European Union Sectoral Emissions Data (EUSED)

— Codebook and Procedures —

Patrick Bayer*

March 29, 2022

Abstract

This codebook describes what data problem the European Union Sectoral Emissions Data (EUSED) addresses and how the data set was constructed. I specify how sectors from the UNFCCC data were matched to sectoral activities under the EU ETS and discuss (graphically and numerically) how well these matches perform. The codebook further explains the folder structure of the data archive and key variables which EUSED provides. EUSED data covers emissions from 33 countries for 7 sectors (5 covered and 2 not covered under the EU ETS) from 1990-2016. This coverage can easily be updated by modifying the freely available code to produce the data.

*School of Government & Public Policy. University of Strathclyde. McCance Building, 16 Richmond Street, Glasgow, G1 1QX, Scotland, UK. Email: patrick.bayer@strath.ac.uk

Acknowledgements: I gratefully acknowledge the financial support of the British Academy that made collecting these data possible as part of awarded funding through their BA/Leverhulme Small Research Grants (SG171349, 1 January 2018 to 30 May 2019). Constantin Brod offered excellent research assistance.

Disclaimer: EUSED is created from freely available data from other sources using freely available statistical software `R`. All code to create the data is available publicly from the author's Harvard Dataverse archive and can be modified by any user assuming the original data sources are acknowledged (see below) and EUSED data/the associated publications are cited properly. EUSED comes with no guarantees or warranty. Data was produced and last tested using `R` version 3.6.1 (2019-07-05) with RStudio version 1.2.1335 on Windows 10, 64 bit on an Intel(R) Core(TM) i7-4650U CPU @ 1.70Ghz 2.30Ghz machine with 8GB RAM.

Citation: If you use the EUSED data, please cite:

- Bayer, Patrick. 2019. "European Union Sectoral Emissions Data (EUSED)." Available at https://dataverse.harvard.edu/dataverse/eused.
- Bayer, Patrick and Michaël Aklin. 2020. "The European Union Emissions Trading System Reduced CO_2 Emissions Despite Low Prices." *Proceedings of the National Academy of the Sciences* 117(16): 8804-8812. Available open access at https://www.pnas.org/doi/10.1073/pnas.1918128117

1 Motivation

A common challenge for policy evaluation is that a rigorous empirical analysis typically requires data about the policy outcome *prior* to the introduction of the policy. These data are however often not available. The same is true for the European Union Emissions Trading System (EU ETS), which was introduced in 2005 to regulate carbon emissions in Europe. While country-level emissions data exist at least since 1990, the EU ETS covers only some sectors (accounting roughly for 50% of EU-wide carbon emissions) and sector-level emissions data is much harder to come by. This is further complicated by the fact that the EU ETS accounting of carbon emissions is by *activity*, not sector. Carbon emissions are therefore process-specific: this means that, for instance, emissions from fuel combustion and emissions for steel production would be classified as different activities even if emissions occur in the same plant and sector. EUSED addresses this problem and creates sector-level emissions data for EU ETS-regulated sectors even before emissions trading started in Europe.

2 A Matching Approach by Sector

In order to address this challenge, I created the European Union Sectoral Emissions Data (EUSED) by matching emissions from two existing data sets. This section describes the used data sources, the conceptual logic of the match, and provides a list of how emissions are matched across the two data sets at the sector level.

2.1 Data Sources

EUSED relies on two, publicly available data sets from the website of the European Environment Agency (EEA):

— Greenhouse gas (GHG) inventory data from the "National emissions reported to the UNFCCC and to the EU Greenhouse Gas Monitoring Emissions," available from the EEA here

— EU ETS inventory data from "European Union Emissions Trading System (EU ETS) data from EUTL," available from the EEA here.

In this codebook, I refer to the first data set as the 'UNFCCC data' or 'UNFCCC emissions' and to the second one as the 'EU ETS data' or 'EU ETS emissions.' EUSED uses the UNFCCC data in its version 21 (created 31/5/2018) and the EU ETS data in its version 31 (created 11/9/2018). These data are updated regularly by the EEA, so I advise using newer versions if necessary, for instance to expand temporal coverage. To ensure the provided code runs correctly, the data I used to create EUSED are part of the data archive and sit in the `./data` folder. The data can also be downloaded from the EEA websites linked above.

2.2 The Matching Logic

The basic idea of matching emissions is straightforward when realizing two features of the data. First, the UNFCCC data has larger time coverage in terms of years (1990-2016) compared to the EU ETS data (2005-2016)—the EU ETS became only operational in 2005. Second, the UNFCCC data has also larger coverage in terms of emissions as it records a country's *total* emissions (broken down by sector), while the EU ETS records only emissions regulated under the policy, mostly emissions from electricity production and pollution intensive industry.

The goal of the matching exercise is to find combinations of sectoral emissions from the UNFCCC data that approximate EU ETS sectoral emissions almost perfectly for those years in which temporal coverage of the two data sets overlaps. With a sufficiently good match, we can then use emissions from the *same* sectors in the UNFCCC data as a proxy for emissions in the EU ETS sector before the introduction of the EU ETS carbon market policy.

To fix ideas, assume a country's total emissions U come from four sectors, $u_1, ..., u_4$ and are recorded in the UNFCCC data as $U = u_1 + u_2 + u_3 + u_4$. Denote EU ETS-regulated emissions for the same country as E, where emissions come from sectors e_1 and e_2. Notably, there is no direct correspondence between sector definitions in the UNFCCC data, which are categorized according to IPCC Common Reporting Framework (CRF) definitions and EU ETS activities. This means,

there is no direct match between, say u_1 and e_1. The value of the EUSED data set is that it matches sectoral emissions from the UNFCCC data to activity emissions from the EU ETS data, based on sectoral similarity (see next section for details).

Assume that we find, for instance, that $u_1 + u_3 \approx e_1$ for the 2005-2016 years when both data sets are available, then EUSED uses the *summed* sector emissions $u_1 + u_3$ from the UNFCCC data as a proxy of what sectoral emissions of e_1 would have been before 2005 when the EU ETS was introduced and data were recorded.

2.3 Overview of Matched Sectors

The obvious challenge then is to find matches between (combinations of) sectoral emissions in the UNFCCC data with (combinations of) activity emissions in the EU ETS data. For this, I rely on the "Annual European Greenhouse Gas Inventory 1990-2016 and Inventory Report 2018" as the EU's official submission to the UNFCCC Secretariat. This report contains a mapping table (EEA, 2018, Table 1.10, p.32) which offers guidance on mapping of emissions across data sets. While this mapping is useful, it often lists the same CRF categories in the UNFCCC data for different EU ETS activities. EUSED resolves this problem by aggregating emissions up into the following five broader, mutually exclusive (to the best degree possible) sectoral categories, i.e., energy; metals; minerals; chemicals; and paper. These groups are created based on the above mentioned mapping table and similarity in outputs. This means that emissions from iron and steel production would not be assigned to emissions from paper production.

Table 1 below provides the mapping of EUSED sectors, EU ETS activities (from the EU ETS data), and UNFCCC CRF categories (from the UNFCCC data). All EU ETS activities from stationary sources (activity codes 20-44) are matched (excluding aviation), except activity code 45 for emissions from greenhouse gas capture under Directive 2009/31/EC, for which there is no clear correspondence in the UNFCCC data. Verified emissions from this sector are 0.05% of total EU ETS-regulated emissions (EEA, 2018, p.31).

TABLE 1: EUSED sector mappings for EU ETS and UNFCCC data.

EUSED sector	EU ETS activity codes	UNFCCC CRF category
Energy	20 Combustion of fuels 21 Refining of mineral oil	1.A.1 Energy industries
Metals	22 Production of coke 23 Metal ore roasting or sintering 24 Production of pig iron 25 Production or processing of ferrous metals 26 Production of primary aluminum 27 Production of secondary aluminum 28 Production or processing of non-ferrous metals	1.A.2.a Iron and steel 1.A.2.b Non-ferrous metals 2.C Metal industry
4 Minerals	29 Production of cement clinker 30 Production of lime, or calcination of dolomite/ magnesite 31 Manufacture of glass 32 Manufacture of ceramics 33 Manufacture of mineral wool 34 Production or processing of gypsum or plasterboard	1.A.2.f Non-metallic minerals 2.A Mineral industry
Paper	35 Production of pulp 36 Production of paper or cardboard	1.A.2.d Pulp, paper, and print
Chemicals	37 Production of carbon black 38 Production of nitric acid 39 Production of adipic acid 40 Production of glyoxal and glyoxylic acid 41 Production of ammonia 42 Production of bulk chemicals 43 Production of hydrogen and synthesis gas 44 Production of soda ash and sodium bicarbonate	1.A.2.c Chemicals 2.B Chemical industry

3 The EUSED Data Set

This section provides an overview of the contents of the data archive, with information about the basic folder structure, the main EUSED data, key variables, and a reference to an application of the EUSED data.

3.1 Folder Structure

The data archive ships with 8 folders and two `R` scripts in the top hierarchy. The two `R` scripts produce two versions of the EUSED data.

— `EUSED_regulated.R` produces a data set that breaks total emissions in any country-year down into emissions that are regulated under the EU ETS and those that are not.

— `EUSED_sector.R` produces a more disaggregated data set that breaks total emissions in any country-year down into emissions by sector. Summing emissions from EU ETS regulated sectors and non-regulated sectors produces the data set referenced above.

Note: Copying the entire data archive onto your system (so that the internal data structure remains intact) and running the code automatically produces all the files in the `./eused` and `./figures_xxx` folders. Re-running the `R` code is however *not* required if you simply want to use the EUSED data set. The ready-to-use versions of the EUSED data sit in the `./eused` folder.

— The `./data` folder contains the data sources for the UNFCCC data (`UNFCCC_v21.csv`) and the EU ETS data (`ETS_Database_v31.csv`). It also comes with a copy of the EU GHG Inventory report (EEA, 2018) for reference.

— All six `./figures_xxx` folders contain output plots that show emissions over time, separately by country. The five sector folders show emission time series of the matched UNFCCC and EU ETS data. The `./figures_regulated` folder contains plots that show total country emissions (for reference) together with total regulated emissions from the UNFCCC and EU ETS data. All plots provide a visualization of how closely UNFCCC and EU ETS data can be matched. These plots are produced automatically by running the respective `R` scripts mentioned above.

3.2 Main Data

Here I describe some basic features and match statistics for the two main EUSED data sets. I will refer to the `EUSED_regulated.csv` data set as the 'aggregated data,' produced by the `EUSED_regulated.R` script, whereas I refer to the `EUSED_sector.csv` data set as the 'sectoral data,' produced by the `EUSED_sector.R` script. Both data sets are also available in `.RData` format.

Both EUSED data sets provide emissions data for 33 countries (all EU ETS countries plus Switzerland and Turkey) and cover 1990-2017 years (UNFCCC data coverage: 1990-2016; EU ETS data coverage: 2005-2017).

3.2.1 Aggregated Data

For each country-year, the aggregated data breaks emissions down into total emissions, (summed) emissions from regulated sectors and (summed) emissions from unregulated sectors. To asses how well matched UNFCCC emissions map onto EU ETS emissions during the years of overlap of the two time series, I calculate a simple ratio measure, ρ to quantify the 'goodness-of-match,' which I define as:

$$\text{Ratio } \rho = \frac{\text{EU ETS emisisons}}{\text{UN emissions}}.$$

A perfect match would yield $\rho = 1$, whereas $\rho < 1$ means that UNFCCC-based emissions are smaller than actual EU ETS emissions and $\rho < 1$ means UNFCCC-based emissions are larger than actual EU ETS emissions. Across all countries and years, we obtain a good match both for unweighted ($\bar{\rho} = 0.981$) and weighted ($\bar{\rho} = 0.975$) averages, where I use country's total emissions as weights. Table 2 below shows averages separately by country. It also reports R^2 statistics for each country from mean regressions of UNFCCC emissions (dependent variable) on EU ETS emissions (independent variable) in an effort to assess similarity in trends, not only levels.

TABLE 2: 'Goodness-of-match' statistics by country, 2005-2016

ID	Country	ρ average	R^2	ID	Country	ρ average	R^2
1	Austria	0.907	0.999	17	Latvia	0.942	0.999
2	Belgium	0.938	0.997	18	Liechtenstein	1.55	0.328
3	Bulgaria	0.954	0.998	19	Lithuania	0.873	0.975
4	Croatia	1.15	1.000	20	Luxembourg	0.848	0.997
5	Cyprus	0.994	1.000	21	Malta	1.000	1.000
6	Czech Republic	0.928	1.000	22	Netherlands	0.921	0.995
7	Denmark	0.984	1.000	23	Norway	0.939	0.984
8	Estonia	0.914	0.999	24	Poland	0.958	1.000
9	Finland	0.959	1.000	25	Portugal	0.970	0.999
10	France	0.950	0.998	26	Romania	0.932	0.999
11	Germany	1.03	0.999	27	Slovakia	0.971	0.999
12	Greece	0.949	1.000	28	Slovenia	0.975	1.000
13	Hungary	0.946	0.999	29	Spain	0.967	0.998
14	Iceland	1.05	1.000	30	Sweden	0.959	0.992
15	Ireland	1.02	1.000	31	United Kingdom	0.993	0.997
16	Italy	0.962	0.999				

Note: The table shows the above defined 'goodness-of-match' statistic by country for EUSED, 2005-2016. It also reports the R^2 statistic from by-country regressions through the mean for EU ETS emissions (dependent variable) on UN emissions (independent variable).

While the 'goodness-of-match' statistic ρ is a useful measure to assess the quality of the match,

I recommend to also visually inspect the time series of the matched UNFCCC and EU ETS emissions because ρs different from 1 (also reported as means in the top-right corner of each by-country figure) can still be compatible with similar trends in the time series, yet differences in levels. In order to quantify similarity in trends, I also report (in Table 2 and all figures) R^2 statistics from a simple regression through the mean for EU ETS emissions (dependent variable) on UN emissions (independent variable).

What an appropriate quality of the match is will depend on each user's application, but carefully assessing the quality of match is important. Differences in levels (i.e., ρs much different from 1) may not matter when using fixed effects models for example, which identify estimated effects from within variation. For other applications, it may be useful to exclude particular countries that produce a fairly poor match. This is particularly true for smaller countries, for which achieving a good match is more difficult.

3.2.2 Sectoral Data

The more disaggregated data set breaks emissions down by sector. In particular, for each country-year, EUSED provides emissions for five regulated sectors (energy, metals, minerals, chemicals, and paper), as defined in Table 1 above as well as for the unregulated transport sector. Total and residual emissions (as the difference between the total and all summed sectoral emissions) are also included in EUSED. The plots to visually assess goodness-of-match by sector and country are in the `./figure_xxx` folders. Overall match quality varies by sector as Table 3 shows, with the largest sectors in terms of emissions, i.e., energy, metals, and minerals, performing well, whereas the smaller chemicals sector performs rather poorly.

TABLE 3: The table shows 'goodness-of-match' statistics by sector, 2005-2016

ID	Sector	ρ average		Mean R^2
		unweighted	weighted	
1	Energy	1.05	1.06	0.961
2	Metals	0.788	0.823	0.906
3	Minerals	0.928	0.907	0.992
4	Chemicals	0.377	0.378	0.735
5	Paper*	0.835	0.881	0.931

Note: The table shows the above defined 'goodness-of-match' statistic by sector in the EUSED data set. It also reports the mean R^2 test statistic from by-country regressions through the mean for EU ETS emissions on UN emissions.
*Estonia, Germany, and Sweden are excluded because of reporting problems in the UN paper emissions data.

The paper sector deserves particular attention as well because of potential reporting problems in the UNFCCC data. Paper emissions in Estonia and Germany are much smaller than what is reported in the EU ETS data. Moreover, Sweden is reported to have had zero emission in 2015,

which is inconsistent with the entire time series and 2014 and 2016 emissions. The tabular information in Table 3 excludes these three countries, which then results in a decent match quality. Several smaller countries (i.e., Croatia, Cyprus, Iceland, Liechtenstein, Luxembourg, and Malta) do not have paper emissions under the EU ETS.

Despite these challenges, the mean R^2 statistic from by-country regressions through the mean for EU ETS emissions (dependent variable) on UN emissions (independent variable) are generally fairly good even for sectors with ρ values different from 1. This means that even though the match performs poorly for reproducing emission *levels*, trends are still approximated fairly well. Whether the achieved quality of the match is sufficient for an analysis depends on the application. As mentioned above, when, for instance, fixed effects models are used, poorly matching levels may not be too problematic. *Users are advised to carefully consider the use of the data.*

3.3 Key variables

Both EUSED data sets have seven key variables:

— iso2: 2-digit ISO country identifier

— year: year variable

— sector: sector variable, where values can either be: Regulated; Unregulated; Total (for `EUSED_regulated.csv`) or Energy; Metals; Minerals; Chemicals; Paper; Residual; Transport; or Total (for `EUSED_sector.csv`).
Note: Values for 'Unregulated' emissions in `EUSED_regulated.csv` are *identical* to 'Residual' emissions in `EUSED_sector.csv`.

— un: emissions from the UNFCCC data base (years: 1990-2016)

— ets: emissions from the EU ETS data base (years: 2005-2017)

— country: country name

— ratio: ρ goodness-of-match variable

3.4 Application

For an application of EUSED , please see Bayer and Aklin (2019), who study the effectiveness of the EU ETS.

References

Bayer, Patrick, and Michaël Aklin. 2019. "The European Union Emissions Trading System Reduced CO_2 Emissions Despite Low Prices." Working Paper.

EEA. 2018. "Annual European Union Greenhouse Gas Inventory 1990-2016 and Inventory Report 2018." Available at https://www.eea.europa.eu/publications/european-union-greenhouse-gas-inventory-2018.