The Rise and Fall of CO_2 Emissions

A data exploration into the best predictors of a country's CO_2 output, the top CO_2 reducing countries & which renewable energies will become the cheapest

A Report by Dolphinity

M.Schotten, June 2023

The Rise and Fall of CO_2 Emissions

A data exploration into the best predictors of a country's CO_2 output, the top CO_2 reducing countries & which renewable energies will become the cheapest

Author: Michiel Schotten
M.Schotten@dolphinity.io

Dolphinity – Data Science Consultancy & Reporting
Alkmaar, The Netherlands
www.dolphinity.io

Version 1.0, June 2023

Citation: M. Schotten (2023). *The Rise and Fall of CO_2 Emissions. A data exploration into the best predictors of a country's CO_2 output, the top CO_2 reducing countries & which renewable energies will become the cheapest.* Technical Report by Dolphinity, the Netherlands, 56 pp. Available at: https://github.com/Dolphinity-io/CO2-emissions.

Feedback and Requests:

Dear reader, if you have any feedback, comments and/or questions about this report or the results therein, I would very much appreciate if you could let me know at M.Schotten@dolphinity.io. Also if you have any requests about data analyses on other topics of interest, literature reviews and/or customized reporting (either in short or longer formats), I invite you to contact me and am happy to discuss options with you. Thank you for your interest.

Executive Summary

The main topic of the analysis presented in this report is global emissions of CO_2, the "molecule of Life" that stands at the basis of photosynthesis and, therefore (together with the number 42, of course[1]), of all organic life on Earth. It's a high-level and somewhat simplified analysis that looks at three aspects of this: which variables are the best predictors of a country's CO_2 emissions, which countries are "best-in-class" with respect to reducing their emissions, and which renewable energy technologies will likely become the cheapest in the near future.

The analysis was carried out as final assignment of the "Data Analytics with Python" degree from Winc Academy in the Netherlands, and exclusively used datasets found at the "Our World In Data" website at https://ourworldindata.org/.

Because of the broad and open nature of the questions (especially the first one) and the many nuances, assumptions and caveats involved in interpreting the results, the report has become quite lengthy to do these full justice. This three-page Executive Summary is therefore meant for those readers with limited time, to briefly take home the key takeaways from each of the three parts of the analysis. If you're such a reader, I would recommend to also take a quick look at the figures (only those in the main text, not the Appendix), as the proverbial picture says more than a 1,000 words. And for full details, you can find my Jupyter Notebooks created for this analysis at https://github.com/Dolphinity-io/CO2-emissions.

- ***Question 1***: *What is the biggest predictor of a large CO_2 output per capita of a country?*

To answer this, I looked at a total of 35 variables within five self-defined categories or sectors: Economy & Prosperity, Energy, Transportation, Food & Agriculture, and Demographic & Sustainability (or: "Other"). For each variable, I used the annual data from all countries in the world, insofar and for whichever years they were available; and then divided the annual CO_2 emissions of each country (as well as most of the other variables) by its annual population size, in order to get the "per capita" values.

CO_2 emissions were then scatter plotted against each of the variables (so yielding 35 different charts, shown in Appendix 2), with linear regression lines fitted through the data. I assumed that the variables with the best cross-correlation statistics, measured as their R^2-values (with 0 meaning no correlation, and 1 a perfect fit), were also the best predictors of CO_2 output – as long as the slope of the regression line was also positive, of course (otherwise they would be predictors of a *small* CO_2 output); which was indeed the case for all statistically significant regressions.

[1] Douglas Adams – *The Hitchhiker's Guide to the Galaxy. A Trilogy in Five Parts.*

The results are summarized in the horizontal bar chart of Figure 2, showing that **the *biggest* predictor** – at least within the context of this simplified analysis – **is a country's per capita electricity *generation* from coal, oil and gas** (*i.e.*, from fossil fuels), measured in Megawatt-hours, with an R^2-value of 0.71 (or a Pearson Correlation Coefficient R of 0.84; see also Figure 1). The next biggest predictors of a large CO_2 output are also in the Energy sector: a country's overall primary energy *consumption*; and the primary energy consumption just from fossil fuels (both with an R^2=0.68). A country's "Gross National Income" (GNI, in US$) is the fourth biggest predictor (R^2=0.66), followed by variables in the Transportation, Economic and Food sectors.

As a caveat: unfortunately, no suitable datasets were available for aviation or international shipping, which I suspect would also be large predictors of CO_2 emissions (although notably difficult to assign to individual countries). Other caveats to consider are that I only looked at linear regressions (results might be different when looking at polynomial, higher order fits through the data); and that because data were pooled together for all countries and years, with some countries and years having more data points than others for certain variables, this may have introduced country, regional and/or period biases in the results.

- ***Question 2****: Which countries are making the biggest strides in decreasing CO_2 output?*

For this question I used the same dataset, albeit only the annual CO_2 per capita values for all countries. A pre-analysis, done to maximize the number of countries with available data, revealed that using the following reference years would give the most well-balanced overall picture: 1964 (with data for ~90% of all countries), 1990 (97% of countries), 2008, and 2021 (both with data for all 219 countries). This in turn led to three reference periods over which to calculate CO_2 reduction: 1964-2021 for the "long-term" reduction performance, 1990-2021 for the "mid-term" performance, and 2008-2021 for the "short-term" performance. In addition, I looked both at the *relative* reduction in 2021, measured as % of the country's baseline CO_2 output in the reference period start year (either 1964, 1990, or 2008), and at the *absolute* reduction, in tonnes of CO_2 per capita, where I subtracted the 2021 output from the start year output.

The results of this multifaceted approach are summarized in Figure 3, which shows the top 7 CO_2 reducing countries for each reference period and for both ways of measuring the reduction. And while none of the top 7 lists are the same, there are quite a few overlaps. A general observation is that the best CO_2 reducers tend to be relatively small countries, and more specifically island nations in the Caribbean are doing well. Overall, measured across the board, **Curaçao is one of the countries that is making the biggest strides in decreasing CO_2 output**, followed by **Sint Maarten (Dutch part) and Luxembourg**. Other countries that are doing well (mostly during specific reference periods) are **Aruba, Democratic Republic of Congo, Moldova, {Bonaire, Sint Eustatius and Saba}, Singapore, Qatar, Estonia, and Ukraine**. A likely cause for the good CO_2 reduction performance of both Curaçao and Aruba is the long-term presence of large oil refineries on these islands, which were scaled down over the years and finally closed.

- ***Question 3**: Which non-fossil fuel energy technology will have the best price in the future?*

A small dataset with the global annual "Levelized Cost of Energy" (LCOE) data for seven renewable energy technologies was used to address this question. LCOE considers the costs of operating an industrial-scale power plant across its lifetime (averaged across the world for each year) and is expressed in U.S. Dollar per kilowatt-hour of generated electricity ($/ kWh), with the US$ expressed as its value in 2021 to account for the effects of inflation. The renewable energies for which such LCOE data were available in the period 1983-2021 were: Bioenergy, Geothermal, Hydropower, Solar Concentrated, Solar Photovoltaic, Wind Offshore, and Wind Onshore (with only the latter having the annual data for all 39 years). To predict future prices for each I used regressions, where I fitted both linear and so-called log-linear regression lines through the LCOE data points (the latter involved taking the logarithm of LCOE-values).

Figure 4 and Table 1 together show the results of this analysis, with all seven renewables plotted in the same chart. It's clear that the log-linear regression lines provided better fits through the data than the regular regression lines, indicating exponential price decay (indicative of so-called "learning rate" technologies). It's also clear that **Solar Photovoltaic showed the fastest annual decline of LCOE prices at -17.6% /year** and will therefore (at least according to this dataset) have **the best future price of all seven renewable energy technologies**. The next cheapest renewables will be Solar Concentrated (-9.7% /year) and Onshore Wind (-4.9% /year). From the log-linear regression line equation, I also made the forecast that Solar PV will **reach a price of 1 cent/ kWh in 2028** (in 2021 US$ prices).

As a big caveat: when comparing my LCOE forecasts for 2023 with the actual real-world LCOE prices today (May 2023) as a sanity check, I was surprised to find that, despite a very high correlation of R^2=0.98 for the 2010-2021 price decline, the LCOE for Solar PV has almost *doubled* from 2021 to 2023, even when corrected for the high 20% inflation in the past two years. This led me to undertake a "mini literature review" (although clearly outside the scope of the main analysis done for Q3) to look for potential reasons for this, the results of which are described in an extended Discussion and 7-point summary in Appendix 1.

It yielded some very valuable insights and **made me realize that the above conclusion may be a premature one**, and that great caution should be taken in relying too much on the forecasts in Table 1. It also serves as a textbook illustration of one of the central challenges of data science, so elegantly summarized in the truism put forth by famous physicist Niels Bohr: "*Prediction is very difficult, especially if it's about the future.*"

Table of Contents

1. Introduction

This analysis was carried out as final assignment of the "Data Analytics with Python" degree from Winc Academy in the Netherlands, for which the certificate of completion was awarded to the report's author M. Schotten in June 2023: https://certificates.wincacademy.com/ee529bbc-e8b9-4fde-a40e-de0113e412ce.

The assignment was to use any dataset from the "Our World in Data" website at https://ourworldindata.org/ (abbreviated to OWID in the remainder of this report) to answer the following three questions regarding CO_2 emissions:

Q1. What is the biggest predictor of a large CO_2 output per capita of a country?

Q2. Which countries are making the biggest strides in decreasing CO_2 output?

Q3. Which non-fossil fuel energy technology will have the best price in the future?

All data analyses to answer these questions were done with custom Python code, using the pandas, NumPy, and Matplotlib libraries, in two Jupyter Notebooks. These Notebooks are supplements to this report and can be found in the same Github folder at https://github.com/Dolphinity-io/CO2-emissions. And while the most relevant charts created in the Notebooks are presented as Figures in the main text below, all the additional charts providing more detail and insights can also be found in the Appendix.

2. Analysis & Results per question

2.1. Q1 – Biggest predictor of CO_2 output

2.1.1. Q1: Method (including general approach) and datasets

General caveat

Because of the broad and open nature of the question – theoretically, *any* measurable variable could be the biggest predictor of a country's CO_2 output – and the large amount of different datasets available from OWID, I chose to focus on what I considered to be the most *likely* variables to correlate with CO_2 emissions.

Therefore, the conclusions of this analysis come with a caveat: it's still theoretically possible – although unlikely – that the biggest CO_2 predictor is another variable than found here; the biggest CO_2 predictor found here, however, is certainly the biggest predictor of all the variables that were *analyzed.*

Classification of variables by sector

I classified the OWID datasets that I thought were the most likely candidates to correlate with a country's CO_2 output under the following self-assigned sectors:

1. **Economy & Prosperity**: Gross Domestic Product (GDP), exports, Human Development Index (HDI, which has a scale from 0 to 1), and Gross National Income (GNI).

2. **Energy**: primary energy consumption, energy consumption per GDP, electricity demand, electricity generation, fossil fuel (*i.e.*, coal, oil and gas) consumption (absolute and % of total primary energy consumption), electricity generation from fossil fuels (absolute and % of total generation), % energy consumption from coal, % electricity generation from coal, nuclear energy consumption, nuclear energy generation, renewable energy consumption, and renewable energy generation.

3. **Transportation**: amount of railway travel, total number of new car registrations, nr. of fossil fuel (*i.e.*, diesel and petrol) car registrations (absolute and % of total), nr. of diesel car registrations (absolute and % of total), and nr. of pure electric (*i.e.*, not hybrid) car registrations (absolute and % of total).

4. **Food & Agriculture**: agricultural land use, meat production, fish production (total of aquaculture and fish captured in the wild), daily caloric supply (*i.e.*, total of carbohydrate, protein and fat intake), protein supply (*i.e.*, animal and plant-based; absolute and % of total caloric supply), and animal protein supply (absolute).

5. **Demographic & Sustainability**: % of the population living in cities and amount of plastic waste. This sector is somewhat of an "Other" category, for those variables that didn't fit into the previous four.

No aviation or shipping data

Please note that in the self-assigned sector “Transportation”, I did not analyze any variables related to aviation or international shipping of products and goods. Both of those are known large contributors to CO_2 emissions, but also fall outside of current international agreements, due to the inherent difficulty to assign them to any individual country – although this is much easier for domestic flights than for international flights.

Regardless, I didn’t find any suitable datasets on the OWID website: none for shipping, and the datasets that I did find for aviation, at https://ourworldindata.org/transport#aviation, were already the per capita CO_2 emissions *attributed* to aviation, calculated by OWID with their own methodology and inherent assumptions. As such, it’s not a suitable variable to do a linear regression on with respect to the per capita CO_2 emissions of a country as a whole; what we could do with this dataset instead is to calculate the % of a country’s per capita CO_2 emissions that could be attributed to aviation. That, however, would make it very difficult to compare to the other variables above that are all analyzed with linear regressions, and thus not help us in answering question 1. Aviation therefore falls outside of the scope of the approach taken in this analysis.

“Per capita” conversion

Furthermore, please note that most of these variables (except for HDI, energy consumption per GDP, and all the “% share of total” variables) were first transformed to their “per capita” equivalents, by dividing them by the country’s population in the respective year, in order to be able to meaningfully compare them to that country and year’s “CO_2 per capita” value (measured in tonnes, *i.e.*, units of 1,000 kg CO_2 emitted per person).

This procedure “normalizes” the data for a country’s population, which varies from year to year and which affects a country’s CO_2 output (*i.e.*, generally speaking, countries with larger populations are likely to emit more CO_2). By normalizing for population size, this potentially confounding factor will not obscure or affect any correlations between the variable in question and CO_2 emissions.

Plotting CO_2 against each variable

For each variable, a separate chart was made, where CO_2 per capita (as the dependent variable) was plotted on the y-axis against the variable in question (taken as the independent variable) on the x-axis. Each plotted data point is therefore the (*variable, CO_2*) combination for a single country[2] and single measurement year. To simplify the analysis and interpretation of the results, I chose to pool all data together for all countries and all years in each single chart

[2] Please note that OWID uses https://www.worlddata.info/countrycodes.php as its list of official country names and codes across its various datasets. However, these “countries” also include overseas and dependent territories that are not technically countries, although they usually have some form of self-governance or autonomous status. An example is “Bonaire, Sint Eustatius and Saba”, also called “Caribbean Netherlands” (ISO country code: BES), which technically consists of three Dutch municipalities that are officially part of the Netherlands. In contrast, the other Netherlands Antillean islands “Aruba”, “Curaçao”, and “Sint Maarten (Dutch part)” each do have the status of independent country.

For consistency, all “countries” in the OWID datasets that have country codes from the list above will also be referred to as countries throughout this report.

(even though correlations might in theory also vary considerably across different time periods and countries), without *e.g.* color-coding data points to indicate different countries or time periods.

Varying sample sizes

Naturally, in order for a data point to be plotted, the combination of CO_2 per capita, the variable in question, and (in most cases) population size – required to calculate the "per capita" values – needs to be available for a specific country and year: that is, if any of the three are missing, it can't be plotted. This resulted in a widely varying number of data points (indicated as the sample size "N" in each chart), varying from N=161 for plastic waste data – with data only available for 2010 – to N=14,838 for % urban population data, and with a median sample size of N=5,039 across all 35 variables. Generally, for variables with a lower sample size N, the correlation strength (expressed as R^2) of the linear regression between the variable and CO_2 tends to be lower as well, which should be kept in mind when interpreting the R^2-values.

Potential country biases

Likewise, the number of countries for which data were plotted also varied substantially across variables (ranging from 19 to 217, with a median of 173 countries), as did the number of plotted years (ranging from just 1 year to 222 and a median of 53 plotted years). For most variables, the most recent year was either 2021 or 2019, while the earliest plotted year was 1800 (only for % urban population) – but most variables had annual data available only from the last few decades, with a median start year of 1965.

As data for some of the variables are only available for a handful of countries (*e.g.*, car registration data are only available for Western European countries and Turkey), and, in addition, for ***all*** the variables the number of years for which data are available varies substantially across countries, this could lead to "country bias" (or even region bias). This means that those countries that have most data – *i.e.*, years – available for a particular variable will weigh most heavily on the results, and thus, on the conclusions for that variable. While this is not ideal, it's the best we have available; but this caveat should be kept in mind when interpreting the conclusions.

Linear regressions for CO_2 vs. each variable

In addition to scatter plotting the (*variable, CO_2*) data points for each variable, as already mentioned a linear regression line was fitted through the data as well. A text box in each chart shows the equation of the regression line for that variable (expressed as: CO_2 = slope*variable + y-intercept), the R^2-value (also known as the Coefficient of Determination, or Pearson Correlation Coefficient [= R] squared – which varies between 0 and 1 and indicates the strength of the linear cross-correlation), the sample size N, and the p-value. The latter is the statistical significance, *i.e.*, the calculated chance of finding a Pearson correlation coefficient of R or even higher in a sample if it would be 0 in the population (*i.e.*, if there would be no correlation). In other words, the smaller p is, the higher the probability (namely, 1.0 – p) that x and y are indeed linearly correlated, by at least R.

Limitations of HDI interpretation

Two of the assumptions (or better: formal conditions) in order to perform a Pearson linear regression between two variables is that the datasets of both variables have a so-called normal distribution and are "scale", *i.e.*, evenly spaced, numerical variables (for more details, please see: https://www.scribbr.com/statistics/pearson-correlation-coefficient/#when and https://www.scribbr.com/statistics/levels-of-measurement/).

Although I didn't specifically test for these conditions, I assumed they are likely to be the case for almost all analyzed variables – except for the Human Development Index or HDI, which is a calculated variable on a 0-to-1 scale, with a methodology that uses four input metrics as development measures.[3] Because of how it's calculated, HDI is a so-called ordinal variable; similar to, for example, the Richter scale to measure earthquakes, which is logarithmic and also ordinal. This means that, *e.g.*, the interval between HDI=0.3 and 0.4 does not represent the same difference in development of a country as the interval between HDI=0.4 and 0.5. An explanation of ordinal variables can be found here:
https://www.youtube.com/watch?v=N74MeeY5KZw

Because HDI is an ordinal variable (while CO_2 per capita is a scale variable), strictly speaking we cannot calculate the Pearson correlation between HDI vs. CO_2, but should calculate the Spearman rank correlation coefficient instead (please see https://peterstatistics.com/CrashCourse/3-TwoVarUnpair/OrdScale/OrdScale3.html). For this analysis I ignored this, however, as the HDI is still somewhat like a scale variable; and also for the sake of simplicity, in order to easily compare HDI to all other variables.

Throwing out Kuwait 1991 data point

Through manual inspection of the data, I discovered that the 1991 data point for Kuwait is an enormous outlier, which affected the vertical scale of most of the plots and thus degraded their readability and usefulness. That year, Kuwait had enormously high CO_2 emissions – both in absolute terms: 492.8 million tonnes; and in relative terms: a 12x increase compared to the previous year 1990 (37.8 tonnes), followed by a stark 17x decrease the next year 1992 again (29.6 tonnes).

For those old enough to remember, there was a famous cause for this: the Iraqi troops set many of the Kuwait oil wells on fire during the first Gulf War, which took months to put out and during which period enormous amounts of CO_2 were released to the atmosphere. I therefore decided to remove this data point, which is justified: it can be considered an artifact and has no relationship to any of the variables being considered in this analysis for Q1.

---

[3] HDI is a composite calculated variable that has four metrics as input: 1) years of life expectancy; 2) expected number of years of schooling at birth; 3) mean years of schooling in the entire population; and 4) Gross National Income or GNI per capita. For more details on the methodology and the exact calculation, please see https://ourworldindata.org/human-development-index#how-is-the-human-development-index-calculated.

<u>Only linear regressions</u>

Finally, please note that only linear regressions were performed in this analysis, to keep it relatively simple and straightforward and also to make comparisons among the many variables possible. If some variables would correlate with CO_2 emissions in a non-linear fashion, however (*e.g.*, logarithmically or with higher degree polynomial fits through the data), that might affect the order of variables that show the highest correlation with CO_2 as well, and thus affect the final conclusions. This is another caveat that should be kept in mind.

2.1.2. Q1: Results

Lowest and highest predictor variables

All of the 35 scatter plots with their respective linear regression lines, *i.e.*, one for each variable against which CO_2 per capita was plotted, can be found in Appendix 2 (Figures A2-1 through A2-5, with subplots for each of the five sectors), for sake of completeness. R^2-values varied widely, from essentially 0 (meaning, no correlation whatsoever) for agricultural land use, to **R^2 = 0.71 for electricity generation from fossil fuels**, *i.e.*, coal, oil and gas (Fig. 1), which was also the highest correlation found among all variables. This translates to a quite high Pearson Correlation Coefficient R = 0.84.

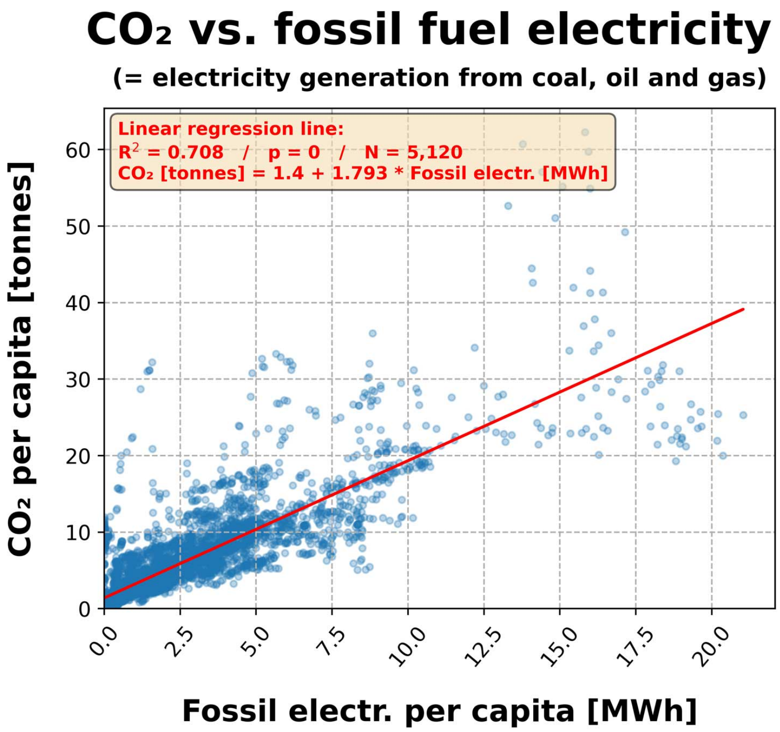


Figure 1. *Scatter plot of per capita CO_2 emissions (in tonnes) as a function of per capita electricity generation from fossil fuels (i.e., coal, oil and gas, in Megawatt-hour), which turned out to be the best predictor of a country's emissions, at least in this analysis. Each data point represents one CO_2 vs. fossil electricity measurement combination, for a single year in a single country (with multiple annual data per country, from the period 1985-2021 and 207 countries). The red regression line shows a good linear fit through the data, with R^2=0.71.*

High statistical significance for most variables

The linear regression shown in Fig. 1 had a statistical significance $p << 10^{-99}$, which was rounded to 0 in the chart. In other words, the chance that there is NO correlation between a country's fossil electricity generation and its CO_2 emissions is almost infinitesimally small. This was actually the case for most analyzed variables, even for those (surprisingly enough) with very low R^2-values, such as per capita nuclear electricity generation (in MWh), with R^2 = 0.016

(meaning that the Pearson R = 0.13), and $p < 10^{-25}$. The interpretation is that, while the linear cross-correlation of this variable with CO_2 was very small, this small correlation in itself is still highly significant.

Only five of the 35 variables did not show any statistical significance, *i.e.*, had no linear cross-correlations with CO_2 emissions whatsoever: the % energy consumption from coal (p=0.18), % diesel of total new car registrations, number of electric car registrations, % electric of total car registrations, and hectares of agricultural land use (the last four variables with p between 0.30 and 0.50). Furthermore, renewable energy consumption (MWh) showed p=0.06, *i.e.*, not statistically significant (for which the threshold is considered to be p<0.05), but almost so.

All significant correlations were positive

Furthermore, out of the 35 analyzed variables, all 30 with (highly) significant p-values for their linear regression lines also had positive slopes of those lines, meaning that increasing the value of those variables in a country also raises that country's CO_2 output. In other words, they are indeed (linear) predictors of a large per capita CO_2 output of a country, just as question Q1 specified.

The remaining five variables that are mentioned above, those with large, non-significant p-values, either had a flat linear regression line (*i.e.*, with a slope close to 0, for two variables) or a *negative* slope (for three variables) – meaning, they could theoretically be predictors of a *small* instead of a large CO_2 output. But because of the non-significant p-values – as well as the fact that all five had an R^2-value of close to 0 ($R^2 < 0.01$) –, those regression lines are essentially meaningless, *i.e.*, there was no linear cross-correlation between any of those variables and CO_2 output.

Low correlations for "% share of total" variables

As a sidenote: please note that, somewhat surprisingly, the high correlation for fossil electricity generation concerned only its *absolute* value, expressed as Megawatt-hour (MWh) per capita, and not its *relative* share of the total electricity generation in a country, expressed as %, with a low R^2 of only 0.05 (shown in Appendix 2, in Fig. A2-2). In general, very low to low R^2-values were also found for other "% share of total" variables.

These are: % coal of energy consumption (R^2 = 0.0), % fossil consumption (R^2 = 0.03), % coal of electricity generation (R^2 = 0.04), % diesel of new car registrations (R^2 = 0.0), % fossil car (*i.e.*, diesel and petrol) registrations (R^2 = 0.02; while on the other hand, the *absolute* number of new fossil car registrations showed a high correlation of R^2 = 0.62!), % electric car registrations (R^2 = 0.0), % of the population living in cities (R^2 = 0.07) and % protein in diet (R^2 = 0.17). This may indicate that variables expressed as relative shares of the total are perhaps not the best predictors for CO_2 emissions, which are themselves measured in absolute values.

Top 14 CO_2 predictors

Figure 2 shows a horizontal bar chart for the top 14 predictors of per-capita CO_2 emissions that were found in this analysis, ordered by their R^2-values and with their overarching sectors indicated by color. It can be seen (perhaps as expected) that variables in the **Energy** sector were the highest predictors of a country's CO_2 output, with the **top predictor – fossil electricity generation** – followed by the **overall primary energy consumption (nr. 2)**, the **primary energy consumption from fossil fuels (nr. 3)**, and the overall electricity demand and generation in positions 9 and 10.

Variables from the **Economy & Prosperity** sector were also major predictors, with GNI (nr. 4), GDP (nr. 8) and HDI (nr. 12), as well as (in the **Transportation** sector) the annual number of new car registrations, with fossil fuel cars, all cars, and diesel cars in positions 5, 6, and 7, respectively. Finally, in the **Food & Agriculture** sector, the daily animal protein intake, overall protein intake and total daily caloric supply conclude the list in positions 11, 13, and 14, respectively. And while the R^2 of 0.28 for daily caloric supply may seem quite low at first sight, this actually translates to a Pearson Correlation Coefficient R = 0.53. A lower threshold of R = 0.5 was in fact used to come up with the list of the top 14 predictors.

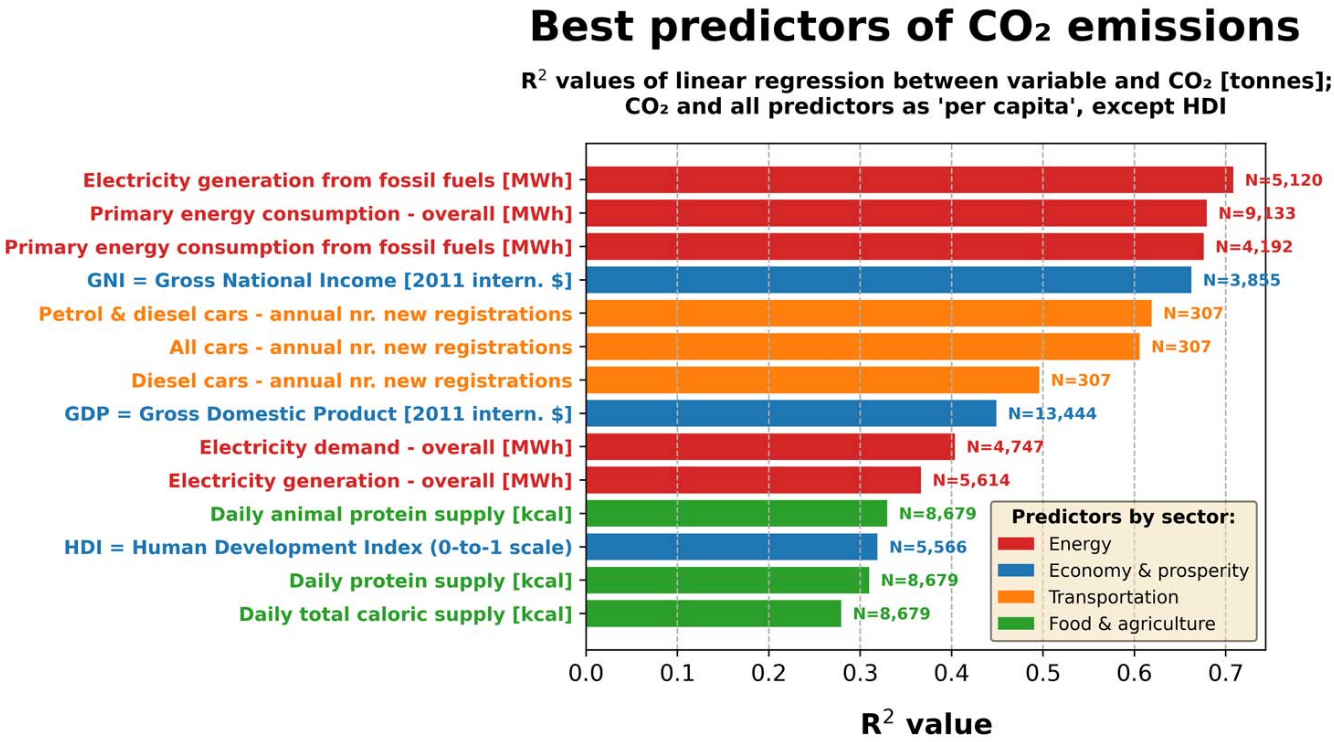


Figure 2. *Horizontal bar chart of the Coefficients of Determination (R^2-values, showing strength of the Pearson cross-correlation on a 0-to-1 scale) for the linear regression lines between each of the analyzed variables and CO_2 emissions (i.e., one regression plot per variable). The top 14 best correlating variables (out of 35 analyzed) are shown, with each variable color coded by its assigned sector. All variables, except for HDI, were plotted as their 'per capita' values, where the annual data of most countries worldwide (where available) were pooled in each single regression plot. Sample size N of each dataset is indicated at the end of each bar.*

Correlation is not causation...

Before ending this section that answers Q1, we should once again remind ourselves of perhaps one of these most quoted (but also often forgotten) adages and foundational pillars of the scientific method, and of data science and statistics in particular: *i.e.*, that correlation between two variables does not necessarily mean that one of the variables *caused* the other one. That is something especially important to remember regarding this analysis, which specifically aimed to find variables *predictive* of a large CO_2 per capita output based on their linear *correlation* with it.

"To predict" does not mean the same as "to cause", as it could of course very well be that both the predictor variable and the CO_2 per capita output are caused by yet a third (set of) variable(s). HDI for example, which is an indicator of how "developed" a country is, is unlikely to directly *contribute* a lot to CO_2 emissions; but we know that the burning of fossil fuels, such as in combustion engines of new cars or in coal power plants for electricity generation, *does* directly emit CO_2. So some of the measured variables are (besides predictors) direct causative variables as well.

In summary…

While acknowledging all the caveats mentioned in section 2.1.1 with regard to how these results should be interpreted – especially the fact that I investigated only a limited number of variables (and no aviation or shipping variables), only looked at their linear regressions with CO_2 emissions (and not, *e.g.*, at higher degree polynomials fits), and that there might be country, regional or period biases in the data –, at least we can conclude that within the limited scope and context of this somewhat simplified analysis, the answer to Question 1 ("What is the biggest predictor of a large CO_2 output per capita of a country?") is: the **per capita electricity generation from fossil fuels [MWh]**.

2.2. Q2: Countries making the biggest strides in reducing CO_2 output

2.2.1. Q2: Method and dataset

Dataset and variables used

First of all, as the combined and edited dataset that I used to answer Q1 already contained the annual CO_2 emissions data for all countries worldwide, I used that same dataset again – and within the same Jupyter Notebook – to answer Q2 as well. Since only the CO_2 data were needed to answer Q2, and none of the 35 other variables that I analyzed for Q1, I did make a smaller subset of the dataset. I also organized it differently for easier handling – with four variables, *i.e.*, a country's annual "CO_2 per capita" emissions (in tonnes of CO_2), a country's annual *overall* CO_2 emissions (in Megatonnes, or Mt), its annual population size, as well as the annual CO_2 growth percentages, displayed in four rows stacked on top of each other per country, with the annual data for each variable in 58 columns, from 1964 to 2021. (A rationale as to why this specific year range was chosen will follow shortly below.)

Similar to the approach for Q1, it makes most sense for Q2 as well to compare CO_2 reduction performance among countries in terms of their "CO_2 per capita" emissions, instead of their overall annual emissions. As already explained in section 2.1.1, this procedure "normalizes" a country's CO_2 output – and therefore also its output reduction efforts – with respect to its population size, which makes it possible to meaningfully compare the CO_2 reduction of large countries with those of small(er) countries. So I ended up using only the "CO_2 per capita" emissions to answer Q2, but kept the other three variables in the dataset as well for reference: those could provide additional context down the line if needed, *e.g.* in case of any unusual findings, in the form of *post-hoc* analyses and/or exploring alternative approaches.

Different approaches may yield very different answers...

With these annual CO_2 data in hand, there are multiple ways to address Q2, depending on how you define "reducing CO_2 output" – and it especially depends on whether you look at the *absolute* amounts of reduced CO_2, expressed as tonnes per capita of CO_2 reduction; or at the *relative* amounts of reduction, *i.e.*, the CO_2 output in a certain year expressed as a percentage of the output in an earlier baseline year. In addition, the period over which the CO_2 reduction is considered will have a major impact on the answer as well: it may very well be that countries that are the top reducers over a longer-term period (say, 50 to 100 years) are not doing so well over the short term, say in the last 10 to 15 years – or vice versa.

Another important factor to consider is that countries may differ significantly in the number of years for which annual CO_2 emissions data and/or population data are available, where data for older years may be missing. So when a longer-term period is considered (where we always take the start year as the reference year to calculate the reductions from), we could potentially be excluding many countries from the analysis, thus biasing the results. We therefore need to carefully consider our approach, as in this day and age we of course want to be as "inclusive"[4] as possible; and for the most complete and balanced picture, we would need to compare the results from multiple approaches with each other.

Pre-analysis to maximize the number of countries

To tackle the last issue first, I did a quick preliminary analysis, where for each year I plotted the number of countries[5] that had annual CO_2 per capita values available for that year. The earliest year in the entire dataset with a CO_2 per capita data point was 1750, but only for a single country: the UK. From there on, the number of countries steadily grew to 104 countries that had annual CO_2 data available for the year 1949, but then the number suddenly jumped to 167 countries in 1950. After another gradual increase, the number briefly stabilized at 197 countries in 1964 for a few years, saw another jump to 213 countries in 1990, and finally stabilized at the final number of 219 countries with CO_2 per capita data from 2008 onward. The two plots showing this can be found in Appendix 2 (Fig. A2-6).

Reference periods over which to calculate CO_2 reductions

Based on these numbers, I decided to take four reference years for the most balanced approach and to maximize the number of countries included:

- ***1964*** – the year where the number briefly stabilized at 197 countries, which is 90% of all 219 countries in the dataset;
- ***1990*** – with 213 countries, or ~97% of all countries;
- ***2008*** – the start year onward from which CO_2 data were available for all 219 countries;
- ***2021*** – the final year in the dataset, and therefore also the year to focus on with respect to answering which countries are (at least as of most recently) the biggest CO_2 reducers.

This led to three periods over which to calculate CO_2 reductions, *i.e.*, **1964-2021** which yields a view on the relatively ***long-term CO_2 reduction*** performance (57-year period); **1990-2021** which shows the ***mid-term reduction*** (31 years); and **2008-2021** for a relatively ***short-term reduction*** view (13 years).

CO_2 reductions both in absolute and relative terms

Finally, as noted in the section above, in order to get the most balanced view on which are the best performing CO_2 reducers, for each period I also compared the top countries with respect to their *absolute* CO_2 reduction (in tonnes of CO_2 reduced per capita) with the list of top countries in terms of reducing their *relative* % of CO_2 emissions. So for each of the three reference periods, to get the absolute reduction (or $\Delta[CO_2$ per capita$]$), I subtracted a country's CO_2 per capita value in 2021 from its value in the reference period start year (either 1964, 1990, or 2008). Likewise, for the relative reduction expressed as %, I divided the 2021 value by the value in the reference period start year – where that start year's CO_2 per capita value served as the 100% baseline (which will be different for most countries).

---

[4] For optimal effect, please imagine this being said with the voice of Mr. Burns from "The Simpsons".

[5] Please refer to footnote 2 again for what OWID defines as countries and note that these include overseas territories that are not technically countries; so these territories could also end up in our list of best CO_2 reducing countries.

As an aside, the assignment text for Q2 recommended using the relative CO_2 outputs to calculate CO_2 reductions, but in a sense the absolute reduction as defined above is also somewhat of a relative measure, as it shows the relative *difference* in CO_2 output between two years. In addition, CO_2 outputs expressed as per capita values (instead of the overall, total emissions) could themselves even be considered a relative measure – that is, relative to a country's population –, as they allow valid comparisons among countries.[6] Regardless, using both measures as defined above will yield the most comprehensive and balanced picture of the top CO_2 reducing countries, from several different perspectives.

[6] For an even broader view on this, it could be argued that the distinction between what is "absolute" and "relative" is an age-old philosophical and metaphysical dilemma that some of the brightest minds in history – such as Albert Einstein – spent a significant part of their lives pondering about.

2.2.2. Q2: Results and discussion

Top 7 countries for each method of measuring CO_2 reduction

Figure 3 shows a summary of this analysis, with each of the six subplots representing a different way of measuring CO_2 reduction, as outlined above. When interpreting these, please note that because of the logarithmic vertical axes, a decreasing line in the top part of each chart represents a much greater CO_2 reduction than a line with a similar decreasing slope in the bottom part – but generally, the steeper a line decreases, the larger that country's CO_2 reduction is between those two reference years. This especially helps to explain the order of countries in the legends of the three subplots on the right, *i.e.*, the ones showing the CO_2 outputs as *absolute* values: those countries with the steepest decreasing lines between two years are generally the greatest CO_2 reducers during that timeframe.

The subplots on the left on the other hand show the *relative* decreases in CO_2 output, expressed as % of the baseline year value: for those plots, the order of the top 7 countries in each legend simply represents the order of smallest percentage values in 2021. Also note that the single horizontal "100% baseline" in those subplots (equalized for all seven countries) actually represents a different *absolute* CO_2 per capita value for each of those countries. That is also the reason why this complementary view of of both relative and absolute CO_2 reductions is so useful. And, for the very alert reader: you now finally understand why 42 is the Answer to Life, the Universe and Everything.[7]

As a sidenote: a parallel and more simplified view of the six CO_2 reduction measures can be found as horizontal bar charts in Appendix 2, for sake of completeness (Fig. A2-7, a through f). These show the reductions in CO_2 only as measured between the two reference years of each period, and not the other years that are visible Fig. 3, which makes it a bit easier visually to compare the relative magnitudes of CO_2 reduction among the different top reducing countries.

[7] As it turns out, 6x7 is the Ultimate Question indeed…

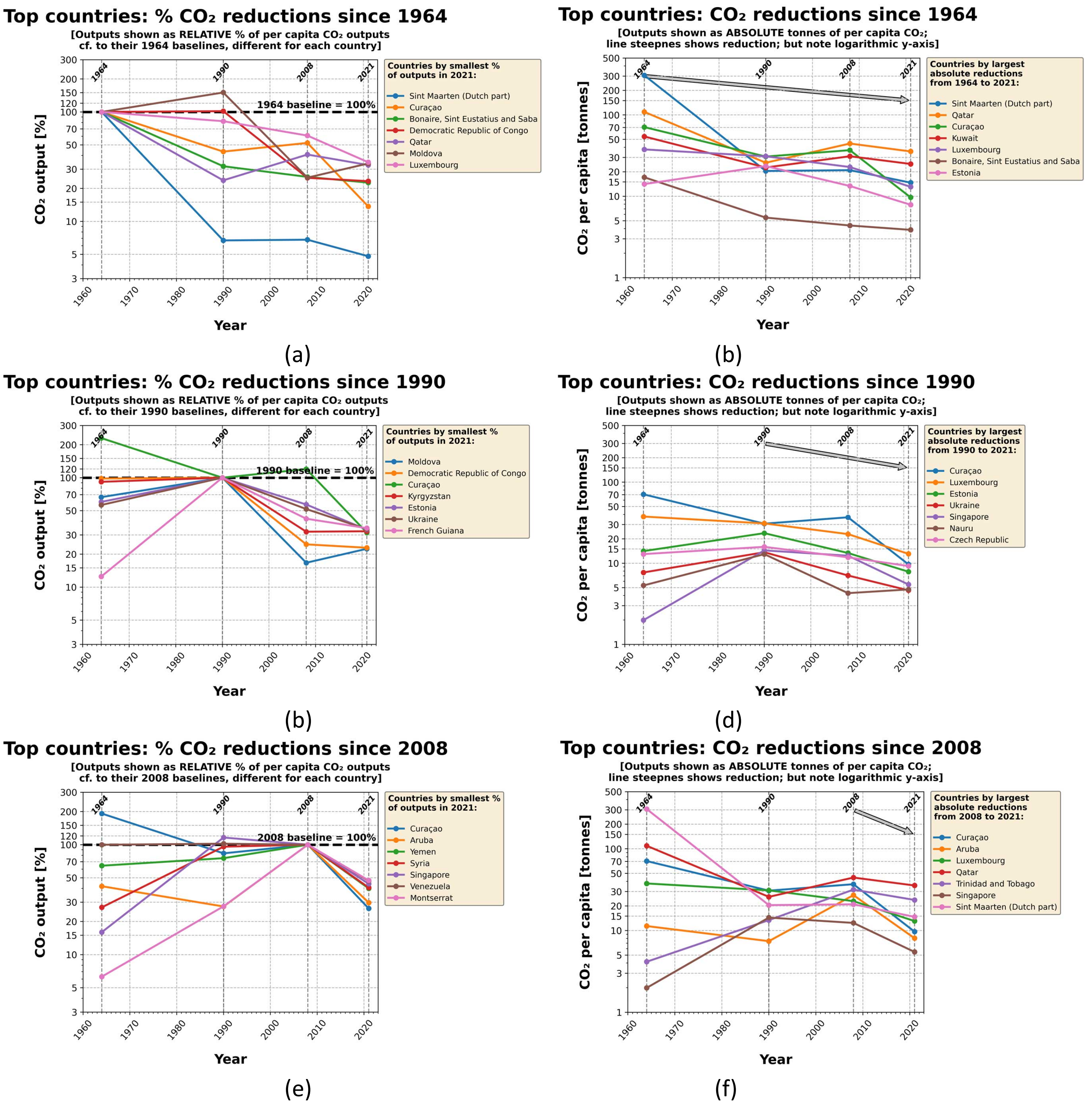


Figure 3. *Time series line charts showing the top 7 countries in reducing their CO_2 per capita emissions, with a different top 7 list for each of six different ways of measuring CO_2 reductions. "Long-term" reductions (1964-2021) are shown in Figs. 3a and 3b, "mid-term" reductions (1990-2021) in Figs. 3c and 3d, and "short-term" reductions (2008-2021) in Figs. 3e and 3f. In addition, for each period the relative CO_2 outputs (as % of the start year baseline values) are shown on the left (Figs. 3a, c, and e), and absolute outputs (in tonnes of CO_2 per capita) on the right (Figs. 3b, d, and f). Please note the logarithmic y-axis in all charts, so that lines with the same steepness may represent different changes, depending on how "high" they are in the chart.*

What about all those Netherlands Antillean island nations?

From a quick glance across the top 7 CO_2 reducing country lists across all six measures in Fig. 3, the first thing that stands out is the disproportionate representation of small countries among them (*e.g.*, Singapore, Luxembourg, Estonia, Kuwait, Moldova); and more specifically, island nations: Sint Maarten (Dutch part), Curaçao, Aruba; Bonaire, Sint Eustatius and Saba (these last three islands together make up one "country"[8], and for clarity I will enclose these with '{ }' brackets from now on), Montserrat, Trinidad and Tobago (these two islands are also a single country), and Nauru.

All of these island nations (except for Nauru in the Pacific Ocean) are in the Caribbean, and the first four of them are part of the Netherlands Antilles: *i.e.*, they were formerly part of the Netherlands[9], and gained their status as independent nations relatively recently. While my home country the Netherlands is itself already one of the smallest countries (at least in terms of land area[10]), these former Dutch colonies are tiny specks by comparison. So the question is whether there is any common thread or theme that connects them and that could explain their world-class performance in terms of CO_2 reduction?

Possible explanations for these small (island) nations as top reducers

While it's outside of the scope of the analysis for Q2 to explore possible reasons for this common thread in any great depth – as Figure 3 by itself pretty much already answers Q2 –, I will dedicate these next few sections to some ideas about factors that *may* explain this observation (or at least partially). An obvious one is that it is potentially much easier for a smaller country to change course, *i.e.*, to reduce their CO_2 output, than for (very) large countries – in terms of adapting their industry and energy infrastructure, *etc.* – similar to how a small speedboat can change direction much more quickly than a large oil tanker.

Another possible explanation is that, because I only looked at the per capita CO_2 outputs, these small countries and island nations may have seen large growths in their population sizes during the reference periods (*i.e.*, the denominator in the "per capita output" calculation), which in turn might have quickly reduced the per capita outputs – even when the countries' *overall* outputs (in Megatonnes of CO_2) may have remained relatively stable or decreased much less dramatically during those periods. And it is of course more likely for small countries that already have small population sizes to begin with, to quickly grow their populations percentagewise than it is for large countries.

---

[8] Actually, technically {Bonaire, Sint Eustatius and Saba} together consist of three Dutch municipalities, see also footnote 2.

[9] Again, except for {Bonaire, Sint Eustatius and Saba}, which still is officially part of the Netherlands.

[10] …but actually quite large in terms of famous painters, speed skating, world-class football players, and participation rate in past World Cup football finals.

Scaled-down oil refineries

Why then do the Netherlands Antillean island nations perform so well in particular? One obvious and likely factor that comes to mind is the presence of the "Isla" oil refinery on Curaçao, constructed by Royal Dutch Shell in 1915 and at one time the largest refinery in the world. In 1985, Shell sold the refinery to Curaçao's government, who in turn leased it to the Venezuelan oil company PdVSA. That contract ended in 2019 and in the years before that, production had already been minimized.[11] This scaling down is probably the biggest contributor to Curaçao topping the lists of best short-term (2008-2021) CO_2 reducers (Figs. 3e and f) – although this might change again, as a restart of the refinery is still being considered.

And regarding Curaçao's impressive CO_2 reductions in the mid- and long-term, prior to that: more efficient and cleaner refinery technologies that likely have been developed over the years have probably played a role in that as well.

Similar to Curaçao, Aruba had the "Lago" oil refinery, but this was scaled down significantly after 1985 as well, and closed in 2009 (and, after a brief restart, closed again in 2020). And indeed there seems to be a relation: Aruba is the second best short-term CO_2 reducer in the world, after Curaçao (Figs. 3e and f).

As for {Bonaire, Sint Eustatius and Saba}, the relation to its long-term CO_2 reduction performance is less clear: Bonaire did have an oil terminal since 1974, owned by Bonaire Petroleum Corporation (BOPEC) which declared bankruptcy in 2021 – but it's unclear to me whether or how simple storage of oil (unlike the active refining of oil, which does produce emissions) factors in to the methodology of how CO_2 emissions are attributed to countries.

The "fast-growing population" explanation

Finally, Sint Maarten (Dutch part) did not take part in any oil industry activities. It turns out, however, and indeed as I already speculated above, that here a fast-growing population is the culprit, with an 11-fold increase from 1964 to 2021 – accompanied by a 20-fold decrease in the CO_2 per capita emissions during that same timeframe (as already shown in Fig. 3a).

To illustrate these different underlying mechanisms of the reductions in CO_2 per capita for Sint Maarten and the other Netherlands Antillean islands, I plotted the population size, the *overall* CO_2 emissions and the CO_2 per capita emissions over time (all as relative % of their 1964 baseline values and plotted together in a single chart), for all four islands. These are shown in Fig. A2-8 in Appendix 2, so that they don't distract too much from the main analysis here.

Indeed it's very clear that for Sint Maarten (Fig. A2-8a), the population and CO_2 per capita lines strongly diverge after 1964, while the line with its overall CO_2 output decreases relatively mildly (to ~50%) and stays pretty much in the middle of the log scale. The same pattern, albeit to a lesser extent, is seen for {Bonaire, Sint Eustatius and Saba} – but for both Curaçao and Aruba it's obvious that the CO_2 per capita line closely follows the line with overall CO_2 output, with only a modest population growth during this timeframe.

[11] See https://mena.nl/artikel/verkoop-raffinaderij-curacao-mislukt-alweer.

In summary…

My answer to Question 2 ("Which countries are making the biggest strides in decreasing CO_2 output?") is:

- ***Long-term (1964-2021)***
  Sint Maarten (Dutch part) had the **best long-term per capita CO_2 reduction**, both the reduction measured in relative % and in absolute tonnes of CO_2. Other countries such as **Curaçao** (nr. 2 and 3, respectively)**, Qatar** (nr. 5 and 2, respectively), **{Bonaire, Sint Eustatius and Saba}, and Luxembourg** also did well regarding their long-term CO_2 reduction and all showed up in the top 7 lists of both the relative and absolute reduction measures.

- ***Mid-term (1990-2021)***
  Curaçao had the **best mid-term per capita CO_2 reduction** measured in absolute tonnes of CO_2, and the third best reduction when measured as relative %. **Moldova and Democratic Republic of Congo** topped the list as numbers 1 and 2 of mid-term relative CO_2 reduction in %, but were not in the top 7 of absolute reducers (and both were also in the top 7 list of *long-term* relative reducers). Furthermore, **Luxembourg** was the second best mid-term reducer in absolute tonnes of CO_2 (but not in the top 7 of mid-term relative reducers), and **Estonia and Ukraine** showed up in the top 7 lists of both mid-term measures.

- ***Short-term (2008-2021)***
  Last but not least, **Curaçao also had the best short-term per capita CO_2 reduction**, both in relative and absolute terms. **Aruba** was the **second best short-term reducer** in both measures as well, and **Singapore** was also in the top 7 lists of both measures (as well as in the top 7 of *mid-term* absolute reducers). **Luxembourg** and **Sint Maarten (Dutch part)** were numbers 3 and 7, respectively, in the list of short-term *absolute* reducers, but did not show up in the top 7 list of short-term *relative* reducers.

Overall per capita CO_2 reduction

Overall, measured across the board as long-term, mid-term and short-term per capita CO_2 reduction performance, and expressed both in relative and absolute terms, we can say that **Curaçao is one of the countries that is making the biggest "strides" in decreasing CO_2 output**. And with "strides" in quotation marks, as this reduction is likely just a side effect of the scaled-down, and in 2019 finally closed oil refinery "Isla" on the island, rather than active CO_2 reduction policies.

Other countries that are doing well across the board are **Sint Maarten (Dutch part) and Luxembourg**, as well as (to a lesser degree, or with a focus on specific reference periods) **Aruba, Democratic Republic of Congo, Moldova, {Bonaire, Sint Eustatius and Saba}[12], Singapore, Qatar, Estonia, and Ukraine**, more or less in this order.

[12] Again, technically this is not a country, please refer to footnote 2.

2.3. Q3: Best future price for non-fossil fuel energy

2.3.1. Q3: Method and dataset

Dataset

The single CSV dataset that I used for this analysis can be found at https://ourworldindata.org/grapher/levelized-cost-of-energy, which in turn has the aggregated data for seven non-fossil fuel energy technologies, obtained from the "International Renewable Energy Agency" (IRENA) 2022 report[13]. The dataset has the so-called annual "Levelized Cost Of Energy" (LCOE) data for these seven technologies, for a number of individual countries as well as aggregated data for the world.

As an aside, please note that the IRENA report itself refers to LCOE as "Levelised Cost Of *Electricity*", but OWID likely borrowed the "Energy" term from research firm Lazard, which also publishes reports on LCOE[14].

Levelized Cost Of Energy (LCOE)

The levelized cost of energy (LCOE) estimates the average cost per unit of energy generated across the lifetime of a new power plant for the technology in question, measured in so-called "constant 2021 US$" (so adjusted for inflation) per kilowatt-hour (kWh). To obtain the LCOE values expressed in US$ per Megawatt-hour (MWh) – *e.g.*, to compare them with LCOE results in other reports –, one can simply multiply these by a factor of 1,000.

The LCOE incorporates most (although not all) of the costs of building and operating a power plant across its lifetime; *e.g.*, for electricity from solar cells, it would include the costs for building and maintaining the facility to produce them. These total costs are then averaged per year and this annual LCOE can of course change from one year to the next. For additional details, please see the excellent overview article by OWID founder and director Max Roser, at https://ourworldindata.org/cheap-renewables-growth#the-price-of-electricity-from-the-long-standing-sources-fossil-fuels-and-nuclear-power.

Limitations of technology classification

The IRENA/ OWID dataset contains annual LCOE data for a total of seven non-fossil fuel energy technologies (collectively also referred to as "renewables"), *i.e.*, Bioenergy, Geothermal, Hydropower, Offshore Wind (*i.e.*, from wind turbines built at sea), Onshore Wind, Solar Concentrated (*i.e.*, heat energy from sunlight concentrated by mirrors), and Solar Photovoltaic ("solar PV", *i.e.*, solar panels that convert light directly to electricity).

---

[13] The IRENA 2022 report is available from https://www.irena.org/publications/2022/Jul/Renewable-Power-Generation-Costs-in-2021.

[14] *E.g.*, Lazard's 2023 LCOE+ report (version 16.0) is available from https://www.lazard.com/research-insights/2023-levelized-cost-of-energyplus/.

Please keep in mind that this rather coarse classification excludes some smaller-scale renewables such as tidal energy, and that quite a few other LCOE reports use a more fine-grained classification, especially for Solar Photovoltaic; *e.g.*, the Lazard 2023 report distinguishes between “Rooftop Residential”, “Rooftop Commercial & Industrial”, “Community”, “Crystalline Utility Scale” and “Thin Film Utility Scale” solar PV cells, with a wide range of LCOE values between them. So the results from this analysis may present a somewhat limited view.

Only World data

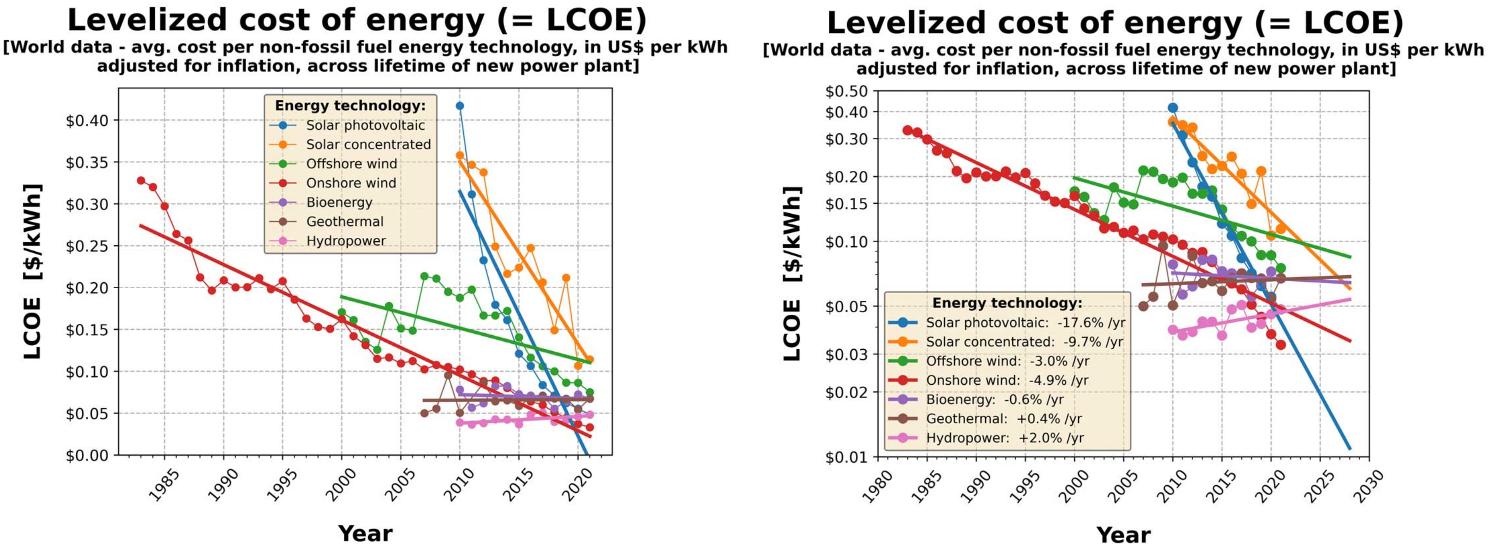
In addition, the dataset has annual LCOE data for a total of 20 countries, including the US, China, India, and quite a few European countries, but each country only has data for either Solar Photovoltaic, Onshore Wind, or both – but not for any of the other five renewables. Therefore I decided to only use World data, which includes LCOE data for all seven technologies, albeit for a different time range for each technology.

2.3.2. Q3: Results

Linear regressions

Similar to the charts made for Q1, and using the same Python helper function, linear regressions were done for the plots of annual LCOE data over time, with regression lines fitted through the LCOE data points for each technology. All seven technologies were plotted in a single chart for easy comparison (Figure 4).

The first plot in Fig. 4a shows the LCOE data on a regular y-axis, while the second plot in Fig. 4b shows them on 10-base logarithmic y-axis (a so-called log-linear plot). As is visually apparent, the regression lines show much better fits through the data for the log-linear plot, which makes sense if one considers that energy prices will not drop to zero (as the regression line in Fig. 4a suggests would have already happened for Solar Photovoltaic in 2021), but instead – if they decrease – tend to decay exponentially. Exponential price decay would show itself as a straight line in a log-linear regression, and indeed that's what we see, especially for Solar Photovoltaic and Onshore Wind.



(a) (b)

Figure 4. *Scatter plot of the annual Levelized Cost Of Energy (LCOE) data (aggregated for the world) for seven renewable energy technologies, with linear regression lines through the data. Figure 4a shows the LCOE values plotted on a regular y-axis, with ordinary linear regression lines showing reasonable fits through the data. Figure 4b shows the same data plotted on a 10-base logarithmic y-axis, with log-linear regression lines (based on the the log-transformed LCOE data) showing much better fits. The time scale on the x-axis has been extended in Fig. 4b in order to make forecasts for future LCOE prices. All dollar amounts are expressed as 2021 US$, to account for the effects of inflation.*

Interpretation of log-linear regressions

In order to accomplish these straight regression lines in a log-linear plot, the LCOE data first needed to be log-transformed themselves in Fig. 4b. How then to interpret the regression line parameters in a log-linear plot, also in order to be able to forecast future LCOE values, is described in various excellent online resources[15].

In short: using the slope "m" and intercept "b" of the linear regression line to first construct the line equation *y = m*x + b* (with x being the year of the LCOE prediction), we then need to raise 10 to the power of this line equation to get the expected LCOE (in US$ /kWh) in a certain year. The reason for this is that the log was 10-base, and 10^x is its reverse operation.

Likewise, to get the annual price decrease (or increase) expressed as a percentage for each technology (which is all we need to answer Q3), we need to raise 10 to the power of slope "m" – which gives the *relative* decrease or increase in LCOE from one year to the next (with 1.0 meaning "no change") – and then subtract 1 and multiply by 100 to transform it into a %-value, with negative percentages indicating price decreases (which we would like to see) and positive values price increases.

LCOE forecasts

Using these log-linear line equations, I made two helper functions: one to predict the LCOE of a certain energy technology in any given year, and one to predict the year in which an energy technology would reach any given LCOE value, with a nice milestone being the year in which LCOE is expected to reach 1 cent per kWh. This predicted year, as well as LCOE forecasts for 2023, 2030 and 2050, can be seen in Table 1 for each renewable.

[15] Two excellent resources on how to interpret the parameters of log-linear regression lines are: https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf and https://medium.com/swlh/log-transformations-in-linear-regression-the-basics-95bc79c1ad35.

Table 1. *Pearson cross-correlation statistics for the log-linear regression lines through the annual LCOE data, for the seven renewable energy technologies plotted in Figure 4b. The font color of each renewable energy in the index column corresponds to its plotted color in Figure 4, for easy cross-reference. Colors in the columns "R^2", "p-value" and "Annual Δ LCOE" carry a different meaning, however, related to their value (please see the main text for details). Columns on the right show LCOE price forecasts for three future years, based on the log-linear regression equations, as well as the year when LCOE is forecasted to reach 1 US$ cent/ kilowatt-hour. All dollar amounts are expressed as 2021 US$.*

	R^2	p-value	N	Log-linear regression equation	Annual Δ LCOE	LCOE in 2023	LCOE in 2030	LCOE in 2050	Year when LCOE = $0.01 /kWh
Solar photovoltaic	R^2=0.98	p=3.40e-10	N=12	LCOE [$/kWh] = $10^{(168.03 - 0.084 * Year)}$	-17.6% /yr	$0.03/kWh	$0.007/kWh	$1.6e-04/kWh	2028
Solar concentrated	R^2=0.84	p=3.04e-05	N=12	LCOE [$/kWh] = $10^{(88.22 - 0.044 * Year)}$	-9.7% /yr	$0.10/kWh	$0.05/kWh	$0.006/kWh	2045
Onshore wind	R^2=0.95	p=2.39e-25	N=39	LCOE [$/kWh] = $10^{(42.67 - 0.022 * Year)}$	-4.9% /yr	$0.04/kWh	$0.03/kWh	$0.01/kWh	2052
Offshore wind	R^2=0.41	p=0.0013	N=22	LCOE [$/kWh] = $10^{(25.48 - 0.013 * Year)}$	-3.0% /yr	$0.10/kWh	$0.08/kWh	$0.04/kWh	2098
Bioenergy	R^2=0.02	p=0.6246	N=12	LCOE [$/kWh] = $10^{(3.87 - 2.50e-03 * Year)}$	-0.6% /yr	$0.07/kWh	$0.06/kWh	$0.06/kWh	2351
Geothermal	R^2=0.01	p=0.7246	N=14	LCOE [$/kWh] = $10^{(-4.83 + 1.81e-03 * Year)}$	+0.4% /yr	$0.07/kWh	$0.07/kWh	$0.08/kWh	N/A
Hydropower	R^2=0.38	p=0.0326	N=12	LCOE [$/kWh] = $10^{(-18.3 + 0.008 * Year)}$	+2.0% /yr	$0.05/kWh	$0.06/kWh	$0.08/kWh	N/A

Table 1, of which the layout and formatting was done entirely within the Jupyter Notebook by using the pandas "styler" functions, also shows the Pearson correlation results for each linear regression line: the R^2-value, statistical significance "p" and sample size N; please refer back to the relevant part in section 2.1.1 for an explanation on how to interpret these parameters. In Table 1, the R^2-values that indicate a very high log-linear correlation, *i.e.*, $R^2 \geq 0.8$, are emphasized in bold green, while those showing hardly any correlation ($R^2 \leq 0.2$) are in bold red. Likewise, p-values that are generally considered as highly statistically significant ($p \leq 0.01$) are in bold green, while those that are not statistically significant ($p \geq 0.05$, which means there's a greater than 5% chance that the correlation is less than indicated or even absent) are in bold red.

This helps us to interpret the LCOE forecasts in Table 1: *i.e.*, we know to take the forecasts for Bioenergy and Geothermal (with very low R^2- and very high p-values) with a grain of salt; we also see in Fig. 4 that their lines are essentially flat or slightly increasing over time. The regression lines for Solar Photovoltaic, Solar Concentrated, and Onshore Wind on the other hand are highly significant, also with very high correlations between LCOE-values and time – so we can safely assume that those LCOE forecasts will likely be quite accurate. For Offshore Wind and Hydropower (the latter with a slightly *increasing* LCOE over time), the correlations are medium-sized.

In addition to the regression results on the left, Table 1 also contains the equations of the log-linear regression lines, which express LCOE as a function of the input year. The annual changes in LCOE (indicated in a background color gradient from green for quickly decreasing, to red for increasing LCOEs), as well as the LCOE forecasts calculated from these equations, are shown on the right side of the Table.

And the winner is...

From both Figure 4 and Table 1, it's very clear that **Solar Photovoltaic** displays the quickest annual drop in LCOE-values of all renewables, with **-17.6%/year** decreasing about twice as fast as **Solar Concentrated** (nr. 2 in the contest) and three to four times as fast as **Onshore Wind** (the bronze medalist). In the extrapolation, Solar Photovoltaic will already reach a LCOE price of $0.01/kWh (albeit expressed as 2021 US$) as early as 2028, and (not shown in Table 1) even a 10x lower LCOE of $0.001 in 2040. For Solar Concentrated and Onshore Wind on the other hand, the $0.01/kWh milestone is expected to be reached only by mid-century, and for Offshore Wind by the end of this century, at least at the *recent* annual rates of price declines. The reason for my emphasis on the word "recent" will become apparent below.

...Or is it?

It's clear from Figure 4b and Table 1 that the annual LCOE data for Solar PV display an almost perfect fit with its log-linear regression line, with an amazingly high R^2=0.98, thus giving a lot of confidence that the equation for this line in Table 1 is very accurate – at least until the latest data point in 2021. It came as quite a surprise, therefore, that when I wanted to compare (as a sanity check) my LCOE prediction for 2023 with the actual LCOE prices today, at the time of writing this in May 2023, I found that those LCOE prices for Solar PV – even when corrected for the high inflation since 2021 – have almost *doubled* over the two-year period from 2021 to 2023.

So this clearly broke the ~18% annual decline trend, putting into doubt all LCOE forecasts in Table 1, as well as portraying my above answer to Q3 as a potentially premature one. In other words, things are not as rosy as our forecasts suggest and we need to add quite a bit of nuance to this idealized picture, and take great caution in relying too much on these forecasts.

2.3.3. Q3: Extended discussion and literature review in Appendix 1

While it's clearly outside of the scope of the main analysis that was performed to answer Q3 – and my answer above to Q3 still stands for that as is – the surprise finding mentioned above made me curious enough to do a "mini literature review", *i.e.*, to take a look at what other reports and experts had to say about this, in search for a potential explanation. Because the conundrum we find ourselves in is that solid data from the 1982-2021 period tell us one thing, with their strong linear regressions that our extrapolated forecasts are based on, but that those findings don't actually match with what current reality tells us, in terms of real-world energy prices – which, as a case study, may point us to one of the central tenets of data science: the inherent limitations of the modeling approach.

My search has led to an extended discussion, in which I relate my findings from Figure 4 and Table 1 to those other reports and concepts, putting them in a wider context and providing caveats for their interpretation. It goes into such topics as "learning rate" and Moore's law, the starkly increased inflation since 2021, a conversion factor of 4.7 that I derived between Solar PV LCOE and solar panel prices, potential causes for the recent trend break in LCOE price declines, whether this trend break is temporary or permanent, global supply chain disruptions, and the possibility that Solar PV has shifted from a so-called “demand-constrained” to a “production-led” industry. It also discusses slight increases in LCOEs in the other renewables since 2021 as well, and reevaluates my answer to Q3 in light of these findings, summarized in a 7-point conclusion.

Because of the length of this discussion, I have moved it to the Appendix 1, so that the interested reader can take a look at it (or at preferred parts of it) there, without distracting too much from the main analysis for Q3 presented here.

Appendix 1

Extended discussion and literature review for section 2.3, "Q3: Best future price for non-fossil fuel energy"

This is a continuation of Section 2.3 in the main report, which contains my main analysis to answer Question 3: "*Which non-fossil fuel energy technology will have the best price in the future?*"

The reason that I undertook the "mini literature review" presented in this Appendix 1 is because the extrapolated forecasts for LCOE prices (= "Levelized Cost of Energy") for the year 2023, based on my analysis of 1983-2021 data for seven renewable energy technologies (summarized in Fig. 4 and Table 1 there), did not match the actual LCOE price for Solar Photovoltaic today, in May 2023. This review paints my answer to Q3 in section 2.3.2 (*i.e.*, that Solar PV will have the best future price) as a potentially premature one, provides necessary context and several caveats in interpreting the results, and urges caution in relying too much on the LCOE forecasts in Table 1.

Because of the length of this extended discussion – which did yield me some very valuable insights into the likely underlying reasons for the mismatch, which I summarized in a 7-point Conclusion –, I decided to move it to this Appendix 1, for those readers who are interested in this in-depth exploration.

Sanity check 1: similarity to OWID chart

A first quick sanity check to ensure the validity of the charts in Figure 4 is that they look the same as the plot of these data on the OWID website itself at https://ourworldindata.org/grapher/levelized-cost-of-energy?yScale=log, both its linear and log format (which you can manually change in the top left), with my Fig. 4 having the added feature of the regression lines.

Sanity check 2: how our 2023 forecasts compare to today's prices

- **The Lazard Report**
  The last year of LCOE measurements from this dataset was 2021, so another good sanity check is to see how well the LCOE forecasts compare with today's actual prices in 2023, two years later. One source of 2023 LCOE data that was published only very recently (at the time of writing this), in April 2023, is research firm **Lazard's "2023 Levelized Cost of Energy+" report** at https://www.lazard.com/research-insights/2023-levelized-cost-of-energyplus/.[16]

[16] *Lazard's 2023 Levelized Cost Of Energy+. Version 16.0.* By George Bilicic & Samuel Scroggins, Lazard, April 12, 2023.

The Lazard report (version 16.0) has LCOE data for four of our seven renewables: Solar PV (with four subcategories, as mentioned previously toward the end of section 2.3.1), Offshore Wind, Onshore Wind (with two subcategories) and Geothermal. The best overviews of the 2023 LCOE data for these are shown in two different charts: a horizontal bar chart on page 2 of their report, showing a *range* of LCOE values in 2023 for each renewable; and a line chart on their page 9, which shows the *mean* LCOE values for each renewable – except for Offshore Wind – from 2009 through 2023. I haven't requested permission to reprint these charts here, but a PDF of the report is freely available for download at the website above.

Their LCOE data are expressed in $/ MWh, so we need to reduce these by a factor of 1,000 to compare them to the LCOE 2023 forecasts in $/ kWh in Table 1. Also, the Lazard report does not mention that their LCOE-values were adjusted for inflation, so I will assume that their 2023 LCOEs are expressed in actual 2023 US$ prices, vs. constant 2021 US$ prices in Table 1. In addition, the Lazard report specifically mentions that their LCOE numbers are the unsubsidized prices, but that seems to be in line with the IRENA report[17] – the source of the OWID dataset used in this analysis, and available as PDF download at https://www.irena.org/publications/2022/Jul/Renewable-Power-Generation-Costs-in-2021 –, which states in their Methodology section on page 180 that their "*(...) analysis excludes the impact of government incentives or subsidies, (...)*".

The first thing that stands out in the Lazard numbers on their page 2 is the wide range of LCOE-values for their different Solar PV subcategories, ranging from $0.024/kWh on the low end to $0.282/kWh on the high end – so by about a factor 10 apart. As a ballpark range, however, this compares reasonably well with our aggregated 2023 LCOE forecast for Solar PV of $0.029/ kWh (rounded to $0.03/kWh in Table 1).

- **Accounting for the high inflation since 2021**
  For a more accurate comparison, we should also take into account the official **US$ annual inflation rates** since 2020/2021, derived from the "Consumer Price Index" (CPI-U, which is provided by the U.S. Department of Labor) – although these official numbers quite possibly significantly underreport the *actual* inflation rates. These official rates are: 7.0% for 2021, 6.5% for 2022, and 5.0% for 2023 to date[18], so overall we would need to add roughly 20% to our LCOE 2023 forecasts (provided as 2021 US$ in Table 1) in order to express them in current 2023 US$.

[17] IRENA (2022). *Renewable Power Generation Costs in 2021.* By Michael Taylor, Pablo Ralon, Sonia Al-Zoghoul, Matthias Jochum & Dolf Gielen, International Renewable Energy Agency, Abu Dhabi, 204 pp. ISBN 978-92-9260-452-3.

[18] See https://www.usinflationcalculator.com/inflation/current-inflation-rates/.

If we do this for Solar PV in Table 1, we get an updated LCOE forecast for 2023 of **$0.035/kWh**, which moves it closer to the median of the Lazard LCOE range. However, the *mean* LCOE in 2023 for their lowest-priced subcategory "Solar PV – Utility Scale", in the line chart on page 9 of the Lazard report, is still **$0.060/kWh**, which is almost twice as high as our 2023 forecast. That same line chart also shows that the mean LCOE in 2021 in their earlier report was, unexpectedly, actually *lower*, at $0.038/kWh. I will discuss potential reasons for that recent *increase* in LCOE prices for Solar PV in the sections below.

For the other three renewables, our 2023 LCOE estimates (and adjusted for the 20% inflation that has occurred since 2021, so expressed as 2023 US$), compare even better with the Lazard report: for Onshore Wind this is $0.053/kWh (Lazard, p. 9: $0.050), for Offshore Wind $0.118/kWh (Lazard, p. 6: $0.106), and for Geothermal $0.080/kWh (Lazard, p. 9: $0.082). All in all, this gives a lot of confidence in the LCOE forecasts in Table 1, as long as we will adjust them for the (high) inflation since 2021.

Solar PV forecasts from the 2015 Fraunhofer ISE report

Another potential benchmark to compare my LCOE forecasts with is the **2015 report** from the Fraunhofer Institute for Solar Energy Systems (ISE), called **"Current and Future Cost of Photovoltaics"**[19].

While the Solar PV LCOE forecasts in that report are from eight years ago and expressed as 2014 Euros/ kWh, we can convert them to 2021 US$/ kWh (or at least an estimate of it) by taking the 2014 average exchange rate of 1.328 $/ € [20] and then applying an aggregate US$ inflation factor of 1.118 for the period 2014-2020 [21].

This gives an inflation-adjusted range of Solar PV LCOE forecasts for the year 2050 (for Germany and Spain, on p. 54 of their report) of $0.027 – $0.065/kWh, in 2021 US$. Note that this is about *169 to 406 times higher* (!) than my 2050 LCOE forecast for Solar PV in Table 1 ($0.00016/kWh), and indeed the high end of their LCOE prediction for 2050 has already been reached today, in 2023. Even though their estimate was, in their words, a conservative one, it goes to show how quickly LCOE prices have declined since 2014.

[19] Johannes N. Meyer, Simon Philipps, Noha S. Hussein, Thomas Schlegl & Charlotte Senkpiel (2015). "*Current and Future Cost of Photovoltaics – Long-term Scenarios for Market Development, System Prices and LCOE of Utility-Scale PV Systems*". Technical Report by the Fraunhofer Institute for Solar Energy Systems (ISE), commissioned by Agora Energiewende (Daniel Fürstenwerth & Mara M. Kleiner), Berlin, Germany. Available as PDF from https://www.researchgate.net/publication/282654082_Current_and_Future_Cost_of_Photovoltaics.

[20] See https://www.exchangerates.org.uk/EUR-USD-spot-exchange-rates-history-2014.html.

[21] See https://www.usinflationcalculator.com/inflation/current-inflation-rates/.

Concept of "learning rate"

As already briefly discussed in section 2.3.2, the very goods fit of the log-linear regression lines with LCOE data over time in Figure 4b – especially for Solar PV, Solar Concentrated and Onshore Wind – indicate exponential decay of LCOE prices over time. Such exponential price decays are characteristic of a phenomenon known as "learning rate" or "Wright's law" – or, when applied to computer chips, also known as "Moore's Law", which is the observation by Intel's co-founder Gordon Moore that the number of transistors on microprocessors tends to double every two years.

The basic idea behind "learning rates" is that technologies that are increasingly mass-produced tend to fall in price as a result, because the production process itself leads to better knowledge and experience – thus, the term *learning* – of how to optimize the production. The lower prices in turn lead to increased demand for the technology and thus more incentives again for mass-production. This can lead to a self-perpetuating cycle of falling prices and increasing demand. This seems to be happening especially for Solar PV modules (often simply referred to as "solar panels") since the last few decades.

Technically, the term "learning rate" (or: "learning curve") applies to a chart of the logarithmic price of a technology plotted against the logarithmic ***cumulative production*** of that technology, which is not the same as time plotted on the x-axis, as in Figure 4b. The actual learning curve for solar panels is reprinted in Figure A1-1, showing a log-log plot of Solar PV module price (in 2019 US$/ **Watt** – so *not* the same as LCOE) against the cumulative installed Solar PV capacity worldwide (in Megawatt). The downward trend line indicates a steady, average **20.2% price decline** in Solar PV ***modules*** for every ***doubling*** of installed capacity since 1976, which is therefore its actual learning rate in the technical definition of the term.

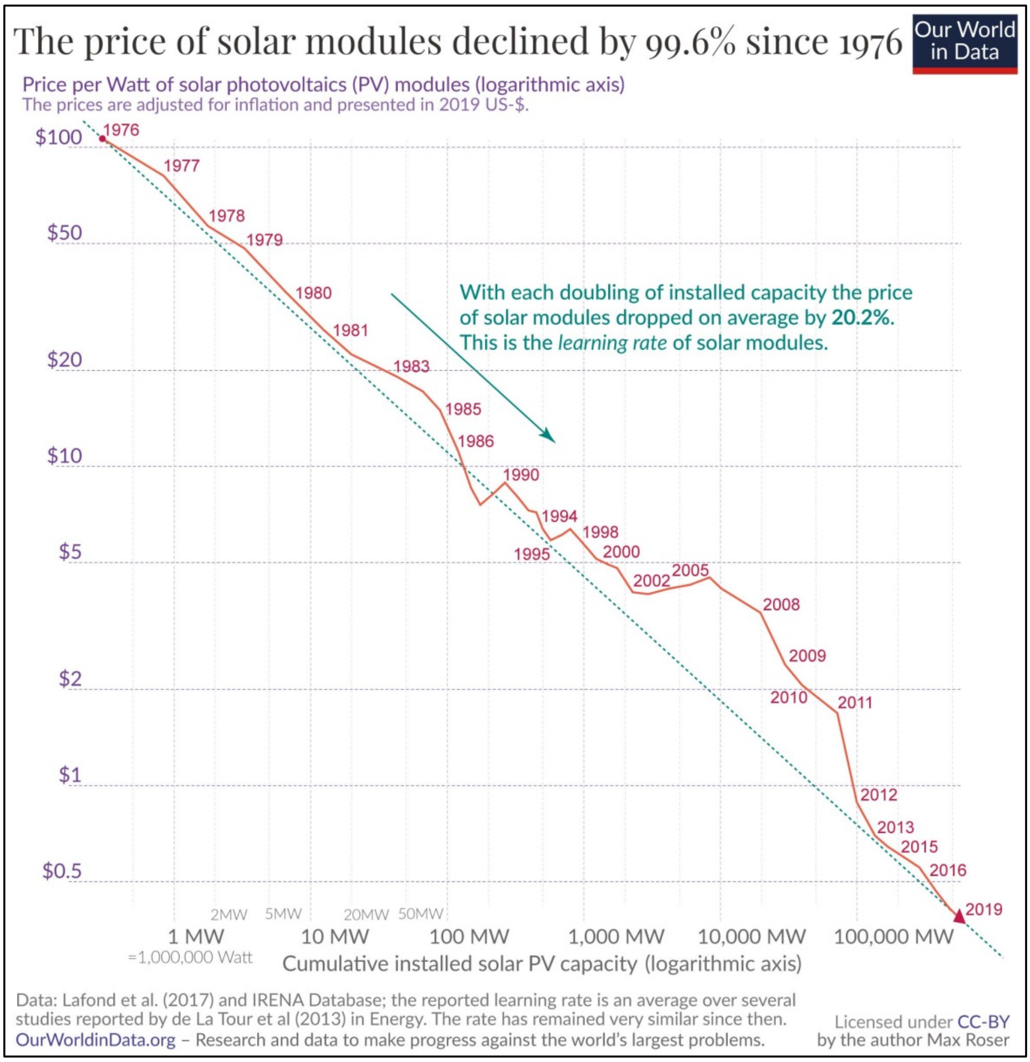


Figure A1-1. *Learning curve for Solar PV modules for the period 1976-2019. Reprinted under a CC-BY license from the excellent article on this topic by Max Roser at https://ourworldindata.org/cheap-renewables-growth#a-short-history-of-solar-from-outer-space-to-the-cheapest-source-of-energy-on-earth.*

Comparing Solar PV learning rate to its annual LCOE decline

How then does this compare to the 17.6% ***annual*** price decline that we found for Solar PV ***LCOE*** prices? First, we should note that LCOE prices are expressed in US$ per kilowatt-*hour*, vs. US$ per *Watt* for Solar PV module prices. The first expresses the ***amount*** of actual energy produced by Solar PV in a unit of time (in a production-scale industrial power plant, with the cost averaged over its entire lifetime); while the latter simply expresses the (potential) ***power*** of a Solar PV ***module*** under ideal conditions (*i.e.*, at a 90 degrees incidence angle of sunlight, in 25° C, assuming a maximum solar power of 1,000 W/m^2, *etc.*), measured in so-called "peak Watts" (often denoted as Wp instead of W)[22].

[22] For a helpful explanation of the difference between the kWh and kWp units for solar panels, please see https://academy.dualsun.com/hc/en-us/articles/360017363999-What-is-the-difference-between-energy-kWh-and-peak-power-kWp-.

Secondly, footnote 19 of Max Roser's OWID article at https://ourworldindata.org/cheap-renewables-growth#note-19 gives a clear and brief explanation as to why the (linear) passage of time and (logarithmic) cumulative installed capacity actually have the same effect, at least if the latter indeed exhibits exponential growth (as has been the case for Solar PV since 1976, as his chart in Fig. A1-1 shows). We may therefore quite safely conclude that the 17.6% annual price decline for Solar PV ***LCOE prices*** (in ***$/ kWh***), found in this analysis, likely stems from the learning rate seen for Solar PV ***module prices*** (in ***$/ Wp***). The question is how to convert between these two measures.

Another chart in Max Roser's article[23], reprinted below as Figure A1-2, actually *does* show the learning curve for LCOE prices (rather than for the module prices) of Solar PV – as well as for a number of other energy technologies. But again, it has the cumulative installed *capacity* (in Megawatt), logarithmically plotted on the x-axis, not *time* in years as in Figure 4b. The Solar PV learning rate that is apparent from his chart is **36%** – *i.e.*, a 36% decline in LCOE prices (also since 2010) for every doubling of cumulative installed Solar PV capacity. This is about twice as much as the 17.6% *annual* LCOE price decline shown in Table 1, which suggests that a doubling of cumulative installed Solar PV capacity happens roughly every two years.

---

[23] See https://ourworldindata.org/cheap-renewables-growth#do-electricity-prices-follow-learning-curves.

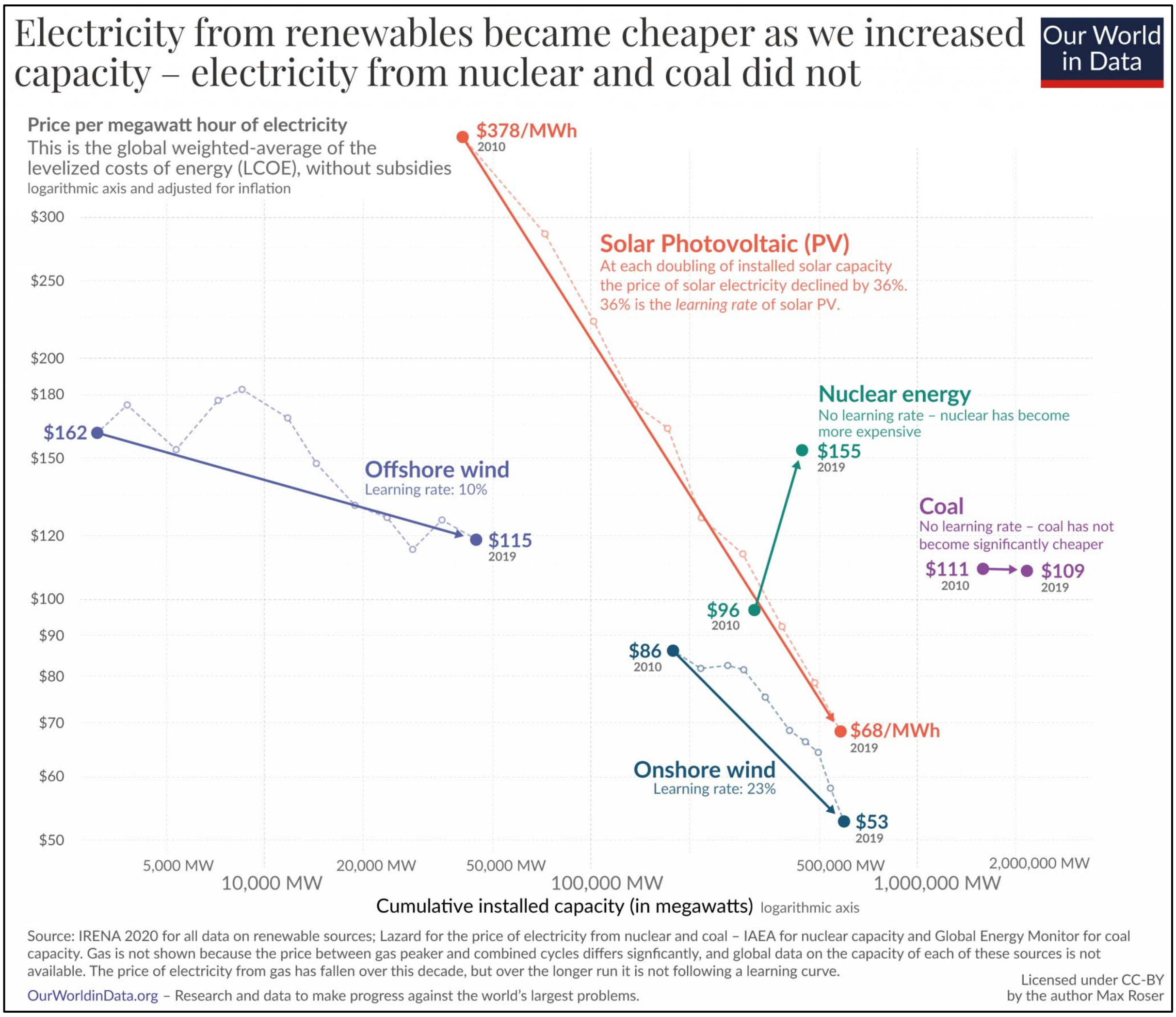


Figure A1-2. *Learning curve for LCOE prices (in 2019 US$/ MWh) of various energy sources, including Solar PV, for the period 2008-2019. Reprinted under a CC-BY license from the excellent article on this topic by Max Roser at https://ourworldindata.org/cheap-renewables-growth#do-electricity-prices-follow-learning-curves.*

Conversion factor between LCOE and Solar PV module prices

Why this extended discussion of learning rates, the different ways of measuring price declines for Solar PV, *etc.*? We need this background in order to understand how to arrive at a conversion factor between LCOE prices (in $/ kWh) and Solar PV module prices (in $/ Wp). And this, in turn, will help to see why the trend of LCOE price declines may recently have been broken, at least for Solar PV. I calculated such a conversion factor from a peer-reviewed paper by Farmer & Lafond (2016)[24], referenced in Max Rosen's OWID article. In the caption of their Figure 1 (on p. 3 of the downloadable 2015 arXiv PDF preprint), it says "*(...) based on recent estimates of levelized costs, we took $0.177/kWh = $0.82/Wp in 2013 (2013$).*"

[24] J. Doyne Farmer & François Lafond (2016). How predictable is technological progress? *Research Policy*, Volume 45, Issue 3, pp. 647-665. https://doi.org/10.1016/j.respol.2015.11.001. A 2015 arXiv preprint PDF of this paper is freely downloadable at https://arxiv.org/pdf/1502.05274.pdf.

From their Figure 1, it's clear that they refer to Solar PV modules here. Indeed, the LCOE data point from our OWID dataset for 2013 was $0.179/kWh, so "good enough for government work", as they say – although our data point is expressed in 2021 US$, and the Farmer & Lafond one (of $0.177/kWh) in 2013 US$. So in order to get an even more accurate conversion factor, *i.e.*, one that will also be valid for our LCOE forecasts and that actually uses the log-linear regression equation from Table 1, we should first *calculate* an LCOE for 2013 based on that equation – which turns out to be $0.198/ kWh expressed in *2021* US$ – and *then* apply the inversed, cumulative inflation factor that has occurred from 2013 through 2020, to express it in 2013 US$ as well. Official inflation rates of the US$ have been relatively mild during this period, at 1-2% per year, or 13.5% over the entire seven-year period.[25]

So our LCOE "forecast" for 2013, based on the log-linear regression equation and expressed in 2013 US$, is 0.198/ 1.135 = $0.174/kWh – *i.e.*, very close to the 2013 LCOE of $0.177/kWh that Farmer & Lafond used. If we then divide the Solar PV module price in 2013 US$, *i.e.*, $0.82/Wp, by our 2013 LCOE estimate, we get **a conversion factor of 4.70**. In other words, in order to convert our LCOE forecasts (in $/ kWh) to Solar PV module prices (in $/ Wp), we need to multiply them by 4.7. This is under the assumption, of course, that the same conversion factor that was valid in 2013 will (more or less) hold into the future – and therefore, by extension, also that roughly the same inflation rates will apply both to solar panel prices and to LCOE prices for industrial-level Solar PV power plants.

Has the trend in LCOE price declines for Solar PV been broken?

We can now use this 4.7 conversion factor to come up with a **forecast for Solar PV module prices in 2023**, and also expressed in 2023 US$/ Wp, so for the actual solar panel prices that we would see today. This is: our 2023 Solar PV LCOE forecast expressed in 2023 US$/ kWh, *i.e.*, $0.035 * 4.7 = **$0.164/Wp**. And expressed in 2023 Euros, taking today's (May 2023) exchange rate of 0.906 € / $, this is: **€0.148/Wp**. Again, please keep in mind that this forecast is based on our very highly correlating (R^2=0.98, $p<1.0*10^{-9}$) log-linear regression equation through the actual 2010-2021 LCOE data for Solar PV in Figure 4b and Table 1, with some reasonable assumptions made along the way.

How does this solar panel price prediction for 2023 compare to today's actual prices? The German solar panel brokerage firm "pvXchange" shows the price development for Solar PV modules for the past twelve months on their website, at https://www.pvxchange.com/Price-Index, with their current April 2023 snapshot reproduced here as Figure A1-3. Their price index shows a range of prices, from low cost panels (currently €0.17/Wp), mainstream "standard" panels (€0.29/Wp), to expensive high-end crystalline panels (€0.37/Wp). If we take the current price of **€0.29/Wp** for standard panels as the "average" solar panel price, it's amazing to see that this is about twice as high as our 2023 solar panel price forecast. And even the low cost panels today are ~15% higher than our forecast of €0.148/Wp.

[25] Official US$ inflation rates were retrieved from https://www.usinflationcalculator.com/inflation/current-inflation-rates/.

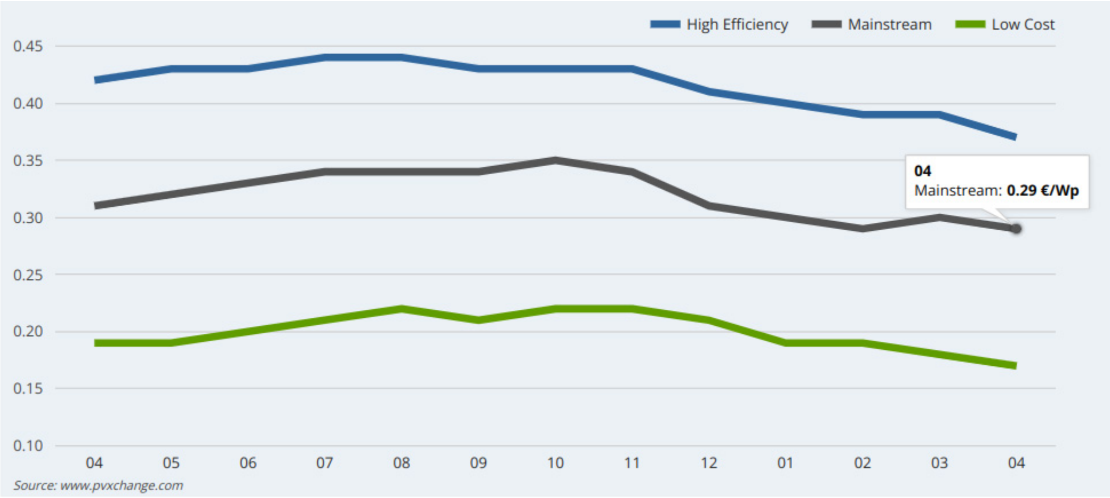


Figure A1-3. *Brokerage firm pvXchange's Price Index for three types of Solar PV modules from April 2022 through April 2023, reprinted from https://www.pvxchange.com/Price-Index (snapshot from May 3rd 2023). Month numbers are plotted on the x-axis, Solar PV module prices (in € /Wp, or Euros per "peak Watt" of power) are on the y-axis.*

This suggests that the strong and highly correlating LCOE price decline trend for Solar PV (declining at 17.6% /per year) has been broken since 2021 – and that this is not just due to inflation, because the effects of the recent strong inflation have already been corrected for in our €0.148/Wp solar panel forecast (*i.e.*, it's already expressed in 2023 Euros). Indeed, the pvXchange Price Index in Fig. A1-3 also shows that Solar PV module prices (which are expected to move in line with LCOE prices, as per my assumption above) have essentially remained flat since the past 12 months, where we would have expected an ~18% decline instead. In fact, prices for the "mainstream panels" went ***up*** from April through October 2022 – thus breaking the LCOE decline trend for Solar PV for the first time since 2005 (visible in the Solar PV module price "learning curve" in Fig. A1-1 above) –, although they did come back down again in the past months, by about the same amount.

To be fair, however, it's unknown whether the pvXchange Price Index chart has been corrected for inflation, *i.e.*, whether panel prices from previous months are expressed in *today's* (April 2023) value of the Euro, or in the actual Euro price in the respective month. If the chart has not been corrected yet for this past year's strong inflation, the relative flatness of the price line may also be a (partial) result of that recent inflation – *i.e.*, in that case the strong inflation would (at least to some extent) have counteracted the strong, longer-term price decline trend.

Nevertheless, if we convert the current and average Solar PV module price of €0.29/Wp back to a current LCOE estimate, by first changing it to 2023 US$ and then dividing by our conversion factor of 4.7, we get a **2023 LCOE estimate of $0.068/kWh** (expressed in 2023 US$). Please note that this is relatively close to the *actual* and averaged 2023 LCOE of **$0.060/kWh** mentioned in the April 2023 Lazard report, which is a confirmation that the conversion factor of 4.7 between LCOE and solar panel prices is still quite accurate today (as per our assumption above).

And also please note that this is still almost twice as high as our 2023 LCOE *forecast* for Solar PV, expressed in 2023 US$, of **$0.035/kWh**. In fact, it's even higher than our most recent *actual* LCOE data point for Solar PV in 2021 as well: $0.048/kWh (expressed in 2021 US$ – or correcting for the ~20% inflation since then, $0.058/kWh in 2023 US$). In other words, it does indeed seem likely that the strongly correlating annual LCOE price decline trend of -17.6% has broken since 2021, with prices having essentially flatlined, or gone slightly up in the past two years.

Possible causes for the trend break

Assuming that we have indeed been able to answer the above question of the recent trend break affirmatively, can we find any potential causes for this? It is outside the intended scope of this report to address this in great detail, but we can find some pointers in a very informative article by Finlay Colville about this topic, at https://www.pv-tech.org/where-is-module-pricing-headed-in-the-next-two-years/. (For context, Dr. Colville is Head of Research at PVTech, one of the premiere online news sources about the Solar PV supply chain worldwide.) It turns out that his prediction – made in March 2022 – that "*(…) no-one should expect to see any meaningful declines in module prices for the next 18 months at least; (…)*"; and "*(…) fundamentally flat-to-moderately-increasing*", has proven very accurate thus far, given the reprinted chart in Fig. A1-3.

According to his analysis, one of the main reasons for the trend break is a much increased demand for solar panels, which started around the "COP26" Climate Change meeting in Glasgow in November 2021 – after the Solar PV industry already reached the subsidy-free point, for the first time, in 2019. With COP26, global corporations started to see *not* jumping on the "net zero targets" train as an existential threat – for themselves commercially, that is, in terms of brand reputation. Because of this much increased demand for PV modules since then, the regular economic laws of prices being driven by demand and supply have kicked in, and the prices have ***increased*** since then as a result, *i.e.*, simply because of demand outstripping supply. In other words, Solar PV shifted from a so-called "demand-constrained" to a "production-led" industry.

According to the article, a second major driver for the increased demand – apart from the renewed "net zero" aspirations – has been the conflict between Ukraine and Russia, which started in February 2022. This has led to a renewed appraisal of the concept of "energy security", *i.e.*, the notion of the importance for countries to be energy-independent, in which Solar PV can play a major role.

Not discussed in the article but related to this, as a general observation: we have also seen soaring energy prices *in general* since the past year – especially for fossil fuel energies, such as natural gas – where, according to various reports I've heard, in some places consumers and businesses even saw their energy bills soar with a factor of about 10. In all likelihood, this has affected the prices for renewables, and thus Solar PV modules, as well.

And finally, as another potential driver, ongoing and intermittent disruptions in global supply lines since 2020-2021 – not just for the energy sector, but for products and goods in general – have likely played a role in the increased demand and prices as well.[26]

Is the trend break temporary or permanent?

One counterargument that could be made to the main thesis in Dr. Colville's analysis, *i.e.*, that a recent, sharply increased demand for solar panels has been the primary driver for the trend break, is that increased demand is actually one half of the earlier discussed self-perpetuating learning cycle; the other half being falling prices. In the learning cycle theory, those falling prices are an indirect result of increased demand, as that leads to higher mass production and thus better optimization of the production process.

Regular economic theory on the other hand prescribes that higher demand for something tends to *raise* prices – so it may all depend on the balance between those two price drivers. That is, if demand becomes *much* higher than the "optimal" demand that drives the learning cycle – and also if there is not much optimization left to be accomplished for the production process, *i.e.*, if it's already been mostly or fully optimized – I would suspect that may break the learning rate price decline and start a phase of price consolidation or even increases.

The question then is whether the observed trend break in Solar PV module and LCOE price decline is just a temporary (albeit multiyear) blip, or whether this represents a permanent change in the industry. The 2002-2005 period saw a similar stalling/ slight increase of Solar PV module prices (Fig. A1-1), but then they started decreasing again at a faster rate and "caught up" to the longer-term trend line again in 2013. It is unknown at this point whether something similar will happen to Solar PV LCOE prices in the mid to late 2020s – but if the conclusion in the above mentioned analysis is correct, *i.e.*, that Solar PV has shifted to a demand-driven industry, that might not be the case and prices might very well stabilize or even increase further into the foreseeable future.

On the other hand, Dr. Colville's 2022 article mentioned plans announced by China for new polysilicon plants in the 2023-2025 period (polysilicon is the key raw material in the Solar PV supply chain). If and when these are realized, they might deliver sufficient production capacity to bring prices down again over the coming years – and the recent decrease in solar panel prices visible in Fig. A1-3, from \$0.35/Wp in October 2022 to \$0.29/Wp in April 2023, is somewhat suggestive of that.

But if this very recent decline indeed persists, it's unknown whether this will be again at the strong annual LCOE price decline rate of -17.6% that I found in Table 1, or at a slower rate: much depends also on how strongly the demand will increase within that timeframe as well. At this point, it's simply too early to tell. What is clear, however, is that because of the recent trend break in annual LCOE decline (at least for Solar PV), the LCOE forecast years in Table 1 are not valid anymore – and, in a best-case scenario, will be delayed by at least several years.

---

[26] See *e.g.* https://viewpoint.bnpparibas-am.com/what-you-need-to-know-about-polysilicon-and-its-role-in-solar-modules/.

What about the other renewables?

I have focused the discussion above on Solar PV – mostly because of the readily available and very recent analyses and price predictions for this renewable; and also because, with its strongest log-linear correlation (R^2=0.98) of all renewables in Table 1, it is the "poster child" for the learning rate concept. Yet the same (likely) underlying causes for the recent trend break in Solar PV price decline may very well apply to the other renewables as well, however. Which are, as discussed: (1) a much stronger demand for renewables because of the renewed "net zero" aspirations and energy security concerns; (2) much higher energy prices in general; and (3) intermittent supply chain disruptions.

Indeed, the line chart on p. 9 of the Lazard report (please see this and this section above) shows that LCOE prices have also gone up for Onshore Wind and Geothermal from 2021 through 2023, suggesting they may have increased for all seven renewables in Table 1. But the Lazard report chart also shows that the mean LCOE for Solar PV has exhibited the *strongest* increase during this two-year period, not just of all renewables but of all energy forms in general: it almost doubled, from $0.036 to $0.060/kWh. This while the mean LCOE for Onshore Wind increased from $0.038 to $0.050/kWh (so once again it's lower than the LCOE for Solar PV), and Geothermal from $0.075 to $0.082/kWh.

Interestingly, Fig. 4b shows that the last few LCOE data points for both Onshore Wind (2017-2021 data) and Offshore Wind (2016-2021) dove *under* their respective log-linear regression lines, and also started decreasing at a faster annual rate than their longer-term regression lines, almost at the same rate as Solar PV. This is not unusual *per se*: Fig. 4b shows that the same happened for Onshore Wind during 1987-1989, and for Offshore Wind during 2000-2003, after which periods the LCOE prices for both renewables flatlined again for a few years.

But their recently accelerated LCOE annual decline rates may also be an indication of new, longer-term trends. (In line with this, please note that for Onshore Wind, the recent 2021-2023 LCOE *increase* simply brought it back to its longer-term and slower 4.9% annual decline rate of the regression line in Fig. 4b – as my 2023 LCOE forecast of $0.053/kWh, expressed as 2023 US$, closely matches its current mean LCOE of $0.050/kWh in the 2023 Lazard report, p. 9. And for Offshore Wind and Geothermal, my 2023 predictions closely match the current LCOEs as well.) The potentially new, accelerated LCOE decline trends for Onshore and Offshore Wind, together with the recent significant increase (almost doubling) of Solar PV LCOE prices, may make the (perhaps prematurely?) declared winner in section 2.3.2 not so clear-cut after all…

In conclusion

As a key takeaway from this extended discussion, we should first and foremost remind ourselves of the wise words that Danish physicist Niels Bohr has famously been quoted as saying: "*Prediction is very difficult, especially if it's about the future.*" This is especially true for data science and extrapolations made from linear regressions, because, as another saying goes, "*Past performance is no guarantee of future results*".

What happened to the annual LCOE prices of Solar PV could serve as a textbook example of this truism: *despite* an amazingly strong log-linear correlation of R^2=0.98 for the 2010-2021 time series, something fundamentally has shifted in 2021 that caused that strong correlation to break.

We can therefore draw the following tentative conclusions, for now:

1. **The learning curve for Solar PV shown in Figure 4b no longer holds** and has – at least temporarily – been halted. Instead of the former 17.6% annual decrease in LCOE prices, the trend has reversed and those prices have almost doubled from 2021 through 2023.

2. It is **unknown at this point whether this is a permanent shift** that will hold well into the foreseeable future, or that it is a temporary blip after which Solar PV will resume its former regular learning curve again (as happened after its 2002-2005 LCOE price consolidation as well) – or perhaps at a slower decline rate, *i.e.*, following a downward trend line that is less steep.

3. Regardless, **the Solar PV forecasted LCOE prices for 2023, 2030 and 2050 in Table 1 are no longer valid**, nor the prediction that it will reach $0.01/kWh in 2028. In the best-case scenario, even if and when Solar PV will resume its same former learning rate of strong, ~18% annual LCOE price declines, those years will be pushed out by a significant delay.

4. On top of this, a **starkly increased official US$ inflation rate since 2021** – with at least 20% inflation over the 2021-2023 period – **will push those forecasted years out even further**, as the US$ has already lost much of its purchasing power since 2021.

5. It's likely that the **observed learning curves for Solar Concentrated, Onshore Wind and Offshore Wind in Fig. 4b have also been temporarily halted** since 2021, as the same (probable) underlying main cause for the Solar PV price consolidation – *i.e.*, a much increased demand for renewable energy – also affects those renewables.

6. Indeed **their LCOE prices also slightly increased in the 2021-2023 period, but not so much as for Solar PV** – for Onshore and Offshore Wind simply moving them back to their former, longer-term decline rates, after recent accelerated declines since 2016-2017.

7. Together, these observations **paint the conclusion drawn in section 2.3.2 – *i.e.*, that Solar PV will have the cheapest price of all renewables in the future – as a potentially**

premature one. Even though that conclusion was based on solid data from 1983-2021 shown in Fig. 4b and Table 1, the global situation and context have changed significantly since then. Therefore **it may very well happen that, in the years to come, Solar Concentrated and/or Onshore or Offshore Wind will surpass Solar PV in having the strongest annual declines of LCOE prices**.

Appendix 2

For sake of completeness, all the additional charts that were created in the Jupyter Notebook that contains the data analyses for Q1 and Q2, available at https://github.com/Dolphinity-io/CO2-emissions, are also shown below – so as bonus material for sections 2.1 and 2.2 of the report, respectively. These extra charts provide additional details and insights into the answers to both questions Q1 and Q2.

Please feel free to share these charts with others as you wish. If and when you do so, however, I respectfully ask you to give proper credit, including the link to this report at https://github.com/Dolphinity-io/CO2-emissions.

Finally, the datasets, detailed comments explaining the process and rationale, and the Python code used to create all of these charts can be found in the Notebook above.

Additional charts for Section 2.1: Biggest predictor of CO_2 output

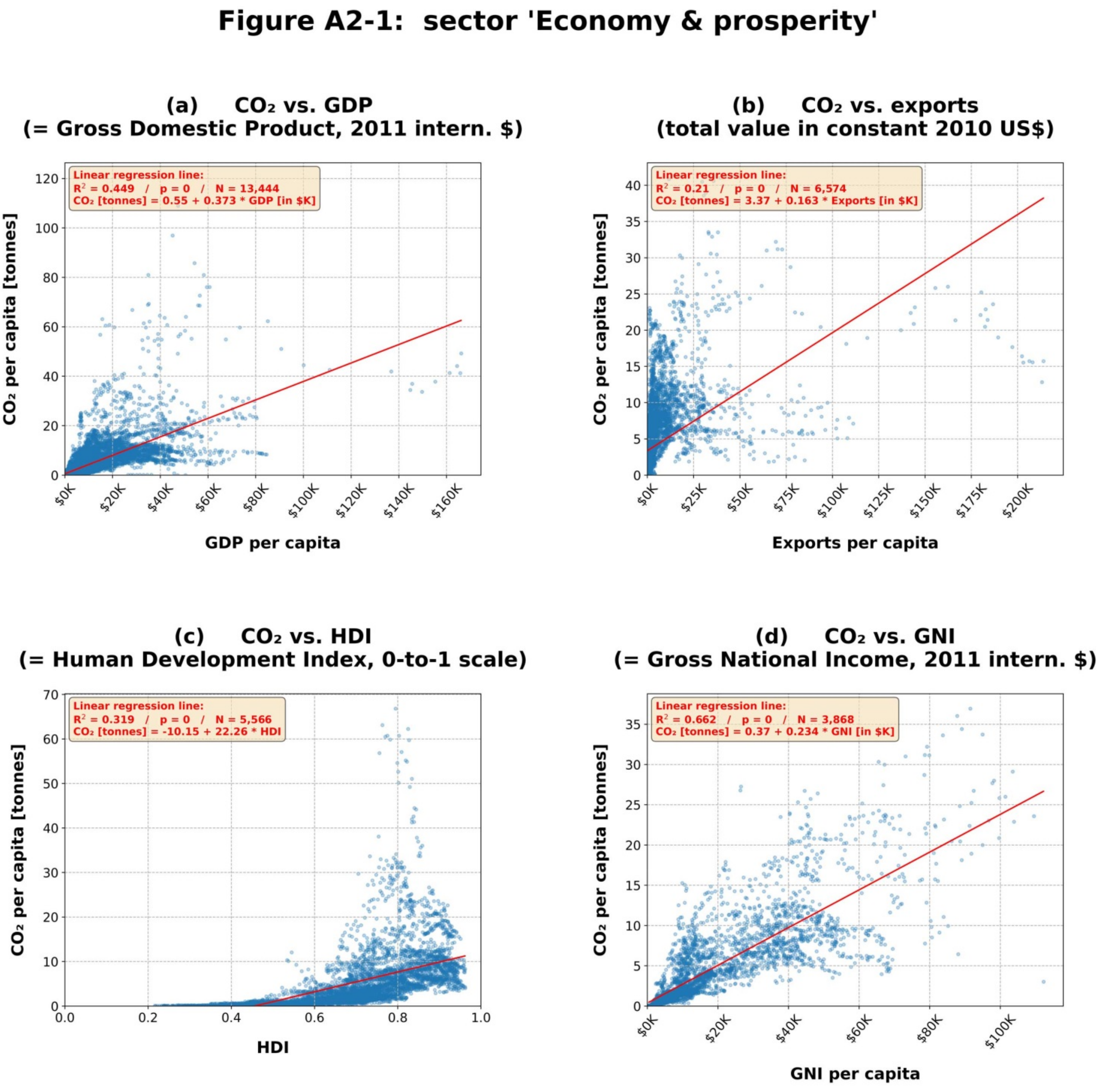


Figure A2-1. *All linear regression plots between CO_2 per capita emissions and each of the four variables in the sector "Economy & Prosperity". For further details, please see this and this explanation in section 2.1.*

Figure A2-2: sector 'Energy'

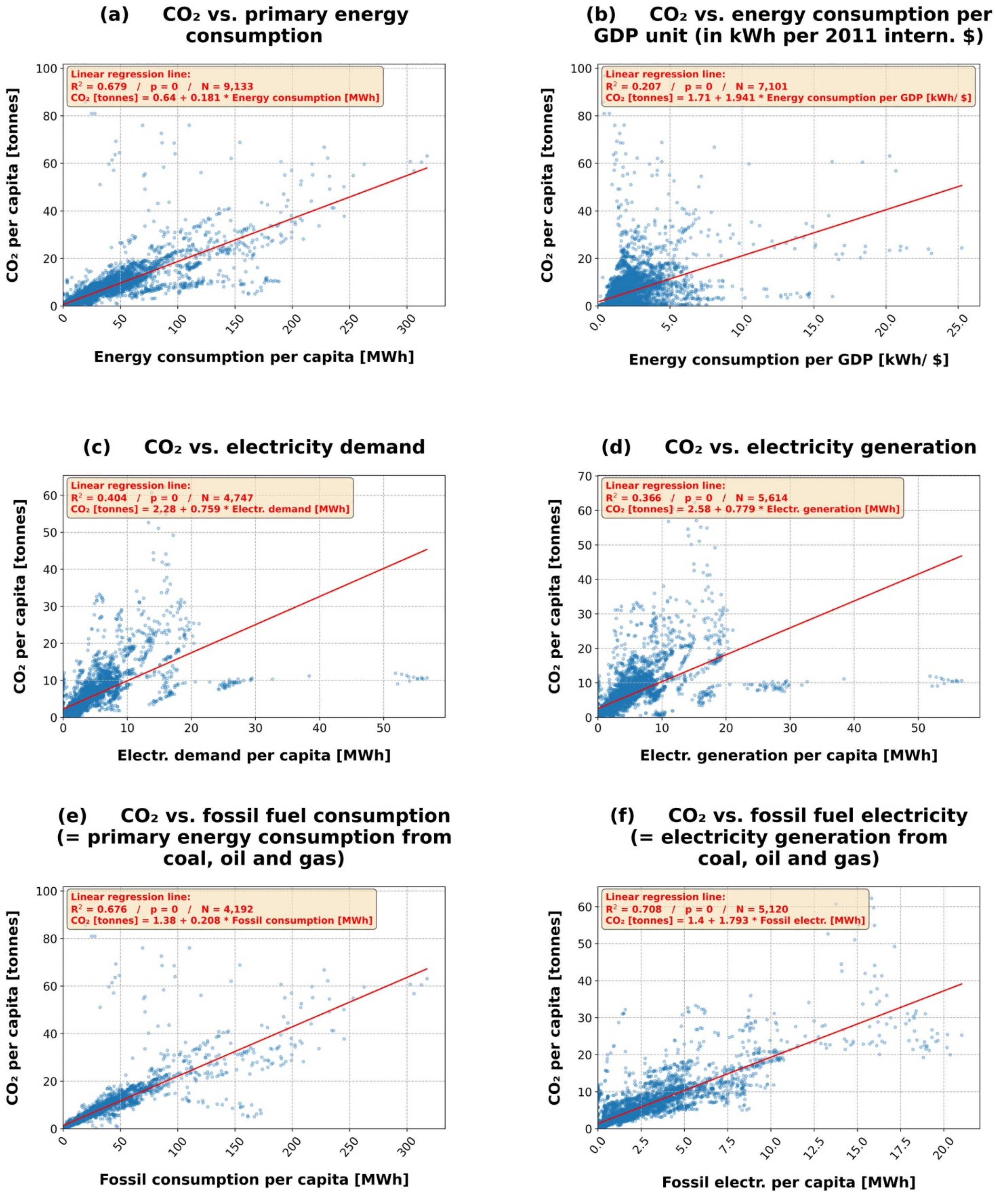


Figure A2-2. *(Part 1 of 3, continued on next two pages.) All linear regression plots between CO_2 per capita emissions and each of the 14 variables in the sector "Energy". For further details, please see this and this explanation in section 2.1.*

Figure A2-2: sector 'Energy'

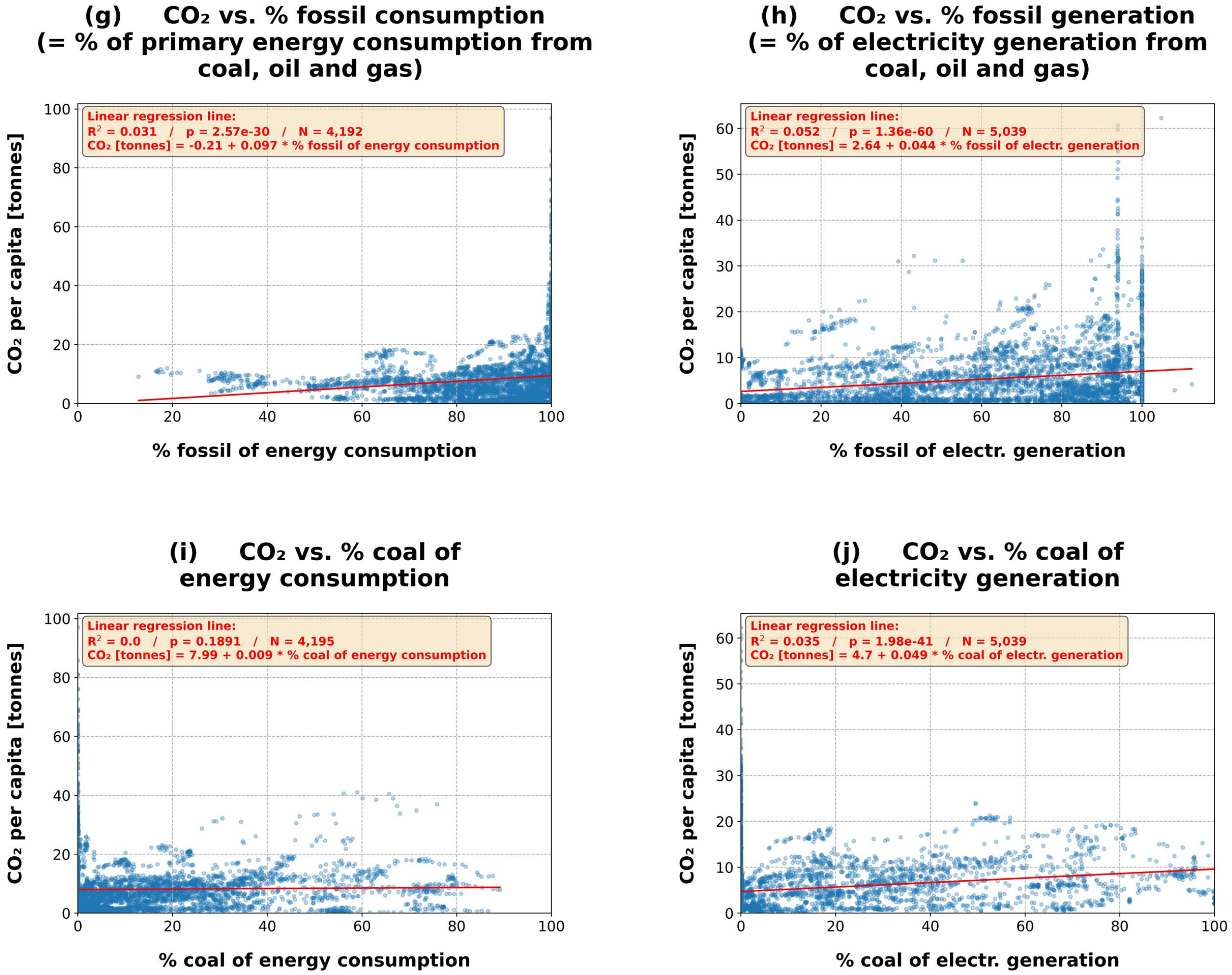


Figure A2-2. *(Part 2 of 3, continued on next page.) All linear regression plots between CO_2 per capita emissions and each of the 14 variables in the sector "Energy". For further details, please see this and this explanation in section 2.1.*

Figure A2-2: sector 'Energy'

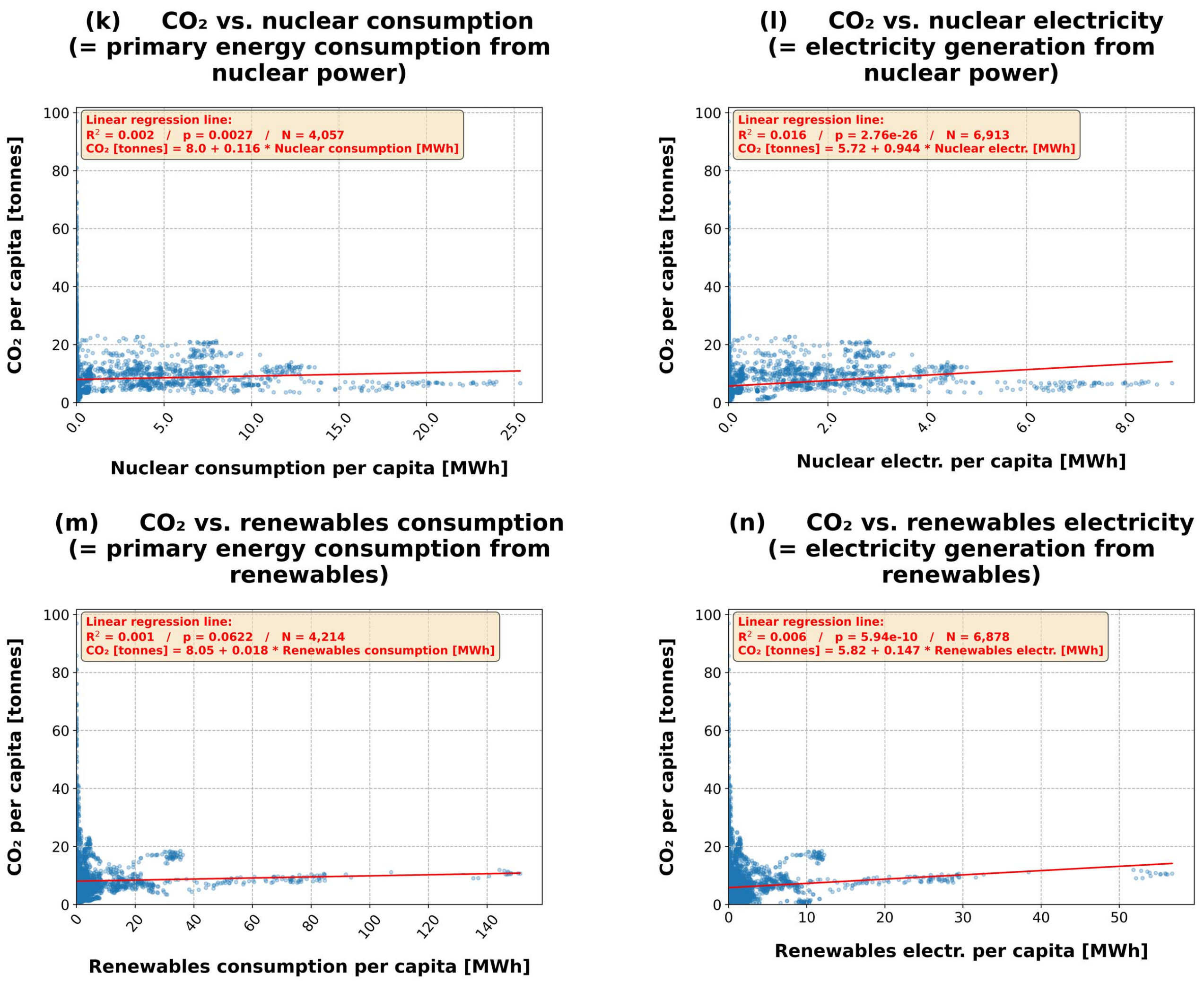


Figure A2-2. *(Part 3 of 3, continued from previous two pages.) All linear regression plots between CO_2 per capita emissions and each of the 14 variables in the sector "Energy". For further details, please see this and this explanation in section 2.1.*

Figure A2-3: sector 'Transportation'

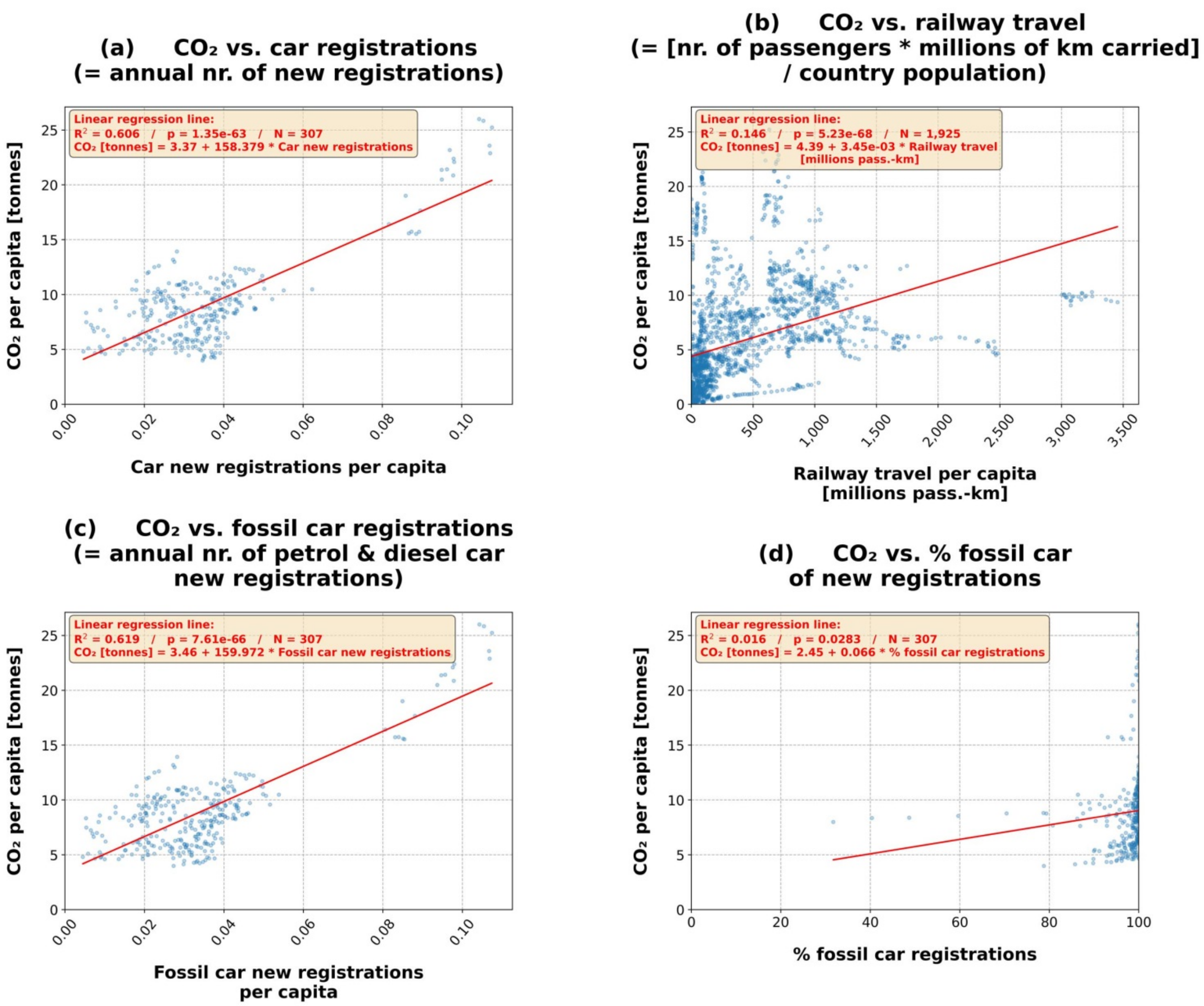


Figure A2-3. *(Part 1 of 2, continued on next page.) All linear regression plots between CO_2 per capita emissions and each of the eight variables in the sector "Transportation". For further details, please see [this](#) and [this explanation](#) in section 2.1.*

Figure A2-3: sector 'Transportation'

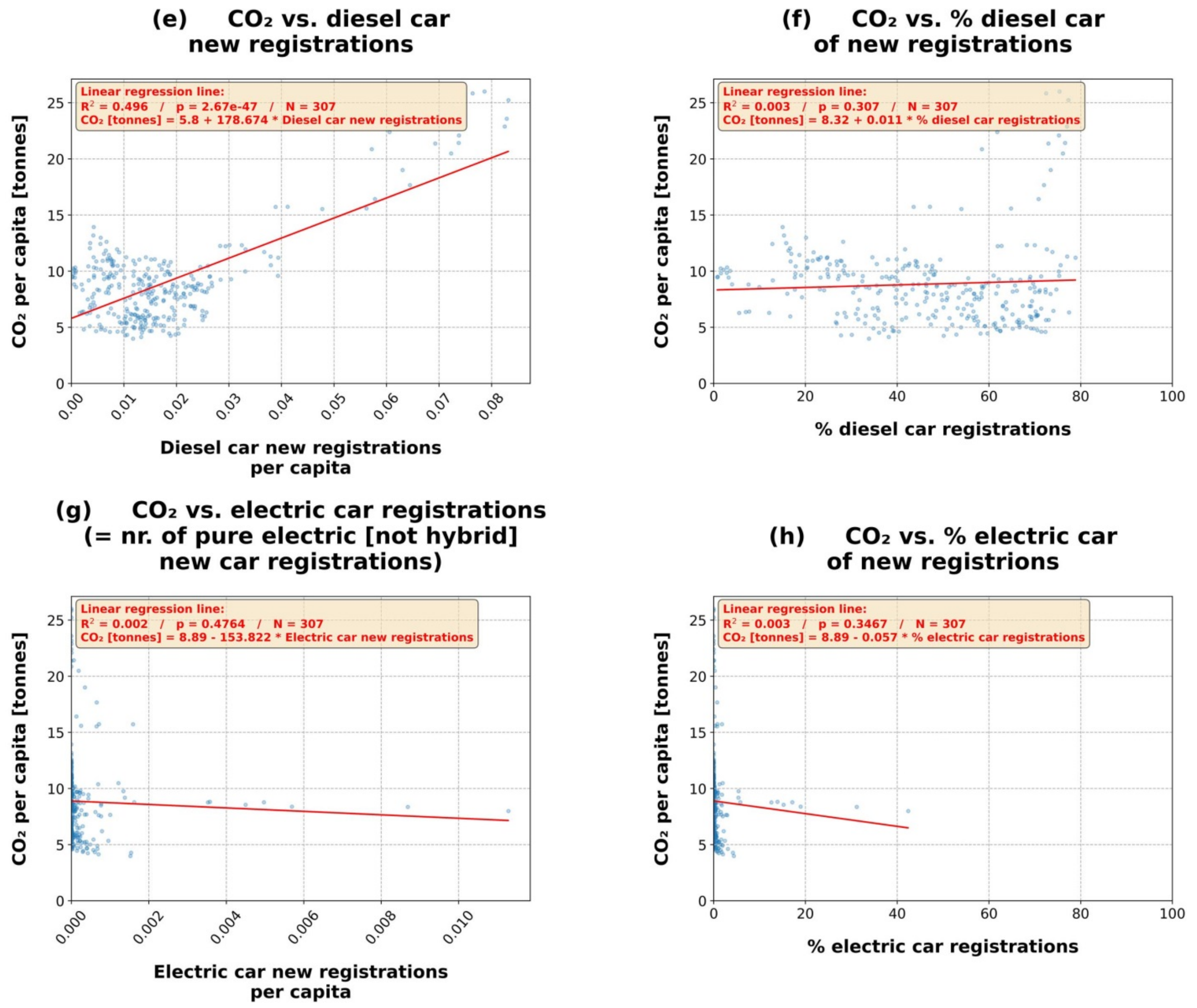


Figure A2-3. *(Part 2 of 2, continued from previous page.) All linear regression plots between CO_2 per capita emissions and each of the eight variables in the sector "Transportation". For further details, please see this and this explanation in section 2.1.*

Figure A2-4: sector 'Food & agriculture'

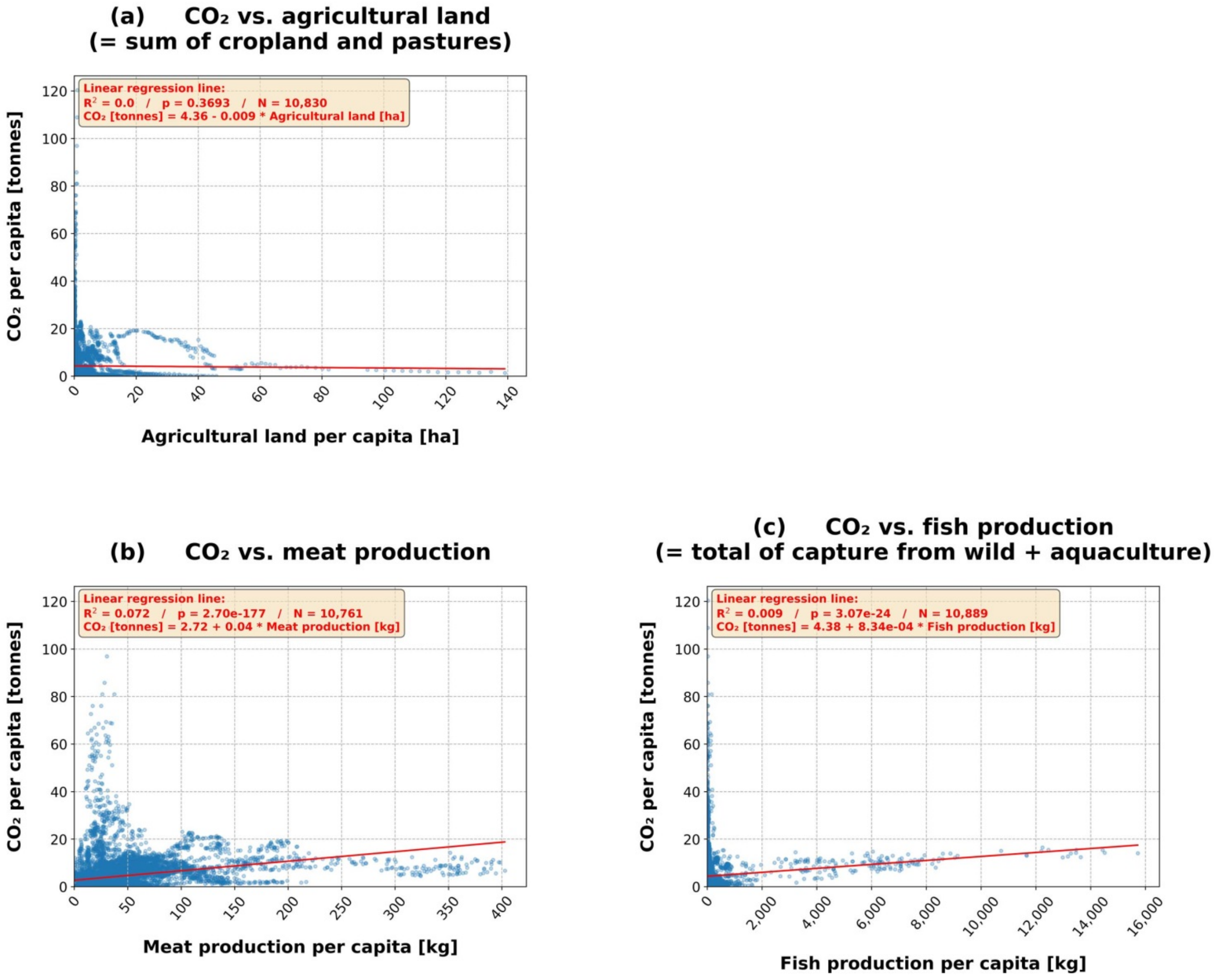


Figure A2-4. *(Part 1 of 2, continued on next page.) All linear regression plots between CO_2 per capita emissions and each of the seven variables in the sector "Food & Agriculture". For further details, please see this and this explanation in section 2.1.*

Figure A2-4: sector 'Food & agriculture'

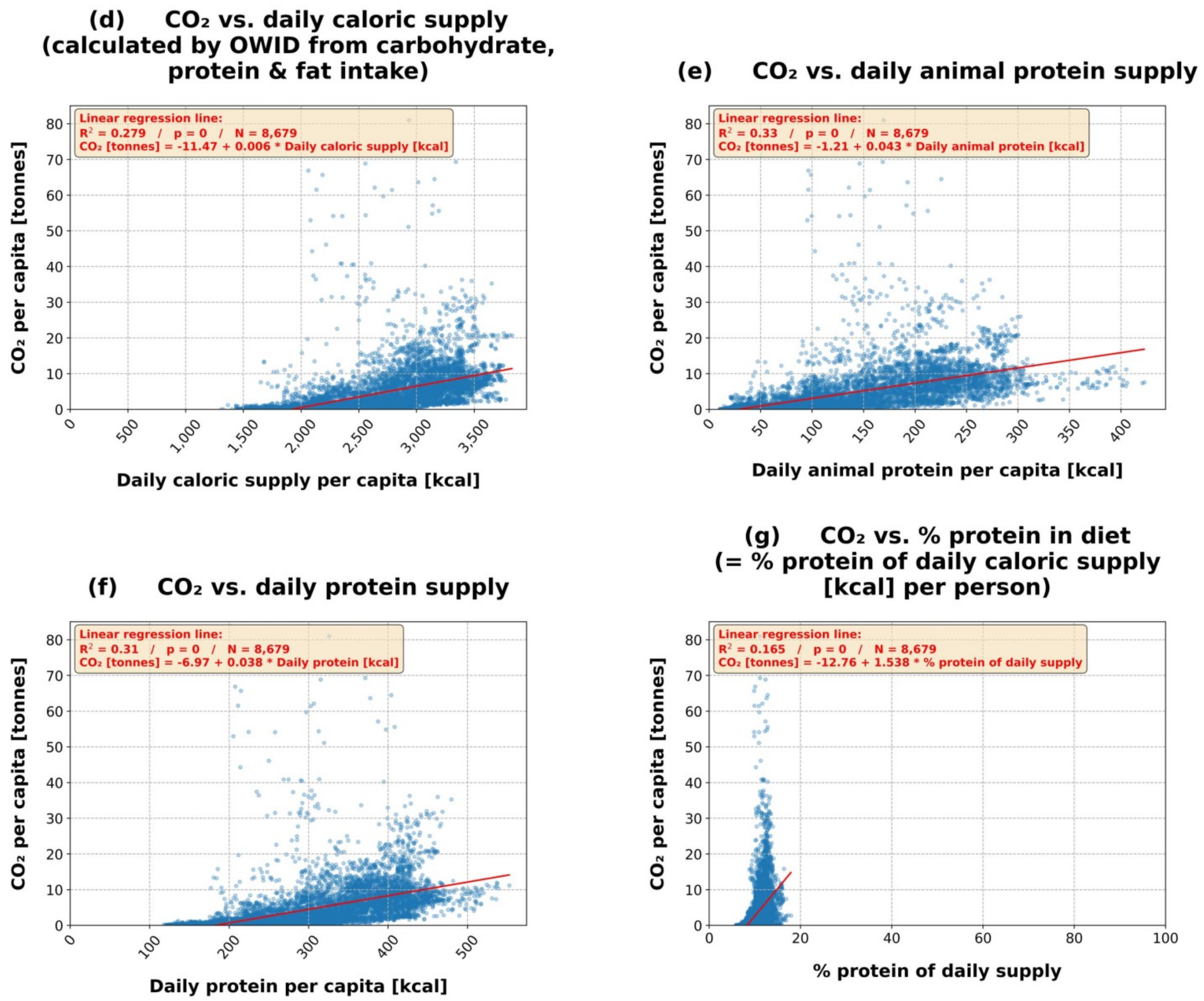


Figure A2-4. *(Part 2 of 2, continued from previous page.) All linear regression plots between CO_2 per capita emissions and each of the seven variables in the sector “Food & Agriculture”. For further details, please see this and this explanation in section 2.1.*

Figure A2-5: sector 'Demographic & sustainability'

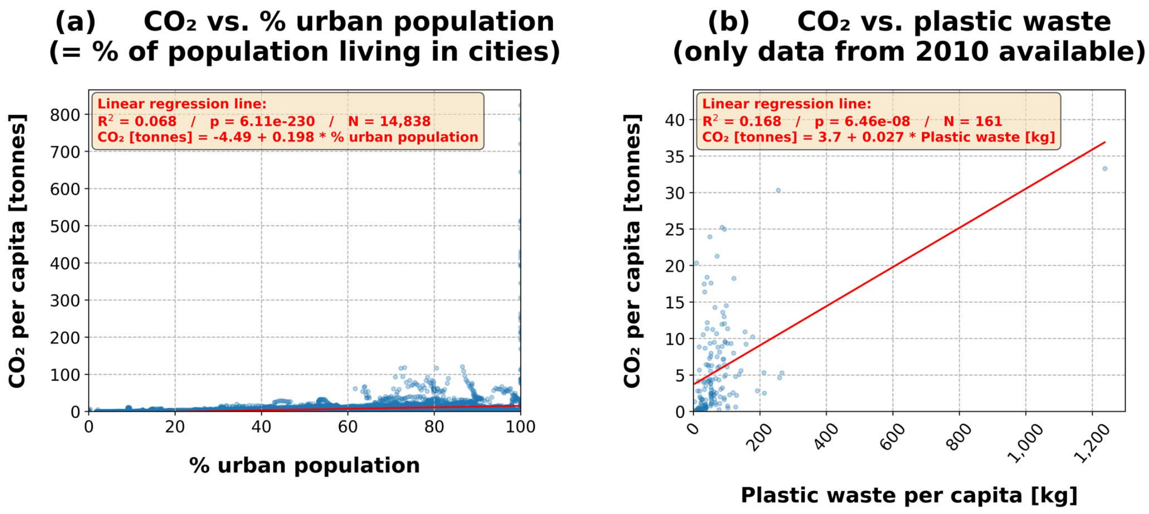
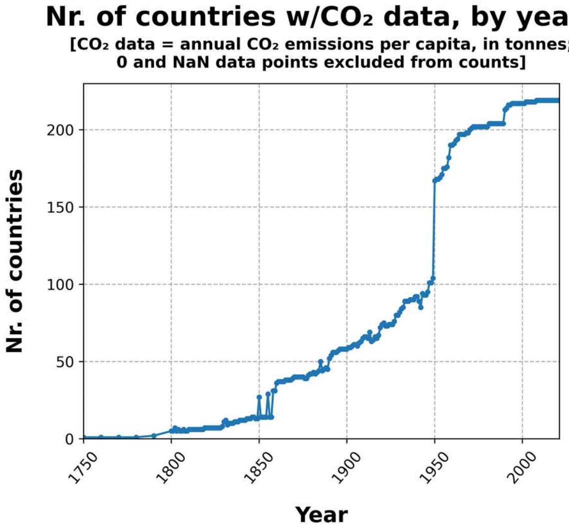
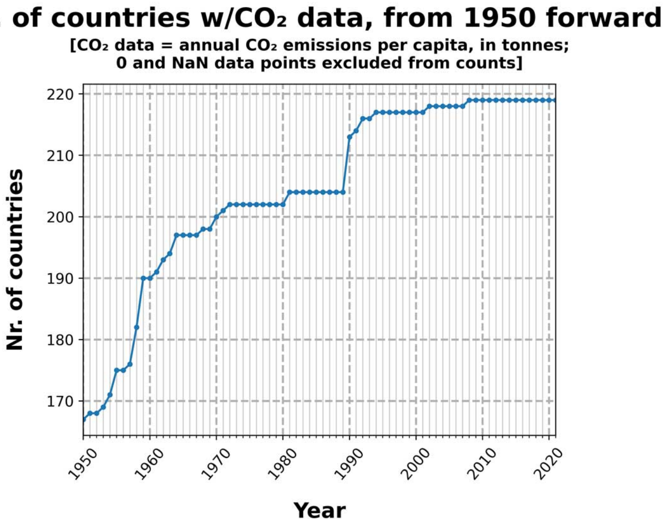


Figure A2-5. *Linear regression plots between CO_2 per capita emissions and both of the variables in the sector "Demographic & Sustainability" (or: "Other"). For further details, please see this and this explanation in section 2.1.*

Additional charts for Section 2.2: Biggest strides in decreasing CO_2 output





(a) (b)

Figure A2-6. *Number of countries that have annual CO_2 per capita data available for each year, (a) for all years in the dataset (1750-2021), and (b) for the period 1950-2021. For additional details, please see the explanation in the relevant part of section 2.2.1.*

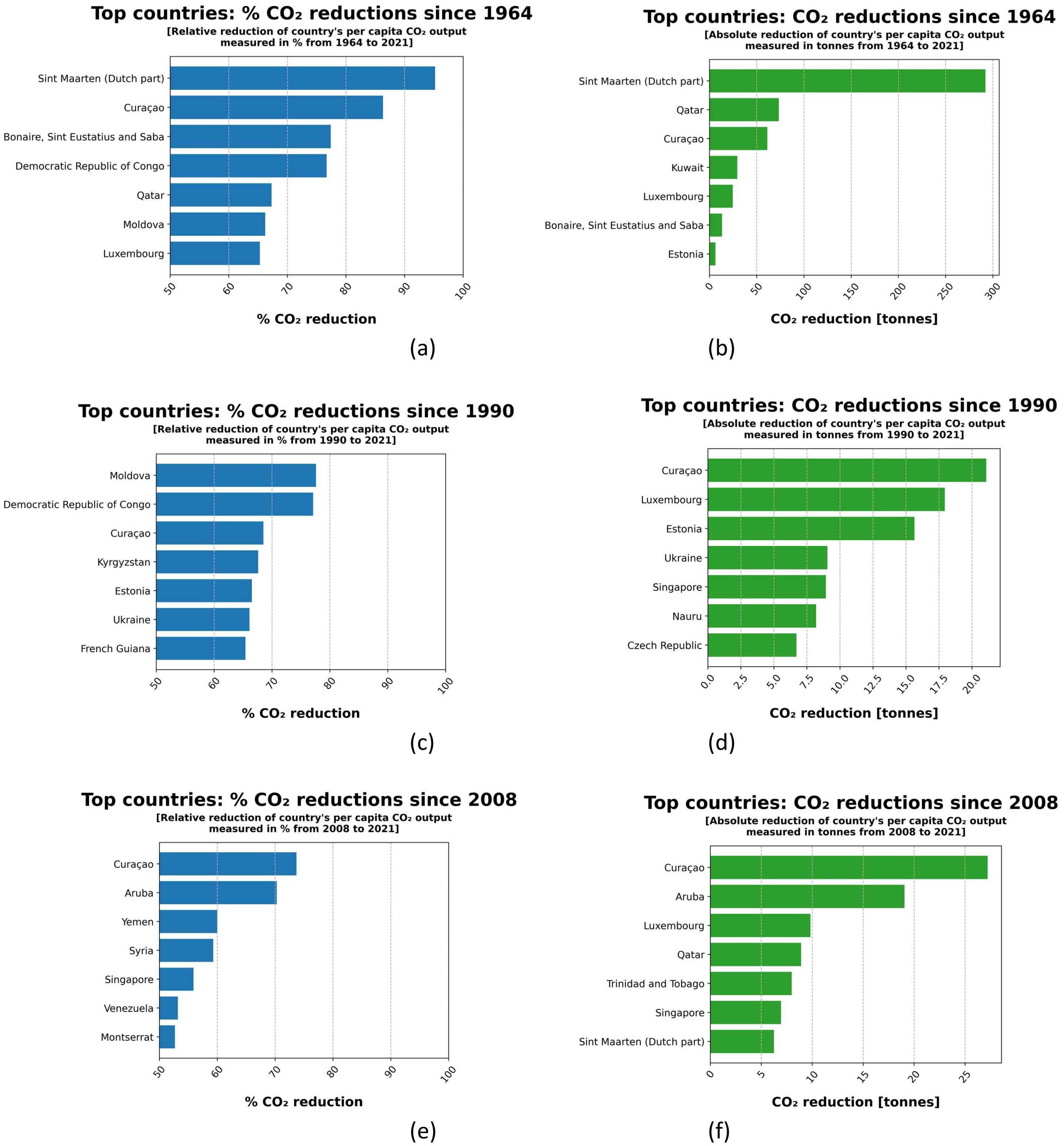
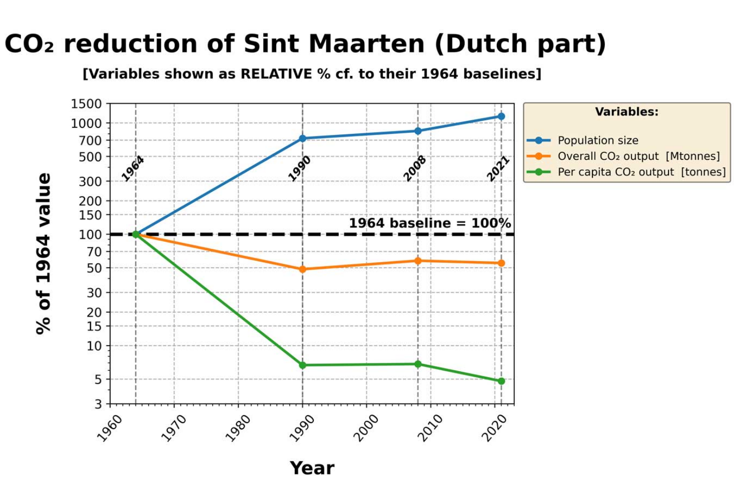
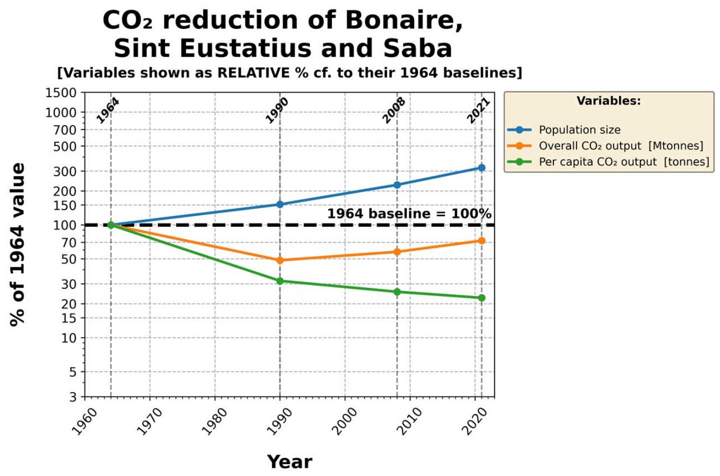


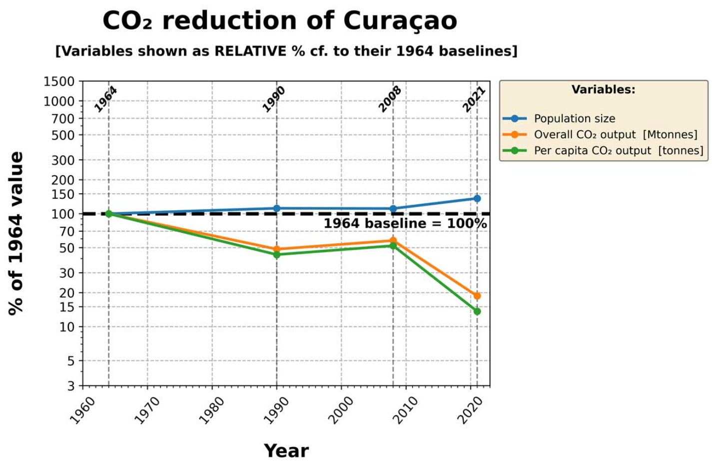
Figure A2-7. *Top seven CO_2 reducing countries for each of six ways of measuring CO_2 reduction, conveying the same information as [Figure 3](#) in section 2.2 of the main text, but plotted as horizontal bar charts instead. These allow easier cross-country comparisons of the relative magnitudes of CO_2 reduction. Relative % reductions (Fig. A2-7a, c, and e) are shown as blue bars, absolute reductions measured as tonnes of CO_2 per capita (Fig. A2-7b, d, and f) are shown as green bars. For additional details, please see the caption of Figure 3.*



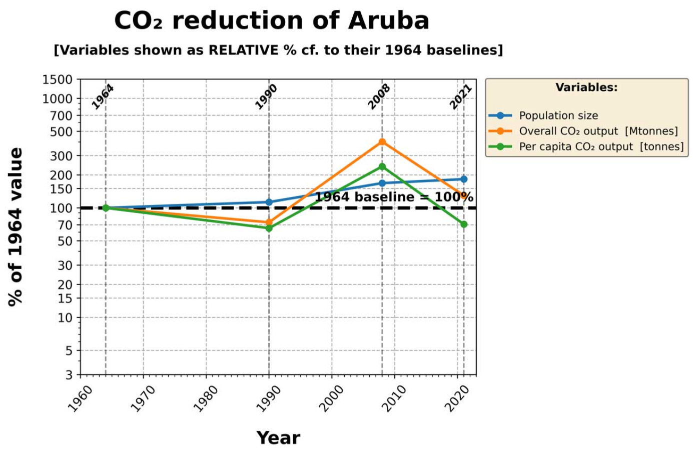
(a)



(b)



(c)



(d)

Figure A2-8. *Different mechanisms behind the CO_2 per capita reductions of four Netherlands Antillean countries from 1964 through 2021. Population size (blue lines), Megatonnes of overall CO_2 output (orange lines) and tonnes of per capita CO_2 output (green lines) are shown as relative % of their 1964 baseline values. For additional details and how to interpret these, please see the explanation in the relevant part of section 2.2.2.*