methodology |
| N | 78 | Employment activities | | | 1,820 | 1,169 | 30,615 | 30,155 | *98.50* | |
| N | 79 | Travel agency, tour operator and other reservation service and related activities | | Y | 57 | - | 8,670 | 8,630 | *99.54* | |
| N | 80 | Security and investigation activities | | Y | 77 | - | 9,910 | 9,835 | *99.24* | |
| N | 81 | Services to buildings and landscape activities | | | 7,841 | 5,277 | 46,525 | 46,225 | *99.36* | |
| N | 82 | Office administrative, office support and other business support activities | | | 4,096 | 2,725 | 116,340 | 116,160 | *99.85* | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| O | 84 | Public administration and defence; compulsory social security | | | | 121 | 76 | 7,695 | 7,295 | *94.80* | |
| P | 85 | Education | | | | 7,866 | 5,419 | 45,490 | 43,935 | *96.58* | |
| Q | 86 | Human health activities | | | | 10,755 | 7,394 | 57,620 | 57,170 | *99.22* | |
| Q | 87 | Residential care activities | | | | 939 | 634 | 10,825 | 10,420 | *96.26* | |
| Q | 88 | Social work activities without accommodation | | | | 6,208 | 3,925 | 36,110 | 35,790 | *99.11* | |
| R | 90 | Creative, arts and entertainment activities | | | | 2,266 | 1,481 | 29,955 | 29,925 | *99.90* | |
| R | 91 | Libraries, archives, museums and other cultural activities | | | | 338 | 218 | 1,830 | 1,780 | *97.27* | |
| R | 92 | Gambling and betting activities | Y | | | - | - | 975 | 940 | *96.41* | |
| R | 93 | Sports activities and amusement and recreation activities | | | | 7,047 | 4,689 | 35,495 | 35,265 | *99.35* | |
| S | 94 | Activities of membership organisations | | | | 1,014 | 483 | 22,340 | 22,265 | *99.66* | |
| S | 95 | Repair of computers and personal and household goods | Y | | | - | - | 9,665 | 9,650 | *99.84* | Missing required industry data for regression input variables |
| S | 96 | Other personal service activities | | | | 12,881 | 8,815 | 76,275 | 76,245 | *99.96* | |
| T | 97 | Activities of households as employers of domestic personnel | Y | | | - | - | - | - | *0.00* | Missing required industry data for regression input variables |
| T | 98 | Undifferentiated goods- and services-producing activities of private households for own use | Y | | | - | - | 5 | 5 | *100.00* | Missing required industry data for regression input variables |
| U | 99 | Activities of extraterritorial organisations and bodies | Y | | | - | - | - | - | *0.00* | Missing required industry data for regression input variables |
| | | | | | | - | - | | | | |
| **Included Firms** | | | | **2,085,365** | **2,731,995** | **246,991** | **167,169** | **2,765,145** | **2,754,605** | | |
| ***Number of SIC Codes*** | | | | **22** | **7** | **66** | **57** | ***85*** | ***85*** | | |
| ***Excluded*** | | | | **24%** | **1%** | | | | | | |
| ***Captured*** | | | | **76%** | **99%** | **76%** | **75%** | | | | |

## Supporting Information S3.2 – Transaction Category Mapping and Conversion Factor Calculations

This supporting information provides details on each transaction categories mapping to Greenhouse Gas Protocol Scope 3 categories, their conversion factor, and description. The figure below illustrates how these spend codes map to Scope 3 categories. We then discuss how each Scope 3 category is estimated, before providing a detailed table of transaction code, description, and conversion factor. The final table contains descriptive statistics of estimated emissions categories by industry.

Figure S4. Sankey plot shows TCS to GHG Protocol Mapping. Diagram created using SankeyMATIC.




**Estimating Category 1. Purchased Goods & Services**

UKMRIO conversion factors are mapped to 522 transaction classifications, relevant to this Scope 3 category. As noted in Table 1 of the main text, we assume that transportation costs are embedded within payments for these purchases, thereby implicitly capturing Category 4 (Upstream Transport and Distribution), with these emissions also reflected in each EEMRIO conversion factor.

**Estimating Category 3. Fuel- and Energy-related purchases**

To calculate upstream emissions from fuel and energy consumption, we isolate two relevant categories of spend: “Energy & Utilities” and “Vehicle Fuelling”. The former refers to 4 sub-categories of expenditure, containing energy providers selling electricity, natural gas, and heating oil, whilst the latter refers to diesel and petrol purchased for mobile combustion.

As shown in Phillpotts et al. (2025), two adjustments to "Energy & Utilities" spending are necessary. In instances where UK utility firms are known to provide electricity, gas, and heating oil to consumers, spending is disaggregated into fuels and electricity. Second, consumption-weighted adjustment factors are required to reflect lower unit costs for higher-usage bands.

To create a set of conversion factors capturing the upstream emissions associated with purchases of different energy and fuel types. We divide well-to-tank emissions per unit figures ($CO_2e$ / kwh (or litre); BEIS and DEFRA, 2022) by average price paid by non-domestic consumers per unit (£ / kwh (or litre); DESNZ, 2024) to return emissions produced per pound spent ($CO_2e$ / £). These reflect the emissions produced during the extraction, processing, and distribution of fuels and energy. As spending within “Vehicle Fuelling” reflects the consumption of both petrol and diesel, we produce a single conversion factor through a weighted average based on the UK’s consumption split between petrol and diesel (DESNZ and BIES, 2022b) and the unit price difference. For gas and oil, we use an average conversion factor given the near zero difference in individual factors. Finally, for electricity, our emission conversion factor is based on the UK grid’s emission intensity (location-based method). We provide this calculation in Table S3.

Table S3. Calculation of WTT emission conversion factors

| | Scope 1 related | | | | Scope 2 related |
|---|---|---|---|---|---|
| | Petrol | Diesel | Gas | Oil | Electricity |
| Emissions per Unit (kg CO2e / Unit) | 0.61 | 0.61 | 0.03 | 0.53 | 0.02 |
| Average Unit Price (Unit / £) | 1.31 | 1.35 | 0.03 | 0.43 | 0.15 |
| Expenditure factors (kg CO2e / £) | 0.47 | 0.45 | 1.02 | 1.23 | 0.12 |
| Consumption Split | 68% | 32% | | | |
| **Conversion Factors (kg CO2e / £)** | **0.457** | | **1.13** | | **0.125** |

## Estimating Category 6. Business travel

To calculate emissions from Business Travel, we use 9 spend sub-categories under 'Travel', including general travel, air, rail, bus, car rental, and hotels. We are required to assume travel costs are captured directly in the business account, excluding employee-paid expenses later reimbursed. Emissions are calculated using UKMRIO-mapped conversion factors for rail, land, water, and air transport, and accommodation services.

## Estimating Category 9. Downstream transportation and distribution

Given upstream transport emissions are assumed to be embedded in purchase costs; all spend with logistics and courier merchants is assumed to reflect downstream distribution. Emissions are calculated using 9 spend sub-categories under "Transportation and Logistics", including freight, delivery, maritime, and rail, mapped to UKMRIO conversion factors for relevant transport and courier sectors.

Table S4. Expenditure codes, headings, emissions factors and associated Scope 3 category

| Scope 3 Category | Transaction Code | Description 1 | Description 2 | Description 3 | Emission Factor |
|---|---|---|---|---|---|
| 1 | code_1 | Adult | | | 0.139712 |
| 1 | code_2 | Arts & Entertainment | | | 0.263201 |
| 1 | code_3 | Arts & Entertainment | Celebrities & Entertainment News | | 0.176778 |
| 1 | code_4 | Arts & Entertainment | Comics & Animation | | 0.173578 |
| 1 | code_5 | Arts & Entertainment | Comics & Animation | Anime & Manga | 0.173578 |
| 1 | code_6 | Arts & Entertainment | Comics & Animation | Cartoons | 0.173578 |
| 1 | code_7 | Arts & Entertainment | Comics & Animation | Comics | 0.195566 |
| 1 | code_8 | Arts & Entertainment | Entertainment Industry | | 0.155141 |
| 1 | code_9 | Arts & Entertainment | Entertainment Industry | Film & TV Industry | 0.155141 |
| 1 | code_10 | Arts & Entertainment | Entertainment Industry | Recording Industry | 0.155141 |
| 1 | code_11 | Arts & Entertainment | Events & Listings | | 0.285517 |
| 1 | code_12 | Arts & Entertainment | Events & Listings | Bars Clubs & Nightlife | 0.302584 |
| 1 | code_13 | Arts & Entertainment | Events & Listings | Concerts & Music Festivals | 0.302584 |
| 1 | code_14 | Arts & Entertainment | Events & Listings | Expos & Conventions | 0.302584 |
| 1 | code_15 | Arts & Entertainment | Events & Listings | Film Festivals | 0.18328 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_16 | Arts & Entertainment | Events & Listings | Movie Listings & Theater Showtimes | 0.18328 |
| 1 | code_17 | Arts & Entertainment | Fun & Trivia | | 0.221568 |
| 1 | code_18 | Arts & Entertainment | Fun & Trivia | Flash-Based Entertainment | 0.221568 |
| 1 | code_19 | Arts & Entertainment | Fun & Trivia | Fun Tests & Silly Surveys | 0.221568 |
| 1 | code_20 | Arts & Entertainment | Humor | | 0.153234 |
| 1 | code_21 | Arts & Entertainment | Humor | Funny Pictures & Videos | 0.153234 |
| 1 | code_22 | Arts & Entertainment | Humor | Political Humor | 0.153234 |
| 1 | code_23 | Arts & Entertainment | Movies | | 0.152584 |
| 1 | code_24 | Arts & Entertainment | Music & Audio | | 0.135089 |
| 1 | code_25 | Arts & Entertainment | Music & Audio | CD & Audio Shopping | 0.135089 |
| 1 | code_26 | Arts & Entertainment | Music & Audio | Classical Music | 0.135089 |
| 1 | code_27 | Arts & Entertainment | Music & Audio | Country Music | 0.135089 |
| 1 | code_28 | Arts & Entertainment | Music & Audio | Dance & Electronic Music | 0.135089 |
| 1 | code_29 | Arts & Entertainment | Music & Audio | Experimental & Industrial Music | 0.135089 |
| 1 | code_30 | Arts & Entertainment | Music & Audio | Jazz & Blues | 0.135089 |
| 1 | code_31 | Arts & Entertainment | Music & Audio | Music Education & Instruction | 0.135089 |
| 1 | code_32 | Arts & Entertainment | Music & Audio | Music Equipment & Technology | 0.135089 |
| 1 | code_33 | Arts & Entertainment | Music & Audio | Music Reference | 0.135089 |
| 1 | code_34 | Arts & Entertainment | Music & Audio | Music Streams & Downloads | 0.135089 |
| 1 | code_35 | Arts & Entertainment | Music & Audio | Music Videos | 0.135089 |
| 1 | code_36 | Arts & Entertainment | Music & Audio | Pop Music | 0.135089 |
| 1 | code_37 | Arts & Entertainment | Music & Audio | Radio | 0.135089 |
| 1 | code_38 | Arts & Entertainment | Music & Audio | Religious Music | 0.135089 |
| 1 | code_39 | Arts & Entertainment | Music & Audio | Rock Music | 0.135089 |
| 1 | code_40 | Arts & Entertainment | Music & Audio | Soundtracks | 0.135089 |
| 1 | code_41 | Arts & Entertainment | Music & Audio | Urban & Hip-Hop | 0.135089 |
| 1 | code_42 | Arts & Entertainment | Music & Audio | World Music | 0.135089 |
| 1 | code_43 | Arts & Entertainment | Offbeat | | 0.201884 |
| 1 | code_44 | Arts & Entertainment | Offbeat | Occult & Paranormal | 0.201884 |
| 1 | code_45 | Arts & Entertainment | Online Media | | 0.201884 |
| 1 | code_46 | Arts & Entertainment | Online Media | Online Image Galleries | 0.201884 |
| 1 | code_47 | Arts & Entertainment | Performing Arts | | 0.212149 |
| 1 | code_48 | Arts & Entertainment | Performing Arts | Acting & Theater | 0.212149 |
| 1 | code_49 | Arts & Entertainment | Performing Arts | Circus | 0.212149 |
| 1 | code_50 | Arts & Entertainment | Performing Arts | Dance | 0.212149 |
| 1 | code_51 | Arts & Entertainment | Performing Arts | Magic | 0.212149 |
| 1 | code_52 | Arts & Entertainment | Performing Arts | Opera | 0.212149 |
| 1 | code_53 | Arts & Entertainment | TV & Video | | 0.152584 |
| 1 | code_54 | Arts & Entertainment | TV & Video | Online Video | 0.152584 |
| 1 | code_55 | Arts & Entertainment | TV & Video | TV Commercials | 0.152584 |
| 1 | code_56 | Arts & Entertainment | TV & Video | TV Shows & Programs | 0.152584 |
| 1 | code_57 | Arts & Entertainment | Visual Art & Design | | 0.162184 |
| 1 | code_58 | Arts & Entertainment | Visual Art & Design | Architecture | 0.156058 |
| 1 | code_59 | Arts & Entertainment | Visual Art & Design | Art Museums & Galleries | 0.163805 |
| 1 | code_60 | Arts & Entertainment | Visual Art & Design | Design | 0.201884 |
| 1 | code_61 | Arts & Entertainment | Visual Art & Design | Painting | 0.201884 |
| 1 | code_62 | Arts & Entertainment | Visual Art & Design | Photographic & Digital Arts | 0.201884 |
| 1 | code_63 | Autos & Vehicles | | | 0.225799 |
| 1 | code_64 | Autos & Vehicles | Bicycles & Accessories | | 0.33606 |
| 1 | code_65 | Autos & Vehicles | Bicycles & Accessories | Bike Parts & Repair | 0.33606 |
| 1 | code_66 | Autos & Vehicles | Bicycles & Accessories | BMX Bikes | 0.33606 |
| 1 | code_67 | Autos & Vehicles | Boats & Watercraft | | 0.302606 |
| 1 | code_68 | Autos & Vehicles | Campers & RVs | | 0.375221 |
| 1 | code_69 | Autos & Vehicles | Classic Vehicles | | 0.375221 |
| 1 | code_70 | Autos & Vehicles | Commercial Vehicles | Cargo Trucks & Trailers | 0.375221 |
| 1 | code_71 | Autos & Vehicles | Motor Vehicles (By Type) | | 0.375221 |
| 1 | code_72 | Autos & Vehicles | Motor Vehicles (By Type) | Hybrid & Alternative Vehicles | 0.375221 |
| 1 | code_73 | Autos & Vehicles | Motor Vehicles (By Type) | Motorcycles | 0.375221 |
| 1 | code_74 | Autos & Vehicles | Motor Vehicles (By Type) | Off-Road Vehicles | 0.375221 |
| 1 | code_75 | Autos & Vehicles | Motor Vehicles (By Type) | Trucks & SUVs | 0.375221 |
| 1 | code_76 | Autos & Vehicles | Vehicle Codes & Driving Laws | | 0.170341 |
| 1 | code_77 | Autos & Vehicles | Vehicle Codes & Driving Laws | Vehicle Licensing & Registration | 0.170341 |
| 1 | code_78 | Autos & Vehicles | Vehicle Parts & Services | | 0.299548 |
| 3 | code_79 | Autos & Vehicles | Vehicle Parts & Services | Gas Prices & Vehicle Fuelling | 0.457 |
| 1 | code_80 | Autos & Vehicles | Vehicle Parts & Services | Vehicle Parts & Accessories | 0.299548 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_81 | Autos & Vehicles | Vehicle Parts & Services | Vehicle Repair & Maintenance | 0.297154 |
| 1 | code_82 | Autos & Vehicles | Vehicle Shopping | | 0.375221 |
| 1 | code_83 | Autos & Vehicles | Vehicle Shopping | Used Vehicles | 0.375221 |
| 1 | code_84 | Autos & Vehicles | Vehicle Shows | | 0.259202 |
| 1 | code_85 | Beauty & Fitness | | | 0.309717 |
| 1 | code_86 | Beauty & Fitness | Beauty Pageants | | 0.222804 |
| 1 | code_87 | Beauty & Fitness | Body Art | | 0.291649 |
| 1 | code_88 | Beauty & Fitness | Cosmetic Procedures | | 0.291649 |
| 1 | code_89 | Beauty & Fitness | Cosmetic Procedures | Cosmetic Surgery | 0.291649 |
| 1 | code_90 | Beauty & Fitness | Cosmetology & Beauty Professionals | | 0.291649 |
| 1 | code_91 | Beauty & Fitness | Face & Body Care | | 0.291649 |
| 1 | code_92 | Beauty & Fitness | Face & Body Care | Hygiene & Toiletries | 0.291649 |
| 1 | code_93 | Beauty & Fitness | Face & Body Care | Make-Up & Cosmetics | 0.291649 |
| 1 | code_94 | Beauty & Fitness | Face & Body Care | Perfumes & Fragrances | 0.291649 |
| 1 | code_95 | Beauty & Fitness | Face & Body Care | Skin & Nail Care | 0.291649 |
| 1 | code_96 | Beauty & Fitness | Face & Body Care | Unwanted Body & Facial Hair Removal | 0.291649 |
| 1 | code_97 | Beauty & Fitness | Fashion & Style | | 0.738472 |
| 1 | code_98 | Beauty & Fitness | Fashion & Style | Fashion Designers & Collections | 0.738472 |
| 1 | code_99 | Beauty & Fitness | Fitness | | 0.287046 |
| 1 | code_100 | Beauty & Fitness | Hair Care | | 0.296838 |
| 1 | code_101 | Beauty & Fitness | Hair Care | Hair Loss | 0.296838 |
| 1 | code_102 | Beauty & Fitness | Spas & Beauty Services | | 0.291649 |
| 1 | code_103 | Beauty & Fitness | Spas & Beauty Services | Massage Therapy | 0.291649 |
| 1 | code_104 | Beauty & Fitness | Weight Loss | | 0.291932 |
| 1 | code_105 | Books & Literature | | | 0.195286 |
| 1 | code_106 | Books & Literature | Children's Literature | | 0.195286 |
| 1 | code_107 | Books & Literature | E-Books | | 0.195286 |
| 1 | code_108 | Books & Literature | Fan Fiction | | 0.195286 |
| 1 | code_109 | Books & Literature | Literary Classics | | 0.195286 |
| 1 | code_110 | Books & Literature | Poetry | | 0.195286 |
| 1 | code_111 | Books & Literature | Writers Resources | | 0.195286 |
| 1 | code_112 | Business & Industrial | | | 0.310656 |
| 1 | code_113 | Business & Industrial | Advertising & Marketing | Public Relations | 0.110192 |
| 1 | code_114 | Business & Industrial | Aerospace & Defense | Space Technology | 0.202416 |
| 1 | code_115 | Business & Industrial | Agriculture & Forestry | | 2.278005 |
| 1 | code_116 | Business & Industrial | Agriculture & Forestry | Agricultural Equipment | 0.423007 |
| 1 | code_117 | Business & Industrial | Agriculture & Forestry | Forestry | 0.327824 |
| 1 | code_118 | Business & Industrial | Agriculture & Forestry | Livestock | 2.314544 |
| 1 | code_119 | Business & Industrial | Automotive Industry | | 0.309141 |
| 1 | code_120 | Business & Industrial | Business Education | | 0.10258 |
| 1 | code_121 | Business & Industrial | Business Finance | | 0.113865 |
| 1 | code_122 | Business & Industrial | Business Finance | Venture Capital | 0.113865 |
| 1 | code_123 | Business & Industrial | Business Operations | | 0.070301 |
| 1 | code_124 | Business & Industrial | Business Operations | Business Plans & Presentations | 0.108621 |
| 1 | code_125 | Business & Industrial | Business Operations | Management | 0.108621 |
| 1 | code_126 | Business & Industrial | Business Services | | 0.150649 |
| 1 | code_127 | Business & Industrial | Business Services | Consulting | 0.15233 |
| 1 | code_128 | Business & Industrial | Business Services | Corporate Events | 0.194539 |
| 1 | code_129 | Business & Industrial | Business Services | E-Commerce Services | 0.19147 |
| 1 | code_130 | Business & Industrial | Business Services | Fire & Security Services | 0.153014 |
| 1 | code_131 | Business & Industrial | Business Services | Office Services | 0.152384 |
| 1 | code_132 | Business & Industrial | Business Services | Office Supplies | 0.425337 |
| 1 | code_133 | Business & Industrial | Business Services | Writing & Editing Services | 0.195513 |
| 1 | code_134 | Business & Industrial | Chemicals Industry | | 1.225531 |
| 1 | code_135 | Business & Industrial | Chemicals Industry | Cleaning Agents | 0.836584 |
| 1 | code_136 | Business & Industrial | Chemicals Industry | Plastics & Polymers | 0.633761 |
| 1 | code_137 | Business & Industrial | Construction & Maintenance | | 0.333827 |
| 1 | code_138 | Business & Industrial | Construction & Maintenance | Building Materials & Supplies | 0.501233 |
| 3 | code_139 | Business & Industrial | Energy & Utilities | | - |
| 3 | code_140 | Business & Industrial | Energy & Utilities | Electricity | 0.125 |
| 3 | code_141 | Business & Industrial | Energy & Utilities | Oil & Gas | 1.130 |
| 3 | code_142 | Business & Industrial | Energy & Utilities | Renewable & Alternative Energy | - |
| 1 | code_143 | Business & Industrial | Hospitality Industry | | 0.307688 |
| 1 | code_144 | Business & Industrial | Hospitality Industry | Event Planning | 0.280247 |
| 1 | code_145 | Business & Industrial | Hospitality Industry | Food Service | 0.31608 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_146 | Business & Industrial | Industrial Materials & Equipment | | 0.656415 |
| 1 | code_147 | Business & Industrial | Industrial Materials & Equipment | Heavy Machinery | 0.421459 |
| 1 | code_148 | Business & Industrial | Manufacturing | | 0.680609 |
| 1 | code_149 | Business & Industrial | Metals & Mining | | 0.380651 |
| 1 | code_150 | Business & Industrial | Metals & Mining | Precious Metals | 0.761028 |
| 1 | code_151 | Business & Industrial | Pharmaceuticals & Biotech | | 0.273856 |
| 1 | code_152 | Business & Industrial | Printing & Publishing | | 0.195286 |
| 1 | code_153 | Business & Industrial | Retail Trade | | 0.11123 |
| 1 | code_154 | Business & Industrial | Retail Trade | Retail Equipment & Technology | 0.111137 |
| 1 | code_155 | Business & Industrial | Small Business | MLM & Business Opportunities | 0.121549 |
| 1 | code_156 | Business & Industrial | Textiles & Nonwovens | | 0.762562 |
| 9 | code_157 | Business & Industrial | Transportation & Logistics | | 1.135895 |
| 9 | code_158 | Business & Industrial | Transportation & Logistics | Freight & Trucking | 0.620678 |
| 9 | code_159 | Business & Industrial | Transportation & Logistics | Mail & Package Delivery | 0.263766 |
| 9 | code_160 | Business & Industrial | Transportation & Logistics | Maritime Transport | 2.406699 |
| 9 | code_161 | Business & Industrial | Transportation & Logistics | Moving & Relocation | 0.620678 |
| 9 | code_162 | Business & Industrial | Transportation & Logistics | Packaging | 0.263766 |
| 9 | code_163 | Business & Industrial | Transportation & Logistics | Parking | 0.339032 |
| 9 | code_164 | Business & Industrial | Transportation & Logistics | Rail Transport | 0.794241 |
| 9 | code_165 | Business & Industrial | Transportation & Logistics | Urban Transport | 0.676149 |
| 1 | code_166 | Computers & Electronics | | | 0.250638 |
| 1 | code_167 | Computers & Electronics | CAD & CAM | | 0.15329 |
| 1 | code_168 | Computers & Electronics | Computer Hardware | | 0.411838 |
| 1 | code_169 | Computers & Electronics | Computer Hardware | Computer Components | 0.411838 |
| 1 | code_170 | Computers & Electronics | Computer Hardware | Computer Drives & Storage | 0.411838 |
| 1 | code_171 | Computers & Electronics | Computer Hardware | Computer Peripherals | 0.411838 |
| 1 | code_172 | Computers & Electronics | Computer Hardware | Desktop Computers | 0.411838 |
| 1 | code_173 | Computers & Electronics | Computer Hardware | Laptops & Notebooks | 0.411838 |
| 1 | code_174 | Computers & Electronics | Computer Security | | 0.153359 |
| 1 | code_175 | Computers & Electronics | Computer Security | Hacking & Cracking | 0.153359 |
| 1 | code_176 | Computers & Electronics | Consumer Electronics | | 0.411838 |
| 1 | code_177 | Computers & Electronics | Consumer Electronics | Audio Equipment | 0.411838 |
| 1 | code_178 | Computers & Electronics | Consumer Electronics | Camera & Photo Equipment | 0.411838 |
| 1 | code_179 | Computers & Electronics | Consumer Electronics | Car Electronics | 0.411838 |
| 1 | code_180 | Computers & Electronics | Consumer Electronics | Drones & RC Aircraft | 0.411838 |
| 1 | code_181 | Computers & Electronics | Consumer Electronics | Game Systems & Consoles | 0.411838 |
| 1 | code_182 | Computers & Electronics | Consumer Electronics | GPS & Navigation | 0.411838 |
| 1 | code_183 | Computers & Electronics | Consumer Electronics | TV & Video Equipment | 0.437809 |
| 1 | code_184 | Computers & Electronics | Electronics & Electrical | | 0.411838 |
| 1 | code_185 | Computers & Electronics | Electronics & Electrical | Electronic Components | 0.411838 |
| 1 | code_186 | Computers & Electronics | Electronics & Electrical | Power Supplies | 0.411838 |
| 1 | code_187 | Computers & Electronics | Enterprise Technology | | 0.152469 |
| 1 | code_188 | Computers & Electronics | Enterprise Technology | Data Management | 0.152469 |
| 1 | code_189 | Computers & Electronics | Networking | | 0.15329 |
| 1 | code_190 | Computers & Electronics | Networking | Data Formats & Protocols | 0.15329 |
| 1 | code_191 | Computers & Electronics | Networking | Network Monitoring & Management | 0.15329 |
| 1 | code_192 | Computers & Electronics | Networking | VPN & Remote Access | 0.15329 |
| 1 | code_193 | Computers & Electronics | Programming | | 0.15329 |
| 1 | code_194 | Computers & Electronics | Programming | Java (Programming Language) | 0.15329 |
| 1 | code_195 | Computers & Electronics | Software | | 0.15329 |
| 1 | code_196 | Computers & Electronics | Software | Business & Productivity Software | 0.15329 |
| 1 | code_197 | Computers & Electronics | Software | Device Drivers | 0.15329 |
| 1 | code_198 | Computers & Electronics | Software | Internet Software | 0.15329 |
| 1 | code_199 | Computers & Electronics | Software | Multimedia Software | 0.15329 |
| 1 | code_200 | Computers & Electronics | Software | Operating Systems | 0.15329 |
| 1 | code_201 | Computers & Electronics | Software | Software Utilities | 0.15329 |
| 1 | code_202 | Finance | | | 0.102513 |
| 1 | code_203 | Finance | Accounting & Auditing | | 0.113888 |
| 1 | code_204 | Finance | Accounting & Auditing | Billing & Invoicing | 0.113888 |
| 1 | code_205 | Finance | Accounting & Auditing | Tax Preparation & Planning | 0.113888 |
| 1 | code_206 | Finance | Banking | | 0.100455 |
| 1 | code_207 | Finance | Credit & Lending | | 0.113865 |
| 1 | code_208 | Finance | Credit & Lending | Credit Cards | 0.113865 |
| 1 | code_209 | Finance | Credit & Lending | Credit Reporting & Monitoring | 0.113865 |
| 1 | code_210 | Finance | Credit & Lending | Loans | 0.113865 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_211 | Finance | Financial Planning & Management | | 0.102753 |
| 1 | code_212 | Finance | Financial Planning & Management | Retirement & Pension | 0.102753 |
| 1 | code_213 | Finance | Grants Scholarships & Financial Aid | | 0.10596 |
| 1 | code_214 | Finance | Grants Scholarships & Financial Aid | Study Grants & Scholarships | 0.10596 |
| 1 | code_215 | Finance | Insurance | | 0.088193 |
| 1 | code_216 | Finance | Insurance | Health Insurance | 0.088193 |
| 1 | code_217 | Finance | Investing | | 0.113865 |
| 1 | code_218 | Finance | Investing | Commodities & Futures Trading | 0.113865 |
| 1 | code_219 | Finance | Investing | Currencies & Foreign Exchange | 0.113865 |
| 1 | code_220 | Finance | Investing | Stocks & Bonds | 0.113865 |
| 1 | code_221 | Food & Drink | | | 0.469376 |
| 1 | code_222 | Food & Drink | Beverages | | 0.552626 |
| 1 | code_223 | Food & Drink | Beverages | Alcoholic Beverages | 0.727946 |
| 1 | code_224 | Food & Drink | Beverages | Coffee & Tea | 0.725362 |
| 1 | code_225 | Food & Drink | Beverages | Juice | 0.42134 |
| 1 | code_226 | Food & Drink | Beverages | Soft Drinks | 0.42134 |
| 1 | code_227 | Food & Drink | Cooking & Recipes | | 0.758549 |
| 1 | code_228 | Food & Drink | Cooking & Recipes | BBQ & Grilling | 0.811989 |
| 1 | code_229 | Food & Drink | Cooking & Recipes | Desserts | 0.735321 |
| 1 | code_230 | Food & Drink | Cooking & Recipes | Soups & Stews | 0.811989 |
| 1 | code_231 | Food & Drink | Food | | 0.758549 |
| 1 | code_232 | Food & Drink | Food & Grocery Retailers | | 0.758549 |
| 1 | code_233 | Food & Drink | Food | Baked Goods | 0.734311 |
| 1 | code_234 | Food & Drink | Food | Breakfast Foods | 0.735321 |
| 1 | code_235 | Food & Drink | Food | Candy & Sweets | 0.81948 |
| 1 | code_236 | Food & Drink | Food | Grains & Pasta | 0.747368 |
| 1 | code_237 | Food & Drink | Food | Meat & Seafood | 0.803618 |
| 1 | code_238 | Food & Drink | Food | Snack Foods | 0.756658 |
| 1 | code_239 | Food & Drink | Restaurants | | 0.31608 |
| 1 | code_240 | Food & Drink | Restaurants | Fast Food | 0.31608 |
| 1 | code_241 | Food & Drink | Restaurants | Pizzerias | 0.31608 |
| 1 | code_242 | Food & Drink | Restaurants | Restaurant Reviews & Reservations | 0.31608 |
| 1 | code_243 | Games | | | 0.219961 |
| 1 | code_244 | Games | Arcade & Coin-Op Games | | 0.222804 |
| 1 | code_245 | Games | Board Games | | 0.222804 |
| 1 | code_246 | Games | Board Games | Chess & Abstract Strategy Games | 0.222804 |
| 1 | code_247 | Games | Board Games | Miniatures & Wargaming | 0.222804 |
| 1 | code_248 | Games | Card Games | | 0.222804 |
| 1 | code_249 | Games | Card Games | Collectible Card Games | 0.222804 |
| 1 | code_250 | Games | Card Games | Poker & Casino Games | 0.094973 |
| 1 | code_251 | Games | Computer & Video Games | | 0.279959 |
| 1 | code_252 | Games | Computer & Video Games | Casual Games | 0.279959 |
| 1 | code_253 | Games | Computer & Video Games | Driving & Racing Games | 0.279959 |
| 1 | code_254 | Games | Computer & Video Games | Fighting Games | 0.279959 |
| 1 | code_255 | Games | Computer & Video Games | Music & Dance Games | 0.279959 |
| 1 | code_256 | Games | Computer & Video Games | Sandbox Games | 0.279959 |
| 1 | code_257 | Games | Computer & Video Games | Shooter Games | 0.279959 |
| 1 | code_258 | Games | Computer & Video Games | Simulation Games | 0.279959 |
| 1 | code_259 | Games | Computer & Video Games | Sports Games | 0.279959 |
| 1 | code_260 | Games | Computer & Video Games | Strategy Games | 0.279959 |
| 1 | code_261 | Games | Computer & Video Games | Video Game Emulation | 0.279959 |
| 1 | code_262 | Games | Family-Oriented Games & Activities | | 0.222804 |
| 1 | code_263 | Games | Family-Oriented Games & Activities | Drawing & Coloring | 0.222804 |
| 1 | code_264 | Games | Family-Oriented Games & Activities | Dress-Up & Fashion Games | 0.222804 |
| 1 | code_265 | Games | Gambling | | 0.094973 |
| 1 | code_266 | Games | Gambling | Lottery | 0.094973 |
| 1 | code_267 | Games | Online Games | Massively Multiplayer Games | 0.279959 |
| 1 | code_268 | Games | Puzzles & Brainteasers | | 0.222804 |
| 1 | code_269 | Games | Roleplaying Games | | 0.222804 |
| 1 | code_270 | Games | Table Games | | 0.222804 |
| 1 | code_271 | Games | Table Games | Billiards | 0.222804 |
| 1 | code_272 | Games | Word Games | | 0.222804 |
| 1 | code_273 | Health | | | 0.258847 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_274 | Health | Aging & Geriatrics | | 0.260433 |
| 1 | code_275 | Health | Health Conditions | | 0.306182 |
| 1 | code_276 | Health | Health Conditions | AIDS & HIV | 0.306182 |
| 1 | code_277 | Health | Health Conditions | Allergies | 0.306182 |
| 1 | code_278 | Health | Health Conditions | Arthritis | 0.306182 |
| 1 | code_279 | Health | Health Conditions | Cancer | 0.306182 |
| 1 | code_280 | Health | Health Conditions | Diabetes | 0.306182 |
| 1 | code_281 | Health | Health Conditions | Ear Nose & Throat | 0.306182 |
| 1 | code_282 | Health | Health Conditions | Eating Disorders | 0.306182 |
| 1 | code_283 | Health | Health Conditions | Endocrine Conditions | 0.306182 |
| 1 | code_284 | Health | Health Conditions | Genetic Disorders | 0.306182 |
| 1 | code_285 | Health | Health Conditions | Heart & Hypertension | 0.306182 |
| 1 | code_286 | Health | Health Conditions | Infectious Diseases | 0.306182 |
| 1 | code_287 | Health | Health Conditions | Neurological Conditions | 0.306182 |
| 1 | code_288 | Health | Health Conditions | Obesity | 0.306182 |
| 1 | code_289 | Health | Health Conditions | Pain Management | 0.306182 |
| 1 | code_290 | Health | Health Conditions | Respiratory Conditions | 0.306182 |
| 1 | code_291 | Health | Health Conditions | Skin Conditions | 0.306182 |
| 1 | code_292 | Health | Health Conditions | Sleep Disorders | 0.306182 |
| 1 | code_293 | Health | Health Education & Medical Training | | 0.275678 |
| 1 | code_294 | Health | Health Foundations & Medical Research | | 0.275678 |
| 1 | code_295 | Health | Medical Devices & Equipment | | 0.299749 |
| 1 | code_296 | Health | Medical Facilities & Services | | 0.299749 |
| 1 | code_297 | Health | Medical Facilities & Services | Doctors' Offices | 0.291932 |
| 1 | code_298 | Health | Medical Facilities & Services | Hospitals & Treatment Centers | 0.291932 |
| 1 | code_299 | Health | Medical Facilities & Services | Medical Procedures | 0.291932 |
| 1 | code_300 | Health | Medical Facilities & Services | Physical Therapy | 0.291932 |
| 1 | code_301 | Health | Men's Health | | 0.306182 |
| 1 | code_302 | Health | Mental Health | | 0.306182 |
| 1 | code_303 | Health | Mental Health | Anxiety & Stress | 0.306182 |
| 1 | code_304 | Health | Mental Health | Depression | 0.306182 |
| 1 | code_305 | Health | Nursing | | 0.291932 |
| 1 | code_306 | Health | Nursing | Assisted Living & Long Term Care | 0.27296 |
| 1 | code_307 | Health | Nutrition | | 0.306182 |
| 1 | code_308 | Health | Nutrition | Special & Restricted Diets | 0.306182 |
| 1 | code_309 | Health | Nutrition | Vitamins & Supplements | 0.306182 |
| 1 | code_310 | Health | Oral & Dental Care | | 0.31269 |
| 1 | code_311 | Health | Pharmacy | | 0.306182 |
| 1 | code_312 | Health | Pharmacy | Drugs & Medications | 0.306182 |
| 1 | code_313 | Health | Public Health | | 0.291932 |
| 1 | code_314 | Health | Public Health | Occupational Health & Safety | 0.291932 |
| 1 | code_315 | Health | Reproductive Health | | 0.306182 |
| 1 | code_316 | Health | Substance Abuse | | 0.306182 |
| 1 | code_317 | Health | Substance Abuse | Drug & Alcohol Testing | 0.288524 |
| 1 | code_318 | Health | Substance Abuse | Drug & Alcohol Treatment | 0.285968 |
| 1 | code_319 | Health | Substance Abuse | Smoking & Smoking Cessation | 0.306182 |
| 1 | code_320 | Health | Substance Abuse | Steroids & Performance-Enhancing Drugs | 0.306182 |
| 1 | code_321 | Health | Vision Care | | 0.31269 |
| 1 | code_322 | Health | Vision Care | Eyeglasses & Contacts | 0.31269 |
| 1 | code_323 | Health | Women's Health | | 0.306182 |
| 1 | code_324 | Hobbies & Leisure | | | 0.327003 |
| 1 | code_325 | Hobbies & Leisure | Clubs & Organizations | | 0.175955 |
| 1 | code_326 | Hobbies & Leisure | Clubs & Organizations | Youth Organizations & Resources | 0.175955 |
| 1 | code_327 | Hobbies & Leisure | Crafts | | 0.749917 |
| 1 | code_328 | Hobbies & Leisure | Crafts | Fiber & Textile Arts | 0.749917 |
| 1 | code_329 | Hobbies & Leisure | Merit Prizes & Contests | | 0.221568 |
| 1 | code_330 | Hobbies & Leisure | Outdoors | | 0.222804 |
| 1 | code_331 | Hobbies & Leisure | Outdoors | Fishing | 0.222804 |
| 1 | code_332 | Hobbies & Leisure | Outdoors | Hiking & Camping | 0.222804 |
| 1 | code_333 | Hobbies & Leisure | Paintball | | 0.222804 |
| 1 | code_334 | Hobbies & Leisure | Radio Control & Modeling | | 0.391751 |
| 1 | code_335 | Hobbies & Leisure | Radio Control & Modeling | Model Trains & Railroads | 0.391751 |
| 1 | code_336 | Hobbies & Leisure | Special Occasions | | 0.250713 |
| 1 | code_337 | Hobbies & Leisure | Special Occasions | Holidays & Seasonal Events | 0.250713 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_338 | Hobbies & Leisure | Special Occasions | Weddings | 0.269648 |
| 1 | code_339 | Hobbies & Leisure | Water Activities | | 0.222804 |
| 1 | code_340 | Hobbies & Leisure | Water Activities | Boating | 0.230317 |
| 1 | code_341 | Hobbies & Leisure | Water Activities | Surf & Swim | 0.222804 |
| 1 | code_342 | Home & Garden | | | 0.162967 |
| 1 | code_343 | Home & Garden | Bed & Bath | | 0.115391 |
| 1 | code_344 | Home & Garden | Bed & Bath | Bathroom | 0.115391 |
| 1 | code_345 | Home & Garden | Domestic Services | | 0.038032 |
| 1 | code_346 | Home & Garden | Domestic Services | Cleaning Services | 0.038032 |
| 1 | code_347 | Home & Garden | Gardening & Landscaping | | 0.099156 |
| 1 | code_348 | Home & Garden | Home & Interior Decor | | 0.057796 |
| 1 | code_349 | Home & Garden | Home Appliances | | 0.074543 |
| 1 | code_350 | Home & Garden | Home Furnishings | | 0.057578 |
| 1 | code_351 | Home & Garden | Home Furnishings | Curtains & Window Treatments | 0.057578 |
| 1 | code_352 | Home & Garden | Home Furnishings | Kitchen & Dining Furniture | 0.057578 |
| 1 | code_353 | Home & Garden | Home Furnishings | Lamps & Lighting | 0.064357 |
| 1 | code_354 | Home & Garden | Home Furnishings | Living Room Furniture | 0.057578 |
| 1 | code_355 | Home & Garden | Home Furnishings | Rugs & Carpets | 0.084417 |
| 1 | code_356 | Home & Garden | Home Improvement | | 0.103608 |
| 1 | code_357 | Home & Garden | Home Improvement | Construction & Power Tools | 0.108453 |
| 1 | code_358 | Home & Garden | Home Improvement | Doors & Windows | 0.103619 |
| 1 | code_359 | Home & Garden | Home Improvement | Flooring | 0.103619 |
| 1 | code_360 | Home & Garden | Home Improvement | House Painting & Finishing | 0.041905 |
| 1 | code_361 | Home & Garden | Home Improvement | Plumbing | 0.103608 |
| 1 | code_362 | Home & Garden | Home Safety & Security | | 0.099156 |
| 1 | code_363 | Home & Garden | Home Storage & Shelving | | 0.1154 |
| 1 | code_364 | Home & Garden | Home Swimming Pools Saunas & Spas | | 0.099169 |
| 1 | code_365 | Home & Garden | HVAC & Climate Control | | 0.103514 |
| 1 | code_366 | Home & Garden | HVAC & Climate Control | Fireplaces & Stoves | 0.103514 |
| 1 | code_367 | Home & Garden | Kitchen & Dining | | 0.133682 |
| 1 | code_368 | Home & Garden | Kitchen & Dining | Cookware & Diningware | 0.061086 |
| 1 | code_369 | Home & Garden | Kitchen & Dining | Major Kitchen Appliances | 0.047058 |
| 1 | code_370 | Home & Garden | Kitchen & Dining | Small Kitchen Appliances | 0.047058 |
| 1 | code_371 | Home & Garden | Laundry | | 0.047058 |
| 1 | code_372 | Home & Garden | Laundry | Washers & Dryers | 0.047058 |
| 1 | code_373 | Home & Garden | Nursery & Playroom | | 0.115475 |
| 1 | code_374 | Home & Garden | Pest Control | | 0.039885 |
| 1 | code_375 | Home & Garden | Yard & Patio | | 0.110967 |
| 1 | code_376 | Home & Garden | Yard & Patio | Lawn Mowers | 0.108356 |
| 1 | code_377 | Internet & Telecom | | | 0.231717 |
| 1 | code_378 | Internet & Telecom | Communications Equipment | | 0.296539 |
| 1 | code_379 | Internet & Telecom | Communications Equipment | Radio Equipment | 0.296539 |
| 1 | code_380 | Internet & Telecom | Email & Messaging | | 0.17321 |
| 1 | code_381 | Internet & Telecom | Email & Messaging | Text & Instant Messaging | 0.17321 |
| 1 | code_382 | Internet & Telecom | Email & Messaging | Voice & Video Chat | 0.17321 |
| 1 | code_383 | Internet & Telecom | Mobile & Wireless | | 0.27693 |
| 1 | code_384 | Internet & Telecom | Mobile & Wireless | Mobile & Wireless Accessories | 0.27693 |
| 1 | code_385 | Internet & Telecom | Mobile & Wireless | Mobile Apps & Add-Ons | 0.27693 |
| 1 | code_386 | Internet & Telecom | Mobile & Wireless | Mobile Phones | 0.27683 |
| 1 | code_387 | Internet & Telecom | Service Providers | | 0.17321 |
| 1 | code_388 | Internet & Telecom | Service Providers | Cable & Satellite Providers | 0.17321 |
| 1 | code_389 | Internet & Telecom | Web Services | | 0.15329 |
| 1 | code_390 | Internet & Telecom | Web Services | Affiliate Programs | 0.15329 |
| 1 | code_391 | Internet & Telecom | Web Services | Web Design & Development | 0.15329 |
| 1 | code_392 | Jobs & Education | | | 0.102602 |
| 1 | code_393 | Jobs & Education | Education | | 0.10258 |
| 1 | code_394 | Jobs & Education | Education | Colleges & Universities | 0.10258 |
| 1 | code_395 | Jobs & Education | Education | Distance Learning | 0.10258 |
| 1 | code_396 | Jobs & Education | Education | Homeschooling | 0.10258 |
| 1 | code_397 | Jobs & Education | Education | Primary & Secondary Schooling (K-12) | 0.10258 |
| 1 | code_398 | Jobs & Education | Education | Standardized & Admissions Tests | 0.10258 |
| 1 | code_399 | Jobs & Education | Education | Teaching & Classroom Resources | 0.10258 |
| 1 | code_400 | Jobs & Education | Education | Training & Certification | 0.10258 |
| 1 | code_401 | Jobs & Education | Education | Vocational & Continuing Education | 0.10258 |
| 1 | code_402 | Jobs & Education | Jobs | | 0.117628 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_403 | Jobs & Education | Jobs | Career Resources & Planning | 0.117628 |
| 1 | code_404 | Jobs & Education | Jobs | Job Listings | 0.113208 |
| 1 | code_405 | Jobs & Education | Jobs | Resumes & Portfolios | 0.117628 |
| 1 | code_406 | Law & Government | | | 0 |
| 1 | code_407 | Law & Government | Government | | 0 |
| 1 | code_408 | Law & Government | Government | Courts & Judiciary | 0 |
| 1 | code_409 | Law & Government | Government | Visa & Immigration | 0 |
| 1 | code_410 | Law & Government | Legal | | 0 |
| 1 | code_411 | Law & Government | Legal | Bankruptcy | 0 |
| 1 | code_412 | Law & Government | Legal | Legal Education | 0 |
| 1 | code_413 | Law & Government | Legal | Legal Services | 0 |
| 1 | code_414 | Law & Government | Military | | 0 |
| 1 | code_415 | Law & Government | Public Safety | | 0 |
| 1 | code_416 | Law & Government | Public Safety | Crime & Justice | 0 |
| 1 | code_417 | Law & Government | Public Safety | Emergency Services | 0 |
| 1 | code_418 | Law & Government | Public Safety | Law Enforcement | 0 |
| 1 | code_419 | Law & Government | Public Safety | Security Products & Services | 0 |
| 1 | code_420 | Law & Government | Social Services | | 0 |
| 1 | code_421 | News | | | 0.1609 |
| 1 | code_422 | News | Business News | | 0.1609 |
| 1 | code_423 | News | Business News | Company News | 0.1609 |
| 1 | code_424 | News | Business News | Financial Markets News | 0.1609 |
| 1 | code_425 | News | Gossip & Tabloid News | Scandals & Investigations | 0.1609 |
| 1 | code_426 | News | Health News | | 0.1609 |
| 1 | code_427 | News | Politics | | 0.1609 |
| 1 | code_428 | News | Sports News | | 0.1609 |
| 1 | code_429 | News | Weather | | 0.1609 |
| 1 | code_430 | Online Communities | | | 0 |
| 1 | code_431 | Online Communities | Blogging Resources & Services | | 0 |
| 1 | code_432 | Online Communities | Dating & Personals | | 0 |
| 1 | code_433 | Online Communities | Dating & Personals | Matrimonial Services | 0 |
| 1 | code_434 | Online Communities | Dating & Personals | Personals | 0 |
| 1 | code_435 | Online Communities | Dating & Personals | Photo Rating Sites | 0 |
| 1 | code_436 | Online Communities | File Sharing & Hosting | | 0 |
| 1 | code_437 | Online Communities | Online Goodies | | 0 |
| 1 | code_438 | Online Communities | Online Goodies | Clip Art & Animated GIFs | 0 |
| 1 | code_439 | Online Communities | Online Goodies | Skins  Themes & Wallpapers | 0 |
| 1 | code_440 | Online Communities | Online Goodies | Social Network Apps & Add-Ons | 0 |
| 1 | code_441 | Online Communities | Photo & Video Sharing | | 0 |
| 1 | code_442 | Online Communities | Photo & Video Sharing | Photo & Image Sharing | 0 |
| 1 | code_443 | Online Communities | Social Networks | | 0 |
| 1 | code_444 | Online Communities | Virtual Worlds | | 0 |
| 1 | code_445 | People & Society | | | 0 |
| 1 | code_446 | People & Society | Family & Relationships | | 0 |
| 1 | code_447 | People & Society | Family & Relationships | Family | 0 |
| 1 | code_448 | People & Society | Family & Relationships | Marriage | 0 |
| 1 | code_449 | People & Society | Family & Relationships | Troubled Relationships | 0 |
| 1 | code_450 | People & Society | Kids & Teens | | 0 |
| 1 | code_451 | People & Society | Kids & Teens | Children's Interests | 0 |
| 1 | code_452 | People & Society | Kids & Teens | Teen Interests | 0 |
| 1 | code_453 | People & Society | Religion & Belief | | 0 |
| 1 | code_454 | People & Society | Seniors & Retirement | | 0 |
| 1 | code_455 | People & Society | Social Issues & Advocacy | | 0 |
| 1 | code_456 | People & Society | Social Issues & Advocacy | Charity & Philanthropy | 0 |
| 1 | code_457 | People & Society | Social Issues & Advocacy | Discrimination & Identity Relations | 0 |
| 1 | code_458 | People & Society | Social Issues & Advocacy | Green Living & Environmental Issues | 0 |
| 1 | code_459 | People & Society | Social Issues & Advocacy | Human Rights & Liberties | 0 |
| 1 | code_460 | People & Society | Social Issues & Advocacy | Poverty & Hunger | 0 |
| 1 | code_461 | People & Society | Social Issues & Advocacy | Work & Labor Issues | 0 |
| 1 | code_462 | People & Society | Social Sciences | | 0 |
| 1 | code_463 | People & Society | Social Sciences | Economics | 0 |
| 1 | code_464 | People & Society | Social Sciences | Political Science | 0 |
| 1 | code_465 | People & Society | Social Sciences | Psychology | 0 |
| 1 | code_466 | People & Society | Subcultures & Niche Interests | | 0 |
| 1 | code_467 | Pets & Animals | | | 0.49728 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_468 | Pets & Animals | Animal Products & Services | Pet Food & Supplies | 0.49728 |
| 1 | code_469 | Pets & Animals | Animal Products & Services | Veterinarians | 0.08693 |
| 1 | code_470 | Pets & Animals | Pets | | 0.08693 |
| 1 | code_471 | Pets & Animals | Pets | Birds | 0.08693 |
| 1 | code_472 | Pets & Animals | Pets | Cats | 0.08693 |
| 1 | code_473 | Pets & Animals | Pets | Dogs | 0.08693 |
| 1 | code_474 | Pets & Animals | Pets | Exotic Pets | 0.08693 |
| 1 | code_475 | Pets & Animals | Pets | Fish & Aquaria | 0.08693 |
| 1 | code_476 | Pets & Animals | Pets | Horses | 0.08693 |
| 1 | code_477 | Pets & Animals | Pets | Rabbits & Rodents | 0.08693 |
| 1 | code_478 | Pets & Animals | Pets | Reptiles & Amphibians | 0.08693 |
| 1 | code_479 | Pets & Animals | Wildlife | | 0.08693 |
| 1 | code_480 | Real Estate | | | 0.084914 |
| 1 | code_481 | Real Estate | Real Estate Listings | | 0.084914 |
| 1 | code_482 | Real Estate | Real Estate Listings | Bank-Owned & Foreclosed Properties | 0.084914 |
| 1 | code_483 | Real Estate | Real Estate Listings | Commercial Properties | 0.084914 |
| 1 | code_484 | Real Estate | Real Estate Listings | Lots & Land | 0.084914 |
| 1 | code_485 | Real Estate | Real Estate Listings | Residential Rentals | 0.084914 |
| 1 | code_486 | Real Estate | Real Estate Listings | Residential Sales | 0.084914 |
| 1 | code_487 | Real Estate | Real Estate Listings | Timeshares & Vacation Properties | 0.084914 |
| 1 | code_488 | Real Estate | Real Estate Services | | 0.084914 |
| 1 | code_489 | Reference | | | 0 |
| 1 | code_490 | Reference | Directories & Listings | | 0 |
| 1 | code_491 | Reference | Directories & Listings | Business & Personal Listings | 0 |
| 1 | code_492 | Reference | General Reference | | 0 |
| 1 | code_493 | Reference | General Reference | Biographies & Quotations | 0 |
| 1 | code_494 | Reference | General Reference | Calculators & Reference Tools | 0 |
| 1 | code_495 | Reference | General Reference | Dictionaries & Encyclopedias | 0 |
| 1 | code_496 | Reference | General Reference | Forms Guides & Templates | 0 |
| 1 | code_497 | Reference | General Reference | Public Records | 0 |
| 1 | code_498 | Reference | General Reference | Time & Calendars | 0 |
| 1 | code_499 | Reference | Geographic Reference | | 0 |
| 1 | code_500 | Reference | Geographic Reference | Maps | 0 |
| 1 | code_501 | Reference | Humanities | | 0 |
| 1 | code_502 | Reference | Humanities | History | 0 |
| 1 | code_503 | Reference | Humanities | Myth & Folklore | 0 |
| 1 | code_504 | Reference | Humanities | Philosophy | 0 |
| 1 | code_505 | Reference | Language Resources | | 0 |
| 1 | code_506 | Reference | Language Resources | Foreign Language Resources | 0 |
| 1 | code_507 | Reference | Libraries & Museums | | 0 |
| 1 | code_508 | Reference | Libraries & Museums | Museums | 0 |
| 1 | code_509 | Science | | | 0.185857 |
| 1 | code_510 | Science | Astronomy | | 0.185857 |
| 1 | code_511 | Science | Biological Sciences | | 0.185857 |
| 1 | code_512 | Science | Biological Sciences | Neuroscience | 0.185857 |
| 1 | code_513 | Science | Chemistry | | 0.185857 |
| 1 | code_514 | Science | Computer Science | | 0.185857 |
| 1 | code_515 | Science | Earth Sciences | | 0.185857 |
| 1 | code_516 | Science | Earth Sciences | Atmospheric Science | 0.185857 |
| 1 | code_517 | Science | Earth Sciences | Geology | 0.185857 |
| 1 | code_518 | Science | Ecology & Environment | | 0.185857 |
| 1 | code_519 | Science | Ecology & Environment | Climate Change & Global Warming | 0.185857 |
| 1 | code_520 | Science | Engineering & Technology | | 0.185857 |
| 1 | code_521 | Science | Engineering & Technology | Robotics | 0.185857 |
| 1 | code_522 | Science | Mathematics | | 0.185857 |
| 1 | code_523 | Science | Mathematics | Statistics | 0.185857 |
| 1 | code_524 | Science | Physics | | 0.185857 |
| 1 | code_525 | Science | Scientific Institutions | | 0.185857 |
| 1 | code_526 | Sensitive Subjects | | | 0.265295 |
| 1 | code_527 | Shopping | | | 0.40695 |
| 1 | code_528 | Shopping | Antiques & Collectibles | | 0.524893 |
| 1 | code_529 | Shopping | Apparel | | 0.72111 |
| 1 | code_530 | Shopping | Apparel | Athletic Apparel | 0.736107 |
| 1 | code_531 | Shopping | Apparel | Casual Apparel | 0.729235 |
| 1 | code_532 | Shopping | Apparel | Children's Clothing | 0.729235 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_533 | Shopping | Apparel | Clothing Accessories | 0.729235 |
| 1 | code_534 | Shopping | Apparel | Costumes | 0.729235 |
| 1 | code_535 | Shopping | Apparel | Eyewear | 0.736107 |
| 1 | code_536 | Shopping | Apparel | Footwear | 0.718746 |
| 1 | code_537 | Shopping | Apparel | Formal Wear | 0.736107 |
| 1 | code_538 | Shopping | Apparel | Headwear | 0.736107 |
| 1 | code_539 | Shopping | Apparel | Men's Clothing | 0.729235 |
| 1 | code_540 | Shopping | Apparel | Swimwear | 0.736107 |
| 1 | code_541 | Shopping | Apparel | Undergarments | 0.736107 |
| 1 | code_542 | Shopping | Apparel | Women's Clothing | 0.729235 |
| 1 | code_543 | Shopping | Auctions | | 0.50949 |
| 1 | code_544 | Shopping | Classifieds | | 0.110192 |
| 1 | code_545 | Shopping | Consumer Resources | | 0 |
| 1 | code_546 | Shopping | Consumer Resources | Consumer Advocacy & Protection | 0 |
| 1 | code_547 | Shopping | Consumer Resources | Coupons & Discount Offers | 0 |
| 1 | code_548 | Shopping | Consumer Resources | Product Reviews & Price Comparisons | 0 |
| 1 | code_549 | Shopping | Entertainment Media | | 0.222358 |
| 1 | code_550 | Shopping | Entertainment Media | Entertainment Media Rentals | 0.136983 |
| 1 | code_551 | Shopping | Gifts & Special Event Items | | 0.508223 |
| 1 | code_552 | Shopping | Gifts & Special Event Items | Cards & Greetings | 0.75558 |
| 1 | code_553 | Shopping | Gifts & Special Event Items | Flowers | 0.193032 |
| 1 | code_554 | Shopping | Gifts & Special Event Items | Gifts | 0.508223 |
| 1 | code_555 | Shopping | Luxury Goods | | 0.627033 |
| 1 | code_556 | Shopping | Mass Merchants & Department Stores | | 0.486517 |
| 1 | code_557 | Shopping | Photo & Video Services | | 0.436448 |
| 1 | code_558 | Shopping | Tobacco Products | | 1.005129 |
| 1 | code_559 | Shopping | Toys | | 0.619265 |
| 1 | code_560 | Shopping | Toys | Building Toys | 0.619265 |
| 1 | code_561 | Shopping | Toys | Die-cast & Toy Vehicles | 0.619265 |
| 1 | code_562 | Shopping | Toys | Dolls & Accessories | 0.619265 |
| 1 | code_563 | Shopping | Toys | Ride-On Toys & Wagons | 0.619265 |
| 1 | code_564 | Shopping | Toys | Stuffed Toys | 0.619265 |
| 1 | code_565 | Sports | | | 0.175955 |
| 1 | code_566 | Sports | Animal Sports | | 0.175955 |
| 1 | code_567 | Sports | College Sports | | 0.175955 |
| 1 | code_568 | Sports | Combat Sports | | 0.175955 |
| 1 | code_569 | Sports | Combat Sports | Boxing | 0.175955 |
| 1 | code_570 | Sports | Combat Sports | Martial Arts | 0.175955 |
| 1 | code_571 | Sports | Combat Sports | Wrestling | 0.175955 |
| 1 | code_572 | Sports | Extreme Sports | | 0.175955 |
| 1 | code_573 | Sports | Extreme Sports | Drag & Street Racing | 0.175955 |
| 1 | code_574 | Sports | Fantasy Sports | | 0.175955 |
| 1 | code_575 | Sports | Individual Sports | | 0.175955 |
| 1 | code_576 | Sports | Individual Sports | Cycling | 0.175955 |
| 1 | code_577 | Sports | Individual Sports | Golf | 0.175955 |
| 1 | code_578 | Sports | Individual Sports | Gymnastics | 0.175955 |
| 1 | code_579 | Sports | Individual Sports | Racquet Sports | 0.175955 |
| 1 | code_580 | Sports | Individual Sports | Skate Sports | 0.175955 |
| 1 | code_581 | Sports | Individual Sports | Track & Field | 0.175955 |
| 1 | code_582 | Sports | International Sports Competitions | | 0.222804 |
| 1 | code_583 | Sports | International Sports Competitions | Olympics | 0.222804 |
| 1 | code_584 | Sports | Motor Sports | | 0.222804 |
| 1 | code_585 | Sports | Sporting Goods | | 0.222804 |
| 1 | code_586 | Sports | Sporting Goods | Sports Memorabilia | 0.222804 |
| 1 | code_587 | Sports | Sporting Goods | Winter Sports Equipment | 0.222804 |
| 1 | code_588 | Sports | Sports Coaching & Training | | 0.175955 |
| 1 | code_589 | Sports | Team Sports | | 0.175955 |
| 1 | code_590 | Sports | Team Sports | American Football | 0.175955 |
| 1 | code_591 | Sports | Team Sports | Australian Football | 0.175955 |
| 1 | code_592 | Sports | Team Sports | Baseball | 0.175955 |
| 1 | code_593 | Sports | Team Sports | Basketball | 0.175955 |
| 1 | code_594 | Sports | Team Sports | Cheerleading | 0.175955 |
| 1 | code_595 | Sports | Team Sports | Cricket | 0.175955 |
| 1 | code_596 | Sports | Team Sports | Hockey | 0.175955 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | code_597 | Sports | Team Sports | Rugby | 0.175955 |
| 1 | code_598 | Sports | Team Sports | Soccer | 0.175955 |
| 1 | code_599 | Sports | Team Sports | Volleyball | 0.175955 |
| 1 | code_600 | Sports | Water Sports | | 0.175955 |
| 1 | code_601 | Sports | Water Sports | Surfing | 0.175955 |
| 1 | code_602 | Sports | Water Sports | Swimming | 0.175955 |
| 1 | code_603 | Sports | Winter Sports | | 0.175955 |
| 1 | code_604 | Sports | Winter Sports | Ice Skating | 0.175955 |
| 1 | code_605 | Sports | Winter Sports | Skiing & Snowboarding | 0.175955 |
| 6 | code_606 | Travel | | | 0.601042 |
| 6 | code_607 | Travel | Air Travel | | 0.947955 |
| 6 | code_608 | Travel | Air Travel | Airport Parking & Transportation | 0.620678 |
| 6 | code_609 | Travel | Bus & Rail | | 0.676149 |
| 6 | code_610 | Travel | Car Rental & Taxi Services | | 0.313692 |
| 6 | code_611 | Travel | Cruises & Charters | | 1.069828 |
| 6 | code_612 | Travel | Hotels & Accommodations | | 0.269648 |
| 6 | code_613 | Travel | Hotels & Accommodations | Vacation Rentals & Short-Term Stays | 0.269648 |
| 6 | code_614 | Travel | Specialty Travel | | 1.042665 |
| 1 | code_615 | Travel | Tourist Destinations | | 0.269648 |
| 1 | code_616 | Travel | Tourist Destinations | Beaches & Islands | 0.250713 |
| 1 | code_617 | Travel | Tourist Destinations | Mountain & Ski Resorts | 0.250713 |
| 1 | code_618 | Travel | Tourist Destinations | Regional Parks & Gardens | 0.250713 |
| 1 | code_619 | Travel | Tourist Destinations | Theme Parks | 0.250713 |
| 1 | code_620 | Travel | Tourist Destinations | Zoos-Aquariums-Preserves | 0.250713 |
| 1 | code_901 | Payroll | | | 0 |
| 1 | code_902 | Industry adjustment | | | 0 |
| 1 | code_903 | Business & Industrial | Energy & Utilities | Water | 0.34626 |
| 5 | code_904 | Business & Industrial | Energy & Utilities | Waste & Recycling | 1.392327 |

## Table S5. Industry interquartile and median emissions estimates by Scope 3 category.

| SIC | Description | Category 1. Purchased Goods and Services | | | Category 3. Fuel and energy-related activities | | | Category 6. Business Travel | | | Category 9. Downstream transportation and distribution | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Lower 25% | Median | Upper 25% | Lower 25% | Median | Upper 25% | Lower 25% | Median | Upper 25% | Lower 25% | Median | Upper 25% |
| | | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* | *kg Co2e* |
| **2** | **Forestry and logging** | **14,775** | **36,494** | **92,710** | **6,582** | **2,812** | **3,551** | **573** | **1,196** | **2,417** | **72** | **533** | **2,342** |
| **3** | **Fishing and aquaculture** | **38,438** | **91,763** | **235,545** | **3,330** | **1,504** | **2,328** | **590** | **1,373** | **3,114** | **20** | **217** | **960** |
| **8** | **Other mining and quarrying** | **61,734** | **142,245** | **460,682** | **-** | **-** | **-** | **771** | **1,511** | **3,557** | **423** | **3,815** | **18,208** |
| **10** | **Manufacture of food products** | **36,631** | **119,011** | **444,999** | **16,261** | **4,960** | **13,469** | **224** | **646** | **1,423** | **227** | **2,125** | **12,001** |
| **11** | **Manufacture of beverages** | **40,461** | **91,025** | **256,189** | **11,316** | **4,529** | **8,309** | **368** | **788** | **1,555** | **533** | **2,579** | **8,908** |
| **13** | **Manufacture of textiles** | **15,526** | **46,066** | **136,682** | **4,089** | **1,914** | **2,889** | **257** | **594** | **1,509** | **62** | **555** | **5,916** |
| **14** | **Manufacture of wearing apparel** | **23,537** | **57,946** | **174,765** | **321** | **168** | **874** | **228** | **443** | **980** | **125** | **677** | **7,893** |
| **16** | **Manufacture of wood products** | **18,194** | **45,484** | **146,029** | **4,562** | **2,195** | **3,406** | **455** | **899** | **1,773** | **32** | **265** | **2,469** |
| **17** | **Manufacture of paper and paper products** | **43,798** | **153,379** | **605,264** | **1,103** | **268** | **3,762** | **355** | **1,105** | **2,084** | **1,050** | **10,550** | **55,218** |
| **18** | **Printing and reproduction of recorded media** | **15,249** | **35,972** | **102,040** | **4,870** | **2,666** | **4,033** | **229** | **563** | **1,498** | **125** | **1,034** | **5,327** |
| **20** | **Manufacture of chemicals and chemical products** | **26,841** | **107,952** | **471,067** | **20,404** | **5,361** | **8,463** | **296** | **836** | **1,926** | **468** | **4,040** | **23,130** |
| **22** | **Manufacture of rubber and plastic products** | **57,893** | **179,101** | **482,234** | **14,978** | **5,208** | **10,213** | **466** | **1,102** | **2,457** | **261** | **2,185** | **12,810** |
| **23** | **Manufacture of other non-metallic mineral products** | **18,655** | **62,304** | **203,942** | **10,465** | **3,013** | **5,311** | **423** | **900** | **2,168** | **72** | **876** | **5,665** |
| **24** | **Manufacture of basic metals** | **33,110** | **107,509** | **270,454** | **14,710** | **5,745** | **8,641** | **555** | **1,184** | **2,430** | **224** | **1,559** | **7,482** |
| **25** | **Manufacture of fabricated metal products** | **22,842** | **63,454** | **189,120** | **7,941** | **3,006** | **5,319** | **491** | **1,072** | **2,189** | **59** | **490** | **3,281** |
| **26** | **Manufacture of computer, electronic and optical products** | **24,177** | **75,631** | **247,600** | **10,374** | **2,787** | **5,336** | **296** | **888** | **2,050** | **159** | **1,794** | **8,294** |
| **27** | **Manufacture of electrical equipment** | **17,797** | **46,999** | **146,190** | **5,938** | **2,293** | **3,496** | **522** | **1,121** | **2,269** | **31** | **395** | **4,160** |
| **28** | **Manufacture of machinery and equipment nec** | **40,104** | **110,142** | **318,705** | **11,343** | **4,454** | **6,916** | **586** | **1,361** | **2,874** | **275** | **1,920** | **8,512** |
| **29** | **Manufacture of motor vehicles, trailers and semi-trailers** | **11,888** | **37,870** | **130,836** | **5,469** | **2,192** | **3,852** | **388** | **1,034** | **2,041** | **84** | **1,045** | **5,443** |
| **30** | **Manufacture of other transport equipment** | **13,365** | **34,844** | **120,090** | **4,386** | **2,457** | **3,969** | **792** | **1,883** | **4,175** | **54** | **662** | **4,850** |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 31 | Manufacture of furniture | 15,254 | 42,819 | 141,293 | 5,973 | 2,485 | 4,151 | 448 | 975 | 2,087 | 65 | 514 | 4,180 |
| 32 | Other manufacturing | 16,921 | 42,163 | 143,456 | 4,065 | 1,786 | 2,898 | 254 | 687 | 1,672 | 170 | 926 | 5,998 |
| 33 | Repair and installation of machinery and equipment | 20,960 | 51,066 | 142,329 | 4,587 | 2,124 | 2,770 | 675 | 1,582 | 2,658 | 29 | 217 | 1,386 |
| 41 | Construction of buildings | 15,364 | 40,698 | 122,924 | 3,286 | 1,590 | 2,700 | 381 | 858 | 1,778 | 25 | 198 | 1,334 |
| 42 | Civil engineering | 18,210 | 48,492 | 139,961 | 4,408 | 2,002 | 2,840 | 598 | 1,239 | 2,409 | 42 | 358 | 2,233 |
| 43 | Specialised construction activities | 13,450 | 28,170 | 73,652 | 2,569 | 1,245 | 1,625 | 488 | 990 | 1,898 | 10 | 59 | 469 |
| 45 | Wholesale and retail trade and repair of motor vehicles and motorcycles | 14,222 | 36,068 | 96,459 | 4,249 | 2,224 | 2,750 | 323 | 774 | 1,768 | 28 | 138 | 971 |
| 46 | Wholesale trade, except of motor vehicles and motorcycles | 52,040 | 180,582 | 548,907 | 10,511 | 4,339 | 5,739 | 393 | 1,038 | 2,423 | 712 | 5,397 | 24,791 |
| 47 | Retail trade, except of motor vehicles and motorcycles | 24,199 | 71,096 | 191,738 | 4,289 | 2,156 | 2,999 | 166 | 437 | 1,023 | 78 | 496 | 3,006 |
| 50 | Water transport | 5,125 | 11,808 | 55,610 | - | - | - | 8,201 | 33,299 | 137,604 | 123 | 671 | 73,971 |
| 52 | Warehousing and support activities for transportation | 5,129 | 16,400 | 61,304 | 5,876 | 2,029 | 3,546 | 354 | 972 | 2,535 | 22,690 | 62,107 | 187,310 |
| 53 | Postal and courier activities | 2,072 | 4,934 | 14,173 | - | - | - | 567 | 1,322 | 2,652 | 4,385 | 9,790 | 33,981 |
| 55 | Accommodation | 4,621 | 11,020 | 33,709 | 6,616 | 3,382 | 5,354 | 4,507 | 10,474 | 26,066 | 7 | 27 | 199 |
| 56 | Food and beverage service activities | 25,801 | 53,762 | 102,561 | 7,507 | 4,154 | 5,788 | 141 | 360 | 863 | 12 | 70 | 439 |
| 58 | Publishing activities | 9,364 | 20,899 | 58,822 | 400 | 162 | 803 | 232 | 585 | 1,550 | 63 | 340 | 3,643 |
| 59 | Motion picture, video and television programme production, sound recording and music publishing activities | 7,122 | 14,088 | 30,434 | 837 | 467 | 1,058 | 502 | 1,124 | 2,286 | 10 | 34 | 137 |
| 61 | Telecommunications | 12,432 | 28,862 | 119,672 | 660 | 189 | 1,735 | 471 | 1,207 | 2,538 | 24 | 140 | 897 |
| 62 | Computer programming, consultancy and related activities | 9,392 | 21,596 | 75,025 | 1,302 | 588 | 1,359 | 219 | 520 | 1,259 | 17 | 78 | 523 |
| 63 | Information service activities | 7,757 | 30,426 | 106,424 | 759 | 353 | 1,572 | 282 | 526 | 1,165 | 24 | 123 | 1,182 |
| 71 | Architectural and engineering activities; technical testing and analysis | 9,002 | 19,284 | 59,104 | 2,710 | 1,333 | 2,014 | 391 | 866 | 1,853 | 14 | 53 | 422 |
| 72 | Scientific research and development | 12,229 | 39,817 | 127,896 | 848 | 365 | 2,221 | 473 | 1,034 | 2,132 | 74 | 459 | 2,430 |
| 73 | Advertising and market research | 9,054 | 20,462 | 59,042 | 1,804 | 943 | 1,524 | 317 | 765 | 1,639 | 18 | 73 | 504 |
| 74 | Other professional, scientific and technical activities | 7,254 | 15,137 | 34,994 | 1,374 | 795 | 1,140 | 298 | 675 | 1,340 | 13 | 38 | 175 |
| 75 | Veterinary activities | 37,349 | 92,700 | 171,194 | 1,834 | 1,005 | 3,405 | 223 | 580 | 1,370 | 43 | 136 | 423 |
| 78 | Employment activities | 13,247 | 37,702 | 122,458 | 2,533 | 1,071 | 1,586 | 370 | 951 | 1,896 | 18 | 73 | 561 |
| 81 | Services to buildings and landscape activities | 3,505 | 7,809 | 19,591 | 2,465 | 1,178 | 1,468 | 385 | 811 | 1,595 | 8 | 35 | 233 |
| 82 | Office administrative, office support and other business support activities | 7,137 | 16,234 | 52,524 | 2,530 | 1,308 | 1,818 | 292 | 707 | 1,570 | 19 | 108 | 1,085 |
| 84 | Public administration and defence; compulsory social security | 6,604 | 17,150 | 53,743 | - | - | - | 774 | 1,368 | 2,666 | 14 | 81 | 905 |
| 85 | Education | 8,996 | 38,942 | 136,825 | 20,921 | 3,077 | 11,874 | 191 | 501 | 1,212 | 16 | 52 | 194 |
| 86 | Human health activities | 20,237 | 75,946 | 181,254 | 3,956 | 1,954 | 3,401 | 144 | 367 | 846 | 16 | 51 | 216 |
| 87 | Residential care activities | 40,211 | 86,473 | 188,570 | 17,127 | 8,492 | 12,196 | 182 | 540 | 1,417 | 16 | 75 | 311 |
| 88 | Social work activities without accommodation | 3,612 | 12,765 | 35,725 | 4,863 | 2,683 | 3,610 | 134 | 350 | 833 | 15 | 68 | 342 |
| 90 | Creative, arts and entertainment activities | 6,238 | 12,782 | 29,861 | 1,622 | 878 | 1,408 | 323 | 762 | 1,536 | 16 | 62 | 364 |
| 91 | Libraries, archives, museums and other cultural activities | 9,956 | 24,551 | 65,550 | 668 | 258 | 2,464 | 184 | 617 | 1,449 | 63 | 265 | 1,666 |
| 93 | Sports activities and amusement and recreation activities | 6,562 | 14,967 | 40,242 | 3,849 | 1,587 | 3,227 | 203 | 513 | 1,246 | 14 | 70 | 369 |
| 94 | Activities of membership organisations | 4,619 | 13,591 | 55,344 | 335 | 149 | 777 | 190 | 557 | 1,480 | 25 | 106 | 396 |
| 96 | Other personal service activities | 6,739 | 11,080 | 19,178 | 1,364 | 883 | 1,139 | 86 | 216 | 487 | 6 | 26 | 114 |

## Supporting Information S3.3 – Expenditure Code Mapping

Excel file containing the mapping of expenditure codes to UKMRIO factors

ee21ap (2026) 'ee21ap/Code-and-Supporting-Files---AP-Thesis: Release 1'. Zenodo. doi:10.5281/zenodo.19034950.

## Supporting Information S3.4 – Stability Results and Binomial Logistic Regression Details

This Supporting Information provides the full test and train results to assess model performance and stability along with binomial model specifications and a summary table of average model performance metrics across all iterations.

Table S6. Performance Metrics for preferred models Test and Train Iterations

| | | Mean | | Median | |
|---|---|---|---|---|---|
| **Fold** | **RSQ** | **AE** | **APE** | **AE** | **APE** |
| **Category 1.** | | | | | |
| 1 | 0.88 | 55.14 | 63.49 | 8.47 | 28.18 |
| 2 | 0.88 | 52.42 | 67.07 | 8.60 | 28.45 |
| 3 | 0.88 | 57.83 | 74.46 | 8.51 | 28.36 |
| 4 | 0.88 | 55.95 | 70.08 | 8.56 | 28.22 |
| 5 | 0.88 | 56.14 | 110.54 | 8.51 | 28.40 |
| **Mean** | **0.88** | **0.06** | **77.13** | **0.01** | **28.32** |
| **Std** | **0.00** | **0.00** | **17.09** | **0.00** | **0.11** |
| **Category 3.1** | | | | | |
| 1 | 0.71 | 4.88 | 77.17 | 1.01 | 48.26 |
| 2 | 0.71 | 4.88 | 76.96 | 1.03 | 49.23 |
| 3 | 0.71 | 4.88 | 75.97 | 1.02 | 48.73 |
| 4 | 0.72 | 4.87 | 75.98 | 1.03 | 48.61 |
| 5 | 0.71 | 4.62 | 76.17 | 1.01 | 47.99 |
| **Mean** | **0.71** | **0.00** | **76.45** | **0.00** | **48.56** |
| **Std** | **0.00** | **0.00** | **0.51** | **0.00** | **0.42** |
| **Category 3.2** | | | | | |
| 1 | 0.74 | 0.86 | 73.34 | 0.21 | 48.93 |
| 2 | 0.73 | 0.87 | 74.30 | 0.21 | 48.94 |
| 3 | 0.73 | 0.84 | 73.74 | 0.21 | 48.86 |
| 4 | 0.73 | 0.85 | 74.78 | 0.21 | 48.45 |
| 5 | 0.73 | 0.85 | 73.49 | 0.21 | 48.11 |
| **Mean** | **0.73** | **0.85** | **73.68** | **0.21** | **48.56** |
| **Std** | **0.00** | **0.00** | **0.60** | **0.00** | **0.37** |
| **Category 6** | | | | | |
| 1 | 0.44 | 1.31 | 128.27 | 0.42 | 59.12 |
| 2 | 0.44 | 1.33 | 129.69 | 0.43 | 59.18 |
| 3 | 0.44 | 1.47 | 131.61 | 0.43 | 59.04 |
| 4 | 0.44 | 1.39 | 129.64 | 0.42 | 59.11 |
| 5 | 0.44 | 1.32 | 134.24 | 0.43 | 59.84 |
| **Mean** | **0.44** | **0.00** | **130.69** | **0.00** | **59.26** |
| **Std** | **0.00** | **0.00** | **2.07** | **0.00** | **0.29** |
| **Category 9** | | | | | |
| 1 | 0.47 | 7.19 | 552.10 | 0.18 | 93.47 |
| 2 | 0.47 | 6.51 | 560.20 | 0.18 | 93.91 |
| 3 | 0.47 | 7.47 | 550.28 | 0.18 | 93.88 |
| 4 | 0.47 | 6.00 | 543.39 | 0.18 | 93.84 |
| 5 | 0.47 | 7.05 | 542.22 | 0.19 | 93.74 |

| | | | | | |
|---|---|---|---|---|---|
| **Mean** | **0.47** | **0.01** | **549.64** | **0.00** | **93.77** |
| **Std** | **0.00** | **0.00** | **6.51** | **0.00** | **0.16** |

Table S6 presents the results of the 5-fold cross-validation across all modelled Categories. The models are stable across folds, with consistent RSQ values and error distributions. For Categories 1, 3.1, and 3.2 the models perform strongly, with average RSQ values of 0.88, 0.71, and 0.73 respectively, and relatively low median AE values (8.5, 1.0, and 0.2 $tCO_2e$) and median APEs between 28-49%. By contrast, Categories 6 and 9 show lower explanatory power, with average RSQ values of 0.44 and 0.47, and substantially higher relative errors, as reflected in median APEs of 59% and 94% respectively. This suggests that while the models are stable across folds, predictive performance is stronger in Categories 1 and 3, with greater uncertainty in Categories 6 and 9.

Model Specifications: Binomial Logistic Models
Model 4: $logit(P(Y_{ij} = 1)) = \alpha_j$
Model 5: $logit(P(Y_{ij} = 1)) = \beta_t \, ln \, turnover_i + \ \alpha_j \, ln \, turnover_i$

Where:

$logit(P(Y_{ij} = 1))$ is the log odds of the probability of the outcome variable Y for firm i in industry j. Here, Y = 1 indicates the presence of a hotspot, while Y = 0 denotes its absence

$\alpha_j$ represents the variable effect for industry j

$\beta_t$ represents the fixed coefficients for firm turnover

Table S7. Model summary metrics for all 624 regression iterations.

| | | **Model 4** | **Model 5** |
|---|---|---|---|
| **AIC** | **Min** | 173.87 | 26.66 |
| | **Average** | 5,126.82 | 4,855.61 |
| | **Max** | 126,246.72 | 125,772.00 |
| **Median Residual** | **Min** | -0.11 | -0.07 |
| | **Average** | -0.00 | -0.00 |
| | **Max** | -0.00 | -0.00 |
| **Log Likelihood** | **Min** | -63,122.36 | -62,882.00 |
| | **Average** | -2,562.41 | -2,423.80 |
| | **Max** | -85.93 | -9.33 |
| **Deviance** | **Min** | 14.68 | 12.51 |
| | **Average** | 4,907.07 | 4,762.27 |
| | **Max** | 125,830.96 | 125,288.32 |

To convert the predicted log odds into probabilities we use the formula below, this allows us to produce Supporting Information S3.5:

$$Probability = \frac{1}{1 + \exp{(-x)}}$$

## Supporting Information S3.5 – Hotspot Matrix

Excel file containing the interactive hotspot matrix.

ee21ap (2026) 'ee21ap/Code-and-Supporting-Files---AP-Thesis: Release 1'. Zenodo. doi:10.5281/zenodo.19034950.

## Supporting Information S3.6 – Benchmarking tool and access the model

Accessing the Model:

To access the model described in this paper, the interested reader is directed to the following public repo:

https://david-leake.github.io/carbonpredict/

As well as the following CRAN package:

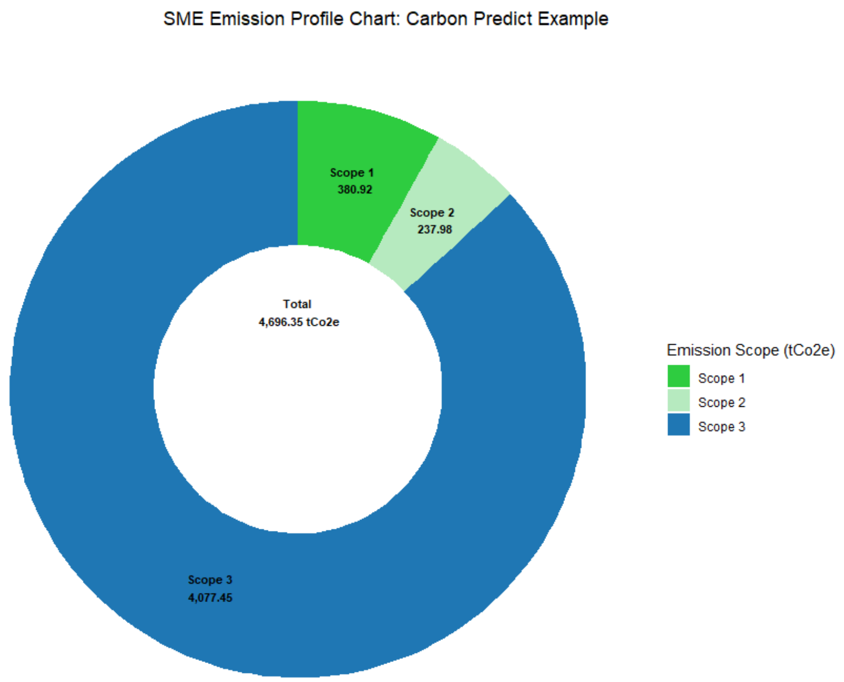
https://cran.r-project.org/package=carbonpredict

Example function use:

```
> sme_emissions_profile(85, 12000000, "Carbon Predict Example")
$scope1
  Predicted Emissions (tCO2e)
1                      380.92

$scope2
  Predicted Emissions (tCO2e)
1                      237.98

$scope3
   Category                                Description Predicted Emissions (tCO2e)
1         1                  Purchased Goods and Services                      349.28
2         2                                Capital Goods                      355.36
3         3           Fuel and Energy-Related Activities                       25.58
4         4     Upstream Transportation and Distribution                      112.73
5         5                  Waste Generated in Operations                      16.07
6         6                              Business Travel                       40.61
7         7                           Employee Commuting                      193.53
8         8                        Upstream Leased Assets                       12.15
9         9   Downstream Transportation and Distribution                      170.07
10       10                 Processing of Sold Products                       17.01
11       11                         Use of Sold Products                     2539.55
12       12            End-of-Life Treatment of Sold Products                  25.53
13       13                      Downstream Leased Assets                      219.97
14    Total                                                                  4077.45
```




Benchmarking to create total Scope 3:

To approximate total Scope 3 emissions, one accurately measured category can be scaled using benchmarking proportions to estimate emissions from the remaining categories (Buchenau et al., 2025). This approach allows us to construct a more complete picture of Scope 3 emissions while avoiding reliance on the weaker predictions typically associated with Categories 6 and 9. For this purpose, we use the underlying data from Figure 2 in Buchenau et al. (2025), which reports the percentage contribution of each Scope 3 category to the total using underlying CDP data. These contributions are disaggregated by firm type (manufacturing vs. services) and by firm size (SMEs vs. large enterprises). Given the SME focus of our model, we use data specific to SMEs only. These are presented in Table S8.

To align the data with our modelling needs, we make two key adjustments to the reported percentages:

Exclusion of Categories 14 and 15 (Franchises and Investments): These categories may only apply to a small subset of SMEs, yet they appear disproportionately large in the SME manufacturing data. To avoid overrepresentation, we exclude them from our scaling approach.

Removal of Category 3 (Fuel- and Energy-Related Activities): This category is estimated separately in our model, using a method aligned with our Scope 1 and 2 emissions calculations.

Following these exclusions, we normalize the remaining category percentages. To estimate emissions for individual Scope 3 categories, we divide our predicted emissions for Categories 1 and 4 by their share of the normalized benchmark. This gives us an estimate of total Scope 3 emissions. We then allocate emissions to each remaining category based on the adjusted percentages.

Table S8. Benchmarking percentages used to generate full emissions profile

| GHG Category | Raw Data | Remaining Share | Used in Model |
|---|---|---|---|
| **Services** | | | |
| category 11 | 51% | 51% | 63% |
| category 15 | 16% | | 0% |
| category 2 | 7% | 7% | 9% |
| category 1 | 7% | 7% | 9% |
| category 13 | 4% | 4% | 5% |
| category 7 | 4% | 4% | 5% |
| category 9 | 3% | 3% | 4% |
| category 3 | 3% | | 0% |
| category 4 | 2% | 2% | 3% |
| category 6 | 1% | 1% | 1% |
| category 12 | 1% | 1% | 1% |
| category 14 | 0% | | 0% |
| category 10 | 0% | 0% | 0% |
| category 5 | 0% | 0% | 0% |
| category 8 | 0% | 0% | 0% |
| **Total** | **100%** | **81%** | **100%** |
| **Manufacturing** | | | |
| category 14 | 75% | | 0% |
| category 15 | 9% | | 0% |
| category 11 | 9% | 9% | 54% |
| category 1 | 2% | 2% | 15% |
| category 10 | 2% | 2% | 14% |
| category 2 | 2% | 2% | 11% |

| | | | |
|---|---|---|---|
| category 12 | 0% | 0% | 2% |
| category 3 | 0% | | 0% |
| category 4 | 0% | 0% | 1% |
| category 13 | 0% | 0% | 1% |
| category 8 | 0% | 0% | 0% |
| category 9 | 0% | 0% | 0% |
| category 5 | 0% | 0% | 0% |
| category 7 | 0% | 0% | 0% |
| category 6 | 0% | 0% | 0% |
| **Total** | **100%** | **16%** | **100%** |

# Supporting Information: Chapter 4

## Supporting Information S4.1 – Exogenous Data for Nowcasting

Excel file containing exogenous data used for nowcasting.

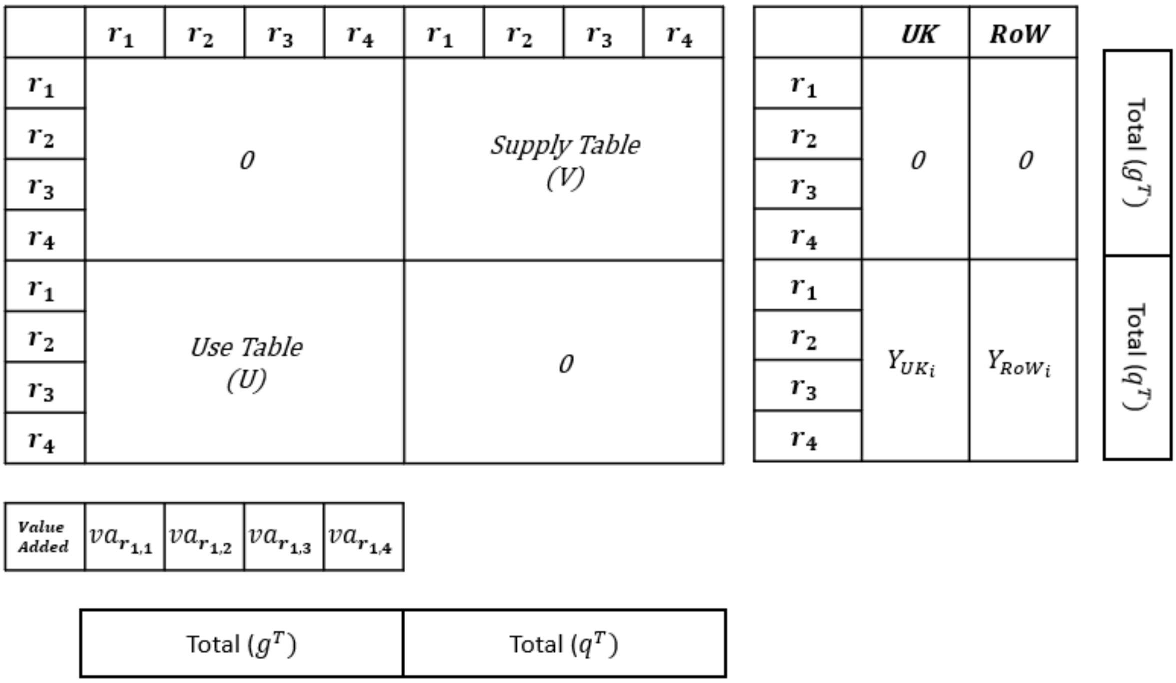
ee21ap (2026) ‘ee21ap/Code-and-Supporting-Files---AP-Thesis: Release 1’. Zenodo. doi:10.5281/zenodo.19034950.

## Supporting Information S4.2 – UKMRIO Adjustments

### Converting the UKMRIO to a symmetric model

The UKMRIO is originally in a supply and use table format, to keep as true to the original ONS data points as possible. Theoretically, this is useful when industries produce more than one product type. The supply table details these as off diagonal products (products made by industries other than their principal industry).

Figure S1 - SUT-MRIO (original UKMRIO structure)




Of the four transformation techniques suggested in the Eurostat manual of supply, use and input-output tables, we adopt Model D - Fixed product sales structure assumption. Here, each product has its own specific sale structure, irrespective of the industry where it is produced (Eurostat, 2008. p296). We present the procedure below.

Let:

V = Supply matrix (industry-by-product)

U = Use matrix for intermediates (product-by-industry)

Y = final demand matrix (product-by-category)

h = Value added matrix (components-by-industry)

We first produce C and D:

C = U $\hat{g}^{-1}$ = the input requirement for producers per unit of output of an industry

D = V $\hat{g}^{-1}$ = the market share coefficients of the supply table

We then make:

B = D C $\hat{g}$ = new transaction matrix

F = D Y = new demand matrix

Figure S2 - SIOT produced for nowcasting

| | $r_1$ | $r_2$ | $r_3$ | $r_4$ |
|---|---|---|---|---|
| $r_1$ | $Z_{r_{1,1}}$ | $Z_{r_{1,2}}$ | $Z_{r_{1,3}}$ | $Z_{r_{1,4}}$ |
| $r_2$ | $Z_{r_{2,1}}$ | $Z_{r_{2,2}}$ | $Z_{r_{2,3}}$ | $Z_{r_{2,4}}$ |
| $r_3$ | $Z_{r_{3,1}}$ | $Z_{r_{3,2}}$ | $Z_{r_{3,3}}$ | $Z_{r_{3,4}}$ |
| $r_4$ | $Z_{r_{4,1}}$ | $Z_{r_{4,2}}$ | $Z_{r_{4,3}}$ | $Z_{r_{4,4}}$ |
| *Value Added* | $v_{r_1}$ | $v_{r_2}$ | $v_{r_3}$ | $v_{r_4}$ |

| | $r_1$ | $r_2$ | $r_3$ | $r_4$ |
|---|---|---|---|---|
| $r_1$ | $Y_{r_{1,1}}$ | $Y_{r_{1,2}}$ | $Y_{r_{1,3}}$ | $Y_{r_{1,4}}$ |
| $r_2$ | $Y_{r_{2,1}}$ | $Y_{r_{2,2}}$ | $Y_{r_{2,3}}$ | $Y_{r_{2,4}}$ |
| $r_3$ | $Y_{r_{3,1}}$ | $Y_{r_{3,2}}$ | $Y_{r_{3,3}}$ | $Y_{r_{3,4}}$ |
| $r_4$ | $Y_{r_{4,1}}$ | $Y_{r_{4,2}}$ | $Y_{r_{4,3}}$ | $Y_{r_{4,4}}$ |

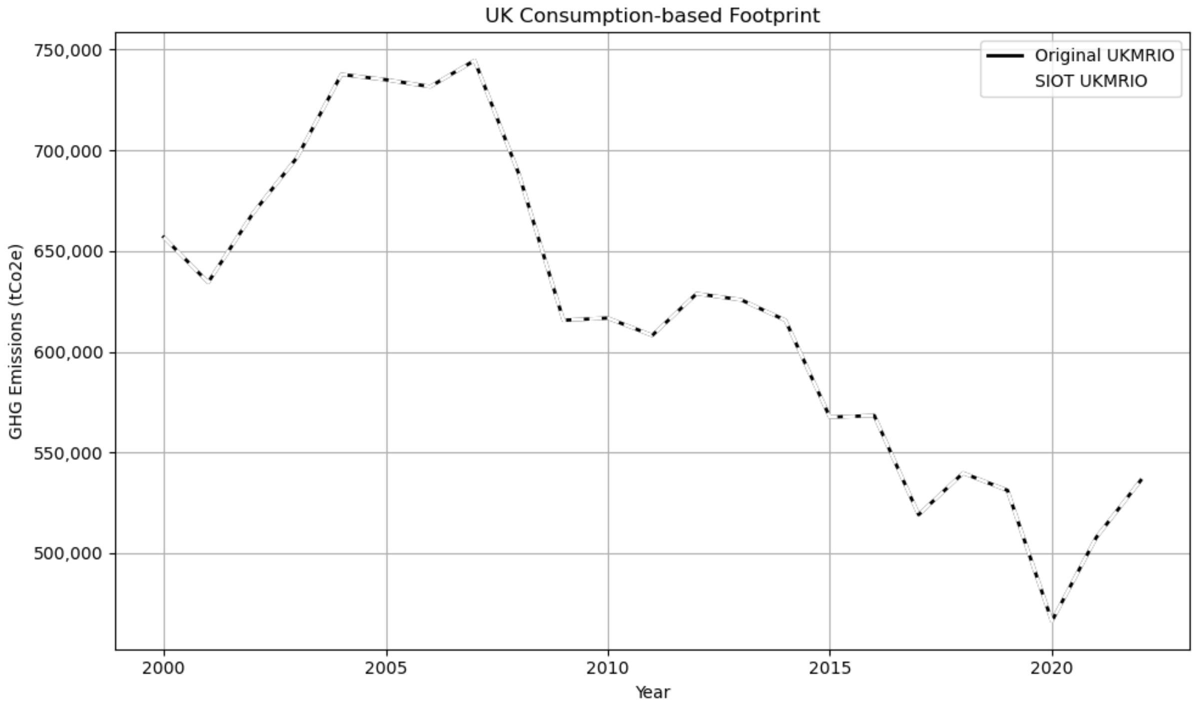
Disaggregating the UKMRIO's Rest of the World demand column

To disaggregate the final demand column for rest of the world, and thus identify final demand export destinations and source of imports, we use FIGARO (Eurostat, 2021). We extract the FIGARO final demand, and aggregate both the origin and destinations of demand into regions (creating a region-by-region matrix), and then further aggregate these regions to match those depicted in the UKMRIO. We convert data into proportions and drop the UK column. This provides the shares for distributing an origin region's final demand across non-UK destination regions.

Comparing the UKMRIO consumption-based estimates it's symmetric counterpart

Before being entirely satisfied with our modified base table, we compare the consumption-based emission estimates produced by the original model, against those produced by the symmetric version. Figure S3 plots these results across the period 2000 to 2022. We observe a near exact match, with a near zero percentage error for all years.

Figure S3 - Comparison of UKMRIO and SIOT consumption-based emission estimates



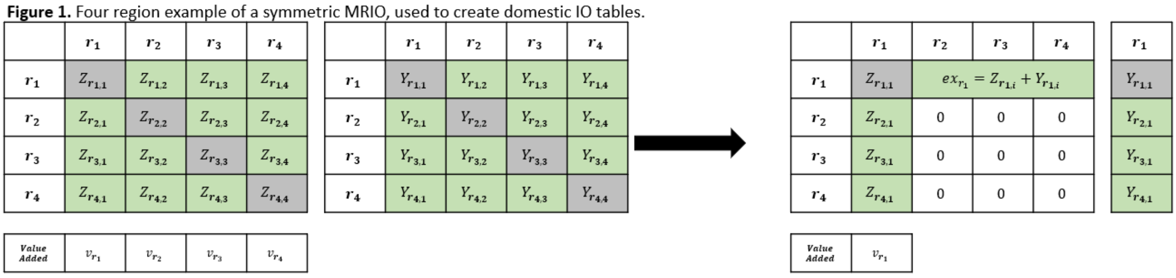

## Supporting Information S4.3 – Nowcasting the core MRIO matrices

This supporting information details the steps involved in projecting the core MRIO matrices. This is the basis for Methods 2 and 3, detailed in the manuscript. The process for projecting the MRIO is developed and presented in Beaufils and Wenz (2022). The key differences between the methods of nowcasting here occur within step 3.

Step 1: Extract domestic input-output tables from MRIO

Figure S4 – Process of splitting a SMIOT into a domestic IOT

**Figure 1.** Four region example of a symmetric MRIO, used to create domestic IO tables.



First, we extract domestic IO tables for each region using the base MRIO table. Each table contains a domestic transaction table, columns for intermediate imports and imports for final demand, rows for domestic industry exports (both to other industry and final demand), and domestic industry value added.

Step 2: Finding base GDP, Export and Import values for each region

Equation 1 demonstrates the income measure of GDP, where regional GDP is equal to regional value added, where value added includes taxes and subsidies:

$$gdp_r = v_r \tag{1}$$

Total imports can be found by summing rows indicating non-domestic sectors in the transaction and final demand matrix (equation 2), whilst total exports can be found by summing non-domestic columns in the transaction and final demand matrix (equation 3).

$$im_r = Z_{\boldsymbol{ir}} + Y_{\boldsymbol{ir}} \tag{2}$$

$$ex_r = Z_{\boldsymbol{ri}} + Y_{\boldsymbol{ri}} \tag{3}$$

Step 3: Project base GDP, Export and Import values to create nowcasting constraints targets

These values are then multiplied by growth rates, externally sourced from high frequency datasets. It is at this point where the level of nowcasting granularity can be determined, where decisions need to be made on whether blanket growth rates applied to all sectors is adequate (Method 2), or if a sector specific projection is required (Method 3). In all nowcasting methods, the overarching growth rates are sourced from the IMF's World Economic Outlook database.

The projected values now form a basis for constructing the projected table. Projected intermediate exports and imports are easily assigned to relevant the row / column constraint. Where values involve domestic production or consumption ($Z_{\boldsymbol{r}_{1.1}}$and $Y_{\boldsymbol{r}_{1.1}}$) further steps are required.For these rows, we must calculate projected aggregate demand for each region, this is the domestic final demand for commodities produced by domestic sectors plus the total exports to trade partners (equation 4).

$$ad_r = y_{\boldsymbol{rr}} + ex_{\boldsymbol{r}} \tag{4}$$

To find these two terms, and subsequently calculate aggregate demand we use a mixture of projected values, and sectoral shares (derived from the base table).

Figure S5 – Domestic IOT with constraints.

**Figure 2.** Domestic IO table generated for each region, with row and column constraints.

| | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_1$ | *Row constraints* |
|---|---|---|---|---|---|---|
| $r_1$ | $Z_{r_{1,1}}$ | $ex_{r_1} = Z_{r_{1,i}} + Y_{r_{1,i}}$ | | | $Y_{r_{1,1}}$ | $x_1$ |
| $r_2$ | $Z_{r_{2,1}}$ | 0 | 0 | 0 | $Y_{r_{2,1}}$ | $im_{r2}$ |
| $r_3$ | $Z_{r_{3,1}}$ | 0 | 0 | 0 | $Y_{r_{3,1}}$ | $im_{r3}$ |
| $r_4$ | $Z_{r_{4,1}}$ | 0 | 0 | 0 | $Y_{r_{4,1}}$ | $im_{r4}$ |

| | |
|---|---|
| *Value Added* | $v_{r_1}$ |

| | | | | |
|---|---|---|---|---|
| *Column constraints* | $x'_1$ | $ex_{r2}$ | $ex_{r3}$ | $ex_{r4}$ |

Export term:

$$ex_r = \sum_i \theta_{ri}^* ex_{ri} \tag{5}$$

The export term corresponds to the sum of exports from region r. The destination regions are given by $\theta_r^*$, which contains market shares of each export destination for region r's products and services. This term is derived from the base table.

Final demand for locally produced commodities:

$$y_{rr} = \varphi_r^* \left( gdp_r - \delta_r - \alpha_r^* \sum_i im_{ri} \right) \tag{6}$$

Final demand for locally produced commodities is derived by first calculating domestic consumption, which is found within the expenditure measure of GDP, where regional GDP is equal to domestic final use plus the regional trade balance (exports minus imports):

$$gdp_r = c_r + \delta_r \tag{7}$$
$$c_r = gdp_r - \delta_r \tag{8}$$

As we are focusing on domestic production, we exclude imports. We do this using the share of final commodities in imports from the base table, add the projected imports value to remove the import share. We then use $\varphi_r^*$ with provides sectoral shares of domestic final demand, derived from the base table.

$$ad_r = \varphi_r^* \cdot \left(gdp_r - \delta_r - \alpha_r^* \sum_i im_{ri}\right) + \sum_i \theta_{ri}^* ex_{ri} \quad (9)$$

To transition from method 2 to method 3, we make changes to the sectoral shares of the UK ($\theta_r^*$ and $\varphi_r^*$). For this we exogenously source data on industry contributions to GDP. With this information, we adjust these terms to increase the proportions of growing industries, whilst keeping the overall change stable through reductions on non-growth industries.

Step 4: Use the Leontief inverse to derive domestic output

The Leontief inverse makes explicit the direct and indirect links between sectors. It is generally used to obtain sectoral shares of production volumes, required to fulfil a given level of demand. Here, we have derived a level of demand, so use the Leontief to calculate the necessary levels of output.

$$L_r^* = (I - Z_{rr}^* \cdot (\widehat{x_r^*})^{-1})^{-1} \quad (10)$$
$$x_r = L_r^* \cdot ad_r \quad (11)$$

Step 5: Two-part balancing

<u>Sub-step 1: Domestic balancing</u>

Regional IO tables are balanced using a GRAS algorithm, allowing for the handling of negative coefficients (subsidies).

Constraints:

1. The domestic inter industry table's margins should meet the gross production requirement x.
2. Export arrays to each trade zone should equal the volume of exports to the corresponding trade zone as prescribed in input.
3. Import arrays plus imported final demand sum up to the volume of imports from the corresponding trade zone.
4. Total value added and domestic final demand are linked to GDP and trade balance in accordance with Equations 11 and 12.

<u>Sub-step 2: International trade</u>

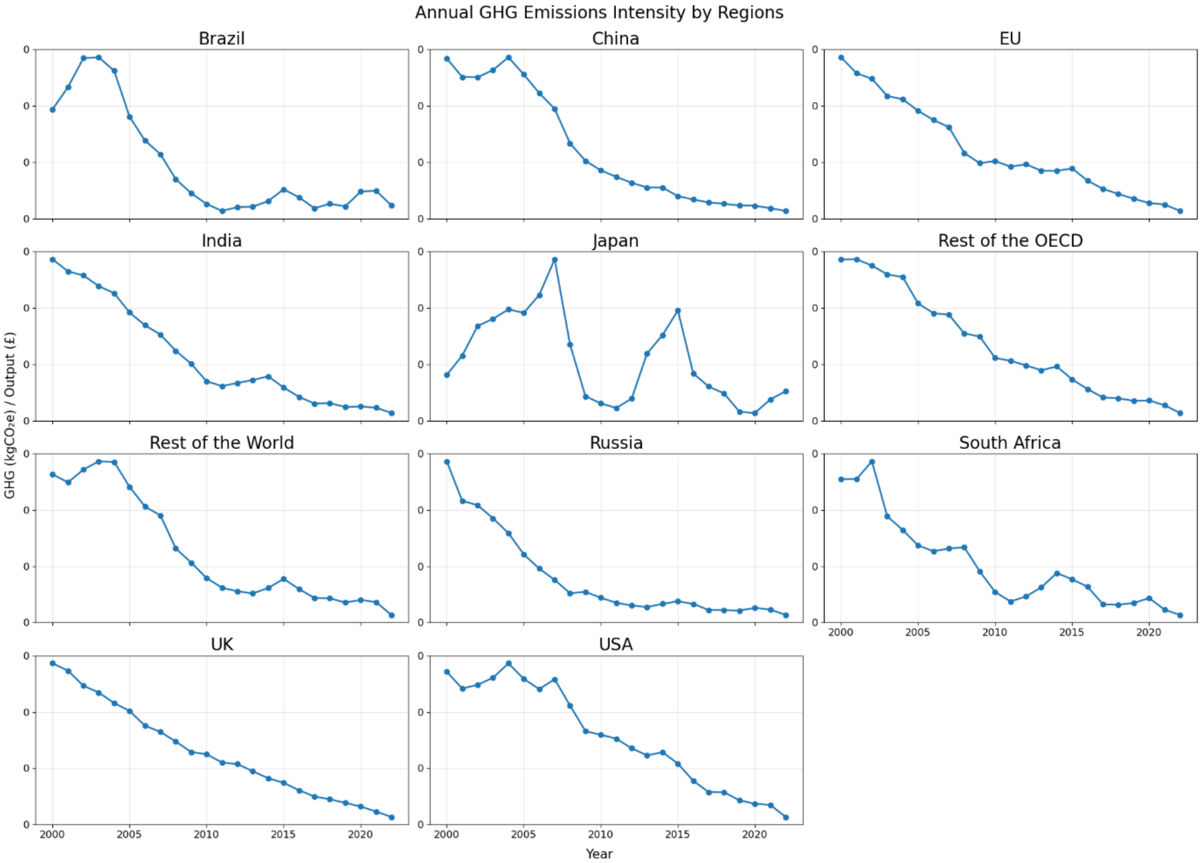
The second step of balancing adjusts international trade with respect to the exports and imports derived at the regional level. Using regional imports and exports constraints determined in sub-step1, each international trade is optimized using the basic RAS algorithm. Once this is done, we input the domestic blocks and create a new MRIO table. This table is balanced and consistent with the prescribed GDP and trade projections.

## Supporting Information S4.4 – Projecting the emissions extension vector

Owing to the time lag in the publication of emissions data, it is necessary to develop a method for estimating emissions at the regional level. As a first step, we examine historical trends in regional emissions. Figure S1 illustrates the annual production-based emissions for each region, as reported in the UKMRIO. We observe a steady increase in emissions across many regions, while the UK, USA, and EU exhibit a consistent decline. For some regions, the trends appear relatively stable and predictable, whereas others display more volatile and irregular patterns. We also investigate how regional emission intensity varies across the period in Figure S2, in all cases the emission intensity has reduced across the period 2000 to 2022, again with some countries have a more stable, consistent decrease.

Figure S6 – Regional emissions as depicted by the UKMRIO

Figure S7 – Regional emission intensity as depicted by the UKMRIO




To project the emissions vector, two key considerations must be addressed. First, it is necessary to capture observable trends in greenhouse gas (GHG) emissions over time. Second, the emissions vector must be linked to projected output since regional output is a direct driver of emissions. Each year modelled within the UKMRIO is in current prices of the year as a standard format. This means that using UKMRIO derived output within the projection methodology is problematic, unless each regions output is deflated with a regional relevant deflator.

We instead use our real GDP, imports and exports growth rates, sourced from the IMF's WEO (IMF, 2022). For each region, we estimate a simple linear regression model over 2001–2022 of year-on-year emissions growth on growth in real GDP, exports, and imports (all in percentage points). This first-pass specification provides a comparable elasticity-style summary across regions.

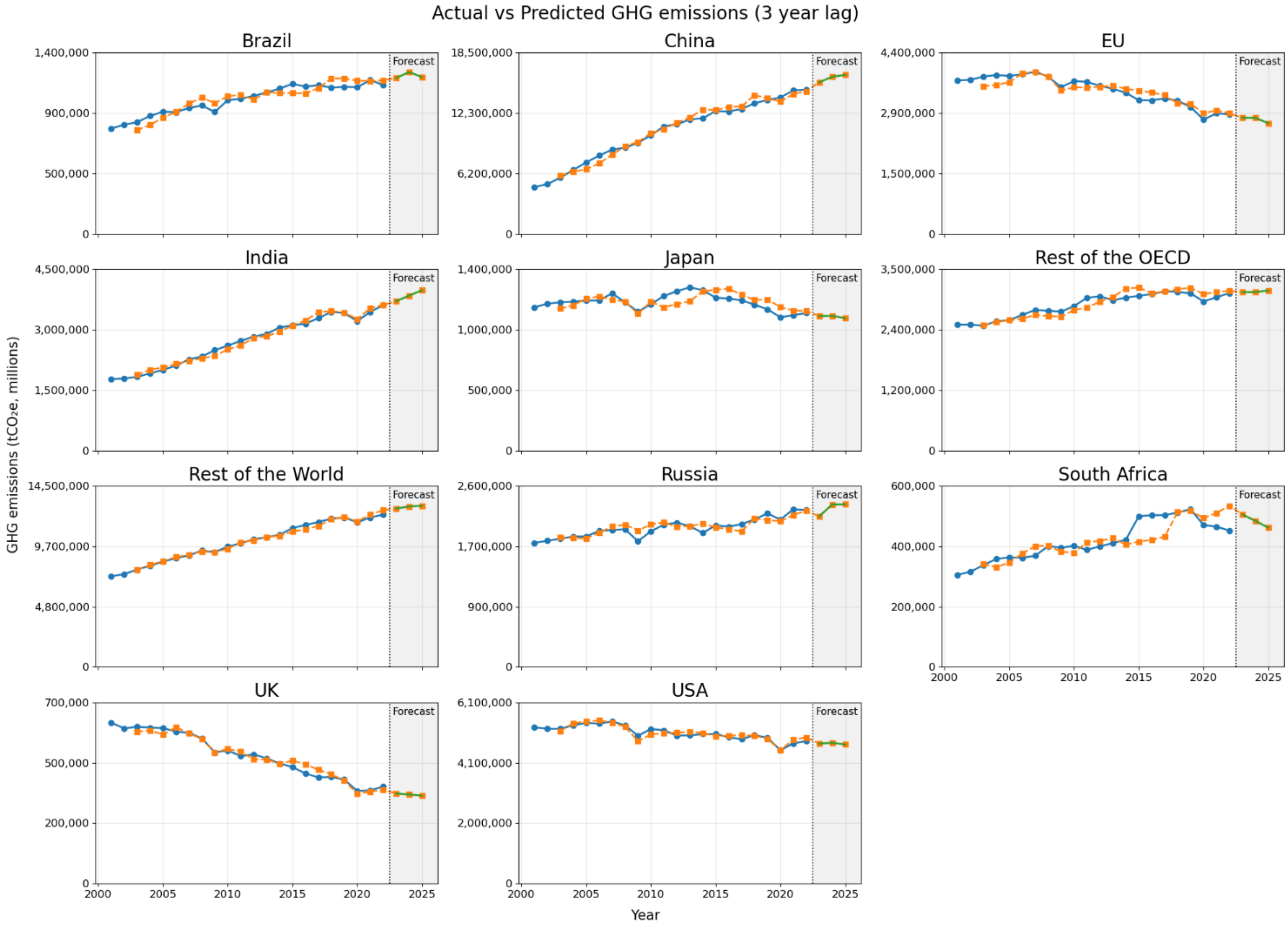
Table S1 – Model summary of regional regressions using GDP and imports to explain emission change.

| Region | Brazil | China | EU | India | Japan | OECD | RoW | Russia | S Africa | UK | USA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 0.01 | 1.39 | -2.88 | -0.06 | -1.24 | -0.34 | 0.46 | 1.01 | 0.35 | -3.70 | -2.34 |
| gdp | 1.14 | 13.42 | 12.91 | 48.29 | 2.12 | 11.47 | 14.82 | -30.15 | 35.57 | 8.05 | 21.30 |
| imports | 12.82 | 14.10 | 22.55 | 20.02 | 30.12 | 18.77 | 15.87 | 12.69 | 13.53 | 43.46 | 37.31 |
| obvs | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 |
| RSQ | 0.53 | 0.59 | 0.62 | 0.69 | 0.42 | 0.51 | 0.72 | 0.39 | 0.10 | 0.58 | 0.81 |
| MAPE | 1.73 | 2.18 | 1.64 | 1.32 | 2.14 | 1.28 | 0.76 | 2.29 | 3.52 | 1.88 | 1.08 |

The regressions show broadly good fit in many large economies, USA (RSQ = 0.80), Rest of World (0.73), EU (0.64), India (0.68), UK (0.58). Import growth consistently returns strong positive correlations, which may be interpreted as higher use of imported fuels and intermediates in domestic production during growth periods. GDP is positive in most regions (notably India 49.1; EU 10.5; USA 21.7; UK 8.3), but negative in Brazil and Russia. Export coefficients are mixed: positive in Brazil/China/Russia/USA/Japan, but negative in the EU/India/UK/South Africa. Intercepts are negative in the EU, UK, and USA, representing an underlying downward trend in emissions net of macro activity (decoupling).

We then combine the exogenous macro-economic variables in the lag period, with the regression predicted GHG rates to predict the emission changes of each region across the three-year lag period. We apply these growth rates to the base year level, to estimate GHG for the nowcasted period. Figure S3 overlays predicted emissions (orange), on the actual emissions (blue). We observe accurate results across most regions, the most accurate being Rest of the World (MAPE = 1.2%), whilst the least accurate is South Africa (MAPE = 6.8%). In addition to predicting known years, we demonstrate the real time application of the approach by extending the prediction for the years 2023, 2024, and 2025, where we have macroeconomic growth rates, but no emissions data.

Figure S8 – Predicted emissions overlayed onto actual emissions (in-sample)




The above results and figures are however OLS regression results, trained on all data post 2000. To fully recreate the time lags of EEMRIO, for each year, we must only include data from 3 years prior for the fitting of the regression. Tp predict GHG levels for 2017 onwards, we rerun the regression with training data up to 2014. We then iteratively include an extra years data for every year predicted after 2017. Applying the regression outputs to base GHG levels we produce Figure x. We observe an increase in error, due the test switching to out-of-sample data, and limited training data. However overall fit is still good. In this case, the most accurate predictions occur in the USA (MAPE = 1.9%) and the least accurate remaining South Africa (15.7%).

Figure S9 – Predicted emissions overlayed onto actual emissions (out-of-sample)