Harnessing the Power of IO for Sustainability

A simulation study based on US data

Ulf von Kalckreuth, Deutsche Bundesbank, DG Statistics

Conference of European Statistics Stakeholders (CESS) 2024, Paris, 16.10.2024

Introduction

- Presentation can be regarded as a complement an extension to session P1b yesterday
- Carbon content measurement: national level, sectoral level, company level, product level
- Indirect emissions (Scope 3 emissions cradle to gate): attributed to production inputs, and their inputs, and…
- Statistical data are averages. How well can averages be used as proxies for company or product level data?
- How should statistical data production develop?

Introduction

- Ideally, we should have company- or product level data containing "true" data on direct and indirect emissions.

- Then we could look and see how aggregate statistics need to be enhanced and developed to yield better proxies

- We do not have that kind of micro data. **But we can simulate it!**

- Von Kalckreuth (2022): simulation for Germany solely on the basis of IO information. Meunier and Charpentier (2024, yesterday): simulation based on IO data for France

- Here: **real world direct emission and Scope 2 intensities for US companies** combined with **simulations of the value chains between those companies**

Direct and indirect emissions and total carbon content

Consider the *bill of material* (BoM) of product *k*, with $a_{k,i}$ being the quantity of good *i* embodied in the production process:

$$\mathbf{a}_k = (a_{k1} \quad a_{k2} \quad \dots \quad a_{kK})'$$

Let d_k be the amount of GHG directly emitted and c_i be the carbon content of input *i*

Then the carbon content of *k* is given as the **sum of direct and indirect emissions**:

direct emissions

indirect emissions

valuation structure of inputs

Carbon content vector

quantity structure of inputs

$$c_k = d_k + \mathbf{c}'\mathbf{a}_k = d_k + \sum_i c_i a_{ki} \qquad (1)$$

If the c_i are known, we can calculate the carbon content of product k directly.

Indirect emissions and total carbon content

If the g_i are unknown, the equation is **recursive**. Equation (1) is an **IO model for production**. We can solve for the GHG value of all products simultaneously. Let

$$\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_K)$$

be the matrix of the BoMs for all produced goods. With **d** the vector of direct emissions for products 1,…, K, we may write:

$$\mathbf{c}' = \mathbf{d}' + \mathbf{c}'\mathbf{A}$$

and solving for **c** yields

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

Carbon contents of all goods

Direct emissions for all goods

Leontief inverse, reflecting production interlinkages

The questions

- Data on sectoral interlinkages exist, from national and international Input-Output tables.
- Can be used to compute proxies for the firm level and the product level
- With finer (and more relevant) sectoral distinctions, the carbon content measurement may get more exact

Focus of this presentation

(1) How useful are aggregates? What is their value in carbon accounting?

(2) How should IO evolve to be of good use for carbon content measurement?

The set-up

This project simulates micro level emissions on the basis of "true" micro level information on (a) direct emissions and (b) electricity use, combined with model based outcomes for indirect emissions, from production interactions

- Macro level database: BEA Input Output data: 405 industries for 2012 (to be replaced by 2017 data), and 71 sectors for 2020
- Micro level database: Trucost company-level data on US economy for 2020
- Aiming at a simplified image of the overall US economy
- Direct emissions and energy consumption **are "real"**
- With **405 sectors**, the BEA Input-Output Tables are **far more detailed** than any conceivable international IO data base.

A laboratory for assessing a large range of measurement questions

Simulating production interactions in the US economy

- Extrapolated detailed level IO matrix for 2020
- Correspondence micro industries -- BEA (both rely on NAICS) -- **hard work** ☺
- 4988 micro level units for 2020 (from 3818 different companies), 97.15% from USA or Canada
- 389 BEA-industries on the "detailed" level and 67 industries on the "summary" level
- Missing: government, priv. households, rel. org. and indep. artists / writers / performers
- Micro level IO table, drawing counterparts from the respective input sectors at random for each unit
- Enhanced by existing data on micro level energy use
- Micro level Leontief matrix combined with micro level data on direct emissions

- Simulation: **carbon content of output for all units**

Some descriptives (1)

Table 1: Regional composition of simulation data

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**	**100.00**	

Some descriptives (2)

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

Chemical industry: heterogeneity on the industry level…

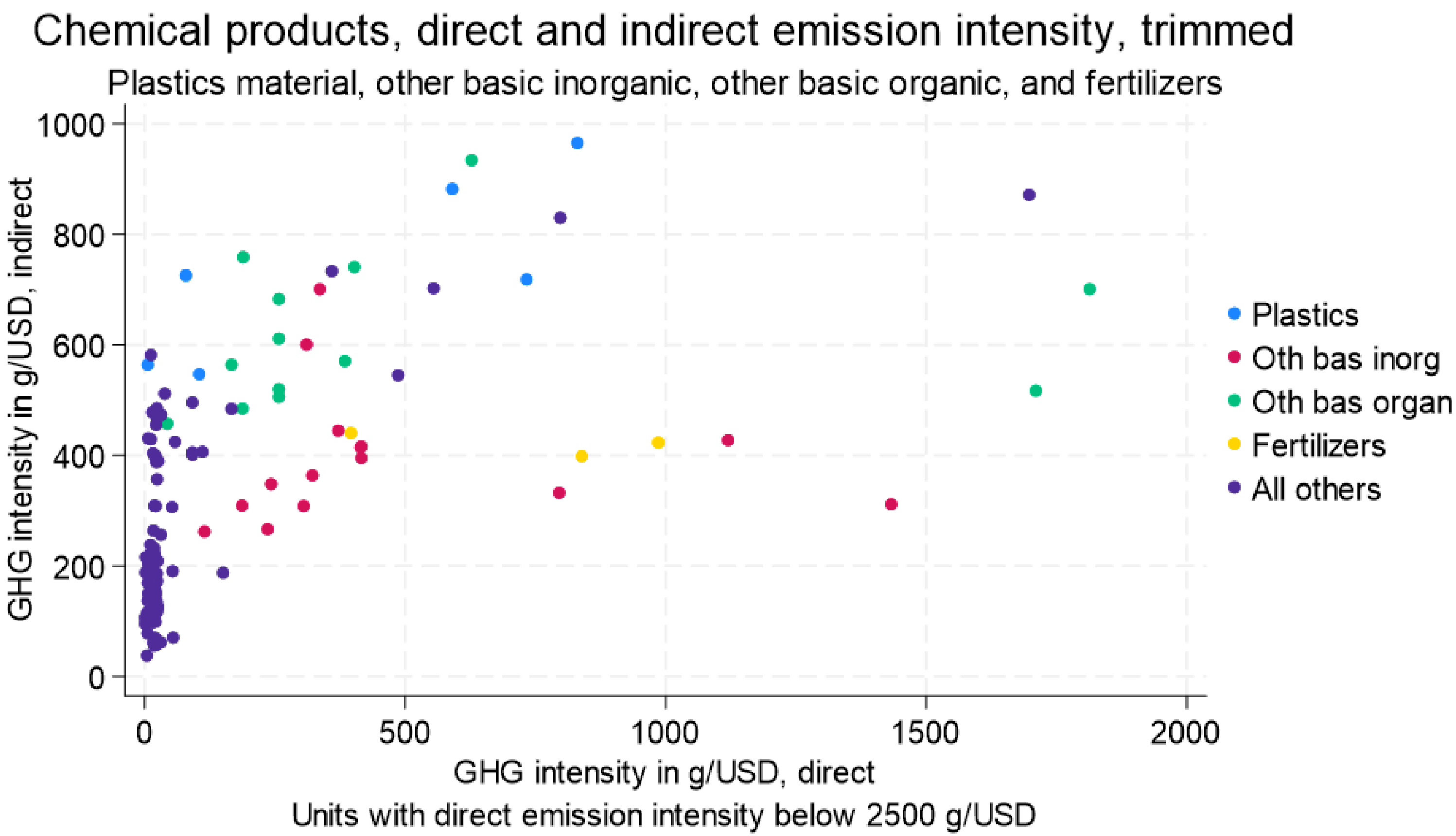
Chemical products

BEA 405 industries

BEA 405 industries	Direct emissions	Carbon content
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 etc.	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 products	33.6	422.5
Total	**168.2**	**455.0**

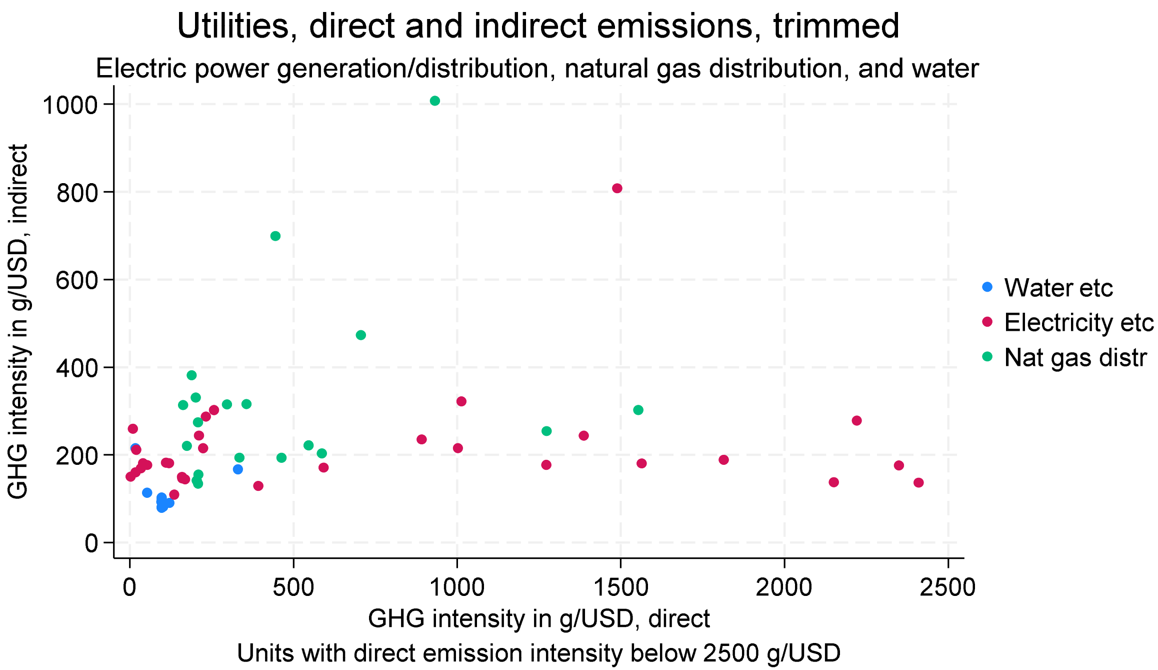
$\underbrace{\text{rows above}}_{\text{BEA 71 industry "Chemical products}}$

… and on the company level



Utilites: Heterogeneity on the industry and on the company level

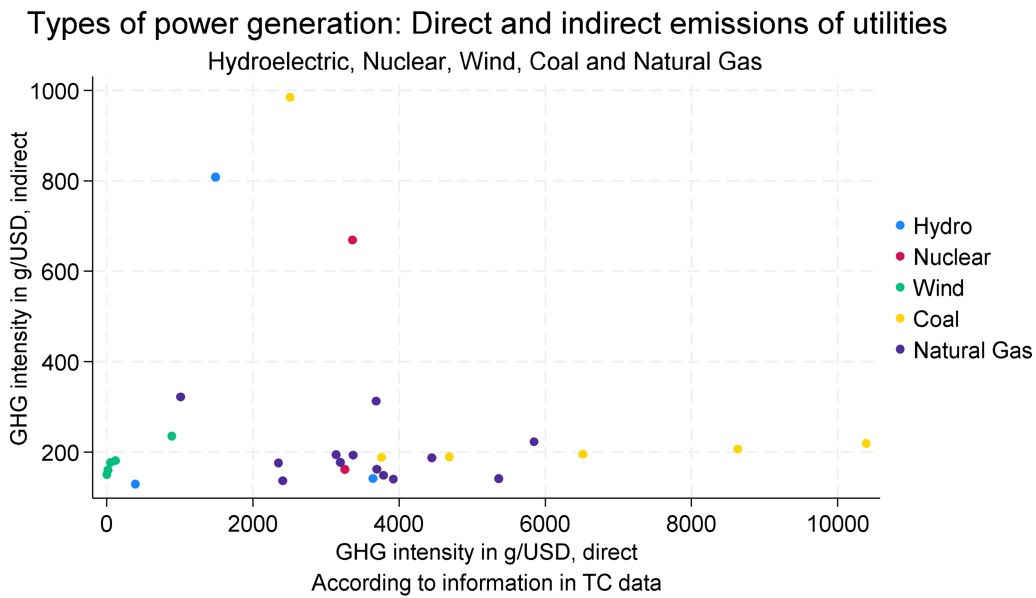
Utilities	Direct emissions	Carbon content
BEA 405 industries		
Electric power generation, transmission, and distribution	2,517.8	2,745.7
Natural gas distribution	809.5	1,231.5
Water, sewage and other systems	99.3	265.2
Total	**2,216.4**	**2,472.6**



Refining industries: the case of electricity production

BEA 405 and BEA 71 do not distinguish between different modes of electricity production. The Trucost data do. Direct intensities vary dramatically. Simulation yields:

Indirect emissions



Direct emissions

Simulation uses identical requirement coefficient for Scope 3 inputs. Provider companies drawn at random

Refining industries: the case of electricity production

Differences will feed back into IO generated carbon intensity estimates. To become more informative, we need to **distinguish between modes of electricity production!**

Similar case: types of agricultural production! Visible in BEA 405, but not in BEA 71

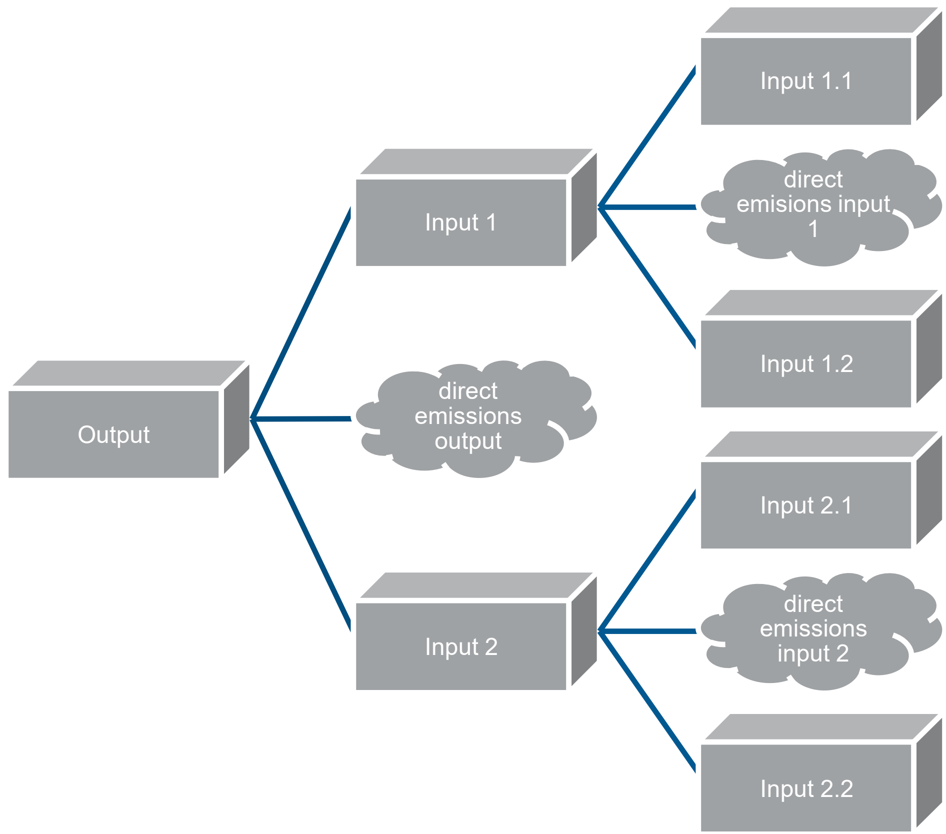
How well do averages as predictors? – Four predictors

With direct emissions, Scope 2 emissions and simulated indisrect emissions at hand, we can look at the value of using averages as predictors.

Four predictors:

1. **BEA 71 averages** of total carbon content (direct and indirect)
2. **BEA 405 averages** of total carbon content (direct and indirect)
3. **"Naïve" carbon accounting**: Direct emissions of producers known. Indirect emissions estimated using BEA 71 averages of total carbon contents
4. **"Advanced" carbon accounting**: Direct emissions both of producers and first tier suppliers known. Indirect emissions of first tier suppliers estimated using BEA 71 averages of total carbon contents.

How well do averages as predictors? – Levels of information



BEA 71 and 405 averages

"Naïve" carbon accounting

Sophisticated carbon accounting

How well do averages as predictors? Results

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

Overall useless! Potential use for homogeneus industries

Better, partly because the heterogeneity of direct emissions is "assumed away"

Zero by definition

Key messages

- Strong heterogeneity of direct emissions: industry averages generally not reliable as predictors on the company level
- Averages of relatively homogeneous industries are informative – eg white collar services
  - Direct emissions from heating and transportation, indirect emissions mostly from electricity
- More granular industry structure will not resolve this issue in general
  - Of help if and as far as homogeneous classes will result
  - Distinguish different **types of agriculture**!
  - Also: Different **modes of energy production** needed!
- In **carbon accounting**, industry averages can be used – specifically for homogeneous industries and supported with analytical data -- as predictors in cases where no direct information is available.
- My bottom line: Give a **full account of your inputs** and use averages or other proxies where there is no direct information -- well knowing how horrible they can be. This can and will be improved upon!

Convergence: Just do it!

Von Kalckreuth (2022) formally shows 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 input structure and direct emission intensities** are used.

However bad the starting values are: **using exact information on direct emissions and the input structure will drive out those bad starting values after a few iterations**.

Important: the **same values must be used throughout the system!** There is a need for regulation and joint data bases. Simulated speed of convergence is reasonably fast and does not depend on initial values.

I simulate the process for the US using the advanced carbon accounting indicators as starting values.

Convergence: Just do it!

Switching to matrix algebra – considering all products

CC estimates

Vector of inital values

$$\mathbf{c}'\big|_1 = \mathbf{d}' + \tilde{\mathbf{c}}'\mathbf{A}$$ 1st round

BoM matrix

$$\mathbf{c}'\big|_2 = \mathbf{d}' + \mathbf{c}'\big|_1 \mathbf{A} = \mathbf{d}' + \left(\mathbf{d}' + \tilde{\mathbf{c}}'\mathbf{A}\right)\mathbf{A} = \mathbf{d}'\left(\mathbf{I} + \mathbf{A}\right) + \tilde{\mathbf{c}}'\mathbf{A}^2$$ 2nd round

⋮

$$\mathbf{c}'\big|_n = \mathbf{d}'\Big(\underbrace{\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \ldots + \mathbf{A}^{\mathbf{n}}}\Big) + \tilde{\mathbf{c}}'\mathbf{A}^{\mathbf{n}+1}$$ n‘th round

$\rightarrow (\mathbf{I} - \mathbf{A})^{-1}$, power expansion of the matrix inverse

Converges to $\mathbf{c} = \mathbf{d}'(\mathbf{I} - \mathbf{A})^{-1}$ if all eigenvalues of **A** smaller 1, otherwise inverse not defined

Simulating the adjustment process