**Econometric long-term forecasting of CO2 emissions**

What is the goal of the task?

The goal is to forecast reliable trajectories of consumption-based emissions per country in order to include them as baseline trajectories in the NICE model. The initial statement is that consumption-based emissions from other CGE models (e.g. REMIND) are not reliable enough in order to model long-term trajectories of emissions per country.

This statement leads to the idea of modelling CO2 emissions using an econometric approach. The goal is to find data that satisfies the following criteria for them to be explanatory variables of interest:

* They should be statistically significant
* The regressions they are in should have a strong forecasting power
* There should be reliable methods of modelling them in the future; e.g. standard baseline scenarios, or widely acknowledged models providing satisfactory forecasts

What is the state of the literature on this topic?

Liddle (2018) uses panel data methods to establish the effect of trade on consumption-based emissions and territorial emissions. Few variables are significant and could be useful to forecast emissions.

Caldeira (2010) details the composition of transfer emissions for the year 2004. They are mostly from intermediate goods, it does not show straight-forward variables that could easily be used as a proxy to estimate consumption-based emissions.

Mahlkow and Wanner (2023) draws stylized facts about the links between consumption-based emissions and trade imbalances.

Bhattacharya et.al (2019) examines the convergence of consumption-based and territory-based carbon emissions intensity across 70 countries. Their work allow to shape three clubs for consumption-based-emissions, for which the carbon intensity of consumption-base emissions should converge by 2030.

Serkan Aras and M. Hanifi Van (2022) uses statistical learning methods to find econometric results about the modelling of carbon emissions.

How can this literature be used ?

The statistical method of Liddle (2018) is a panel data approach using common correlated effect mean group (CCEMG) regressions to estimate the significance of parameters in their ability to estimate emissions.

This method is very relevant in this concept because the data used is cross-sectional correlated and is not stationary. Therefore a single OLS regression would provide biased output and the significance tests would not be true. This method adds the cross-sectional mean of variables in the regression in order to capture common shocks and trends. Then it computes regression coefficients for each country, means them and assesses the significance of these coefficients.

In this task we use the same approach. CCEMG regressions are made on the 1997-2010 period. The coefficients per country are used to estimate data from 2020.

Because there are only 14 time periods, and always at least 4 explanatory variables in the regression, R-squared are almost always over 0.99. If not one can already eliminate the regression. Hence the fitness of a regressions is measured with the following indicators:

* Mean absolute error of the predicted data
* Cross-dependence tests of the residuals of the regressions : residuals should be independent

Data:

The data used up to now is from the Word Development Indicators databank. Liddle (2018) used the same source. Emissions data are from the Global Carbon Atlas.

It is restricted to about 100 countries spanning from 1997 to 2020.

The parameters are: GDP per capita, share of fossil fuel shared used, share of the industry in the GDP, trade balances of goods, trade balances of services, imports and exports per country.

Depending on the regressions, these parameters can be logged.

First results:

Reproducing a regression from Liddle (2018) representing log per capita emissions based on log GDP p.c., industry share, fossil fuel share, imports and exports shares, we get a mean absolute error(MEA) of 75% when it comes to territorial emissions, and 118% when it comes to consumption-based fossil emissions.

The best regressions for now are the following:

* Territorial emissions using GDP\*fossil\_fuel\_share: 17% of MEA ten years ahead. The variables are logged and per capita
* Territorial emissions using GDP (logged and per capita) : 30% of MEA ten years ahead

Other regressions are tested and better results should be incoming.

Leads and next steps:

Improving the methodology:

* For now the cross-sectional means are computed using the means of the data in the tables, but they do not necessarily make sense from an economic point of view (e.g. the mean of trade balances should be zero by definition, but then it is defined as the algebraic mean of trade balances for the available countries)
* The coherence of data used is also to be questioned. GDP is in 2021 constant USD PPP, but trade balances are in nominal constant 2021 USD. The same goes for the industry share of GDP, and imports and exports shares.

Leads for other econometric approaches: “within” estimation could be a possibility, maybe some statistical learning approaches

Searching back in literature for *long-term forecast* of carbon emissions using econometric approaches in order to find ideas for regressions.