

IEooc_Methods5_Exercise4b: Uncertainty Analysis of the income-specific footprints

Goal: Learn how to use Monte-Carlo simulations to estimate the uncertainty of income specific footprints for Germany. Learn how to visualize uncertainties.

Prerequisites

- Completed “IEooc_Methods5_Exercise4a: Calculating income specific footprints for Germany”
- Basic knowledge on error propagation and sensitivity analysis (e.g. check out the IEooc video lecture on data uncertainty and sensitivity of results in MFA system models.
<https://youtu.be/VpK2NgY5FIQ>)
- Basic knowledge on Python (see here for help
https://simschul.github.io/python_basics/python_cheatsheet.html)

Different levels

- Expert level: Only read the instructions provided in this file and solve the exercise by yourself
- Intermediate level: Open the file “IEooc_Methods5_Exercise4b_calculate_income_specific_footprints_TODO.ipynb” and complete all code lines marked by #TODO
- Basic level: Open the file “IEooc_Methods5_Exercise4b_calculate_income_specific_footprints_SOLUTION.ipynb” and go through it step-by-step

Background

In Exercise IEooc_Methods5_Exercise4a we calculated income-specific footprints for German households. You might still remember that we used *average* consumption data for 10 income groups. In this exercise we will use not only average data, but also their standard deviation. The standard deviation (SD) is a measure of the amount of variation or uncertainty of a set of values. High SDs mean we have a high uncertainty of this value, low SDs mean we are rather certain about a value.

In this exercise, we are now using both, the average and the standard deviation to estimate the uncertainty of the income-specific footprints.

<http://www.teaching.industrialecology.uni-freiburg.de/>

Exercise

Step 1: Load data on final demand by income group

Load both data sets “IEooc_Methods5_Exercise4ab_Final_demand_by_income_avg.xlsx” (containing the average values) and “IEooc_Methods5_Exercise4ab_Final_demand_by_income_std.xlsx” (containing the standard deviations). Make sure that the indices (row names) are set correctly.

Step 2: Generate samples

Now, we have two pieces of information for each data point: the average (or mean), and the standard deviation. We use this information to draw a sufficiently large number of random samples for each data point. Thus, we create a sufficiently large number of random variants of the income-specific final demand matrices.

For doing so, we define a function that takes two matrices (average and SD) and the number of samples N as inputs and delivers a list of length N containing random variants of the income-specific final demand matrix.

Since, there is no negative consumption we cannot use the `normal` distribution to draw the samples (because we would get negative values in our sample). Instead we use the `lognormal` distribution: The lognormal distribution is implemented in numpy as `np.random.lognormal`. Check out the help page to learn more:

<https://numpy.org/doc/stable/reference/random/generated/numpy.random.Generator.lognormal.html#numpy.random.Generator.lognormal>

We apply this function to generate a list of length N (start with $N = 100$) of which each element is a slightly different final demand matrix.

Step 3: Plot the distributions of individual data values

Look at some individual data points to check if everything has worked like intended. For example, take the consumption of “Wheat” from “AT” by “Group00_Average_HH” (or any other non-zero entry). Plot a histogram along with the specified mean value and the standard deviation around the mean. Repeat for a few other data entries.

Step 4: Import EXIOBASE and calculate missing elements

As we already did in IEooc_Methods5_Exercise4a we load (parse) the 2013 version of EXIOBASE and calculate all missing elements using “`exio.calc_system()`”.

Methods Part 5 (Input-output analysis)

<http://www.teaching.industrialecology.uni-freiburg.de/>

Step 5: Calculate total footprints by income

We take the sample of final demand matrices which is stored in a list (see step 2), We iterate through each element of that list and calculate footprints using the default S and L matrices from EXIOBASE. From the S matrix we only use the GWP100 GHG emissions '['GHG emissions (GWP100) | Problem oriented approach: baseline (CML, 2001) | GWP100 (IPCC, 2007)']'

Step 6: Process, analyse and plot results

1. Plot the footprints by income group using a bar plot with an error bar covering the 95% uncertainty interval.
2. Plot the distributions of the footprints by income group using a boxplot or violin plot.
3. What is the GHG-footprint of the lowest and highest income group, respectively? Provide the average (mean) number and the uncertainty range.
4. Which income group has the footprint with the highest relative uncertainty?

Acknowledgment:

This exercise was compiled by Simon Schulte, using Material from the Uni Freiburg Supply Chain Management Class co-developed with Dr. Johan Vézé