Received 2 February 2023
Revised 5 June 2023
Accepted 7 July 2023

Corporate carbon emissions data for equity and bond portfolios

Laurens Swinkels
Erasmus Universiteit Rotterdam, Rotterdam, The Netherlands and
Robeco Institutional Asset Management BV, Rotterdam, The Netherlands, and
Thijs Markwat
Robeco Institutional Asset Management BV, Rotterdam, The Netherlands

Abstract

Purpose – To better understand the impact of choosing a carbon data provider for the estimated portfolio emissions across four asset classes. This is important, as prior literature has suggested that Environmental, Social and Governance scores across providers have low correlation.

Design/methodology/approach – The authors compare carbon data from four data providers for developed and emerging equity markets and investment grade and high-yield corporate bond markets.

Findings – Data on scope 1 and scope 2 is similar across the four data providers, but for scope 3 differences can be substantial. Carbon emissions data has become more consistent across providers over time.

Research limitations/implications – The authors examine the impact of different carbon data providers at the asset class level. Portfolios that invest only in a subset of the asset class may be affected differently. Because "true" carbon emissions are not known, the authors cannot investigate which provider has the most accurate carbon data.

Practical implications – The impact of choosing a carbon data provider is limited for scope 1 and scope 2 data for equity markets. Differences are larger for corporate bonds and scope 3 emissions.

Originality/value – The authors compare carbon accounting metrics on scopes 1, 2 and 3 of corporate greenhouse gas emissions carbon data from multiple providers for developed and emerging equity and investment grade and high yield investment portfolios. Moreover, the authors show the impact of filling missing data points, which is especially relevant for corporate bond markets, where data coverage tends to be lower.

Keywords Carbon data, Carbon emissions, Climate, Finance, Investing, Portfolio management

Paper type Research paper

1. Introduction

Companies are required to report on their financial accounts and risks, such that stakeholders are aware of the health of the company. With this information, investors in stocks and corporate bonds can make informed investment decisions. More recently, corporations are encouraged to also publish sustainability-related metrics. As early as in 2000, the Carbon Disclosure Project (CDP) was founded by investors to influence corporate disclosure on their environmental impact. More recently, in 2015, the Taskforce for Climate-related Financial Disclosures (TCFD) was created. It tries to improve and increase climate-related financial information. This is important work, as Krueger *et al.* (2020) find that long-term institutional investors consider climate risks as an important risk factor for their portfolios. Moreover, Bolton and Kacperczyk (2021) find that investors are already requiring an additional risk premium when pricing corporate assets exposed to climate risk.

JEL Classification — G10, G11, P18, Q51, Q54, Q58

The authors would like to thank David Blitz and Emily Homer for valuable feedback. The views expressed in this paper are not necessarily shared by Robeco.

Managerial Finance
Vol. 50 No. 1, 2024
pp. 118-139
Emerald Publishing Limited
0307-4358
DOI 10.1108/MF-02-2023-0077

Recently, new regulation in the European Union (EU) for manufacturers of financial products and financial advisors about sustainability-related disclosures ("Sustainable Finance Disclosures Regulation") came into force. Its aim is to give end-investors better information about the sustainability profile of investment products and prevent asset managers from greenwashing their products [1]. With respect to climate-change risk, the EU developed minimum standards for Climate Transition Benchmarks and Paris-aligned benchmarks (Regulation 2020/1818). Needless to say, these efforts make it crucial for asset management companies and portfolio managers to use the best possible data on greenhouse gas emissions to manage their equity and corporate bond portfolios.

Corporate disclosures are also used as inputs for third parties to evaluate the company's risks or non-financial performance. Examples are credit rating agencies that determine the risks associated with default of bonds issued by the company. More recently, sustainability rating agencies have started rating corporate Environmental, Social and Governance (ESG) performances. While credit rating agencies tend to be well aligned when it comes to estimating the probability and severity of corporate defaults, this is much less the case when it comes to ESG ratings. Berg *et al.* (2022), Gibson Brandon *et al.* (2021) and Dimson *et al.* (2020) show that there is "aggregate confusion" among ESG-rating agencies, as cross-sectional correlations of ESG-ratings range only between 0.38 to 0.71 in Berg *et al.* (2022), 0.23 to 0.75 in Gibson Brandon *et al.* (2021), and 0.30 to 0.59 in Dimson *et al.* (2020). Hain *et al.* (2022) compare three model-based and three language-based *physical* climate risk measures and find that these data points have low correlation among each other. These differences make it challenging for practitioners to manage ESG and physical climate change risks in their investment portfolios.

The papers closest to ours are Busch *et al.* (2022), Kalesnik *et al.* (2022) and Papadopoulos (2022). They compare the corporate emissions data from six, four and three different data providers respectively. The former two use (rank) correlation analyses to draw conclusions, while the latter one examines the distributions of provider pairwise emissions ratios. They mainly focus on company-by-company comparisons of emissions data, while Kalesnik *et al.* (2022) also examine to what extent forward looking information predicts future emissions. All three studies find high consistency for direct emissions (scope 1), and less for indirect emissions, where particularly Scope 3 results in low consistency [2]. Busch *et al.* (2022) and Kalesnik *et al.* (2022) also show that reported data is more consistent between the data providers than estimated data, with correlations around 0.97 for reported and 0.85 for estimated scope 1 emissions. Even these lower correlations of estimated data are still markedly higher than average correlations of ESG scores. Papadopoulos (2022) also investigates the sources of discrepancies between providers in reported data, which can vary from using different organizational boundaries to human errors like typing mistakes. Interestingly, Busch *et al.* (2022) and Papadopoulos (2022) both show that the pairwise comparability between data providers does not necessarily increase over time. This is to a certain extent surprising, as especially after the Paris Agreement in 2015, carbon data has been increasingly scrutinized by the financial industry.

These studies do not, however, analyze the impact of data providers on equity investment *portfolio* carbon reporting and do not include the corporate bond market. Our study aims to fill these gaps.

Our contribution to the literature is three-fold. First, we are the first to compare portfolio carbon accounting metrics on scopes 1, 2 and 3 of corporate greenhouse gas emissions carbon data from multiple providers at the developed and emerging market equity portfolio level. This adds to the company-level comparisons in Busch *et al.* (2022), Kalesnik *et al.* (2022) and Papadopoulos (2022). Second, we extend this comparison to the investment grade and high yield bond portfolio managers [3]. While the sustainable finance literature mostly focuses on equity investing, the investment universe of corporate bonds is markedly different with many private issuers and ultimate debt guarantors that can be different from the issuing

entity. Third, we check how much of the emissions data has been historically reported by companies or, alternatively, estimated by the data providers to increase data coverage.

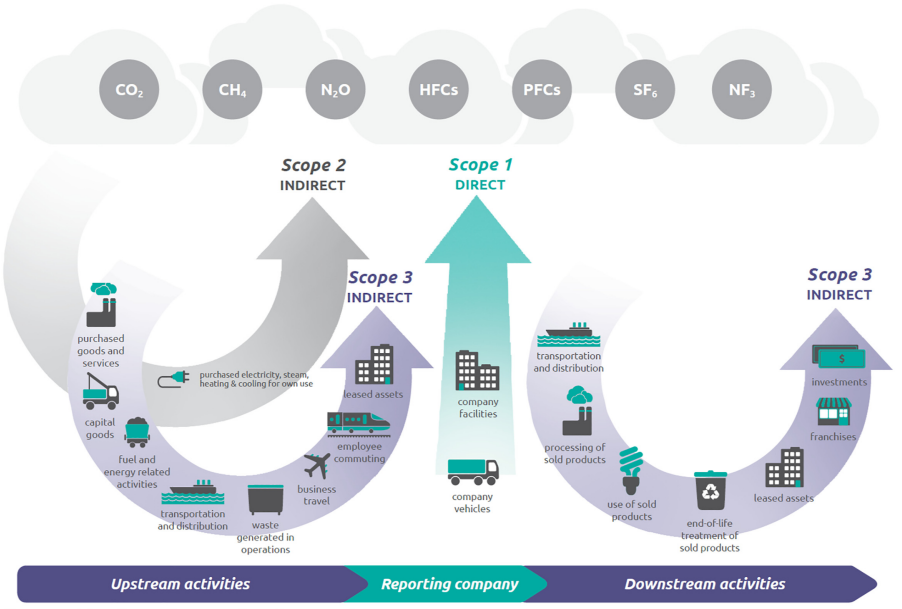
Our findings can be summarized as follows. First, data on direct (scope 1) and indirect energy-related (scope 2) greenhouse gas emissions is similar among the four data providers that we compare. Even though the methodology of some data providers is different, this does not seem to materially affect the aggregate estimate of greenhouse gas emissions, nor does it lead to low correlation of the carbon footprint among data providers. This is contrary to findings on a comparison of *physical* climate risk measures by Hain *et al.* (2022) and ESG metrics as reported in Berg *et al.* (2022), Gibson Brandon *et al.* (2021) and Dimson *et al.* (2020). It is also not consistent with the example presented in the European Central Bank (2022), where 13 banks have reported substantially different scope 1 and scope 2 estimates of one anonymous counterparty. Second, data on greenhouse gas emissions along the supply chain (scope 3) needs to be estimated using many uncertain parameters, leading to much more dispersion among data providers. Third, we show that carbon emission data has become more uniform over time, with an increasing number of companies self-reporting their carbon emissions. This is positive for equity and bond portfolio managers who need to manage climate change risks going forward. More consistent data among providers improves the reliability of carbon reporting of mutual funds. However, for historical analyses on carbon emissions data from before 2011, when less than 60% of the largest thousand companies reported their emissions, the choice of carbon data provider could materially influence the results [4]. The consistency of both reported and estimated emissions has improved over time, and still gets better every year. This contrasts with the findings of Busch *et al.* (2022) and Papadopoulos (2022). A possible reason why Busch *et al.* (2022) do not find increased consistency may be that their sample ends in 2016, just a year after the Paris Agreement was reached and before the financial industry started adopting carbon data at scale. Papadopoulos (2022) uses a relatively small sample of (<1,000) companies. This may lead to a relatively large influence of idiosyncratic noise and might be less robust to changes in the universe composition. In addition, companies with either larger portfolio weights or higher footprint have more impact on the total portfolio carbon footprints, which is a relevant feature for investors not captured by the pairwise evaluation metrics used by Busch *et al.* (2022) and Papadopoulos (2022). Our sample extends both important papers both in time and in the number of companies in the cross-section and gives higher weight to larger companies.

2. Methodology

2.1 Corporate emissions basics

To make it possible for companies to measure and report their greenhouse gas emissions in a complete and comparable manner, the Greenhouse Gas Protocol was formed. Along the lines of the Greenhouse Gas protocol a company is exposed to three different types of emissions: scope 1, scope 2 and scope 3.

These scopes are graphically represented in Figure 1. Scope 1 are the direct emissions that come from the company itself by burning fossil fuels during their business operations. Scope 2 are the emissions that stem from purchased electricity from utility providers. Depending on how this electricity was generated, the emissions can be higher or lower. Scope 3 emissions are all the other indirect emissions from the value chain. This is further divided into up- and downstream emissions, where upstream includes all indirect emissions from creating products/services, such as the capital goods a company uses and the employee commuting. Downstream includes all indirect emissions from using products/services, such as distribution and transportation of the end products, but also the emissions associated with investments a company makes. In total scope 3 is split into 15 different categories that a company needs to report on.



Source(s): Greenhouse Gas Protocol

Figure 1. Graphical representation of the Greenhouse Gas Protocol

Finally, we note that along the lines of the Greenhouse Gas protocol all emissions are measured in ton CO_2 equivalents (tCO_2e). Other greenhouse gases like methane and dinitrogen oxide have a different potency and by converting all greenhouse gases to CO_2 potency these can be compared and added up to one final measure.

2.2 Carbon accounting

The previous subsection concerned the corporate reporting standards, but ultimately, portfolio managers and end-investors are interested in the carbon emissions inside investment portfolios instead of emissions of individual companies. The first step here is to make companies comparable, because larger companies have larger emissions ceteris paribus. To accomplish this, the emissions are usually divided by metrics that proxy the size or business productivity of a company. Revenues and enterprise value including cash (EVIC) are the two most used metrics to adjust the emissions for company size [5]. Therefore, we consider these, but note that other metrics, such as the number of employees, could also be used as a numeraire [6].

The next step is to aggregate the normalized emissions from the constituents of an investment portfolio into an aggregated portfolio figure. If revenues are used as normalizing variable this aggregated figure is usually called WACI (weighted average carbon intensity) or just simply intensity. When on the other hand EVIC is used as denominator to scale the emissions the resulting metric is usually called footprint. These metrics are calculated as:

$$Intensity = \frac{\sum_{i=1}^{N} w_i \times \frac{Emissions_i}{Revenues_i}}{\sum_{i=1}^{N} w_i \times \mathrm{I}\left[\frac{Emissions_i}{Revenues_i}\right]} \text{ and } Footprint = \frac{\sum_{i=1}^{N} w_i \times \frac{Emissions_i}{EVIC_i}}{\sum_{i=1}^{N} w_i \times \mathrm{I}\left[\frac{Emissions_i}{EVIC_i}\right]}$$

where w_i are the portfolio weights and i denotes asset i and where I[x] is 1 if x is a number and 0 otherwise. This terminology can be a bit confusing as footprint (with EVIC used) is also an intensity metric, and intensity (with revenue used) is also describing a certain aspect (emissions) of a company's total climate footprint.

Some data providers have missing data. To assess the data coverage for a portfolio, we use the formulas above, where missing observations are excluded. Typically, differences in data coverage arise because of differences in availability of carbon data. However, a different data coverage between intensity and footprint for the same provider can arise because both denominators are not always available. This can happen for instance for private companies, where the equity market capitalization, and hence EVIC, is not available.

3. Data

3.1 Data providers

We consider four different data providers: Trucost, ISS, MSCI and CDP. This subsection reviews the main data delivery characteristics from these providers. For scope 1 we can see in Table 1 that there is one difference between the data providers. In total there are 7 GHG protocol gasses. ISS, MSCI and CDP stick to the company's reported GHG protocol gasses, whereas Trucost also adds projected emissions from the GHG protocol gasses that are not reported by the company. For instance, suppose a company only reports emissions from CO_2, CH_4 and N_2O, but Trucost expects the company to also emit some HFCs. Trucost then increases scope 1 emissions with an estimated amount, while the three others do not. Note that these additions are generally rather small compared to the reported numbers, which may be a motivation for the three other providers to ignore them altogether.

For scope 2 emissions there are two options: location-based approach vs market-based approach. The location-based approach uses the average intensity on the electricity grid to calculate the carbon emissions. The market-based approach on the other hand uses the contractual agreements made between the companies and energy providers to calculate the carbon emissions. For instance, if a company would settle a contract for 100% green energy, the market-based scope 2 would generally drastically decrease, but the location-based scope 2 remains the same. The reason to prefer market-based is that it is based on the actual electricity contract the company has in place. More demand for green electricity will increase its supply and therefore companies should be rewarded for signing green electricity contracts, especially since they may be more expensive than regular electricity contracts. The argument against it is that they consume the same electricity as anybody else connected to the same grid. The green electricity contract rewards the company without directly reducing actual physical carbon emissions. It merely shifts the accounting of carbon emissions to other electricity users who do not need to report on them.

Around 40% of the companies report both market- and location-based scope 2, while the other 60% only reports location-based scope 2. Table 1 also shows the different approaches

	Trucost	ISS	MSCI	CDP
Scope 1	All GHG protocol gasses	Only reported GHG protocol gasses	Only reported GHG protocol gasses	Only reported GHG protocol gasses
Scope 2	Location-based	Market-based if available, location-based otherwise	Location-based	Market- and location-based
Scope 3	Upstream and Downstream	Aggregated total	All 15 categories separated	All 15 categories separated

Source(s): Authors, Trucost, ISS, MSCI, CDP

Table 1. Characteristics from data providers

with respect to scope 2 emissions. Trucost and MSCI only report the location-based scope 2 figures. ISS uses market-based scope 2 if available and location-based otherwise. We only obtain this merged series, and ISS does not provide an indicator on whether market- or location-based scope 2 is used. CDP is the only provider that reports both emissions separately. Trucost has announced that it will report both approaches separately from fiscal year 2021 onwards.

For scope 3 there are also some profound differences between data providers. ISS is the only company that only reports scope 3 on a total/aggregated level. Trucost provides scope 3 data on a bit more granular level, namely the separation between upstream and downstream. MSCI and CDP both provide estimates for all 15 individual scope 3 emission categories. This can then be easily aggregated to upstream, downstream and total scope 3 emission.

All data providers have their own data cleaning methodology, which makes it possible that for one reporting company all four providers report different footprint figures. However, in general these data-cleaning-related differences are small. Larger reporting errors also occur, such as being of exactly a factor 1,000, but these are often adjusted by all data providers. Despite this, Papadopoulos (2022) indicates that carbon data is still plagued with errors.

3.2 Estimation and modeling

Not all companies report GHG emissions data and hence the coverage will be insufficient for many investment portfolios to align with climate objectives. For this reason, data providers fill these gaps by estimating and modeling the emissions for non-reporting companies. Although reported data is considered of the best quality one can get in carbon accounting, the use of modeled data is generally accepted [7]. Modeled data can be divided in two categories: physical activity-based or economic activity-based. The physical activity-based models use real production data, such as barrels of oil sold or tons of aluminum produced, to calculate the emissions of a company. The economic activity-based is generally constructed by taking the companies revenue and multiplying that by a certain emission factor. The physical activity-based approach is often considered to be of higher quality than the economic activity-based approach although both are in fact estimated data. In general, we see that the physical activity-based approach is more prevalent in the carbon intensive industries.

The different data providers all have their unique and proprietary models, but most adhere to the two approaches described above. For instance, the data providers slightly differ for which (sub)sectors they apply a physical activity-based approach instead of an economic activity-based approach. A data provider may use different models for the three emissions scopes for a single company as input data needed for one of the scopes might not be available. In the empirical part, we do not use the distinction between physical and economic activity-based models. However, understanding the similarities and potential differences in carbon accounting methodology is instructive as it may explain the differences encountered among data providers. Alternatively, if we do not find large differences, the methodological differences can be considered minor from a portfolio management perspective.

3.3 Combining the data

A particular challenge in getting the results is to match the data from the different data providers. First, the data providers use their own corporate identifiers which are not directly comparable. Therefore, the data needs to be mapped to unifying identifiers before we can do our analysis. Although all the data providers provide an International Securities Identification Number (ISIN), this cannot directly be used to combine all the data. ISINs are provided at the security level while emissions are reported at the corporate level. It can occur that for the same company one data provider uses the ISIN from the common stock traded in the United States, while the other data provider uses the ISIN from the stock traded in Germany. For private companies only issuing bonds this problem is even larger as one

company usually has multiple outstanding bonds, each with a unique ISIN. We solve this problem by mapping all different ISINs to Bloomberg company identifiers, and then we combine the datasets from the different data providers based on these Bloomberg company identifiers. It also occurs that there are typos in the ISINs or that an ISIN is simply unknown, which increases the difficulty of the mapping. In those cases, manual mappings are applied, with company names being a major input.

Second, corporate structures are far from straightforward and in some cases even very complex. One company can have different legal entities, such as *plc.*, *ltd.*, *group* or *holding*. It does occur that different data providers use a different entity for reporting. As in many cases the different entities also have different Bloomberg company identifier numbers, this makes the direct comparison based on these identifiers impossible. These issues needed to be resolved manually. Finally, we note that this issue is not only relevant for the mapping of the footprint data but also for corporate revenues and EVIC.

3.4 Index data

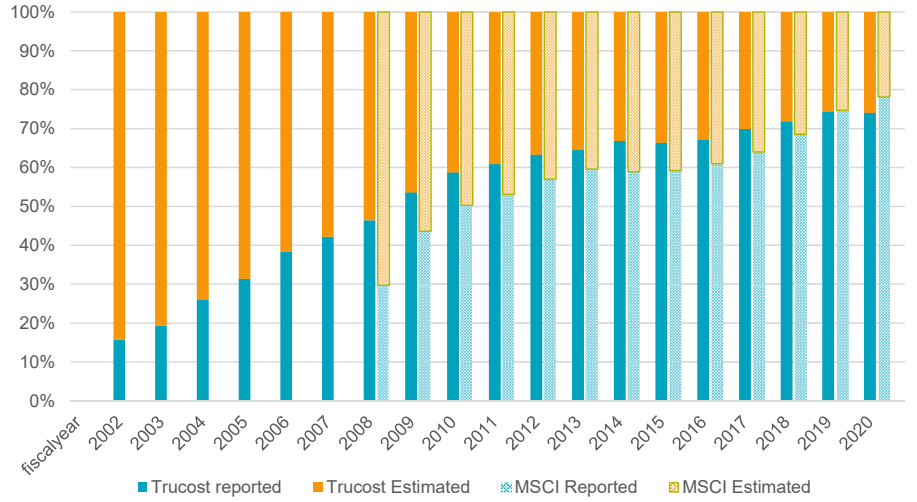
The focus of our paper is to examine what the impact of these differences and similarities are on the reporting of carbon intensity and carbon footprint of investment portfolios. Even though there may be methodological differences among data providers, they may partially cancel out and have little impact at the portfolio level. To proxy for different investment portfolios, we use four popular indices for four different asset classes: the MSCI World Index for developed equity markets, the MSCI Emerging Markets Index for emerging equity markets, the Bloomberg Global Aggregate Corporate Index for investment grade corporate bonds and the Bloomberg Global High Yield Corporate Index for below investment grade corporate bonds. In this way we cover important differences between equity and fixed income, between developed and emerging markets, between investment grade and high yield. These indices cover the vast majority of investable securities for institutional investors; see Doeswijk *et al.* (2014, 2020). The benchmark weights used are from August 31, 2022.

4. Investment portfolio carbon emissions

4.1 Coverage

Figure 2 contains a time series of the share of reported and estimated carbon footprint data for the 1,000 companies with the largest revenues in each year in the Trucost database, so not directly related to a specific investment universe. The chart is based on Trucost and MSCI data [8]. The Trucost carbon data starts in 2002 while the MSCI carbon data starts in 2008. The main observation is the strong growth in emissions reporting by these large companies. In 2002 only 15% of these large companies reported their carbon emissions, and 85% had to be estimated by Trucost. In 2008, when both providers have carbon data, the reported coverage for Trucost has increased to about 45%, where MSCI's coverage is at 30%. In 2011, Trucost crosses the 60% mark for reported data, while this is just over 50% for MSCI. Since then, the percentage has creeped up further to around 75% for both providers. The increased reporting may be partially driven by stakeholders putting pressure on the companies they invest in to start reporting their emissions; see Liesen *et al.* (2015).

Table 2 shows detailed results for the coverage of the different investment portfolios that we consider, using carbon data from fiscal year 2020. The coverage figures in the table are based on the intensity metric. For developed markets, all data providers cover 99% of the market capitalization. The percentage of reported emissions data is between 85 and 88%, and the remainder is estimated. This implies that for a developed large cap index the majority of the market capitalization is covered with reported emissions data, and that there are almost no missing data.



Source(s): Authors, Robeco, Trucost, and MSCI. Size of the company is measured based on company revenues of companies from the Trucost database. The reported percentage is of the number of companies and ignores company size

Figure 2.
Trucost and MSCI reporting types of 1,000 largest companies

Universe	Provider	Reported	Estimated	Missing
Equity	CDP	88%	11%	1%
Developed markets	ISS	85%	14%	1%
	MSCI	87%	12%	1%
	TC	85%	14%	1%
Equity	CDP	62%	34%	4%
Emerging markets	ISS	77%	22%	1%
	MSCI	69%	30%	1%
	TC	62%	36%	2%
Corporate bonds	CDP	57%	4%	39%
Investment Grade	ISS	59%	5%	36%
	MSCI	60%	7%	33%
	TC	56%	20%	24%
Corporate bonds	CDP	26%	5%	69%
High yield	ISS	29%	11%	60%
	MSCI	28%	13%	60%
	TC	26%	22%	52%

Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets), MSCI Emerging Markets Index (equity emerging markets), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade), Bloomberg Global High Yield Corporate Index (corporate bonds high yield). The column "Reported" contains the fraction of the index weight with company-reported greenhouse gas emissions. The column "Estimated" contains the fraction of data provider estimated greenhouse gas emissions, and the column "Missing" the fraction without greenhouse gas emission data. The constituents of the universes are from 31 August 2022. The greenhouse gas emissions data are from 2020

Table 2.
Greenhouse gas emissions data coverage and reporting characteristics

When considering emerging markets larger differences emerge. First, we see that the coverage of CDP starts to lag the three other data providers. Second, ISS shows a substantially higher percentage of reported data. We have aggregated the disclosures into

reported, estimated and missing data to make the disclosures comparable between the providers, but the disclosures are delivered with more granularity by the data providers. Trucost's coverage is a bit lower as they state in the raw disclosures that they have overruled some reported data where the quality was deemed too low, possibly because of incentives to underreport emissions (see for instance Hoepner and Rogelj, 2021).

If we turn to fixed income we see lower coverages, with missing data for investment grade and high yield varying from 24 to 39% and 52 to 69%, respectively. Difference between coverage from data providers is thus larger for fixed income than for equity portfolios. Trucost shows the largest coverage for the fixed income indices, which is attributable to their larger share of estimated footprint data. Conversely, CDP does not seem to actively cover large part of the fixed income universe. If we look at whether companies themselves report emissions, we see that for investment grade this figure is 60% of market cap or just below and for high yield it is even just below 30%.

From now on, we no longer distinguish between whether the coverage from a data provider is via reported or estimated data.

4.2 Base case emission results

The next step is to look at the actual carbon accounting metrics, footprint and intensity, which are shown in Table 3 together with their coverage. The intensity and footprint vary a lot for the different investment universes, with emerging equities and high yield clearly higher than developed equities and investment grade. For our study it is particularly relevant to gauge how the footprint differs between the providers *within* a universe rather than *across* universes.

		Carbon accounting metric		Coverage of metric	
Universe	Provider	Intensity	Footprint	Intensity	Footprint
Equity	CDP	170	54	99%	98%
Developed markets	ISS	170	54	99%	99%
	MSCI	161	53	99%	99%
	TC	170	56	99%	99%
Equity	CDP	372	167	96%	96%
Emerging markets	ISS	366	152	99%	98%
	MSCI	354	154	99%	99%
	TC	375	169	98%	97%
Corporate bonds	CDP	133	49	61%	61%
Investment Grade	ISS	133	47	64%	61%
	MSCI	229	48	67%	61%
	TC	233	48	76%	61%
Corporate bonds	CDP	409	188	31%	31%
High yield	ISS	354	141	40%	39%
	MSCI	333	136	40%	38%
	TC	311	148	48%	38%

Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets), MSCI Emerging Markets Index (equity emerging markets), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade), Bloomberg Global High Yield Corporate Index (corporate bonds high yield). The two carbon accounting metrics are intensity (emissions per unit of revenue) and footprint (emissions per unit of enterprise value including cash). Missing values are excluded from the calculation, not set to zero. The coverage is the total fraction of the index weight with data on greenhouse gas emissions, either company-reported or data provider estimated. The index weights are from 31 August 2022. The greenhouse gas emissions data are from 2020

Table 3.
Carbon accounting metrics and coverage across four asset classes

Starting with developed equity markets, each of the data providers come up with comparable carbon intensity (range 161–170) and footprint (range 53–56). This narrow range is expected, as in this universe the largest part is covered by reported carbon emissions data [9]. However, this similarity still shows that also the estimated part of the index does not lead to large deviations. MSCI has a slightly lower intensity and footprint than the three others. If we turn to emerging market equities, the differences become somewhat larger, but the carbon intensity (range 354–375) and footprints (range 152–169) are still all in the same order of magnitude. Again, MSCI is on the lower side. In general, the different data providers show similar carbon accounting figures for equity markets, confirming the high cross-sectional rank correlation as reported by Kalesnik *et al.* (2022).

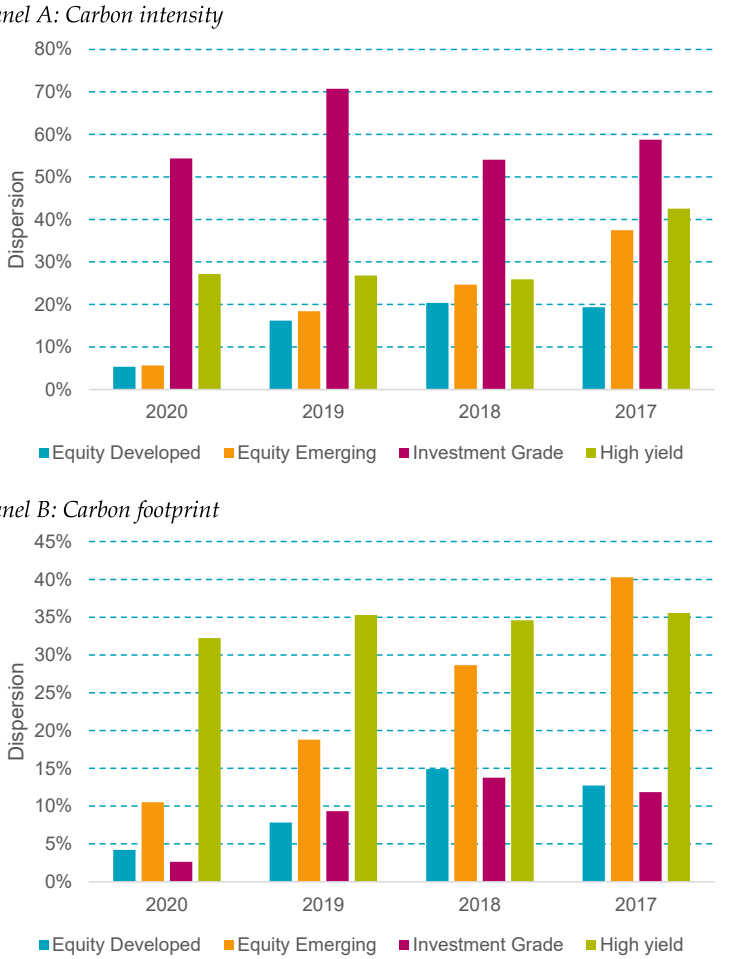
For investment grade the differences for intensity are quite large (range 133–233), with MSCI and Trucost having around double the values of CDP and ISS. Interestingly, the footprint metrics (range 47–49) are actually very similar for the different data providers. This inconsistent pattern between intensity and footprint can be explained using the coverage figures. For each of the four providers the footprint metric has a coverage of 61%, but for the intensity we see that MSCI and particularly Trucost have a higher coverage. This higher coverage partially regards private companies that operate in the energy and utility sector. As these companies are among the companies with the highest footprint, the impact can become as large 100 tCO2e. This is the first occurrence where the choice of data provider has a substantial impact on total carbon emissions.

For high yield we also observe a variation between the different data providers for both intensity (range 311–409) and footprint (range 136–188). Coverage for this asset class is much lower for all providers compared to the other asset classes, hence the effect of estimated data becomes more important, and the carbon accounting metrics start to diverge more. In addition, the different coverage figures may also imply that some sectors are better covered by some data providers, which will also lead to enlarged differences between carbon accounting metrics as the sector a company operates in is an important determinant of the company's footprint [10].

Figure 3 shows the dispersion between carbon intensity and footprint over the last four years for the companies that currently covered in the four asset classes [11]. Keeping the companies and their weights the same allows for a fair comparison over time. Index constituents may change over time, which may result in variation that is not due to developments in data coverage and consistency. The dispersion measure used here is the maximum intensity (Panel A) or footprint (Panel B) minus the minimum value over the four data providers, divided by the average of the maximum and minimum. For both developed and emerging equity markets the trend is clear: dispersion among data providers decreases over time. For intensity from 19% to 5% for developed markets and from 37% to 6% for emerging markets and for footprint from 13 to 4% and 40 to 11%, respectively. Indeed, the consistency of carbon reporting for global equity markets has improved considerably, leading to lower aggregate dispersion across data providers. The pattern is less clear for both fixed income asset classes. Dispersion of carbon intensity for investment grade remains large, while dispersion reduces for high yield. The opposite is true for the carbon footprint measure, where dispersion for high yield remains large, but has decreased for investment grade. These differences are due to the earlier mentioned important differences in coverage between data providers for both fixed income asset classes.

4.3 Sector differences

Figure 4 shows how for each asset class how the percentage distribution of the footprint over each of the sectors. For comparison, the figure also shows the market cap distribution of the different sectors. We focus on the intensities for this analysis as these reveal the most interesting differences. Starting with developed markets we observe that there is not much

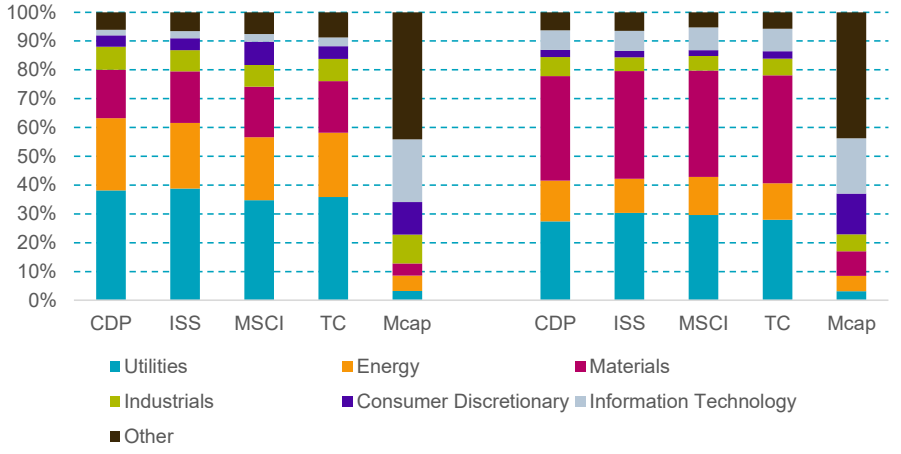


Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets; DM), MSCI Emerging Markets Index (equity emerging markets; EM), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade; IG), Bloomberg Global High Yield Corporate Index (corporate bonds high yield; HY). Dispersion is measured by taking the difference between the largest and smallest value and divide that by the average of the two. Panel A contains carbon intensity and Panel B carbon footprint data. The index weights are from 31 August 2022, and the greenhouse gas emissions data from 2020, 2019, 2018, and 2017. Scope 1 and scope use are used in both panels

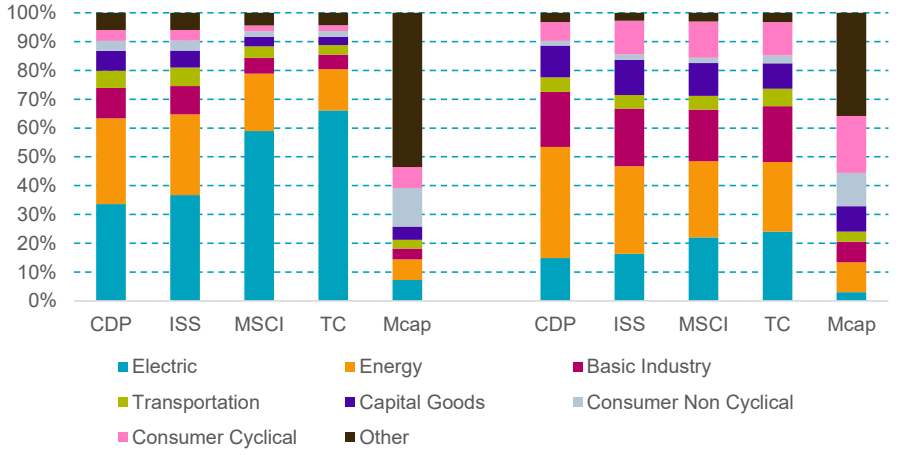
Figure 3.
Dispersion of carbon data between providers over time

difference between the sector contributions, which was also not to be expected given the similarities on the overall figures. MSCI seems to have a slightly smaller contribution from utilities than the other providers. This could for instance be caused by a more conservative

Panel A: Equity markets. Left developed, right emerging.



Panel B: Fixed income markets. Left investment grade, right high yield.



Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets; DM), MSCI Emerging Markets Index (equity emerging markets; EM), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade; IG), Bloomberg Global High Yield Corporate Index (corporate bonds high yield; HY). Panel A contains on the left-hand side data for developed markets and on the right-hand side for emerging markets. Panel B contains on the left-hand side investment grade and on the right-hand side high yield. The index weights are from 31 August 2022. The greenhouse gas emissions data are from 2020. The analysis uses the scope 1 plus scope 2 intensity metric

Figure 4.
Carbon emission sector differences between data providers

estimation model of the emissions of utility companies. Another interesting insight from this figure, but not directly related for the data provider comparison, is that Utilities, Energy and Materials are responsible for up to 80% of the total carbon intensity while only covering 12% of the market capitalization. Emerging equities markets show a rather different pattern in

sector footprint contributions, in the sense that the Materials sector has a larger contribution and Energy has a smaller contribution compared to developed markets. The larger contribution of Materials is mainly the result of mining activities in emerging markets countries. But again, the differences between the data providers are quite small.

Next, we turn to the sector effects in the fixed income indices. Please note that here we use the Bloomberg Classification System (BCLASS) sector classification instead of the MSCI's Global Industry Classification Standard (GICS) [12]. For both investment grade and high yield we see quite some differences between the different data providers. As for fixed income a larger part of the data is estimated, we also expect larger differences than for equities. Starting with investment grade, the most obvious is the much larger weight on Electric, which are the electricity generating utilities, for MSCI and Trucost. This can be explained by MSCI and Trucost covering the footprint of quite a large amount of private utility companies, mainly from North America. As these electric utilities have the largest footprint of all sectors, this dominates the footprint profile. The other sectors contribute less for MSCI and Trucost, but the relative contributions of the other sectors seem to be quite proportional amongst the data providers. For high yield, we see also some differences, with the most relevant differences for the electric and energy sector. Also, here we see that the contribution of electric utility companies is the largest for Trucost. CDP shows a particularly large contribution from the Energy sector. This is explained by CDP mainly relying on reported data for the fixed income universes, and energy companies have both a high intensity and often report their emissions. We also see a smaller figure for CDP for Consumer cyclical, which is caused by CDP not estimating data for quite some home construction companies. In conclusion, as soon as coverage decreases, the sector contributions to carbon intensity start to diverge, possibly leading to divergence in overall carbon intensity.

4.4 Increasing coverage by imputing missing emissions data

For corporate bonds, we have seen that large part of the market is not directly covered data providers. Until now we have not taken any action to improve the coverage. In this subsection, we start by using "carbon inheritance" to increase the coverage. When applying an inheritance methodology to a company without emission data, we search in the corporate hierarchy to find the parent company that has footprint data. If so, the company with missing data can then inherit the data from its parent. In this way, the coverage of the asset class increases, which can lead to more informed portfolio management.

Table 4 shows the results when this inheritance methodology is applied. There are very small differences for developed and emerging equity markets, which makes sense as there was hardly any missing data to start with. More interesting is to focus on the corporate bond markets, where we see quite some differences from Table 3. To start with investment grade, the intensity coverage increasing from figures between 61% (CDP) and 76% (TC) to figures between 89% (CDP) and 94% (TC). Although this is still smaller than for equity markets, it is a considerable improvement. For the footprint metric, all the coverages go from around 61–90%, adding around an extra one-third of the market cap covered by the inheritance. The intensities and footprints have also changed substantially by the increased coverage. In Table 3 we have seen quite some differences in intensities for fixed income between CDP and ISS on the one hand and MSCI and Trucost on the other hand, but with the implied inheritance methodology the carbon intensities have become more similar. The difference was caused by Trucost and MSCI covering more private electric companies, which are now also covered for CDP and ISS via inheritance. Thus, although there was no direct coverage for some companies, the sector contribution on Electric utilities, the most polluting sector, has been aligned and the main differences are resolved. The footprint increases substantially from approximately 48 for all data providers to approximately 74 for all data providers. This

Universe	Provider	Carbon accounting metric		Coverage of metric	
		Intensity	Footprint	Intensity	Footprint
Equity	CDP	170	54	99%	98%
Developed markets	ISS	170	54	99%	99%
	MSCI	161	53	99%	99%
	TC	170	56	99%	99%
Equity	CDP	373	168	97%	97%
Emerging markets	ISS	366	152	99%	99%
	MSCI	354	154	99%	99%
	TC	374	169	98%	98%
Corporate bonds	CDP	232	75	89%	89%
Investment Grade	ISS	237	74	93%	90%
	MSCI	252	74	93%	89%
	TC	247	74	94%	90%
Corporate bonds	CDP	406	193	61%	61%
High yield	ISS	380	153	79%	76%
	MSCI	345	152	77%	74%
	TC	346	163	78%	73%

Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets), MSCI Emerging Markets Index (equity emerging markets), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade), Bloomberg Global High Yield Corporate Index (corporate bonds high yield). The two carbon accounting metrics are intensity (emissions per unit of revenue) and footprint (emissions per unit of enterprise value including cash). Missing values are excluded from the calculation, not set to zero. The coverage is the total fraction of the index weight with data on greenhouse gas emissions, either company-reported or data provider estimated. To increase coverage, a company without data on emissions receives the carbon intensity of its parent company. The index weights are from 31 August 2022. The greenhouse gas emissions data are from 2020

Table 4.
Carbon accounting metrics and coverage after applying inheritance

difference seems large, but this can be largely ascribed to more private utility companies getting a carbon footprint via inheritance from a public mother company. The range for the carbon intensity differences between providers reduces from 100 tCO2e to 20 tCO2e.

For high yield we see an impressive increase in coverage as well. The coverage for intensity from for instance CDP almost doubles from 32% to 63%, and for Trucost increases from 49% to 80%. The coverage of the footprint metric approximately doubled for all providers. When we examine the carbon intensities, we see that the differences between the providers have also become smaller, albeit not as strong as for investment grade. The range is still 61 tCO2e, while it was 98 tCO2e. As stated before, because a large part of high yield is estimated, the differences between the data providers can become relatively large. Although inheritance can slightly alleviate this, some differences remain after applying the inheritance.

The inheritance improved the comparison between the data providers. However, for corporate bonds there were still quite some differences. The coverages are around 90% for investment grade and around 70% for high yield. If the relative coverages of the sectors are not equal between providers (e.g. provider A covers more of sector x than provider B), this might explain the large remaining differences between the carbon accounting metrics from different data providers.

To mitigate this difference in coverage, and the consequential sector imbalances, we enhance the coverage figures to 100% by applying sector defaults. Sector defaults are based on the median intensity or footprint metric within a sector or industry group. As sector classification we primarily use GICS, but for private companies issuing bonds with no GICS values, we use the BCLASS sector classification of Bloomberg. For both classifications we take the most granular levels. The missing intensity or footprint metrics are then filled with the corresponding sector default. For a more sophisticated machine-learning model to estimate unreported data, see Heurtebize *et al.* (2022).

The results are reported in Table 5, where the coverages are not reported as these are by definition 100% after applying the sector defaults. The coverage for the equity markets was already close to 100%, hence the impact of sector defaults on these markets is negligible. However, if we look at the fixed income markets the data providers have converged. This confirms the impact that sector imbalances can have. For high yield we clearly see that both the intensities and footprints have converged. For investment grade this is effect is less present, but for this market the results were already quite similar after applying inheritance. Interestingly, adding sector defaults decreases the overall footprint. Apparently, the majority of the companies with missing footprint data are from sectors with lower footprints.

4.5 Adding scope 3 emissions

While investors have become more and more familiar with scope 1 and scope 2, scope 3 is in many cases still unexplored territory. A notable exception is Furdak *et al.* (2022). The quality of scope 3 data is generally considered lower than that of scope 1 and 2. We therefore also expect larger differences between the data providers than for scope 1 and 2. The distinction between estimated and reported also diminishes for scope 3 disclosures. As mentioned in the methodology section, scope 3 exists of 15 different categories which all separately need to be reported. In many cases a company only reports some of the categories, and in those cases a data providers might estimate the missing categories they deem relevant. The disclosure will then effectively be a mixture of estimated and reported. Also, the data cleaning from the data providers becomes more relevant as scope 3 contains more erroneous reporting than scope 1 or 2. MSCI takes the stand that scope 3 reporting is of such low quality that they use only estimated values, which makes them distinct from the other providers.

Table 6 shows the results for scope 3. We have not reported the coverages anymore as these are equal to those in Table 3 and Table 4. The magnitude of the metrics has

		Scope 1 + 2	
Universe	Provider	Intensity	Footprint
Equity	CDP	171	55
Developed markets	ISS	170	54
	MSCI	161	53
	TC	169	55
Equity	CDP	366	164
Emerging markets	ISS	368	152
	MSCI	356	154
	TC	372	167
Corporate bonds	CDP	222	73
Investment Grade	ISS	230	70
	MSCI	245	72
	TC	242	72
Corporate bonds	CDP	334	154
High yield	ISS	342	134
	MSCI	317	134
	TC	313	143

Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets), MSCI Emerging Markets Index (equity emerging markets), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade), Bloomberg Global High Yield Corporate Index (corporate bonds high yield). The two carbon accounting metrics are intensity (emissions per unit of revenue) and footprint (emissions per unit of enterprise value including cash). Sector median carbon intensities or footprints are applied for missing data, increasing coverage to 100% for each provider and asset class. The index weights are from 31 August 2022. The greenhouse gas emissions data are from 2020

Table 5.
Carbon accounting metrics supplemented with sector medians

Universe	Provider	Plain		With inheritance		With sector median	
		Intens.	Footp.	Intens.	Footp.	Intens.	Footp.
Equity	CDP	1,213	481	1,212	481	1,207	478
Developed markets	ISS	1,213	466	1,213	466	1,213	466
	MSCI	998	371	998	371	997	370
	TC	1,045	420	1,044	420	1,046	421
Equity	CDP	2,041	792	2,051	797	2,015	783
Emerging markets	ISS	1,675	801	1,682	808	1,680	805
	MSCI	1,633	704	1,638	707	1,638	703
	TC	1,834	732	1,837	735	1,819	726
Corporate bonds	CDP	1,024	383	1,326	499	1,241	472
Investment Grade	ISS	1,041	381	1,344	491	1,303	470
	MSCI	1,064	316	1,166	425	1,145	408
	TC	924	318	1,011	418	993	402
Corporate bonds	CDP	2,192	1,234	2,864	1,153	2,225	907
High yield	ISS	2,241	1,131	2,115	1,069	1,931	939
	MSCI	1,901	813	1,751	823	1,663	746
	TC	1,664	989	1,747	874	1,639	777

Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets), MSCI Emerging Markets Index (equity emerging markets), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade), Bloomberg Global High Yield Corporate Index (corporate bonds high yield). The two carbon accounting metrics are intensity (emissions per unit of revenue) and footprint (emissions per unit of enterprise value including cash). Missing values are excluded from the calculation, not set to zero. The two columns ("Intens." Refers to "Carbon intensity" and "Footp." to "Carbon footprint") under "Plain" contain the data without applying inheritance (as in Table 3), the two columns under "With inheritance" contain the data with applying inheritance (as in Table 4), and the two columns under "With sector median" contain the data with applying sector medians (as in Table 5). Scope 3 emissions are included. The index weights are from 31 August 2022. The greenhouse gas emissions data are from 2020

Table 6.
Carbon accounting metrics with scope 3 emissions included

tremendously increased. Intensities for scope 1 and 2 were ranging from 161 for developed equities to 409 for high yield. Now we see these figures have increased from 998 to 2,241. In other words, the intensities have on average increased by more than a factor 5. This implies that the largest part of the total scopes 1, 2 and 3 metrics is determined by scope 3 only, and the impact of scope 1 and 2 will be very small; see also Furdak *et al.* (2022). Hence, if scope 3 differs much per provider we expect this to be directly visible in the results.

Table 6 shows that there are substantial differences for developed equities market, where for scopes 1 and 2 the data providers largely agreed. However, CDP and ISS also here have approximately the same footprint and MSCI again has the lowest footprint. The footprint from CDP is 31% (481/371-1) larger than that of MSCI. Thus, even for developed equity markets, where the reporting rate is the highest, the choice of data provider becomes important even at the aggregate level. For emerging markets, we see structurally larger differences than we observed for scopes 1 and 2. Adding inheritance and sector defaults does, as expected, hardly alter the results as almost all listed companies already have direct emission figures attached.

For the plain values of investment grade corporate bonds, the results are relatively close, but after applying inheritance and sector defaults to increase coverage, the differences between the intensities increase. This is not as expected, because applying inheritance and sector defaults would reduce the differences originated from different coverage number. Thus, the similarity for the plain scope 3 data seems to be a coincidence and the more representative differences became visible after applying inheritance and sector defaults. Tables 3–5 all showed similar footprints for investment grade, but Table 6 shows that when

scope 3 is included, substantial differences emerge. For instance, for the plain figures MSCI has the lowest value (316) and CDP the highest (383), a non-negligible difference of 20%. Inheritance and sector defaults do not make the different data providers converge.

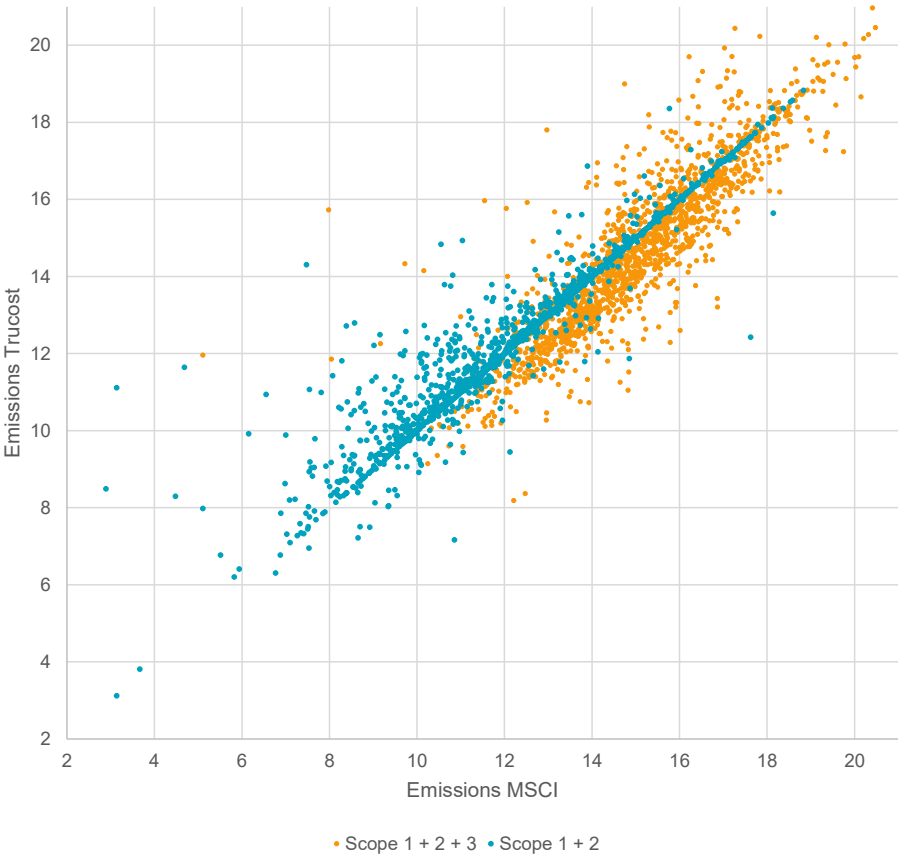
For high yield the differences become even larger, with is in line with the results from scopes 1 and 2. If we look at the plain intensities, the largest difference is 35% (between ISS and Trucost). Applying inheritance would increase the maximum difference (now between CDP en TC) to even 64%. For the plain footprint the maximum difference is 52% (1,234/813-1), which is reduced to 40% by applying inheritance. These large differences indicate that the choice of a data provider can influence business and portfolio management decisions. While for scopes 1 and 2 the addition of sector defaults led to more convergence between the data providers, we see that with scope 3 included even including sector defaults only limits the differences between data providers to a small extent.

To obtain better insight in the larger differences we have observed when including scope 3 we perform an extra analysis by comparing the absolute emissions from different providers with each other. This approach makes no use of any index weights and hence it can provide different insights in the effects of adding scope 3. We start an intuitive example displayed in Figure 5, which shows the log emissions in tons CO2e from MSCI and Trucost in a scatter plot. The blue dots represent scope $1 + 2$ emissions and the orange dots scope $1 + 2 + 3$ emissions. Different observations emerge. First, as seen before, scope 3 is clearly larger, but in some cases the scope $1 + 2$ can also become very large. Second, for scope $1 + 2$ a straight line is visible, which is not the case for scope $1 + 2 + 3$. This is caused by scope 1 and 2 being much more dependent on reported data than scope 3. Most of the disagreement in scope 1 and 2 emissions between these two data providers is in the left-bottom, where emissions are lower and less important when aggregated. Lastly, the dispersion becomes larger when scope 3 is included. Particularly on the high side we see larger outliers for scope 3. Given the log scale, the impact of these observations is large in a portfolio management setting, when absolute numbers are measured without log scale. In summary, for scope $1 + 2$ the data from providers is more in line than when scope 3 is added, which highlights the uncertainty around current scope 3 data.

The example above was limited to 1 universe and 1 data provider combination. To draw some more general conclusions, Table 7 shows a full overview of the dispersion between data providers, for all possible provider combinations and for all universes. The numbers in the table are the cross-sectional correlations between the log emissions from the different providers. For each universe these correlations are computed for both scope $1 + 2$ as well as for scope $1 + 2 + 3$. The final column shows the average over the different data providers. For instance, for the correlation between scope $1 + 2$ from CDP and the scope $1 + 2$ from ISS for the developed equity markets universe, we obtain 93%, but when we include scope 3 the correlation decreases to 87%. In fact, this table summarizes the information from the scatterplot in a single number. Only for emerging equity markets and the combination ISS and Trucost we see a higher correlation when scope 3 is included. As this occurs in only 1 out of 24 possible combinations, the results from this table can be regarded as additional evidence that scopes 1 and 2 are more consistent than scope 3. If we look at the average column, we also see that the difference with and without scope 3 for emerging equity markets is the smallest, which might be caused by emerging markets having relatively many companies in sectors where scope 3 can be estimated relatively easily.

4.6 Location- and market-based scope 2 emissions

In section 3.1 we discussed the difference between location- and market-based scope 2. In this subsection, we shed light on the impact of using one or the other. We only do this for the intensity metric for brevity reasons, but the footprint metric gives similar results. We do not consider scope 3 here, as the large size of scope 3 would dominate all other effects. Because only CDP delivers both scope 2 metrics, we can only perform this analysis with the data from



Source(s): Authors, MSCI, Trucost, Robeco. The constituents are from the MSCI World Index. The chart shows the greenhouse gas emissions in log tons CO2e greenhouse gas emissions data from 2020

Figure 5.
Scatterplot of carbon emission for developed equity constituents

this provider [13]. Lastly, we do not add inheritance here to have the cleanest view on the effect of location- vs market-based scope 2.

Table 8 shows the impact of preferring either one of the two metrics. The coverages of the two metrics are somewhat different. Clearly, location-based scope 2 always has a higher coverage than market-based scope 2. Almost all companies that report scope 2 report the location-based carbon emissions. In addition, some of these companies also report the market-based scope 2 emissions. There are only a handful of companies that only report the market-based figure. The second reason for the better coverage of location-based scope 2 is that market-based scope 2 cannot be estimated. It is really a company specific metric and cannot be proxied by sector averages. Thus, we know that the 72% of coverage for developed market equities is all reported data. The additional 26% to arrive at the total of 98% can both be reported or estimated location-based numbers.

The impact from the choice of the scope 2 method on intensities is summarized in the last two columns. The column scope 1 + 2 uses market-based if available and location-based otherwise. The last column always uses location based. The intensities show only small

Universe	Scope	CDP ISS	CDP MSCI	CDP Trucost	ISS MSCI	ISS Trucost	MSCI Trucost	Average
Equity DM	1&2	93	93	92	95	93	95	94
	1, 2 & 3	87	82	89	87	90	88	87
Equity EM	1&2	87	90	89	91	88	93	90
	1, 2 & 3	85	84	88	88	90	90	87
Corporate bonds IG	1&2	95	95	94	96	95	96	95
	1, 2 & 3	86	81	90	85	90	85	86
Corporate bonds HY	1&2	92	91	90	92	90	93	91
	1, 2 & 3	85	80	84	87	87	87	85

Source(s): Authors, CDP, ISS, MSCI, Trucost, Bloomberg, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets), MSCI Emerging Markets Index (equity emerging markets), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade), Bloomberg Global High Yield Corporate Index (corporate bonds high yield). Numbers are in percentages and represent the correlations of the log absolute emissions in tons CO2e between providers. The last column shows the average correlation over the different combinations of data providers. The index weights are from 31 August 2022. The greenhouse gas emissions data are from 2020

Table 7.
Correlation between data providers' absolute emissions

		Scope 2		Scope 1 + 2	
	Universe	Location	Market	Location	Market
Coverage	Equity DM	98%	72%	99%	99%
	Equity EM	96%	23%	96%	96%
	Corporate bonds IG	61%	48%	61%	61%
	Corporate bonds HY	31%	16%	31%	31%
Intensity	Equity DM			176	170
	Equity EM			375	372
	Corporate bonds IG			138	133
	Corporate bonds HY			413	409

Source(s): Authors, CDP, Robeco. The indices representing the asset classes are: MSCI World Index (equity developed markets; DM), MSCI Emerging Markets Index (equity emerging markets; EM), Bloomberg Global Aggregate Corporate Index (corporate bonds investment grade; IG), Bloomberg Global High Yield Corporate Index (corporate bonds high yield; HY). The four top rows contain the coverage for location-based (Location) and market-based (Market) scope 2 greenhouse gas emissions. The left two columns only contain scope 2 emissions, and the right two columns include both scope 1 and scope 2 emissions. For the market-based numbers in the two right columns, location-based emissions are used if market-based emissions are not available. The index weights are from 31 August 2022. The greenhouse gas emissions data are from 2020

Table 8.
The impact of location- and market-based scope 2

differences in aggregate. The largest differences are found for developed equity markets and investment grade corporate bonds, which have relatively high fractions of market-based reporting. Thus, whether a data provider uses location- or market-based scope 2 has relatively low impact on the carbon intensity of the asset class, even though for individual companies there may be substantial differences.

5. Conclusion

We have analyzed the impact of the choice of the data provider on carbon accounting metrics for four large global asset classes. Even though methodologies across data providers differ, the effects are rather small for developed equity markets, and only slightly higher for emerging equity markets, when limited to scopes 1 and 2.

Coverage is substantially lower for investment grade and high yield corporate bonds, mainly because many corporate bond issuers do not have their shares listed on public equity markets. Coverage can be increased by searching for the parent company of the bond issuer and inheriting the carbon intensity or footprint reported by the parent. Another, cruder, method to increase coverage is to apply the median industry carbon intensity or footprint to companies with missing data. These methods to improve coverage reduce differences in carbon intensity and carbon footprint at the asset class level.

Scope 3 intensities and footprints need to be (partly) estimated more often and are therefore noisier and differ more across data providers. Since their magnitude is typically five times as large as scope 1 and 2 emissions, the differences in scope 3 estimates dominate total carbon intensity and footprint data.

All in all, contrary to the choice of an ESG-score data provider, the choice of a carbon emission data provider is relatively small for portfolio managers and other stakeholders, at least when equity market investments are analyzed for scope 1 and scope 2 emissions. This suggests that investors can measure this metric relatively accurate, which is beneficial for creating investment portfolios with a climate angle. For scope 3, the choice of data provider may have a substantial effect on the portfolio carbon statistics. Measuring value chain emissions are thus less mature, which could result in different optimal portfolios when a different data provider is used. For corporate bonds, plain data coverage is lower than for equity markets, which increases differences among data providers. Imputing missing data by applying inheritance from the parent company or applying sector defaults may increase or decrease differences, depending on the sector distribution of the imputed values. Therefore, continuous improvements on reporting and data quality are indispensable for portfolio managers and other stakeholders, as greenhouse gas emissions are such an important cause for climate change.

Our final note regards the use of carbon data for actively managed high conviction portfolios. Particularly when the number of securities in such portfolio is small, the impact of the choice of the data provider may become larger than in broad, well diversified, indices that we have used in our analyses. In case a portfolio is managed against a certain carbon objective, the choice of the data provider might become even more important as the data on carbon is then a direct input to portfolio construction. Reducing the carbon intensity or carbon footprint of such portfolio may then end up overweighting those securities where the carbon data is poorly estimated to be lower than it in reality is.

Notes

1. For more detailed information and amendments to the regulation, see here.
2. Kalesnik *et al.* (2022) do not consider scope 3 as they deem the sample of companies reporting scope 3 too small.
3. Equity investment strategies dealing with climate risk can for example be found in Andersson *et al.* (2016) and Kolle *et al.* (2022), among others. Mastouri *et al.* (2022) highlight the importance of climate change risk in corporate bond portfolios. Kumar and Firoz (2018) analyze the effect of carbon emissions on the cost of debt for Indian companies.
4. For example, Bolton and Kacperczyk (2021) use greenhouse gas emissions data from Trucost over the period 2005 to 2017.
5. A common definition of EVIC is the market cap of ordinary and preferred shares and the book value of minority interest and total debt.
6. Revenues and EVIC are also advocated by the Task Force on Climate-Related Financial Disclosures Task (TCFD) and the Partnership for Carbon Accounting Financials (PCAF).
7. There may be incentives to underreport carbon emissions, so that the company seems less polluting than it is. However, initial underreporting complicates reducing carbon emissions over time. See,

e.g. Andrew and Cortese (2011) for a discussion on the potential challenges associated with voluntary carbon disclosures and Haque and Islam (2015) for a discussion on carbon accounting fraud. Beauchamp and Cormier (2022) discuss the effect of reporting on embedded CO_2 in proven reserves by oil and gas companies on their equity market value.

8. We were not able to obtain historical data for ISS and CDP.
9. Due to different cleaning approaches data labeled as reported can also differ between data providers.
10. A sector imbalance could also be solved by only considering the index constituents for which all data providers have data. If we apply this approach the main conclusions do not change, although the difference for investment grade bonds diminishes as the private company issue is not relevant anymore. The largest drawback of this approach is that it discards valuable carbon accounting data, with joint coverage decreasing to 97% (Equity DM), 95% (Equity EM), 59% (Corporate bonds IG), 29% (Corporate bonds HY).
11. Changes at the index level may be due to a change in the index composition from year to year or because of changes in company carbon disclosures over time. We prefer to keep the same set of companies, to isolate the effect change at the company level. The longer we go back in history, the less representative a fixed set of companies is for the investment universe at the time. We find four years a reasonable compromise.
12. To keep the charts readable, we have added the categories Utility Other, Industrial Other and Natural Gas to the categories Electric, Basic Industry and Energy.
13. ISS reports market-based if available, location-based otherwise. However, they do not provide both market-based and location-based for one company. Hence, using their data, we cannot compare the effect of switching between market-based and location-based.

References

Andersson, M., Bolton, P. and Samama, F. (2016), "Hedging climate risk", *Financial Analysts Journal*, Vol. 72 No. 3, pp. 13-32.

Andrew, J. and Cortese, C. (2011), "Accounting for climate change and the self-regulation of carbon disclosures", *Accounting Forum*, Vol. 35 No. 3, pp. 130-138.

Beauchamp, C. and Cormier, D. (2022), "Corporate disclosure of CO_2 embedded in oil and gas reserves: stock market assessment in a context of global warming", *Managerial Finance*, Vol. 48 No. 12, pp. 1689-1706.

Berg, F., Kölbel, J. and Rigobon, R. (2022), "Aggregate confusion: the divergence of ESG ratings", *Review of Finance*, Vol. 26 No. 6, pp. 1315-1344.

Bolton, P. and Kacperczyk, M. (2021), "Do investors care about carbon risk?", *Journal of Financial Economics*, Vol. 142 No. 2, pp. 517-549.

Busch, T., Johnson, M. and Pioch, T. (2022), "Corporate carbon performance data: Quo vadis?", *Journal of Industrial Ecology*, Vol. 26 No. 1, pp. 350-363.

Dimson, E., Marsh, P. and Staunton, M. (2020), "Divergent ESG ratings", *Journal of Portfolio Management*, Vol. 47 No. 1, pp. 75-87.

Doeswijk, R., Lam, T. and Swinkels, L. (2014), "The global multi-asset market portfolio, 1959-2012", *Financial Analysts Journal*, Vol. 70 No. 2, pp. 26-41.

Doeswijk, R., Lam, T. and Swinkels, L. (2020), "Historical returns of the market portfolio", *Review of Asset Pricing Studies*, Vol. 10 No. 3, pp. 521-567.

European Central Bank (2022), ECB report on good practices for climate stress testing, Frankfurt am Main.

Furdak, R., Nilsen-Ames, T. and Wang, F. (2022), "Carbon emissions: under the microScope 3", *Journal of Impact and ESG Investing*, Vol. 3 No. 2, pp. 20-35.

Gibson Brandon, R., Krueger, P. and Schmidt, P. (2021), "ESG rating disagreement and stock returns", *Financial Analysts Journal*, Vol. 77 No. 4, pp. 104-127.

Hain, L., Kölbel, J. and Leippold, M. (2022), "Let's get physical: comparing metrics of physical climate risk", *Finance Research Letters*, Vol. 46B, 102406.

Haque, S. and Islam, M. (2015), "Carbon emission accounting fraud", in Schaltegger, S., Svezdov, D., Alvarez Etxeberria, I., Csutora, M. and Günther, E. (Eds), *Corporate Carbon and Climate Accounting*, Springer, Cham.

Heurtebize, T., Chen, F., Soupé, F. and Leote de Carvalho, R. (2022), "Corporate carbon footprint: a machine learning predictive model for unreported data", *Journal of Impact and ESG Investing*, Vol. 3 No. 2, pp. 36-54.

Hoepner, A. and Rogelj, J. (2021), "Emissions estimations should embed a precautionary principle", *Nature Climate Change*, Vol. 11, pp. 638-640.

Kalesnik, V., Wilkens, M. and Zink, J. (2022), "Do corporate carbon emissions data enable investors to mitigate climate change?", *Journal of Portfolio Management*, Vol. 48 No. 10, pp. 119-147.

Kolle, J., Lohre, H., Radatz, E. and Rother, C. (2022), "Factor investing in Paris: managing climate change risk in portfolio construction", *Journal of Investment Management*, Vol. 20 No. 4, pp. 35-51.

Krueger, P., Sautner, Z. and Starks, L. (2020), "The importance of climate risks for institutional investors", *Review of Financial Studies*, Vol. 33 No. 3, pp. 1067-1111.

Kumar, P. and Firoz, M. (2018), "Impact of carbon emissions on cost of debt-evidence from India", *Managerial Finance*, Vol. 44 No. 12, pp. 1401-1417.

Liesen, A., Hoepner, A., Patten, D. and Figger, F. (2015), "Does stakeholder pressure influence corporate GHG emissions reporting? Empirical evidence from Europe", *Accounting, Auditing and Accountability Journal*, Vol. 28 No. 7, pp. 1047-1074.

Mastouri, A., Mendiratta, R. and Giese, G. (2022), "Corporate bonds and climate change risk", *Journal of Portfolio Management*, Vol. 48 No. 10, pp. 78-97.

Papadopoulos, G. (2022), "Discrepancies in corporate GHG emissions data and their impact on firm performance assessment", European Commission, JCR Working Papers in Economics and Finance 2022 No. 12.

Corresponding author
Laurens Swinkels can be contacted at: lswinkels@ese.eur.nl