PRELIMINARY -- THIS VERSION: 12. FEBRUARY 2024

Harnessing the Power of Input-Output Analysis for Sustainability[1]

A simulation study based on US data

Ulf von Kalckreuth
Deutsche Bundesbank

Abstract

Measuring carbon contents reliably, for products, firms and industries, is key for identifying transition risks. Phase 3 of the G20 Data Gaps Initiative asks for collecting emission data and multiregional IO tables to enable the calculation of aggregate carbon contents. What sectoral distinctions do we need, what level of granularity? What is the role of international linkages? Do we need information on technology? How can statistical data be used in carbon accounting? Based on IO tables and company level data from the United States, I build up a micro simulation environment that can act as a laboratory for answering these questions. The data base consists of almost 5000 units located in the United States and Canada (with few exceptions) and enables a rather complete tracking of private economy value chains. The analysis takes a focus on indirect emissions and carbon contents.

First results indicate that, for levels of disaggregation typical for real world IO data, the within-sector heterogeneity of carbon contents is very high in some industries. Exclusive use of aggregate IO data is not warranted on the company level. However, statistical data can be very useful in providing starting values for inputs from industries with low heterogeneity, such as many service industries, in cases where direct information is missing. They may also be used to approximate indirect emissions, when company level information on direct emissions is available. With the upcoming reporting requirements in place, this will be the standard case.

Keywords: greenhouse gas intensities, carbon accounting, green finance

JEL classification: Q56, Q51, C81

[1] Contact address: Ulf von Kalckreuth, Deutsche Bundesbank, Postfach 100602, 60006 Frankfurt am Main, Germany. Phone: +49 (0)69 9566 36010. Email: ulf.von-kalckreuth@bundesbank.de. The author is Principal Economist-Statistician at the DG Statistics of the Deutsche Bundesbank. He is grateful for comments from Karthik Ramanna, Robert Kirchner, Matthew Chambers and Fabienne Fortanier, as well as from participants of presentations at the World Statistics Congress in Ottawa, on July 18, 2023 and at the Bundesbank In-house Workshop on Climate Research in Frankfurt, on March 10, 2023. This paper represents the author’s personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff.

Harnessing the Power of Input-Output Analysis for Sustainability

A simulation study based on US data

1 Introduction

Carbon contents are a key input for all sorts of allocation decisions, for consumers, investors and government agencies, and for the reliable identification of transition risks. Carbon content disclosures and estimations can come on various levels: national, sectoral, group and single company, installations and – without a time dimension – the product level. Quite generally, a major problem for estimating carbon contents are Scope 3 emissions: the carbon dioxide emitted for the production of intermediate inputs. Producers may know their inputs, but they still need good estimates of the carbon contents of these inputs unless there is direct information from providers. In trade policy, it is extremely important to reliably assess the carbon content of imports, in order to avoid carbon leakage.

Input-Output (IO) models provide the natural basis for organising the available information. On a sectoral basis, they take account of all production interlinkages – at least conceptually – using data that is available in most countries, often in a harmonised way. Combining the Input-Output matrix with industry level information on direct emissions, one can readily track those emissions over the entire value chain. Statisticians spend considerable resources to make this information available and to keep it up to date.

Within the framework of Phase 3 of the G 20 Data Gaps Initiative (DGI),[2] Recommendation 1 on greenhouse gas emission accounts and national carbon footprints asks countries and International Organisations for enhancing IO tables and emission statistics in such a way that consistent data is available for all major economies: regarding sector definitions, interlinkage information, information on import and export of intermediary inputs and direct emission statistics for industries.

The project presented here looks for how aggregate measurement and IO tables can best be used and developed as an important source for firm level and product level estimates. Specifically, I investigate the case of the USA. The Bureau of Economic Analysis (BEA) works out extremely refined IO tables. Roughly every 5 years, benchmark statistics with no less than 405 industries and product groups are produced, in addition to the annual tables for 71 industries. In addition, the coverage of US companies in micro level databases on carbon emissions is generally much better than for any other country. Information on trade interlinkages and emissions are available for a reduced set of industries from OECD IO tables.

The paper is structured as follows. After the Introduction, Section 2 describes the idea of setting up a simulation lab as a tool for designing and evaluating aggregate statistics. Section 3

[2] See Ducharme (2022).

shows how the simulation is set up, using micro data from information providers and combining these with rather disaggregated industry level information on production interaction. This can be used to study the information content of more aggregated statistical information. The main challenge is to fit the micro level data into the structure of the existing information on interactions. Section 4 then introduces the relevant measurement concepts and the framework needed for computing carbon contents from data on direct emissions and production interactions. Section 5 gives a descriptive view of the data. The preliminary look has a focus on direct and indirect emissions and the resulting carbon contents. For many industries, companies are very heterogeneous in their direct emissions even at the lowest level of aggregation.

Ultimately, Section 6 gives a first and as yet incomplete attempt to assess the predictive use of aggregate level statistical information for assessing the carbon content of output. I distinguish the direct use of industry level data as a predictor from the use in a company level carbon accounting framework. The former means, for example, using industry level information for evaluating the carbon footprint of asset portfolios. The latter indicates the use of industry level information as a substitute of missing direct information on elements of the value chain. Given the high within-sector heterogeneity of carbon contents in some industries, the exclusive use of aggregate IO data is not warranted on the company level, not even as starting value for iterative carbon accounting procedures. On the other hand, statistical data can be useful in providing starting values for inputs from industries with low heterogeneity, such as many service industries, or with a low share in total input, when direct information is missing. Ideally, accountants have direct information on Scope 1 emissions of suppliers, e.g. from ESG reporting, and only need to fill up information gaps regarding indirect emissions.

2 A simulation lab

As an information source on company level emissions, I use Trucost Environmental Data. These data set has industry classifications for the companies that is closely related to the BEA industry divisions for Input Output tables. Both are based on the NAICS system of company classification. The micro database contains information on direct emissions and energy use, together with information on the industry and revenues. The data base consists of about 5000 units, almost exclusively from the United States and Canada. It allows a rather complete tracking of private economy value chains. Micro level carbon contents are calculated using the methodology developed and explained in von Kalckreuth [2022a, 2022b].

This micro-simulation is **a laboratory to assess various questions of measurement** related to the DGI Recommendation mentioned above, specifically:

- How important are granular sectoral distinctions in areas of activity where emissions are heterogeneous and/or high?
- How important is an explicit account of international interlinkages?
- How well can sectoral data serve as proxies for the carbon content of company level output or products?

- How informative are they as inputs in carbon accounting?

Based on the knowledge of the simulated “truth”, it is possible to compute the average error associated with any measurement method. This amounts to setting up an infrastructure that will enable us to discuss measurement issues consistently and on a quantitative basis.

3 Building the data base

The principal goal in setting up the data base is to reconstruct and simulate the value chains of production, making use of the detailed BEA Input Output Table with its 405 industries. To this end, it is important to find micro level representation for as many BEA 405 industries as possible. I start by working out a correspondence table between the Trucost classification and BEA 405. Both are based on NAICS. Not for all BEA 405 industries there are counterparts in the Trucost classification and vice versa. In cases where a BEA 405 industry has no counterpart in the Trucost data, I assign companies from closely related classes within the same BEA 71 grouping. Therefore, a given company may be used as representative for more than one BEA 405 class. Thus, for a suitable assignment, it is important to capture the structure and heterogeneity of direct emission intensities and the use of energy (Scope 2 emission intensities). The structure of production interactions will be borrowed from IO tables and will be different according to BEA 405 industry.

I concentrate on observations from 2020. If for a given company there is no observation for 2020, I take the latest observations from the period 2016 and after. The sectors are filled using companies from the United States and Canada. Only if there is no such company available, companies from other parts of the world are being used as sector representatives, with a preference for European firms. The resulting micro level database consist of 4,988 units, with the following regional representation:

Table 1: Regional composition of simulation micro data base

Region	**Freq.**	**Percent**	**Cum.**
Europe	69	1.38	1.38
Asia / Pacific	68	1.36	2.75
Africa / Middle East	4	0.08	2.83
USA and Canada	4,846	97.15	99.98
Latin America and Caribbean	1	0.02	100.00
Total	**4,988**		

The data on direct emissions and Scope 2 emissions for the 4,988 units come from 3,818 different companies. The difference is due to using data given sets of Trucost companies for the simulation of more than one detailed level industry when this was necessary for completing the value chains. The resulting data set has representatives for 389 out of 405 detailed level industries and 67 out of 71 summary level industries. The micro data is on listed companies, thus it misses all government activity, private households, religious organisations and independent artists, writers and performers. Apart from these, the coverage is complete.

This information is linked to Input Output Accounts data from BEA[3]. First, I generate symmetric industry by industry direct requirement matrices from the Commodity by Industry direct requirement matrix and the Industry by Commodity transformation matrix[4]. For every industry, this matrix indicates the value of inputs from any other industry needed to generate one dollar's value of output. The last detailed level direct requirement matrix dates from 2012. This detailed level matrix is extrapolated to 2020 using summary level matrices for 2012 and 2020.[5] In a few cases, adjustments were made to prevent the value added becoming too small.[6]

The resulting input-output matrix is "blown up" to the micro level by randomly assigning to each company one representative from each sector from which it receives inputs. This is done in order to preserve heterogeneity. Two units in the same sector may thus be linked to companies with rather different carbon intensity. The resulting 4988 x 4988 interaction matrix is modified by making direct use of data on energy use. The micro data has information on Scope 2 (first tier) carbon intensity. Assuming that this type of indirect emission comes mostly from electricity, the Scope 2 data is converted to unit specific input coefficients for electricity, using the weighted average direct carbon intensity of electricity producers in the Trucost data.[7] For consistency reasons, I set up a notional electricity distribution agent ("the grid") that buys all electricity produced among the units in the data base and sells it to the users of electricity. With this mechanism, the resulting Scope 2 emissions in the simulation are equal to the data provided by Trucost by definition.

4 Measurement concepts: indirect emissions and carbon content[8]

Carbon content is defined recursively: it is the sum of direct emissions attributed to a product and the carbon content of all inputs, covering indirect emissions. Indirect emissions are the result of direct emissions in a chain – or rather a fabric – of other production processes. Those production interlinkages are key for the consistent treatment of indirect emissions. IO analysis is designed for this type of interlinkages, and in fact it has been used in tackling the issue of attributing resource consumption to final output at the sectoral level since the 1970s.

4.1 An IO view

To fix ideas, consider the following. In production planning, every process is defined by a *bill of material* (BoM) that specifies all inputs, plus a *route sheet* that explains how to combine them. A complex production process may be decomposed into several stages. Consider the BoM of product *k*,

$$\mathbf{a}_k = \begin{pmatrix} a_{k1} & a_{k2} & \dots & a_{kK} \end{pmatrix}' ,$$

[3] See BEA Input Output Accounts Data, downloaded 17.03.2023 and before.
[4] See the note Mathematical Derivation of the Total Requirements Tables for Input-Output Analysis, BEA 2017
[5] A detailed level matrix for the year 2017 is about to be published by the BEA and will allow a better approximation.
[6] After extrapolating, value added was lower than 10% in six industries. In these cases, input coefficients were decreased proportionately to obtain a value added of exactly 10%. In the case of public transport, no adjustment was made, as this industry has a significant negative value added in the original requirement matrix of 2012 already.
[7] As above, manipulating the requirement coefficients requires adjustments. The electricity requirement coefficient derived from the micro level information on energy use were transformed by a nonlinear moderator function that prevents the resulting value added becoming too small.
[8] The content of this section is adapted from Section 2.2 in von Kalckreuth [2022a].

with a_{ki} being the quantity of good *i* that enters the production process. There are entries for all input goods in the economy, most of them with a value of zero, of course. Let the amount of GHG emitted directly be given as d_k. Let scalar c_i be the carbon content of good *i*, the quantity of GHG that is emitted in the production of one unit. List the carbon contents of all input goods in a vector as well:

$$\mathbf{c} = \begin{pmatrix} c_1 & c_2 & \dots & c_K \end{pmatrix}' .$$

The carbon content of product *k* is then given as the sum of direct and indirect emissions. Importantly, we do not add a definition for indirect emissions, but simply define them recursively as the carbon content of inputs:

$$c_k = d_k + \mathbf{c}'\mathbf{a}_k = d_k + \sum_i c_i a_{ki} \quad . \tag{1}$$

Indirect emissions are the direct emissions at earlier stages of the value chain. The equation is both perfectly general and encompassing. It relates to products and activities and – for a given time span – to enterprises and sectors as well.

As it stands, the equation is a definition. It helps us understand the problems associated with gathering and processing information. For actual computation, all the c_i corresponding to the BoM of product *k* are required. If these are known, we can calculate the carbon content of product *k* in a straightforward way from direct emissions and the BoM. This is like computing the energy content of food: it is enough that producers know the composition of their product and the energy content of the ingredients. How can carbon contents of outputs be calculated in a world where not all inputs carbon contents are known? Product carbon contents are interdependent – the value for any product will depend on the value of all inputs.

4.2 A reduced form for product carbon contents

If the relevant elements of **c** are unknown, we can use equation (1) recursively and try to compute the carbon content involved, going up the value chain from more complex intermediate inputs down to primary and primitive inputs. The structure is well known from linear production planning and IO analysis, pioneered by Wassily Leontief, and it was indeed the same author who first proposed using IO models for analysing pollution generation associated with inter-industry activity.[9] Conceptually, we can solve for the carbon content of all products simultaneously. Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_K \end{pmatrix}$$

be the matrix of the BoMs for all output goods, 1,…,*K*. With $\mathbf{d}$ being the column vector of the associated direct emissions, one may write:

$$\mathbf{c}' = \mathbf{d}' + \mathbf{c}'\mathbf{A} . \tag{1'}$$

Reordering and postmultiplying the "Leontief inverse" $\left(\mathbf{I} - \mathbf{A}\right)^{-1}$ yields:

[9] Wassily Leontief was awarded the 1973 Nobel Prize for the development of IO analysis. Leontief [1986] covers much of his work. Leontief [1970] himself introduced pollution by augmenting the technology matrix to include a row of pollution generation coefficients, see Qayum [1994] for a consistent reformulation. The direct approach taken here, postulating a proportional relationship between output and pollution, was first advanced, on a sectoral basis, by Just [1974] and Folk and Hannon [1974]. The formulations are equivalent. For IO analysis in general, see Miller and Blair [2022], and specifically Chapter 10 for environmental IO analysis. Suh [2010] is a collection of extensions and applications in the field of industrial ecology.

$$\mathbf{c}' = \mathbf{d}'(\mathbf{I} - \mathbf{A})^{-1}. \tag{2}$$

The carbon contents (product *k* and all the others) result from their own direct emissions and the direct emissions of all the intermediate goods used for their production by intermediation of a matrix derived from the BoM that reflects the interlinkages in production. If the coefficients in the carbon content equation refer to empirical production technologies actually being used to produce goods, 1, …, *K*, it can be taken for granted that the inverse exists and all its elements are non-negative.

As simple and beautiful as this relationship is, it is not possible to use it directly. Matrix $\mathbf{A}$ comprises the BoMs for all products in the economy, including those that have been imported, and if a certain input is produced using two different technologies, it should actually have two separate entries. Meanwhile, vector $\mathbf{d}$ collects the direct emissions that characterise all of these processes. Except for simple cases, this cannot be dealt with at the micro level. Von Kalckreuth [2022a] shows that this is not necessary. Producers do not have to be aware of all the stages of the value chain – they only need to know their own technology and (preliminary) values of the carbon contents of the inputs as provided by their immediate suppliers. If these values are not correct, the iterative use will make them so. Just as the price mechanism is able to process an enormous amount of information in a decentralised way, the exchange of information between producers do the rest of the work. With the E-Liability carbon accounting approach, Kaplan and Ramanna [2021a, 2021b] have suggested a process that enables the necessary information exchange. One question this paper tries to answer is where initial values for an encompassing system of carbon accounting may come from.

5 A look at the data

At the time of writing this draft, the construction work for the data base is not yet finished. Specifically, the micro information have yet to be linked to aggregate BEA data. However, it is very interesting to look at the heterogeneity on the micro level. Using relationship (2) on the micro data on direct emission intensities and the micro level requirement coefficients in the simulation universe, I calculate the “true” unit specific carbon contents, with the associated indirect emission intensities. The same can be achieved by using relationship (1)’ iteratively. Table 2 gives some descriptive statistics: on revenues and on direct emissions, indirect emissions and carbon contents – the latter three both weighted and unweighted.

To convey an idea of industry heterogeneity, Table 3 gives weighted means of direct emissions and carbon content (the sum of direct and indirect emissions) for three BEA 71 industries: 11CA ‘Farms’, 22 ‘Utilities’ and 325 “Chemical products”. The table renders the averages of direct emissions and carbon contents on the level of the BEA 71 aggregate and the BEA 405 industries. The averages are weighted by sales revenues. These tables can by no means interpreted as valid statistical information. Indirect emissions are simulated, the assignment of suppliers to producers is random and in many of the BEA 405 cells there are not more than one or two units. However, they give an impression of the type of heterogeneity involved. The direct emission intensities for the various modes of running a farm are surprisingly diverse. It is interesting to see how in much of the chemical industry direct emissions are dominated by indirect emissions. The example of "utilities" as a compound of electricity, gas distribution and water, sewage and other systems shows how bad a coarse sectoral classification may be geared to the need of assessing emission intensities.

Table 2: Descriptive Statistics

a) Unweighted

Variable	Mean	Std dev	Min	Max
Sales Revenue (k US$)	4,782.3	21,313.7	0.0	523,964.0
Dir emission int. CO_2e, g/US$	119.4	598.8	0.0	22,366.0
Indir emission int, CO_2e, g/US$	180.5	214.4	4.5	2,343.5
Carbon content, CO_2e, g/US$	299.9	679.3	5.2	23,598.3

b) Weighted by sales

Variable	Mean	Std dev
Dir emission int. CO_2e, g/US$	113.3	476.9
Indir emission int, CO_2e, g/US$	168.6	201.1
Carbon content, CO_2e, g/US$	281.9	553.5

4,988 Observations on all variables

To gain an impression of the variability on the micro level, we may first look at BEA 71 industry 22 'utilities', with its three constituent BEA 405 industries: 'Electric power generation, transmission, and distribution, 'Natural gas distribution', and 'Water, sewage, and other systems'. Graph 1 is a scatterplot of direct emissions from individual level data and simulated indirect emissions for the utilities industry. It is obvious that knowledge of the detailed industry confers important information on the order of magnitude of direct and indirect emissions, but that there is important heterogeneity not accounted for by information on detailed industry.

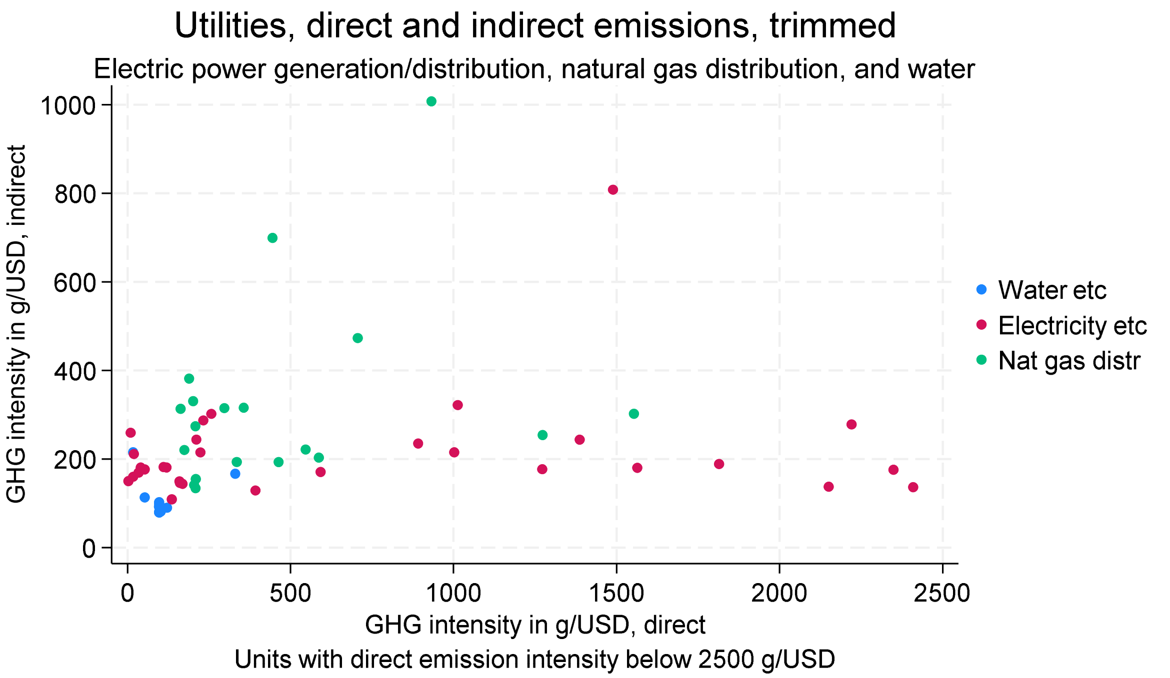
Graph 2 does a similar decomposition for the BEA 71 industry 325 'Chemical Products'. Table 1 shows the strong heterogeneity on the level of detailed industries, and Graph 2 gives an impression of the underlying micro level dispersion. Because of outliers, the scatter plot is trimmed at a value of 2500 g/$ for direct emission intensity. It is visible that the high intensity units are concentrated in a small subset of the detailed level industries that comprise BEA 71 'Chemical Products".

Table 3: Weighted averages of direct emission intensities and carbon contents in three BEA 71 industries

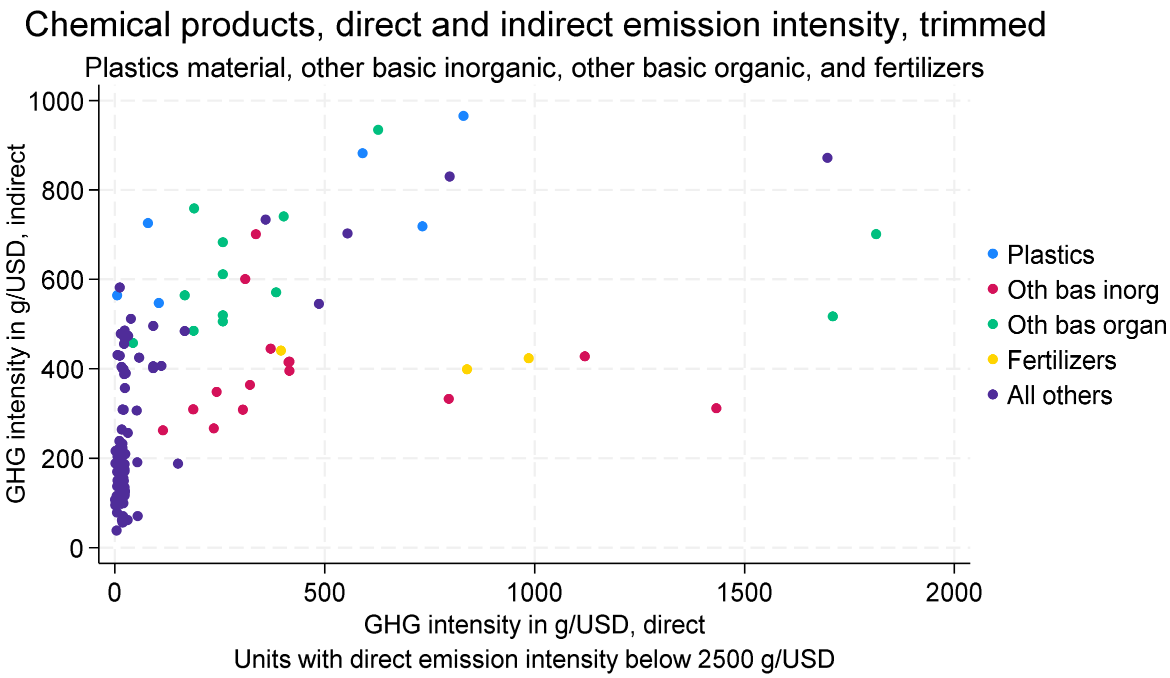
BEA 71 industries	**Emission intensities (CO_2**	
Farms	**direct em.**	**carbon conte**
BEA 405 industries		
Oilseed farming	1,604.2	1,941.1
Grain farming	1,096.6	2,025.3
Vegetable and melon farming	2,056.7	2,456.7
Fruit and tree nut farming	1,642.8	1,940.2
Greenhouse, nursery, and floriculture production	1,811.3	3,444.4
Other crop farming	578.4	1,041.4
Dairy cattle and milk production	662.5	1,506.0
Beef cattle ranching and farming, including feedlots and dual purpose ranching and farming	662.5	1,877.1
Poultry and egg production	1,715.3	2,807.4
Animal production, except cattle and poultry and eggs	1,040.2	1,306.0
Total	**843.2**	**1,636.2**
Utilities		
BEA 405 industries		
Electric power generation, transmission, and distribution	2,517.8	2,745.7
Natural gas distribution	809.5	1,231.5
Water, sewage and other systems	99.3	265.2
Total	**2,216.4**	**2,472.6**
Chemical products		
BEA 405 industries		
Petrochemical manufacturing	554.3	1,256.9
Industrial gas manufacturing	1,697.5	2,569.3
Synthetic dye and pigment manufacturing	797.7	1,627.8
Other basic inorganic chemical manufacturing	533.4	1,001.2
Other basic organic chemical manufacturing	670.2	1,355.1
Plastics material and resin manufacturing	653.3	1,417.1
Synthetic rubber and artificial and synthetic fibers and filaments manufacturing	407.8	1,069.5
Medicinal and botanical manufacturing	23.3	151.3
Pharmaceutical preparation manufacturing	17.0	153.5
In-vitro diagnostic substance manufacturing	20.5	164.2
Biological product (except diagnostic) manufacturing	9.4	73.1
Fertilizer manufacturing	1,595.3	2,043.5
Pesticide and other agricultural chemical manufacturing	74.9	458.7
Paint and coating manufacturing	19.3	490.2
Adhesive manufacturing	103.6	508.8
Soap and cleaning compound manufacturing	26.2	279.9
Toilet preparation manufacturing	6.5	220.8
Printing ink manufacturing	34.4	531.7
All other chemical product and preparation manufacturing	33.6	422.5
Total	**168.2**	**455.0**

This type of heterogeneity does not prevail everywhere. In large parts of the service sector, such as trade or where office work is predominating, direct and indirect intensities are low and uniform, the first mainly due to commuting and travel, the second to heating and electricity. Other services, such as transportation, are heterogeneous and in parts highly carbon intensive. Appendix 1 gives an overview of unweighted industry averages and standard deviations for direct emissions intensity and carbon contents according to BEA 71 industries. It readily appears that heterogeneity is enormous for some industries, while quite moderate for others.

Graph 1



Graph 2



6 Using industry level data for micro level predictions: first results

It is natural to attempt using industry averages as predictors or estimates for individual level outcomes. Actually, the European Commission is doing so on a large scale. The EU taxonomy for sustainable activities is simply a binary classification relying on industry as predictor. In the following, I will start by computing the Root Mean Squared Error (RMSE) of the weighted average using (1) the BEA 71, and (2) the BEA 405 industries as basis.[10] In both cases, only companies in BEA 405 industries with at least 3 units will be considered. The first is what can typically be achieved using statistical information based on the System of Environmental Economic Accounts (SEEA)[11]. Most regularly published national level IO Tables feature a comparable number of industries. The second is, so to speak, the best possible sectoral predictor, at least for indirect intensities: in our idealised world, it is BEA 405 information that underlies the simulated production interlinkages.

In addition, I will consider two predictors that use the rather coarse BEA 71 intensity information in combination with micro level information on input composition. These predictors, labelled "carbon accounting predictors", use correct unit level information on direct emissions and evaluate indirect emission intensity on the basis of equation (1). Estimates for input carbon contents are derived with the help of BEA 71 industry averages. This is the type of computation producers themselves can do: they know their own production routines well, but may not have first-hand information about the emission intensities of their suppliers. I distinguish two versions. A *naïve carbon accounting* solution uses industry averages of total carbon contents (direct and indirect) for the valuation of their inputs. The *advanced carbon accounting* version goes one step further: it evaluates inputs using a composite indicator as the sum of (true) direct emissions intensity and BEA 71 industry averages of indirect emissions intensities of input providers. That is, the producer is assumed to know their own direct emissions as well as the direct emission intensities of their suppliers, relying on statistical information only for evaluating the indirect emissions of suppliers.

Table 4: Predictors for emission intensities – comparing RMSEs

Predictor	RMSE direct emission intensity	RMSE indirect emission intensity	RMSE total carbon content
BEA 71 weighted average	349.5	101.9	363.9
BEA 405 weighted average	311.5	51.7	318.0
Naïve carbon accounting: valuation of inputs using BEA 71 weighted average	(0)	72.9	72.9
Advanced carbon accounting: valuation of inputs using composite indicator	(0)	21.1	21.1

Notes: RMSEs are roots of weighted mean squared prediction errors. They are calculated for units with an industry representation of 3 units at least. For carbon accounting estimators, RMSEs for direct emissions are zero by definition. The composite indicator for evaluating inputs in carbon accounting combines true direct emission intensities with weighted BEA 71 industry averages for indirect estimates.

[10] Strictly speaking, the sector level predictors need to be calculated on the basis of the Leontief inverses for industry aggregates, instead of averaging unit level results on the basis of a micro level Leontief matrix. Inverting a matrix is a non-linear operation, and due to the aggregation bias the results will not be identical. This will be completed at a later stage.

[11] The SEEA is a standard maintained by the United Nations, following similar accounting structures as the Standard of National Accounts (SNA), see System of Environmental Economic Accounting.

The results are collected in Table 4. The first two lines show the weighted Root Mean Squared Errors (RMSE) of predictors that directly use industry level data. For comparison: the overall weighted average of direct and indirect emission intensities is 113.3 g/$ and 168.6 g/$, respectively, see Table 1 above. The high RMSE make clear that sectoral estimates for direct emission intensity are rather useless as predictors on the micro level, at least unconditionally. This is true for both the coarse BEA 71 average and the much more sophisticated BEA 405 average. Any use of industry level information on the level of micro entities will have to be selective.

The RMSE from the sectoral estimations are clearly smaller for indirect emissions than for direct emissions. This may reflect some amount of averaging, as indirect emissions come from many inputs. Furthermore, indirect emissions, reflecting the nature of the inputs, may indeed be stronger conditioned by industry than direct emissions. In addition, of course, the simulation might yet be missing important sources of variation for indirect emission.

For the two carbon accounting indicators, the errors for indirect emission intensities and total carbon content are identical by definition. The “naïve” carbon accounting estimate using the coarse BEA 71 information to evaluate inputs, combined with using adequate information on production technology, is right in the middle between the outcomes for the BEA 71 estimator and the sophisticated BEA 405 estimator. App. 2 shows detailed industry level results for the BEA 71 and the composite indicator.

With the upcoming reporting requirements in the EU, there will often be exact and reliable information on Scope 1 and Scope 2 emissions on the company level. Useful information on Scope 3 emissions is much harder to obtain, as the relevant guidelines[12] leave many options, and data availability is a big concern for accountants. In those cases, industry level statistical information on indirect emissions may be a very useful complement for unit level information on direct emissions – much of the within-sector heterogeneity of carbon contents is due to the direct emissions component.

For a composite indicator, the error dispersion for indirect emission intensity and overall carbon contents are equal by definition. Using it as valuation vector for inputs will bring down the RMSE of carbon accounting down to 21.1. This is the advanced carbon accounting indicator Again, App. 2 shows the details by industry. In most cases, RMSEs for carbon accounting using a composite indicator are very low. For some industries, however, using industry level information to guide the evaluation of inputs are clearly insufficient. This is specifically true for farms, plastic and rubber products, textile mills, fabricated metal products and some other manufacturing industries.

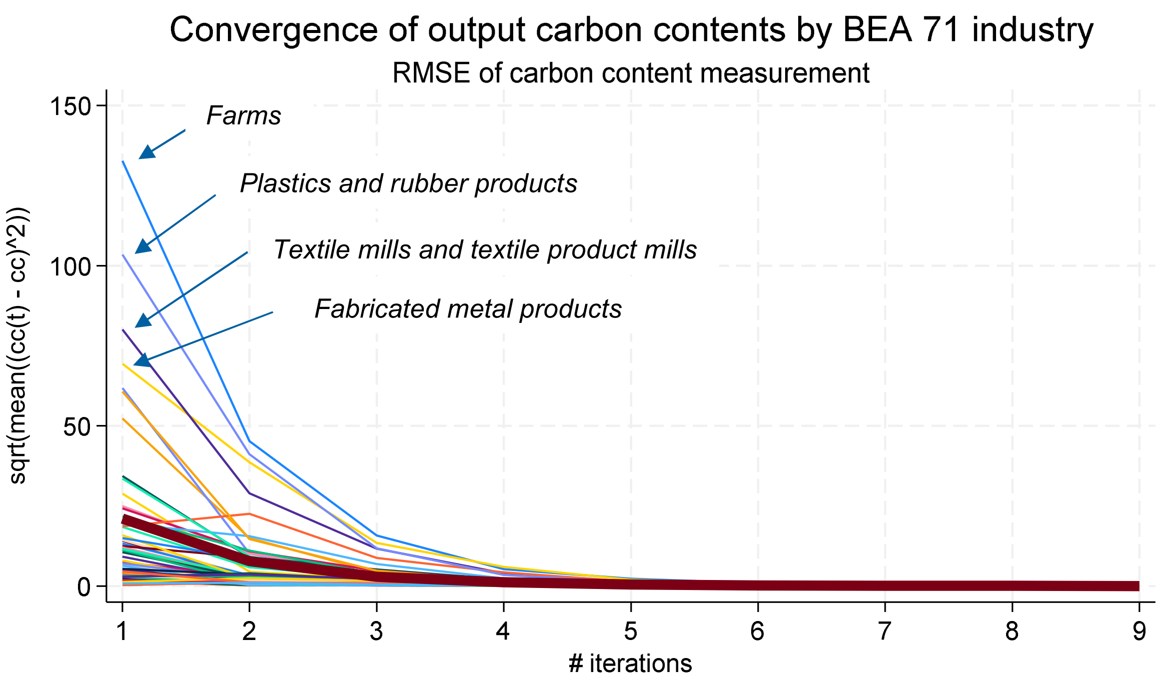
Carbon accounting by definition uses the right composition of inputs – valuations will deviate from true values if the associated input carbon contents are wide off the mark. It has been

[12] For direct emissions and the use of energy, see the standards for disclosure of GHG Scope 1 and 2 emissions: WRI and WBCSD [2004]. For Scope 3 (indirect) emissions, see the two closely related standards for enterprise-level and product-level disclosure: WRI and WBCSD [2011a and 2011b]. Further, see the Technical Guidance for Calculating Scope 3 Emissions in WRI and WBCSD [2014].

shown formally that utilizing the carbon account evaluations of companies as an *input for the next stage of evaluations* will make the estimates converge to the true values, provided that the correct structure of inputs and the correct direct emission intensities are used, see von Kalckreuth (2022a). To investigate this process and to show the rather heterogeneous outcomes from taking statistical aggregates as initial value of carbon accounting, I simulate the social learning process that consists in using carbon accounting methods iteratively.

The results are shown in Graph 3, for the case of the advanced carbon accounting, using a composite indicator for evaluating inputs. Again, the details for the first iteration of advanced carbon accounting can be found in Appendix 2. Speed of convergence is high in most cases. For some industries, RMSEs are unacceptably high at the beginning and quite some way into the future, others are well aligned from the beginning. It appears that aggregate carbon contents can well be used as initial values for some sectors, but not for others. Obviously, using more precise (direct) information for industries with large heterogeneity will improve measurement for all industries, not just the industries affected.

Graph 3



Summarising these preliminary results, it appears that using industry averages of carbon contents directly as proxies for micro level outcomes is not warranted, except in industries with little heterogeneity, such as service industries with a strong focus on office work. Refining sector distinctions will not change this result in an overall sense, though further evaluation work may show that it will be helpful for certain industries. This is a very interesting outcome, given attempts by regulators to identify certain types of activities as sustainable or non-sustainable. On the other hand, industry averages are useful building blocks in micro level computations, to make up for missing information on the value chain. In this respect, much is gained if unit level information on the direct emissions of input providers can be used, and statistical information is needed only to fill information gaps on indirect emissions.

7 References

DUCHARME, Louis Marc. Setting the Scene: the need for a new Data Gaps Initiative. Speech at Side Event of 53rd Session of the UN Statistical Commission „Closing Climate Change Data Gaps: A New G20 Data Gaps Initiative“, February 23, 2022

FOLK, Hugh and Bruce HANNON. An Energy, Pollution and Employment Policy Model. In: Michael Macrakis (ed): Energy: Demand Conservation and Institutional Problems. Cambridge, MA: MIT Press, 1974, 159-173.

JUST, James. Impacts of New Energy Technology Using Generalised Input-Output-Analysis. In: Michael Macrakis (ed): *Energy: Demand Conservation and Institutional Problems*. Cambridge, MA: MIT Press, 1974, 113-128.

VON KALCKREUTH, Ulf, Pulling ourselves up by our bootstraps: the greenhouse gas value of products, enterprises and industries. Deutsche Bundesbank Discussion paper 23/2022, July 2022a.

VON KALCKREUTH, Ulf, Product level greenhouse gas contents – how to get there? SUERF policy note, issue no 288, published September 2022b.

KAPLAN, Robert S. and Karthik RAMANNA. How to Fix ESG Reporting. Harvard Business School Working Paper 22-005 and BSG (Blavatnik School of Government at the University of Oxford) Working Paper 2021-043, July 2021. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3900146

KAPLAN, Robert S. and Karthik RAMANNA. Accounting for Climate Change. *Harvard Business Review* 99(6), 2021b: 120-131.

LEONTIEF, Wassily. *Input-Output Economics*. 2nd ed. New York: Oxford University Press, 1986.

LEONTIEF, Wassily. Environmental Repercussions and the Economic Structure: An Input-Output Approach. *Review of Economics and Statistics* 52 (1970): 262-271.

MILLER, Ronald E. and Peter D. BLAIR. *Input-Output Analysis. Foundations and Extensions.* 3rd ed. Cambridge, etc.: Cambridge University Press, 2022.

QUAYUM, Abdul. Inclusion of Environmental Goods in National Income Accounting. *Economic Systems Research* 6 (1994): 159-169.

SUH, Sangwon (Ed.). Handbook of Input-Output Economics in Industrial Ecology. Dordrecht etc., Springer, 2010.

WRI and WBCSD [World Resources Institute and World Business Council for Sustainable Development]. *The Greenhouse Gas Protocol. A Corporate Accounting and Reporting Standard*. Revised Edition, WRI, 2004.

WRI and WBCSD [World Resources Institute and World Business Council for Sustainable Development]. *Corporate Value Chain (Scope 3) Reporting Standard.* Supplement to the GHG Protocol Corporate Accounting and Reporting Standard, WRI, 2011a

WRI and WBCSD [World Resources Institute and World Business Council for Sustainable Development]. *Product Life Cycle Accounting and Reporting Standard: A standardized methodology to quantify and report GHG emissions associated with individual products throughout their life cycle.* WRI, 2011b.
WRI and WBCSD [World Resources Institute and World Business Council for Sustainable Development]. *Technical Guidance for Calculating Scope 3 Emissions.* Supplement to the Corporate Value Chain (Scope 3) Accounting and Reporting Standard. WRI, 201

Appendix 1: Descriptive statistics, by BEA 71 industry[1)]

Unweighted statistics: Revenue, direct and indirect emissions, carbon content as CO_2 equivalents

	Companies	Rev (k US)	Dir em (g/USD)		Indir em (g/US$)		Carbon content (g/US$)			
BEA 71 industry	**#**	**Mean**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**
Farms	22	368.0	1085.6	775.5	810.5	432.3	1896.1	799.8	841.5	3550.5
Forestry, fishing, and related activities	7	756.0	78.2	52.9	107.2	46.1	185.4	24.4	152.0	212.9
Oil and gas extraction	92	1643.3	965.7	1191.5	453.2	152.2	1418.9	1240.8	408.7	9980.5
Mining, except oil and gas	88	1205.5	506.9	985.2	434.4	156.4	941.2	1081.2	278.7	7553.7
Support activities for mining	39	1526.9	111.5	246.9	133.3	13.1	244.8	246.2	133.1	1663.1
Utilities	87	5192.8	1748.5	2167.8	246.7	212.2	1995.2	2188.3	153.2	10609.4
Construction	94	3564.0	79.0	336.6	205.0	64.9	284.0	351.3	125.5	2609.2
Wood products	17	2262.9	168.6	264.2	280.4	86.5	449.0	249.7	238.6	1364.8
Nonmetallic mineral products	29	1282.5	361.1	594.3	367.0	146.5	728.0	678.8	239.6	3835.0
Primary metals	39	3595.4	748.4	1159.0	754.6	255.2	1503.0	1235.1	451.5	6109.0
Fabricated metal products	65	1976.5	39.0	37.3	687.4	481.1	726.4	484.3	302.4	2379.6
Machinery	108	3743.7	16.0	10.7	330.1	104.2	346.1	104.7	163.7	748.6
Computer and electronic products	210	4698.1	27.4	39.1	58.5	33.7	85.9	62.2	17.9	595.1
Electrical equipment, appliances, and components	47	1882.5	61.8	219.3	343.0	141.5	404.8	240.3	170.2	1697.2
Motor vehicles, bodies and trailers, and parts	234	4644.8	13.4	8.7	536.9	242.1	550.3	242.6	285.7	1833.5
Other transportation equipment	37	8374.2	20.8	9.1	220.4	125.2	241.3	124.2	122.2	540.0
Furniture and related products	26	1790.4	15.8	4.6	393.6	190.2	409.3	189.5	217.7	1020.0
Miscellaneous manufacturing	95	2424.5	18.5	29.0	196.9	71.4	215.4	86.6	127.9	809.7
Food and beverage and tobacco products	80	6675.3	58.1	72.8	646.2	286.7	704.3	311.4	165.4	1909.3
Textile mills and textile product mills	13	1271.3	138.9	54.5	426.0	165.6	564.9	177.5	344.9	1019.1
Apparel and leather and allied products	13	2806.7	34.6	30.0	262.4	40.6	297.0	55.6	239.8	426.4
Paper products	17	5630.0	699.3	1511.5	479.7	76.7	1179.0	1499.0	460.2	6964.6
Printing and related support activities	3	.	.	.	.	.	.	.	.	.
Petroleum and coal products	24	23816.0	576.2	705.6	679.0	149.3	1255.2	748.8	654.9	3826.1
Chemical products	213	4098.9	246.3	1575.8	253.3	216.4	499.6	1677.3	42.4	23598.3
Plastics and rubber products	26	2471.0	61.5	61.4	497.1	137.6	558.6	143.1	334.5	875.7
Wholesale trade	176	14368.4	40.3	98.3	89.7	25.0	130.0	117.4	31.4	1153.9
Motor vehicle and parts dealers	20	7946.2	83.8	20.7	178.6	22.8	262.4	42.5	120.8	287.0
Food and beverage stores	20	24836.9	11.4	4.6	108.6	19.2	120.0	21.3	76.1	153.3
General merchandise stores	11	32008.9	9.4	4.1	121.2	16.0	130.5	18.4	99.0	158.9
Other retail	139	15260.1	65.0	88.0	236.3	60.2	301.3	115.1	71.3	1234.3
Air transportation	21	4056.2	1242.9	604.2	128.8	33.1	1371.7	608.9	769.8	3763.7

Rail transportation	5	11191.7	447.0	71.1	196.2	29.3	643.1	54.4	579.9	723.1
Water transportation	13	1315.9	1508.4	831.6	148.2	27.7	1656.6	829.7	513.7	3237.4
Truck transportation	20	2492.2	162.3	93.4	156.8	19.5	319.0	95.1	248.6	663.4
Transit and ground passenger transportation	2	.	.	.	.	.	.	.	.	.
Pipeline transportation	16	6036.4	860.2	670.2	149.3	113.7	1009.5	671.6	181.8	2508.2
Other transportation and support activities	17	12887.8	63.8	77.4	142.0	17.3	205.8	76.0	104.5	366.4
Warehousing and storage	3	.	.	.	.	.	.	.	.	.
Publishing industries, except internet (includes software)	162	2105.3	4.0	1.5	26.1	7.2	30.1	7.1	18.4	63.9
Motion picture and sound recording industries	15	974.9	3.1	2.5	74.8	13.3	77.9	15.1	40.7	111.2
Broadcasting and telecommunications	118	14439.4	5.2	3.6	78.9	26.4	84.1	27.9	37.2	188.0
Data processing, internet publishing, and other information services	114	4313.9	3.3	1.3	79.3	31.0	82.6	31.6	40.9	402.9
Federal Reserve banks, credit intermediation, and related activities	404	1720.1	0.9	0.5	31.3	5.7	32.2	5.9	24.4	73.6
Securities, commodity contracts, and investments	178	6572.7	23.3	213.1	53.6	13.0	76.9	215.4	35.4	2097.9
Insurance carriers and related activities	186	18078.6	3.4	25.3	34.9	7.3	38.4	28.5	23.1	307.4
Funds, trusts, and other financial vehicles	37	1866.8	0.4	0.1	71.3	75.5	71.7	75.5	53.1	517.7
Housing	414	990.8	20.6	35.2	79.8	69.1	100.4	79.5	5.2	822.5
Other real estate	207	990.8	20.6	35.2	180.1	55.5	200.8	67.3	106.0	722.8
Rental and leasing services and lessors of intangible assets	139	728.9	22.0	93.9	74.7	51.4	96.7	116.0	32.2	992.0
Legal services	1	.	.	.	.	.	.	.	.	.
Computer systems design and related services	78	3393.5	8.1	4.1	34.2	18.2	42.3	19.2	15.1	114.3
Miscellaneous professional, scientific, and technical services	335	1058.6	15.2	55.7	82.5	16.8	97.7	62.6	41.9	1181.0
Management of companies and enterprises	1	.	.	.	.	.	.	.	.	.
Administrative and support services	46	2666.7	10.8	26.0	87.2	48.8	97.9	68.7	34.4	461.4
Waste management and remediation services	9	4071.8	542.3	484.1	200.7	40.3	743.1	490.2	291.8	1455.7
Educational services	21	1137.8	17.9	5.0	127.4	42.4	145.3	45.6	71.2	217.6
Ambulatory health care services	124	7999.4	18.0	4.2	88.4	27.4	106.5	27.7	63.2	195.9
Hospitals	13	8228.2	16.6	1.7	99.6	3.8	116.2	4.1	108.5	123.3
Nursing and residential care facilities	18	1390.7	17.5	0.6	127.7	10.0	145.3	9.7	128.3	159.7
Social assistance	3	1515.1	17.7	0.0	145.7	55.6	163.4	55.6	128.4	227.6
Performing arts, spectator sports, museums, and related activities	8	1079.6	9.2	0.3	74.9	8.1	84.1	8.2	68.6	91.6
Amusements, gambling, and recreation industries	22	395.3	9.4	1.7	131.3	53.3	140.8	53.4	67.5	239.7
Accommodation	28	1327.8	48.8	84.0	202.7	195.1	251.5	277.2	112.0	1582.2
Food services and drinking places	99	2432.9	22.5	8.1	163.6	23.4	186.0	26.8	127.3	290.0
Other services, except government	18	1738.9	20.7	3.7	95.4	29.2	116.0	30.7	68.2	167.4
State and local government enterprises	2	.	.	.	.	.	.	.	.	.
Total	**4988**	**4782.3**	**119.4**	**598.8**	**180.5**	**214.4**	**299.9**	**679.3**	**5.2**	**23598.3**

1) For wholesale and retail trade sectors, direct and scope 2 emission intensities have been rescaled to match National Accounts definitions which are relative to output.

Tabulation only for cells with 5 or more observations

Appendix 2: Carbon contents (CO_2 equiv) and RMSEs for three estimators, by BEA 71 industry[1)]

Weighted statistics: RMSEs of BEA 71 averages, naïve carbon accounting and advanced carbon accounting

BEA 71 industry	#	Mean carbon content (g/US$) Weighted	RMSE BEA 71 averages Weighted	RMSE composite carbon accounting indicator Weighted[2)]	RMSE advanced carbon accounting indicator Weighted[3)]
Farms	22	1636.2	562.0	383.1	132.8
Forestry, fishing, and related activities	7	191.5	20.2	33.9	4.5
Oil and gas extraction	92	1281.0	587.7	104.7	11.2
Mining, except oil and gas	88	910.2	922.3	186.3	28.9
Support activities for mining	39	244.5	180.2	17.4	9.1
Utilities	87	2472.6	2051.7	171.5	11.9
Construction	94	204.1	85.0	62.0	5.0
Wood products	17	497.3	310.9	78.0	14.2
Nonmetallic mineral products	29	972.2	676.4	140.1	18.4
Primary metals	39	1639.2	1332.4	216.8	52.4
Fabricated metal products	65	719.5	356.7	343.9	61.8
Machinery	108	338.9	85.3	83.2	14.4
Computer and electronic products	210	61.6	71.9	39.4	1.6
Electrical equipment, appliances, and components	47	393.5	189.7	134.3	25.0
Motor vehicles, bodies and trailers, and parts	234	518.3	230.5	229.8	34.3
Other transportation equipment	37	173.3	84.6	83.4	13.2
Furniture and related products	26	372.7	163.3	165.1	24.3
Miscellaneous manufacturing	95	195.7	56.1	55.5	21.7
Food and beverage and tobacco products	80	731.2	366.4	340.8	69.5
Textile mills and textile product mills	13	626.6	142.7	118.4	80.1
Apparel and leather and allied products	13	276.9	36.3	25.7	18.7
Paper products	17	832.1	566.0	65.5	19.5
Printing and related support activities	3	.	.	.	.
Petroleum and coal products	24	1191.2	321.7	122.5	33.6
Chemical products	213	455.0	623.8	223.3	60.8
Plastics and rubber products	26	591.8	131.3	111.7	103.5
Wholesale trade	176	120.4	105.8	23.2	3.9
Motor vehicle and parts dealers	20	255.0	48.4	25.1	2.7
Food and beverage stores	20	108.1	27.8	22.4	12.1
General merchandise stores	11	120.3	8.0	7.5	10.6
Other retail	139	261.6	79.6	52.1	3.6
Air transportation	21	1299.5	340.2	44.2	7.6

Rail transportation	5	637.1	40.4	25.2	11.3
Water transportation	13	1264.9	516.6	27.1	8.0
Truck transportation	20	340.7	111.7	23.0	5.1
Transit and ground passenger transportation	2	.	.	.	.
Pipeline transportation	16	1022.5	534.9	104.8	2.0
Other transportation and support activities	17	295.4	84.5	12.8	4.2
Warehousing and storage	3	.	.	.	.
Publishing industries, except internet (includes software)	162	38.7	10.1	10.7	1.4
Motion picture and sound recording industries	15	65.7	20.2	19.1	1.6
Broadcasting and telecommunications	118	102.0	26.1	25.5	7.0
Data processing, internet publishing, and other information services	114	93.0	37.2	36.9	2.1
Federal Reserve banks, credit intermediation, and related activities	404	35.8	4.7	4.4	1.2
Securities, commodity contracts, and investments	178	95.8	290.7	11.0	1.6
Insurance carriers and related activities	186	72.5	94.8	10.0	3.0
Funds, trusts, and other financial vehicles	37	63.0	43.6	43.6	3.7
Housing	414	87.7	97.9	94.3	2.9
Other real estate	207	186.6	80.1	75.6	15.8
Rental and leasing services and lessors of intangible assets	139	178.8	257.0	75.9	1.8
Legal services	1	.	.	.	.
Computer systems design and related services	78	59.1	27.3	28.2	6.4
Miscellaneous professional, scientific, and technical services	335	82.0	38.0	22.6	4.4
Management of companies and enterprises	1	.	.	.	.
Administrative and support services	46	86.1	54.6	43.7	4.2
Waste management and remediation services	9	1182.1	363.3	34.8	7.1
Educational services	21	163.8	46.5	43.5	3.3
Ambulatory health care services	124	85.2	24.7	23.6	5.4
Hospitals	13	116.3	2.5	2.5	1.7
Nursing and residential care facilities	18	146.5	9.0	9.4	2.9
Social assistance	3	.	.	.	.
Performing arts, spectator sports, museums, and related activities	8	80.8	7.3	7.3	1.9
Amusements, gambling, and recreation industries	22	142.9	55.9	55.8	3.2
Accommodation	28	306.7	330.4	232.2	1.3
Food services and drinking places	99	173.6	39.3	31.8	3.0
Other services, except government	18	94.1	31.8	27.7	4.6
State and local government enterprises	2	.	.	.	.
Total[4)]	**4988**	**281.9**	**363.9**	**72.9**	**21.1**

Notes:

1) RMSEs computed for cells with at least 3 observations, tabulation only for cells with at least 5 observations
2) Direct emissions of suppliers as observed, indirect emissions as averages over BEA 71 industries
3) Direct emissions of producers and first tier suppliers as observed, second tier indirect emissions estimated using industry averages
4) RMSE for the total sample, not weighted means of sectoral RMSEs