# Company Estimated Greenhouse Gases (GHG) Emissions

**DATA AND METHODOLOGY**

February 2024

SUSTAINABLE GROWTH

LSEG

# Contents


**INTRODUCTION** ..... 3

**Definitions** ..... 3

**How are these emissions obtained?** ..... 4
Reported GHG emissions ..... 4
Estimated GHG emissions ..... 5

**DATA SOURCES** ..... 6

**As-reported data** ..... 6

**Revenue, sectoral & regional classification data** ..... 6

**Fossil fuel production data** ..... 6

**Input-Output dataset: Exiobase 3** ..... 6

**Focus on the Financial sector** ..... 6

**METHODOLOGY** ..... 7

**Model overview: Hierarchical multi-model approach** ..... 7

**Level 1: Reported data** ..... 7

**Level 2: Extrapolation estimate** ..... 8

**Level 3: Production-based estimate** ..... 9
Fossil fuel production-based estimation (Scope 3 downstream only) ..... 9
Electricity production-based estimation (Scope 1 only) ..... 9

**Level 4: Carbon Ensemble « Best Estimate »** ..... 10
Training Data ..... 10
Multiple model approach ..... 10
Model 1: Sector Median ..... 10
Model 2: Input-Output model ..... 11
Model 3: The Inverse Distance Weighted (IDW) interpolation model (Scope 1 and 2 only) ..... 11
Model 4: Linear Regression (Scope 3 upstream and downstream only) ..... 11
Aggregation strategy ..... 11

**DATA QUALITY SCORE** ..... 12

**APPENDIX** ..... 13

**Table A: Conversion rate of Fossil Fuel data** ..... 13

**Table B: IPCC Emissions coefficients** ..... 13

# Introduction

A company's carbon footprint is a key metric for quantifying its contribution to climate change, as well as its vulnerability in case of a rapid decarbonization trajectory. With increased reporting by corporates, its incorporation into the TCFD recommendations in 2017 marked a further significant milestone, followed by the 2023 inclusion of Scope 3 emissions disclosure requirements in the IFRS S2 standard by the ISSB.

As more investors are relying on GHG emissions assessments to drive and measure their Climate strategy, filling the disclosure gap has become critical.

This document describes the methodology and data used to create our estimated carbon emissions dataset included in the LSEG Data Platform Climate bulk data feed package. A release to LSEG Workspace, our desktop offering, will follow.

More information is available via our research papers below:

- On Scope 1 & 2: Mind the gaps: Clarifying corporate carbon, May 2022
  https://www.lseg.com/en/ftse-russell/research/mind-gaps-clarifying-corporate-carbon
- On Scope 3: Scope for improvement: Solving the Scope 3 conundrum, January 2024
  https://www.lseg.com/en/ftse-russell/research/solving-scope-3-conundrum

## DEFINITIONS

A company's emission can be separated into three complementary categories:

- **Scope 1 emissions**: Direct emissions from owned or controlled sources by the reporting company. For example, direct emissions from a coal-fired power plant.
- **Scope 2 emissions:** Indirect emissions from the consumption of electricity, heat, steam and cooling. For example, indirect emissions from the electricity consumption of company-owned computer servers.
- **Scope 3 emissions:** Indirect emissions from upstream and downstream activities in the company's value chain or it's products life cycle.
  - **Upstream emissions**: indirect emissions from upstream processes in the value chain that contribute to a company's products or services. For example, emissions embedded in steel for a car manufacturer.
  - **Downstream emissions:** indirect emissions from the downstream processes (e.g. distribution to the client) or final use of a company's products and services. For example, emissions from the fuel consumption of an automotive.

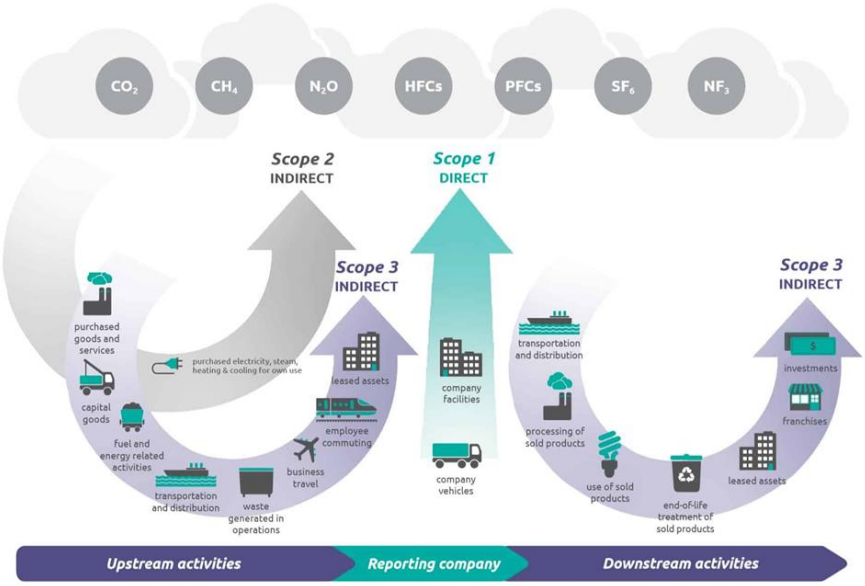


***Figure 1:*** *Corporate carbon emissions classification (Source: GHG Protocol)*

The reporting of Scope 3 emissions is broken out into **15 mutually exclusive categories**, also known as "components". Although companies are encouraged to disclose emissions for all these categories, the quality of the reporting can vary widely, in depth, coverage and accuracy. See Table A in the appendix for the complete list of categories.

Despite issues with reporting, the consideration of Scope 3 emissions is important as they often can account for most of a company's emissions (see Figure 2 below). Companies with large downstream emissions footprint may come under pressure from regulators or face obsolescence risk from the introduction of low carbon alternatives. On the other hand, companies relying on carbon intensive inputs may face significant future cost inflation (e.g. from carbon price).

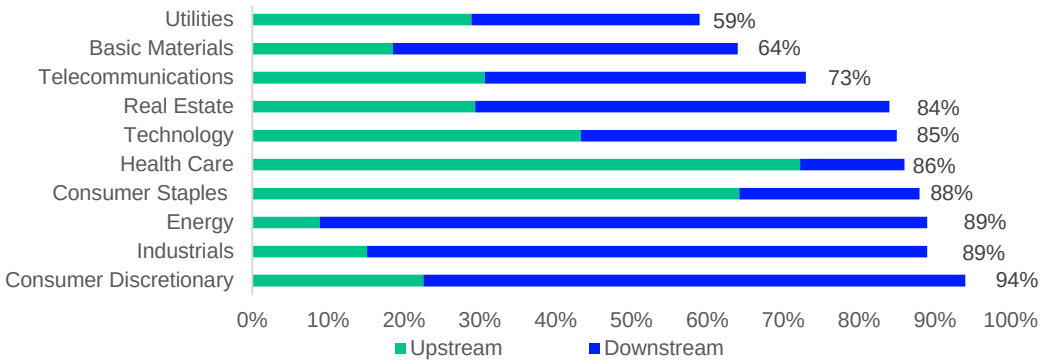


***Figure 2**: Average share of Scope 3 in total emissions*
*(Source: Scope for improvement: Solving the Scope 3 conundrum; FTSE Russell; January 2024)*

# HOW ARE THESE EMISSIONS OBTAINED?

## Reported GHG emissions

There are several ways to obtain Scope emissions for a company:

- Collect information directly from the company disclosures (e.g. Sustainability Reports, or data reported on Company Websites)
- Use third-party databases that work together with companies to gather value chain emissions data (e.g. CDP)

These data points need to pass a robust quality check process to be considered valid.

While the coverage and quality of disclosures has increased in recent years, overall progress is still modest, especially for the more complicated Scope 3 emissions, as can be seen in Figure 3. This is due to several factors:

- The difficulty and cost for some companies to precisely assess their Scope 3 data
- Methodological discrepancies when evaluating the Scope 3 data for specific categories
- The relative recentness of Scope 3 emissions disclosures

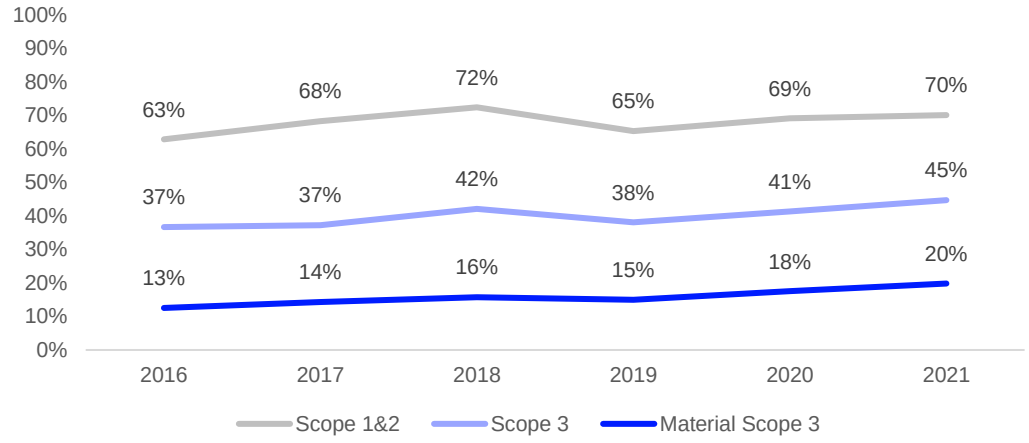


***Figure 3**: Proportion of FTSE All-World constituents disclosing Scope 1&2 and 3 emissions*
*(Source: Scope for improvement: Solving the Scope 3 conundrum; FTSE Russell; January 2024)*

Furthermore, there are significant size, sector and regional biases: smaller companies have fewer incentives and means to disclose their emissions, and historical carbon-intensive sectors such as Utilities or Industrials receive higher scrutiny to disclose than other sectors.

## Estimated GHG emissions

When a company has not disclosed its emissions (or when the disclosed emissions did not pass the quality checks), we use a proprietary algorithm described below to estimate the emissions based on multiple company attributes such as sectorial, regional, revenue and, for some sectors, production data. The algorithm uses LSEG D&A and CDP data which passed the quality check process as training material to derive its estimates for non-disclosing companies.

To avoid biases, the algorithm uses a hierarchical multi-modal approach which selects more reliable information over weaker sources and combines multiple sub-models, using different inputs, taxonomies, and assumptions, in an ensemble method to improve robustness.

# Data Sources

## AS-REPORTED DATA

Scope 1, 2 and 3 as-reported data is captured by LSEG D&A from documents that are publicly disclosed by the company such as annual reports, sustainability reports, company websites and filings. The as reported data is collected at a company-level by a team of more than 700 content specialists, and in the case of scope 1, 2 and 3, according to the specifications from the GHG Protocol, since 2002.

We also use CDP data (Scope 1, 2 and 3), with a history back to 2015.

When multiple as-reported data points are available for a given company, we use a proprietary algorithm to select the more robust datapoint.

## REVENUE, SECTORAL & REGIONAL CLASSIFICATION DATA

Our models are built using sectoral and geographical information for each company, as well as its revenues, including revenue breakdown when available.

## FOSSIL FUEL PRODUCTION DATA

Fossil Fuel Production data is provided by LSEG D&A. Fossil fuel assessments include yearly resource-level production data for each company with reserves exposure, history back to 1996. Resources considered currently include Coal, Oil and Gas.

## INPUT-OUTPUT DATASET: EXIOBASE 3

Exiobase 3 is the data source we use for the purpose of our top-down approach to estimate corporate GHG emissions. This is a Multi-Regional Input Output Environmentally Extended (MRIO-EE) database. A MRIO-EE gathers both monetary quantities exchanged between each country/industry and detailed direct GHG emissions generated by each country/industry during the production process.

## FOCUS ON THE FINANCIAL SECTOR

We do not currently include Scope 3 emissions data for the financial sector; **however, we do for Scope 1 and 2**. A new module will include the financial sector scope 3 emissions in H2 2024, based on a dedicated methodology.

# Methodology

## MODEL OVERVIEW: HIERARCHICAL MULTI-MODEL APPROACH

**Three years window approach**

While calculating model coefficients for each year, we use data from a three-year window. For example, when estimating model coefficients for the year 2022, we use data from 2020 until 2022.

Including multiple years increases samples size for certain peer-groups and dampens volatility and impact from data quality issues. At the same time, having a limited window allows the model to adjust as disclosures increase in number and align in methodology over time.

**Multiple model approach**

The proprietary algorithm uses the following hierarchical set-up:

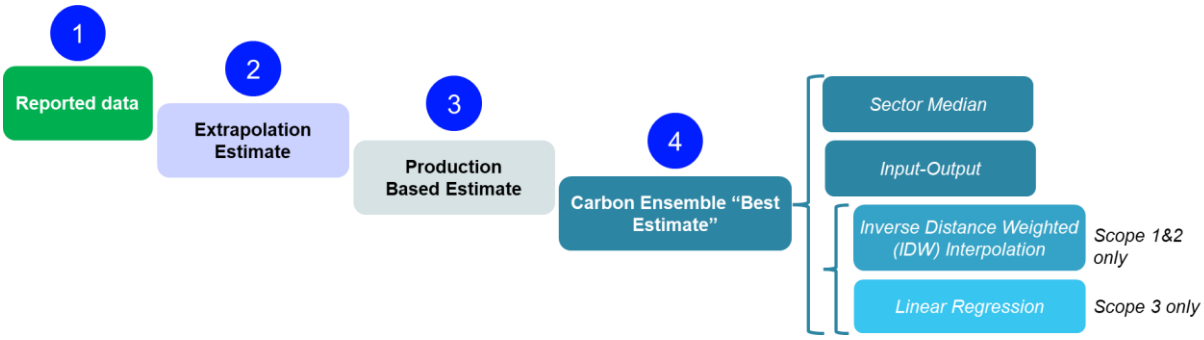


***Figure 4**: Model overview*

When a reliable data point for a company is not available on a higher model level, the model steps down to a lower, less fine-grained estimate.

Please note while the overall model structure is the same, Scope 1, 2 and 3 are modelled independently inside the algorithm. Similarly, for Scope 3 all estimations are modelled for upstream and downstream separately.

## LEVEL 1: REPORTED DATA

Reported data, i.e. values disclosed by the companies themselves are given the highest priority but must pass in-house coverage and quality checks.

First, to qualify as a data point for Scope 3, reported data must pass a materiality filter: we check if the material categories for a given industry are reported. If it's not the case, the reported value is treated as not valid and replaced with an estimate.

A valid data point will have the value "**Reported**" as its source attribute in the dataset.

Second, extreme values in the reported data are corrected by winsorization, i.e. outliers are set to specified percentile of the distribution. For each year and ICB3 category, the reported values within three years window intensities ($year$-2, $year$-1 and $year$) are gathered to estimate the 5th (resp. 10th for Scope 3) and 95th percentiles of the reported values. Reported values falling outside of these ranges will be set to these specified percentiles.

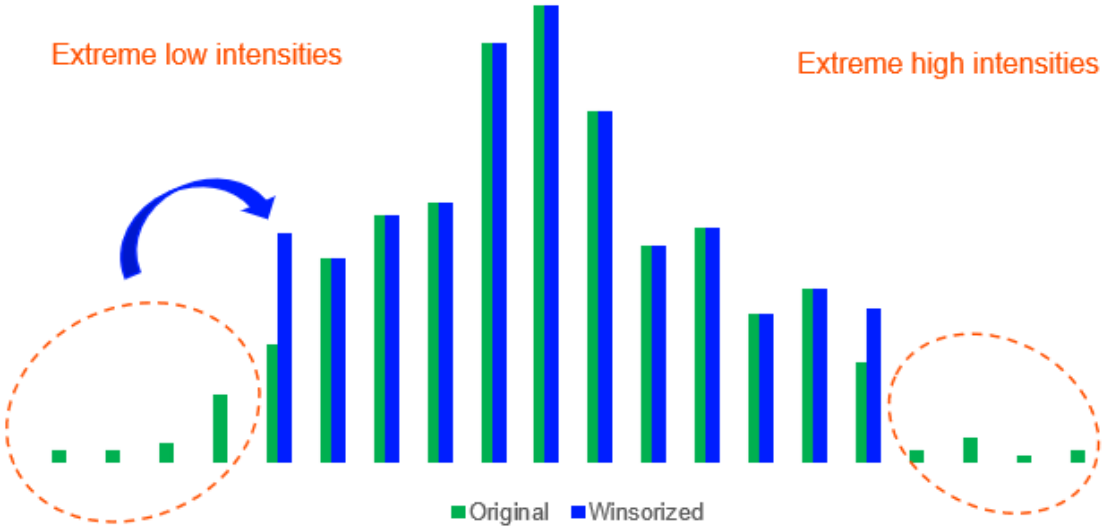


***Figure 5**: Illustration of the Winsorization process*

Note that with this asymmetric winsorization for Scope 3, we target underreporting, which is a more severe issue than overreporting.

A thus modified data point will have the value "**Winsorized**" as its source attribute in the dataset.

## LEVEL 2: EXTRAPOLATION ESTIMATE

If a company reported its emissions in the past but does not have a reported data point for a given year, the last reported intensity value is treated as an estimate of its emission intensity, for up to two years since the last reporting. This is based on the assumption that company-level carbon intensities are stable through time.

For example, if the last reported value for a company is in 2016, the same values for the given company are extrapolated only for 2017 and 2018 but the value for 2019 will remain missing. The carbon intensity is then multiplied by the revenue for the current year to estimate the emissions value of the current year.

A data point that was estimated with this procedure will have the value "**Extrapolated**" as its source attribute in the dataset.

# LEVEL 3: PRODUCTION-BASED ESTIMATE

## Fossil fuel production-based estimation (Scope 3 downstream only)

For **companies that are identified as fossil fuel producing** (Coal, Oil and Gas), a dedicated model is used to estimate its **downstream Scope 3 emissions from Use of Sold product** only.

Company-disclosed annual production figures (i.e., coal, oil and gas) from the LSEG D&A Reserves dataset, an in-house quality-controlled dataset, are multiplied against their respective emission factor (i.e., emissions per unit volume or weight combusted from IPCC guidelines – see Table C in the appendix, named "IPCC emissions coefficients").

The emissions from coal, oil and gas are then summed to calculate the total emissions, which in turn are used to derive the intensities per year.

Outliers are then removed from the set of intensities, by calculating the 25$^{th}$ and 75$^{th}$ percentiles of the final total intensity by year and subsector. The emissions from the outliers below *25$^{th}$ quantile / 5* and above *75$^{th}$ quantile * 5* are discarded.

Please note that this model also allows a carry forward of carbon intensity based on revenues for up to 2 years.

A data point that was estimated with this procedure will have the value "**Fossil fuel production**" as its source attribute in the dataset.

## Electricity production-based estimation (Scope 1 only)

For **companies that are identified as Power Utilities**, a separate model is used to estimate its **Scope 1 emissions**.

Company-disclosed annual power generation data by fuel data (i.e. nuclear, coal) are multiplied against their respective emission factor (i.e., emissions per unit volume or weight combusted). The emissions for the different fuel types are then summed to calculate total emissions, which in turn are used to derive intensities per year.

The algorithm follows the same steps as described in the Fossil fuel production-based estimation (Level 3a), and outliers are identified using intensity at ICB4 level. The emissions from the outliers below *25$^{th}$ quantile - (1.5 * Inter Quartile Range)* and above *75$^{th}$ quantile + (1.5 * Inter Quartile Range)* are discarded.

A data point that was estimated with this procedure will have the value "**Energy model**" (if production data is available) or "**Energy extrapolated**" (if intensity was extrapolated from a given year) as its source attribute in the dataset.

# LEVEL 4: CARBON ENSEMBLE « BEST ESTIMATE »

## Training Data

Sector Median model (Scope 1, 2 and 3), IDW interpolation model (Scope 1 and 2 only) and Linear Regression model (Scope 3 only) are trained using quality checked reported data (i.e. Reported, Winsorized and Extrapolated) on a universe defined as all companies that were at least once in the FTSE All Cap Index between 2012 and the year of the assessment.

## Multiple model approach

To attenuate the biases of any individual estimation technique, we rely on multiple models to produce our final estimates, which use alternative estimation methods, sources of data, peer group classifications, and statistical assumptions.

The table below outlines key strengths and weaknesses of different estimation techniques used in our multi-model approach.

| | Overview | Key strengths | Limitations |
|---|---|---|---|
| **Sector median** | Calculates median carbon intensity for individual sector 'peer groups' as defined by sector and region | Simple and interpretable<br>Granularity can be adjusted to focus on specific sectors or regions | Attributing company to a single sector risks oversimplifying business models<br>Reliant on accurate and granular industry classification system |
| **Linear regression** | Quantifies relationship between firm attributes (sector, multiple financial variables) and reported carbon intensity | Highly flexible, allowing users to include or omit predictive variables across peer groups<br>Based on well-known statistical principles and benefits from set of established techniques to improve output (e.g., variable transformation or regularization) | More complex implementations hinder contribution analysis of emissions results<br>Decisions in data preparation have a significant impact on the value of the eventual predictive coefficients<br>Highly sensitive to the underlying distributions of variables |
| **Interpolation** | Estimates carbon emissions intensity for specific business segments based on reported data by assigning a heavier weight to 'pure-play' firms | Generates more nuanced estimates for complex, diversified firms with multinational exposures than simpler models<br>'Pure-play' or specialized companies have a greater impact on activity carbon intensities than diversified companies | Complex, with multiple computations that can be difficult to communicate.<br>Accuracy depends on numerous, specialized firms to generate intensity estimates on each industrial activity<br>Highly dependent on quality of segment mapping |
| **Input-Output** | Derives carbon intensities for individual business segments from Environmentally Extended Input-Output (EEIO) tables | Transparent methodology and easily auditable<br>Generates nuanced estimates for complex, diversified firms with multinational exposures<br>Consistent boundary conditions for emissions estimates. | Outputs highly dependent on EEIO table selected and quality of segment mapping, leading to large variation between models<br>EEIO tables are infrequently updated and do not reflect year-on-year trends in industry emissions levels |

***Table 1**: Overview of key emissions estimation strategies*
*(Source: Mind the gaps: Clarifying corporate carbon; FTSE Russell; May 2022)*

## Model 1: Sector Median

In the Sector Median model, a company's emission estimate is based on the median carbon intensity of its peer-group. The individual steps of the algorithm are as follows:

- For a given $company_i$, the algorithm selects all peers in the same ICB category for ICB level 4, ICB level 3 and ICB level 2, respectively.
- For each ICB level, all available reported values within a 3-year window ($year_{k-2}$, $year_{k-1}$, $year_k$) of the respective peer group are gathered to calculate a median intensity.
- The final estimated median intensity for $company_i$ is selected at ICB level with the highest graduality (ICB4 → ICB3 → ICB2), which meets the minimal requirement of number of reported observations within the peer group.

## Model 2: Input-Output model

In the Input-Output model a company's estimated carbon emissions are based on the macro-level output and emissions data for each country and industry.

For Scope 1, the industries' total direct emissions are used to derive their Scope 1 carbon intensities, reflecting the emissions associated with one dollar of gross output.

For Scope 2, total direct emissions of sectors related to energy generation are used to derive intensity associated with one dollar spend in those sectors. These intensities are then applied within each sector to derive the emissions related to energy generation with one dollar of gross output.

For Scope 3 upstream, we use the Leontief inverse matrix to get the amount of inputs that are directly and indirectly required to produce a unit of product, from which we ultimately derive the GHG emissions embodied in the product, reflected in the upstream emissions associated with one dollar of gross output.

These intensities are then mapped and converted to intensity factors for each of our own sector Category, Country, and Year.

The GHG Emissions for each scope (Scope 1, Scope 2 and Scope 3 upstream) are estimated separately as the sum of the applicable intensity factors multiplied by the company's adjusted revenue breakdown.

## Model 3: The Inverse Distance Weighted (IDW) interpolation model (Scope 1 and 2 only)

The interpolation model uses the sales by activity data and derived from reported data activity intensity to estimate Scope 1 and Scope 2 emissions.

For each reporting company *i* and each sector *j*, we assign a heavier weight to 'pure-play' firms by considering a weight which is:

$$weigth_{j,i} = \left(\frac{sales_{j,i}}{total\ sales_i}\right)^2$$

Each sector's carbon intensity is then:

$$carbon\ intensity_j = \frac{adjusted\ emissions\ _j}{adjusted\ sales_j}$$

Where:

$$adjusted\ emissions\ _j = \sum_{i \in n_j} total\ emissions_i \times weigth_{j,i}$$

$$adjusted\ sales\ _j = \sum_{i \in n_j} total\ sales_i \times weigth_{j,i}$$

We then estimate final emission by multiplying sales activity with estimated intensity.

## Model 4: Linear Regression (Scope 3 upstream and downstream only)

Similarly, to median estimation, the linear regression estimation is performed using data within a 3-year window ($year_{k-2}$, $year_{k-1}$, $year_k$).

Linear regression is fitted using reported data at ICB4, ICB3 and ICB2 levels separately with region and revenue data used as additional predictors.

Similarly to median estimation, the final estimated median intensity is selected at ICB level with the highest graduality (ICB4 → ICB3 → ICB2), which meets the minimal requirement of number of reported observations within the peer group.

## Aggregation strategy

Scope 1 and Scope 2 emissions estimates are calculated as the median of the Sector Median, Inverse Distance Weight Interpolation, and Input/Output estimates, for any given year of interest.

Scope 3 upstream emissions estimates are a weighted average of the Sector Median, Linear Regression and Input/Output estimates, while Scope 3 downstream is an average of Sector Median and Linear Regression only.

A data point that was estimated with this procedure will have the value "**Aggregated Estimate**" (for Scope 1 and 2) or "**Aggregated model**" (for Scope 3 upstream and downstream).

# Data Quality Score

With banks and investors encouraging greater measurement and disclosure of Financed emissions following the Partnership for Carbon Accounting Financials (PCAF) methodology, our transparent multi-model approach enables them to derive the Data Quality Score for the GHG emissions associated with listed equities and corporate bonds.

We can derive PCAF Data Quality Score from the source of each datapoints with the following guidelines:

| PCAF Data Quality Score | **Scope 1** | **Scope 2** | **Scope 3 upstream** | **Scope 3 downstream** |
|---|---|---|---|---|
| Score 1 | / | | | |
| Score 2 | Reported | | | |
| Score 3 | Energy Model | / | / | Fossil fuel production |
| Score 4 | Energy Extrapolated<br>Winsorized<br>Extrapolated | Winsorized<br>Extrapolated | Winsorized<br>Extrapolated | Winsorized<br>Extrapolated |
| Score 5 | Aggregated Estimate | Aggregated Estimate | Aggregated model | Aggregated model |

***Table 2**: PCAF Data quality score derived from model outputs*

An example on how to use PCAF Quality score with our data can be found in our article Assessment of greenhouse gas emissions in a portfolio (https://developers.lseg.com/en/article-catalog/article/assessment-of-greenhouse-gas-emissions-in-a-portfolio).

# Appendix

## TABLE A: CONVERSION RATE OF FOSSIL FUEL DATA

| Description | Source Unit | Target Unit | Conversion |
|---|---|---|---|
| Coal Production Total | Tonnes/ Metric Tons | Tonnes | * 1 |
| Coal Production Total | US Tons | Tonnes | * 0.90718474 |
| Natural Gas Production (Cubic Feet) | Barrel | KCF | * (5.614583335876 / 1000) |
| Natural Gas Production (Cubic Feet) | BOE (Barrels of oil equivalent) | KCF | * (5658.53 / 1000) |
| Natural Gas Production (Cubic Feet) | BTU (British Thermal Units) | KCF | * (0.00097561 / 1000) |
| Natural Gas Production (Cubic Feet) | Cubic Feet | KCF | / 1000 |
| Natural Gas Production (Cubic Feet) | Cubic Meters | KCF | * (35.314666721489 / 1000) |
| Gas Liquids Production (Barrels) & Crude Oil Production (Barrels) | Barrel | Barrels | * 1 |
| Gas Liquids Production (Barrels) & Crude Oil Production (Barrels) | Barrel of Oil Equivalent | Barrels | * 1 |
| Gas Liquids Production (Barrels) & Crude Oil Production (Barrels) | British Thermal Unit | Barrels | * 0.000000172414 |
| Gas Liquids Production (Barrels) & Crude Oil Production (Barrels) | Cubic Foot | Barrels | * 0.178107606598 |
| Gas Liquids Production (Barrels) & Crude Oil Production (Barrels) | Cubic Meter | Barrels | * 6.289810767584 |
| Gas Liquids Production (Barrels) & Crude Oil Production (Barrels) | Tonne of Oil Equivalent | Barrels | * 7.33 |
| Gas Liquids Production (Barrels) & Crude Oil Production (Barrels) | Metric Tons (Crude Oil) | Barrels | * 7.33 |

## TABLE B: IPCC EMISSIONS COEFFICIENTS

| Fossil fuel | Emission factor |
|---|---|
| Natural Gas | 53.566 kg$CO_2$/kcf |
| Crude Oil | 425.994 kg$CO_2$/bbl |
| Coal | 2458.663 kg$CO_2$/tonne |