RESOURCES *for the* FUTURE

World Carbon Pricing Database: Sources and Methods

Geoffroy Dolphin and Qinrui Xiahou

Working Paper 22-5
January 2022

About the Authors

Geoffroy Dolphin is a post doctoral fellow at Resources for the Future and a PhD graduate from the University of Cambridge Judge Business School where he is an affiliate of the Energy Policy Research Group. Geoffroy's research interests span political economy, energy & environmental economics and climate policy. His research aims to inform the development of climate policy designs that are politically sustainable while being environmentally effective.

Qinrui Xiahou is a research analyst at Resources for the Future. Prior to joining RFF, she was a Student Research Fellow at the Yale Hixon Center for Urban Ecology and a Student Associate of the Yale Climate Change and Health Initiative. Qinrui's research focuses on climate change economics and policy. Her work involves employing econometrics and energy modeling methods to analyze a wide array of policies, such as the Solarize Campaign in the state of Connecticut, the low-carbon transition pathways in China, and the Clean Energy for America Act in the United States. She graduated from Yale University in May 2021 with a master's degree in Environmental Science.

Acknowledgments

This dataset is an extension of one developed while lead author Geoffroy Dolphin pursued his PhD within the Energy Policy Research Group (EPRG) at the University of Cambridge. Its existence owes much to support from EPRG, the Cambridge Judge Business School, and the UK Economic and Social Research Council. The current extension was supported by Resources for the Future.

About RFF

Resources for the Future (RFF) is an independent, nonprofit research institution in Washington, DC. Its mission is to improve environmental, energy, and natural resource decisions through impartial economic research and policy engagement. RFF is committed to being the most widely trusted source of research insights and policy solutions leading to a healthy environment and a thriving economy.

Working papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review. The views expressed here are those of the individual authors and may differ from those of other RFF experts, its officers, or its directors.

Sharing Our Work

Abstract

This note describes the sources and methods used to construct the World Carbon Pricing Database (WCPD). This database contains a harmonized record of the sectoral coverage and prices associated with carbon pricing mechanisms implemented worldwide at the national and subnational levels over the period 1990–2020. The dataset follows IPCC 2006 sectoral disaggregation, which allows for a straightforward integration with other datasets following the same structure.

Contents

1. Introduction

Over the last 30 years, the number of jurisdictions that have implemented a carbon pricing mechanism has grown significantly. Today, 43 national and 32 subnational jurisdictions have such a mechanism in at least one sector. However, a standardized and centralized record of the sectors covered and prices applied to CO_2 emissions by these mechanisms is lacking.[1]

This dataset provides an essential contribution to filling that gap. It covers mechanisms introduced since 1990 at the national and subnational levels and is the most comprehensive attempt at providing a systematic description of carbon pricing mechanisms in terms of their sectoral (and fuel) scope and the associated price signal. It should prove of interest to a wide range of parties, including academic researchers, policy analysts and interested civil society organizations.

A key feature of this dataset is that it provides information structured by territorial jurisdiction, not carbon pricing mechanism. This is achieved by mapping information available for each mechanism onto jurisdictions. This mapping accounts for the possibility that multiple mechanisms apply to the same emissions sectors and, in such instances, presents information separately for each mechanism (see details in section 4). It also covers a long period (1990–2020) and, hence, allows for (re)constructing time series of prices applied to emissions in the jurisdictions that had such prices. In addition, its disaggregation by IPCC 2006 sectors (see section 2) allows for a straightforward integration with other data sources following the same disaggregation (see section 3).

These features make the dataset a valuable tool to track the development of carbon pricing mechanisms. It also provides enough data to analyze their impact in a broad range of social, technological, and sectoral contexts. For instance, for the United States, it provides information on carbon pricing mechanisms in force in states arising from regional initiatives, such as the Regional Greenhouse Gas Initiative (RGGI), or state-level policies, such as the California cap-and-trade mechanism. Similarly, for national jurisdictions, it records the implementation of both regional initiatives, such as the EU emissions trading system (ETS), and national mechanisms targeting the same or other sectors.

This version of the dataset is an extension of a data curation effort undertaken while pursuing my PhD within the EPRG at the University of Cambridge. Its existence owes

[1] The initiative closest to what this dataset provides is the data collection effort led by the World Bank and made available through its Carbon Pricing Dashboard. However, it presents two shortcomings. First, the information available is structured around carbon pricing mechanisms (not jurisdictions). Despite a typical one-to-one mapping between a carbon pricing mechanism and a jurisdiction, this is not always the case (as with, for example, the EU ETS or the Regional Greenhouse Gas Initiative). Second, it is not disaggregated enough to allow for use in—and integration with—detailed analytical work.

much to support from EPRG, the Cambridge Judge Business School, and the UK Economic and Social Research Council. The current extension was supported by Resources for the Future.

This dataset has served as the backbone for two academic research projects.[2] It is now made available as an open-source resource with the hope that it can be useful to many more, as well as benefit from feedback and contributions from a wider community. I also hope that this standardized reporting and assessment of carbon pricing mechanisms will provide easier access to information about them and improve transparency around their implementation.

The source code, written in Python 3, raw data files, and formatted dataset files and scripts are available at https://github.com/g-dolphin/WorldCarbonPricingDatabase under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

The remainder of this technical note is organized as follows. Section 2 describes the database scope and content. Section 3 presents the sources, and Section 4 discusses the methodology. Sections 5 and 6 discuss integration of the dataset with other data and future extensions, respectively. Section 7 concludes.

[2] The initial publication associated with it is Dolphin et al. (2020). It was then updated and expanded, notably for use in Rafaty, Dolphin, and Pretis (2020).

2. Description

The database records information on the institutional design and sectoral coverage of mechanisms creating an explicit price on CO_2 emissions **at the sector-fuel level**.[3] It also records price(s) associated with each of these mechanisms. Information about both carbon taxes and emissions trading (cap-and-trade) mechanisms is recorded and presented separately.

2.1 Scope

The dataset currently covers 198 national and 94 subnational jurisdictions (50 US states, 13 Canadian provinces and territories, 31 Chinese provinces, autonomous regions and municipalities).[4] The data is disaggregated at the level of IPCC sectors following the IPCC 2006 guidelines for national greenhouse gas (GHG) emission inventories (Buendia et al., 2006). For sectors in IPCC category 1A (Fuel combustion activities), the dataset distinguishes between three aggregate fuel types: coal, oil, and natural gas.

The information in the dataset pertains to policy instruments targeting primarily CO_2 emissions.[5] In some instances, these instruments also cover other Kyoto GHG (and IPCC sectors). This coverage is accounted for in the present version. However, it does not incorporate information about carbon pricing mechanisms that primarily target non-CO_2 Kyoto gases. For instance, Spain's tax on HFCs introduced in 2014 is not included. A future iteration of the dataset will integrate such information.

A related observation is that the price information provided (expressed per ton of CO_2 equivalent) is best understood as applying to CO_2 emissions within each IPCC sector. Usually, but not always, that same price also applies to emissions of non-CO_2 Kyoto GHG.

2.2 Variables

The database is structured as follows. It has one *data* file and one *sources* file per jurisdiction, containing the actual data on carbon pricing mechanisms and a citation key that links directly to an entry in the bibliography (providing the details of the

[3] This includes only pricing instruments whose design is directly related to the carbon content of fuels. Hence, it excludes energy taxes and duties and so differs from the structure of the data collected by the OECD to calculate the Effective Rates on Carbon (OECD, 2021a).

[4] Not all of these jurisdictions have a carbon pricing mechanism in force. The database could, however, accommodate information about pricing mechanisms in these jurisdictions, should one be implemented.

[5] This version of the dataset does not provide a systematic record of the GHG covered by each instrument. Such a record is ongoing and will be included in a future update (see section 6).

source from which the information was retrieved), respectively. The full list of references is available under the `References` directory of the GitHub repository. Sources are grouped into five categories (see section 3), so this directory contains five `csv` files (one for each category) with the bibliographic details.

The data files are structured as follows. The first five columns are the "keys" of each database entry. The corresponding column titles are `Jurisdiction, Year, IPCC_cat_code,` and `product`. `Jurisdiction` contains the jurisdiction's name, `Year` the year, `IPCC_cat_code` the code of the IPCC sector, and `Product` the name of the fuel product. Each row can be identified by a unique combination of values for these keys.

The remaining columns in this file are the variables, described in Table 1 below.

Table 1. Dataset Content

Variable Name	Description	Values	Variable Type
tax	A binary value indicating whether the sector(-fuel) is covered by at least one tax instrument.	{0,1}	integer
ets	A binary value indicating whether the sector(-fuel) is covered by at least one emissions trading system.	{0,1}	integer
tax*_id	Identifier of the applicable tax instrument.	-	string
tax*_rate_excl_clcu	Tax rate in current local currency unit (LCU) per ton of CO_2 equivalent.	[0,)	float
tax*_ex_rate	Rate of exemption applicable (e.g., 0.1 if a 10 percent reduction on the full rate applies); 0 if no exemption applies to that entry.	[0,1]	float
tax*_rate_incl_ex_clcu	Net tax rate (accounting for exemption) in current LCU per ton of CO_2 equivalent.	[0,)	float
tax*_curr_code	ISO code of the currency in which the tax rate is recorded (e.g., EUR for euro).	-	string
ets*_id	Identifier of the applicable emissions trading system	-	string
ets*_price	Price of an emissions allowance in current local currency unit per ton of CO_2 equivalent.	[0,)	float
ets*_curr_code	ISO code of the currency in which the allowance price is recorded (e.g., EUR for euro)	-	string

Note: The * in the variable names is a wildcard substituting for "" or "_II". As such, it allows to refer to identical columns for different implemented mechanisms at once.

As the table suggests, the dataset records information separately for carbon taxes and ETSs. Further, the database provides information about *all* pricing mechanism(s) applicable to a sector at the same point in time within a given jurisdiction; that is, the dataset includes as many tax- or ETS-relevant columns as there are applicable mechanisms, as some mechanisms are applicable to the same sector at the same time within a given jurisdiction. Across all jurisdictions with carbon pricing mechanisms, not more than two mechanisms have so far applied to the same sector at the same time within a given jurisdiction.

An objective of this dataset is to provide time series of coverage and prices applicable to emissions. Therefore, when, for a given row in the dataset, a new tax instrument (or ETS) is substituted for an existing one, the information (e.g., price) about this new mechanism will be recorded within the same column. The corresponding `tax*_id` or `ets*_id` value, however, will change.

For carbon taxes, the dataset also records separately "price-based," sector-fuel specific exemptions: whether a given carbon tax regulation contains provisions for some sectors and/or fuels to be subject to a different tax rate. In practice, exemptions implying a different price of CO_2 across fuels are rare; more common are exemptions set at the sector level and implying a different price of CO_2 across sectors. See section 4.1.3 for more details.

Finally, an "NA" value in a *key* column of a given row means that the key is not applicable to that row, such as in the *`Product`* column: for all noncombustion IPCC sectors, a distinction between fuel types is not applicable. The columns recording price level and exemptions have an "NA" value if no pricing mechanism is in place (i.e., if the value of the corresponding coverage binary variable is set to 0).

3. Sources

The primary source of information on institutional design and coverage is legislative acts or related administrative acts from the competent jurisdictions. When such documents could not be retrieved (at this point), we relied on established secondary sources, such as official government publications or publications from international organizations, including the *State and Trends of Carbon Pricing* series published by the World Bank.

The price data is also primarily collected from legislative acts or related administrative acts. When such documents could not be retrieved or a more structured and harmonized data source was available, we relied on secondary sources. For emissions allowance prices, for instance, we used price series made available by the International Carbon Action Partnership (ICAP) through its Allowance Price Explorer (ICAP, 2021).[6] As indicated in section 2.2, prices are recorded in current local currency units (LCU).[7]

To retrieve legislative and administrative acts, we made extensive use of information available in the Climate Change Laws of the World database (www.climate-laws.org) and the OECD Database on Policy Instruments for the Environment (https://pinedatabase.oecd.org/) (OECD, 2021b).

The dataset provides a reference to the source of each data point by including *sources* files for each jurisdiction that follow the same structure as the *data* files and include a citation key (and sometimes a comment) in each cell corresponding to a cell with information in the *data* files. All sources belong to one of the following source types: academic publication (`journal`), book (`book`), dataset (`db`), legislation (`leg`), official government publication (`gvt`), report (`report`), or web page (`web`). The details of each source are recorded in the references `csv` file associated with the source category to which it belongs. Each source is assigned a citation key. The referencing structure combines the source type, citation key, and publication year, as follows: SourceType(CitationKey[Year]). For instance, the `Sweden 1997 1A4C1 ABFLOW036 Coal` entry of the `tax` variable contains the citation `report(SMF-CT[2011])`. The cell might also include a comment, separated from the reference by a semicolon. In our example, it is `; underlying principle of the Swedish` CO_2 `tax is that it applies to motor and heating fuels.`

All referenced documents have been accessed and made part of the dataset's library.

---

[6] ICAP also provides an extensive library of documents related all existing (and future) emissions trading systems' design (https://icapcarbonaction.com/en/ets-library).

[7] For some mechanisms, the legislation expresses the carbon price in a different currency (e.g., USD) than the currency of the jurisdiction to which it applies; if so, the price is recorded in that currency.

4. Methodology

The database is created by constructing a mapping of data on carbon pricing mechanisms to the national and subnational jurisdictions in which they are in force. It is constituted of two essential building blocks: (i) data on jurisdiction, sector, and fuel coverage of each mechanism, and (ii) the tax rates and allowance prices at which these emissions are priced. Importantly, this data had not previously been systematically recorded using a standardized framework. As a result, dataset construction requires three steps:[8]

> **Data collection:** information on each mechanism's institutional design, sectoral coverage and associated prices is collected from official government or secondary sources.
>
> **Data encoding:** the information collected is structured and encoded. Coverage information (jurisdiction, sector, fuel) is recorded in a Python script. Other institutional design features and price information are recorded in ad hoc `csv` files.
>
> **Dataset compilation:** the material created is used to generate the final dataset.

A list of all the mechanisms included in the dataset, as well as the mechanisms' identifiers, is provided in the file `scheme_identifiers.csv` and reproduced in appendix A.1.

4.1 Data Encoding

4.1.1 Scope

Scope dimensions: jurisdiction, year, sector, fuel, GHG.

The information about each mechanism's coverage is encoded as a set of Python lists and dictionaries recorded in the `taxes_scope*.py` and `ets_scope*.py` files, respectively, for carbon taxes and ETSs. The `*` is a wildcard substituting for either `_CO2`, `_CH4`, `_N2O` or `_Fgases`, the four (group of) gases covered in the dataset.[9] Since the sectoral scope of a mechanism may vary by type of GHG it is easier to maintain separate records of sector-year scope of a mechanism for each GHG individually. Thus, the dataset records scope information separately for each GHG.

---

[8] A visual representation of the workflow is provided in the appendix A.4.

[9] Extension of the dataset to non-CO_2 gases is ongoing.

Each `taxes_scope*.py` or `ets_scope*.py` file offers lists containing the following: the jurisdictions to which the mechanism applies, IPCC sectors to which it applies, and fuels to which it applies. Each scope dimension (jurisdiction, sector, fuel) has as many lists as there has been changes in the scope of the mechanism over its lifetime. These lists are respectively assigned to relevant years using dictionaries where "years" are dictionary keys, and the corresponding value is either the appropriate sector or fuel list. A description of the structure of the coverage encoding is presented in appendix A.3.

4.1.2 Prices

The prices associated with each carbon pricing mechanisms are recorded in individual, mechanism-specific, `csv` files. The file naming rule is `[scheme_identifier]_prices.csv`. One price is recorded for each GHG separately. Usually (but not always), the same price applies across fuels and sectors for each GHG. The recorded price is the full price of emissions. Sector- or sector-fuel-level departures from this price (i.e., "price-based" exemptions, see next section) are recorded in separate files.

For carbon tax mechanisms covering emissions from IPCC Energy sectors, the records distinguish between the rate applied to three aggregate fossil fuel categories (coal, oil, and natural gas). The rate, expressed in LCU/tCO_2e, typically does not differ by category, with some exceptions (e.g., Norway). For ETSs (cap-and-trade), only one price value is recorded for each year, as the scope of the ETS is set at the sector level and associated emissions allowances cover emissions within covered sectors, regardless of the fuel from which they originate. The recorded value is either the yearly average of daily allowance prices or the allowance-weighted average of clearing prices in all auctions held within that year. Whether one or the other is recorded is primarily determined by the information publicly available, and which value is recorded is clearly indicated by a comment in the corresponding `comment` column of the files.

Finally, note that the carbon prices recorded in this database reflect the marginal, not average, price of emissions (OECD, 2021). Typically, if the price applies to the entire emission base, the average and marginal carbon prices do not differ. However, when tax-free emissions allowances or emissions permits are granted for free to sectors covered by a pricing mechanism, then a wedge between marginal and average price arises.

4.1.3 Price Rebates

A rebate on the full price of emissions may be granted to particular sectors or fuels falling under the scope of a pricing mechanism. We denote these exemptions as price rebates because they grant a rebate on what is otherwise the full price on emissions, but apply to all emissions within the scope of the mechanism. Such rebates occur in

carbon taxes, not ETSs. In the latter case, all sectors and gases within the scope of the system face the same marginal price on their respective emissions.[10]

Price rebates are currently manually recorded in separate `csv` files that follow the same structure (jurisdiction, year, sector, product) as the main data files.

4.1.4 Scope Exemptions

Scope exemptions are regulatory provisions exempting some of a sector's emitters (and associated emissions) from the scope of a particular instrument, and they include the following:

- *Compliance thresholds*, which exempt some emitters from the scope of a given mechanism based on their total yearly emissions, as for Chile's carbon tax, or rated thermal input, as for the EU ETS; and
- *Administratively set exemptions if covered by another mechanism*, which may occur when the emissions might be covered by two mechanisms and the liability is waived for one of the two.

We record within-sector scope exemptions using a `coverage factor.`[11] This factor is an initial attempt to account for administrative rules of implemented carbon pricing mechanisms affecting the scope of emissions covered within sectors.

The calculation (or direct encoding) of coverage factors is based on administrative data or information about the carbon pricing mechanisms. Our primary approach is to compare administrative emissions data (e.g., EU ETS registry data) to inventory data. In theory, the former is strictly smaller than the latter, as administrative data only includes emissions from covered entities. Two issues may arise, however. First, administrative data is not available for all mechanisms.[12] Second, when it is available, methodological differences in the construction of administrative and inventory data may create inconsistencies rendering a comparison difficult, at least at the sector level. Below, we describe how the coverage factors for various jurisdictions and sectors are set.

Emissions trading: We calculate an adjustment based on the ratio of registry emissions to inventory emissions. As methodological differences in the construction of

[10] In an emissions trading system, industries may be granted some emissions allowances for free, such as in the early phases of most ETSs introduced so far. Such free allocation reduces the average price on emissions faced by covered emitters but not the marginal value of avoided emissions (OECD, 2021a).

[11] Coverage factors are preliminary, and their development is ongoing. As such, they are included in the final dataset.

[12] For ETSs, such records exist as part of the emissions registries maintained. For carbon taxes, similar administrative records may exist showing which entity paid the carbon tax and for which amount of emissions. However, these are not publicly available.

registry and inventory data currently preclude a consistent calculation at the sector level, we calculate this ratio at the jurisdiction level.[13]

Taxes: The assumption is usually that taxes cover 100 percent of a sector's emissions, because taxes apply to all entities consuming the fuel within that sector. Only in some specific cases is the coverage less than 100 percent:

- Singapore: compliance threshold, and
- Colombia: compliance threshold.

When a specific coverage factor could not be determined, 100 percent coverage was assumed.

4.2 Dataset Compilation

The compilation of the dataset happens in 5 steps, which are all contained in the `db_build.py` script:

1. Instantiate a dataframe containing the entire structure of the dataset; that is, the keys columns (`Jurisdiction, Year, IPCC_cat_code,` and `Product),` and all rows.
2. The coded scope information contained in `ets_scope.py` and `taxes_scope.py`, as well as the price information, extracted from the relevant raw `csv` files where they are recorded using the `ets_prices.py` and `tax_rates.py` scripts, are used to generate the following columns: `tax, ets, tax*_id, tax*_rate_excl_ex_clcu, tax*_ex_rate, tax*_curr_code, ets*_id, ets*_price, ets*_curr_code`.
3. Calculate `tax*_rate_incl_ex_clcu` by using files containing information on price rebates.[14]
4. The dataset includes one additional step, calculating mechanism scope values for aggregate IPCC sectors based on the value for subsectors; these take the value 1 if and only if all subsectors are covered.
5. Finally, all variable entries for which the corresponding tax or ETS indicator value is 0 are set to "NA."

[13] The calculations are ongoing. The current version of the dataset does not reflect this development.

[14] This step also ensures that jurisdiction-year-sector(-fuel) specific scope and price entries that cannot be accounted for within the main data flow are recorded, by introducing a limited number of code lines to modify these entries after the dataset is generated.

5. Using the Dataset and Integration with Other Databases

The structure of the dataset allows for a straightforward integration with other data sources that follow IPCC 2006 sectoral disaggregation[15] One such integration is with jurisdictions' GHG emissions inventories, such as reported through the UNFCCC process and available through the UNFCCC data portal (https://di.unfccc.int/time_series) or estimated and compiled by institutions such as the Joint Research Centre of the European Commission and available in its Emissions Database for Global Atmospheric Research (EDGAR, https://edgar.jrc.ec.europa.eu/).[16]

In addition, for IPCC Energy sectors, the dataset provides the corresponding *Flow* (i.e., sector) key of the International Energy Agency's CO_2 emissions from fuel combustion data (IEA, 2019). This facilitates integration with IEA data.

Combining this dataset with GHG emissions data provides several opportunities, including calculating the share of emissions covered by pricing mechanisms within national and subnational jurisdictions. It also allows for calculating emissions-weighted average carbon prices, at the sector or jurisdiction level. Such calculations were performed as part of a separate but related undertaking and resulted in emissions coverage figures and average prices at the jurisdiction level for 1990–2020. A companion paper describes the data sources and methodology related to these metrics.

[15] It also allows for integration with data using different disaggregation, if used in combination with appropriate concordance tables.

[16] For further details about EDGAR methodology, see https://edgar.jrc.ec.europa.eu/methodology.

6. Continuous Update and Future Extensions

The dataset is under continuous development. Although every precaution has been taken to accurately record coverage and price information for each carbon pricing mechanism, the magnitude of the effort has been such that some inaccuracies might remain. Suggestions to update existing records and contributions to the extension of the dataset to other features of carbon pricing mechanisms are welcome; please refer to the guidelines available on the GitHub repository.

In addition, this dataset would benefit from the following extensions.

- Update of data to the latest year to reflect institutional design and price changes pertaining to mechanisms, as well as information on mechanisms established since the last release.
- Integration of information on the carbon pricing mechanisms in force in two Japanese municipalities: Saitama and Tokyo.
- Extension to mechanisms covering other Kyoto GHGs than CO_2 and recording of Kyoto gases covered by each mechanism.
- Integration of information about tax-free allowances and free allocation of emission permits (by sector).

The next update of the dataset will focus on these extensions, with specific attention given to (3) and (4). Implementing the former will allow to provide a comprehensive summary of pricing mechanisms of all GHGs whereas implementing the latter will allow to calculate sector- or sector-fuel-level *average* – not *marginal* – carbon prices and account for tax-free emissions allowances (as in the case of, e.g., the South Africa carbon tax) and freely allocated emissions allowances in emissions trading systems in the calculation of the economy-wide average.

7. Conclusion

Over the last three decades, the number of jurisdictions with a carbon pricing mechanism covering at least part of their GHG emissions has grown significantly. As jurisdictions around the world continue to strengthen their policy response to climate change, existing mechanisms may be expanded, and new mechanisms may be introduced.

Furthermore, within the context of the Paris Agreement, countries and subnational jurisdictions move unilaterally and implement mechanisms differing in sectoral scope and price level. Several of these mechanisms, especially carbon taxes, introduce price rebates and sector-level scope exemptions that lower the effective stringency of the mechanism.

However, the global nature of the climate change problem requires that the stringency of policies in force at the national and subnational levels be easily compared with each other. A vital step in that direction is to record and provide information about these mechanisms within a detailed and standardized framework. We believe that this dataset provides an instrumental contribution. The mechanisms included, the sectoral disaggregation according to IPCC 2006 guidelines, and the period covered make this undertaking unique in its depth and breadth of coverage.

This, in turn, should allow for increased transparency about a key policy instrument to reduce GHG emissions and an easier assessment of the stringency of carbon pricing mechanisms among national and subnational jurisdictions. For example, one can easily assess the sectoral coverage of certain sectors among a group of jurisdictions.

Finally, the structure of the dataset—by territorial jurisdictions and disaggregation at the level of IPCC sectors—allows for a straightforward integration with other datasets.

8. References

Dolphin, G., Pollitt, M., and Newbery, D. 2020. The political economy of carbon pricing: A panel analysis. Oxford Economic Papers.

International Energy Agency. 2019. *CO_2 emissions from fuel combustion* database. https://www.iea.org/data-and-statistics/data-product/greenhouse-gas-emissions-from-energy.

Buendia, L., Miwa, K., Ngara, T., and Tanabe, K. (eds). IPCC 2006, 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Japan: IGES.

ICAP. 2021. Allowance Price Explorer. https://icapcarbonaction.com/en/documentation-allowance-price-explorer

OECD. 2021a. *Effective Carbon Rates 2021: Pricing Carbon Emissions through Taxes and Emissions Trading*, OECD Publishing, Paris, *https://doi.org/10.1787/0e8e24f5-en*.

OECD. 2021b. Database of Policy Instruments for the Environment. https://pinedatabase.oecd.org/

The World Bank. 2021. "State and Trends of Carbon Pricing 2021" (May), World Bank, Washington, DC. Doi: 10.1596/978-1-4648- 1728-1. License: Creative Commons Attribution CC BY 3.0 IGO

Appendix

A.1 Pricing Mechanisms Identifiers

scheme_name	scheme_id
Alberta TIER	can_ab_ets
Alberta carbon tax	can_ab_tax
Argentina carbon tax	arg_tax
Australia CPM	aus_tax
Australia ERF Safeguard Mechanism	aus_sm
BC GGIRCA	can_bc_ggirca
BC carbon tax	can_bc_tax
Beijing pilot ETS	chn_bj_ets
California CaT	us_ca_cat
Canada federal OBPS	can_obps
Canada federal fuel charge	can_tax_I
Canada federal fuel charge	can_tax_II
Chile carbon tax	chl_tax
China national ETS	chn_ets
Chongqing pilot ETS	chn_cq_ets
Colombia carbon tax	col_tax
Denmark carbon tax	dnk_tax
EU ETS	eu_ets
Estonia carbon tax	est_tax
Finland carbon tax	fin_tax
France carbon tax	fra_tax
Fujian pilot ETS	chn_fj_ets

Germany ETS	deu_ets
Guangdong pilot ETS	chn_gd_ets
Hubei pilot ETS	chn_hb_ets
Iceland carbon tax	isl_tax
Ireland carbon tax	irl_tax
Japan carbon tax	jpn_tax
Kazakhstan ETS	kaz_ets
Korea ETS	kor_ets
Latvia carbon tax	lva_tax
Liechtenstein carbon tax	lie_tax
Luxembourg carbon tax	lux_tax
Massachusetts ETS	us_ma_ets
Mexico carbon tax	mex_tax
Mexico pilot ETS	mex_ets
New Brunswick carbon tax	can_nb_tax
New Brunswick OBPS	can_nb_ets
New Zealand ETS	nzl_ets
Newfoundland and Labrador PSS	can_nl_ets
Newfoundland and Labrador carbon tax	can_nl_tax
Northwest Territories carbon tax	can_nt_tax
Norway carbon tax	nor_tax_I
Norway carbon tax	nor_tax_II
Nova Scotia CaT	can_ns_ets
Poland carbon tax	pol_tax
Portugal carbon tax	prt_tax
Prince Edward Island carbon tax	can_pe_tax

Quebec CaT	can_qc_ets
RGGI	us_rggi
Saitama ETS	jpn_sai_ets
Saskatchewan OBPS	can_sk_ets
Shanghai pilot ETS	chn_sh_ets
Shenzhen pilot ETS	chn_sz_ets
Singapore carbon tax	sgp_tax
Slovenia carbon tax	slo_tax
South Africa carbon tax	zaf_tax
Spain carbon tax	esp_tax
Sweden carbon tax	swe_tax
Switzerland ETS	che_ets
Switzerland carbon tax	che_tax
Tianjin pilot ETS	chn_tj_ets
Tokyo CaT	jpn_tok_ets
UK carbon price floor	gbr_tax
Ukraine carbon tax	ukr_tax

A.2 Coding Rules

Price differences within fuel type aggregates: Different tax rates may apply to different fuels within the main fuel categories (i.e., coal/peat, natural gas, oil). This comes in the form of differentiated applicable tax rates or varying exemption rates, such as for Mexico. In such cases, the value in the dataset is the highest rate applicable to the fuel category. The fuel- and/or sector-fuel-specific rates are recorded in the *country notes* for each country.

Encoding of partial and full sectoral exemptions: If a sector is fully exempted from a given pricing mechanism, the coverage variable value is set to 0 for the sector (and all fuel categories, when relevant). If only a specific fuel type within a sector is exempted (partially or totally), then the coverage variable value is set to 1 for the sector (and all fuel categories, when relevant) and the tax_exemption_rate variable value is set to the value corresponding to the level of the exemption (which is 1 if exemption is total).

Price rebates exemptions: If the exemption varies within the sector-fuel level, the highest exemption rate is recorded. For example, farmers in Alberta can receive up to 100 percent carbon levy rebates, depending on their eligibility. In such cases, the coverage variable in the corresponding row of the dataset and the exemption rate are set to 1.

Currency of price record in multijurisdiction mechanisms: When a jurisdiction (or some sectors therein) is part of a multijurisdiction mechanisms, the price is expressed in the common currency of the mechanism and the recorded currency code is consistent with that currency. For instance, for the California-Quebec(-Ontario) cap-and-trade system, the price of allowances is expressed in US dollars (USD).

Recording prices in incomplete years: if a mechanism was not in place for the entirety of a given year, a price (tax rate or allowance price) "pro-rata tempore" is calculated; the recorded price is a weighted average of all prices to which emissions were subjected throughout the year. This applies to all years apart from years in which a mechanism was first introduced. In that case, the recorded price and coverage are recorded "as if" it had been applied from January 1 of that year.

Multiple mechanisms: when two mechanisms apply to the same sector, information on both mechanisms is recorded through additional columns in the dataset.

Coverage records of IPCC sectors that are exclusively sources of non-CO_2 GHG emissions: if a policy instrument that targets CO_2 emissions also covers other Kyoto GHG, then all IPCC sectors covered by that instrument are recorded. However, information about instruments that exclusively target non-CO_2 GHG emissions has not yet been recorded.

Partial price rebates on certain emissions within sectors: In some cases (e.g., the Ireland carbon tax), the tax does not cover all emissions of a sector and a rebate is granted on emissions covered by the tax. If so, a combination of (i) the share of emissions it applies to (`em_share`) and (ii) the exemption rate allows for accurately recording the case.

A.3 Data Raw Files Description

A.3.1 Raw Data

Coverage: coverage information in encoded in the `ets_scope*.py` and `taxes_scope*.py` files. The structure of these records is described below:

```
----------------------------scheme_name----------------------

# List(s) of jurisdictions

scheme_id_jur_* = []

# List(s) of IPCC sectors covered

scheme_id_ipcc_* = []

# List(s) of fuel categories

scheme_id_fuel_* = []


# Coverage dictionaries

scheme_id_jur_cov = {year1: scheme_id_jur, …,2021:
scheme_id_jur}

scheme_id_ipcc_cov = {year1: scheme_id_ipcc, …,2021:
scheme_id_ipcc}

scheme_id_fuel_cov = {year1: scheme_id_fuel, …,2021:
scheme_id_fuel}

------------------------------------------------------------
```

Prices: When price data is manually encoded, it is saved in individual (i.e., one per pricing mechanism) `csv` files. The file naming convention is `[scheme_id]_prices.csv`. Each file follows the same structure:

- *Carbon taxes*: `scheme_id, year, ghg, product, rate, currency_code, source, comment`
- *ETSs*: `scheme_id, year, ghg, allowance_price, currency_code, source, comment`

When price data is retrieved from structured datasets, it is extracted directly from there and reformatted in the Python script.

Price (tax) exemptions: Exemptions are recorded manually *by jurisdiction* (not by pricing mechanism, as exemptions are, for the most part, jurisdiction rather than mechanism specific) in individual `csv` files. The file naming convention is [tax_ex_jurisdictionName].

A.3.2 Auxiliary Files

- `IPCC2006-IEA-category-codes.csv` provides a mapping between IPCC sector names, their associated code and the corresponding International Energy Agency sector code. This latter file is particularly useful to update the dataset, as its .csv files only include IPCC sector codes.
- `scheme_identifiers.csv` provides a correspondence between the names of pricing mechanisms' and their internal database identifier.

A.4 Dataflow Representation