Exploring the Relation Between Sustainability Research and Sustainable Development Goal 7

DS 143 - Final Project

Maha Mapara, Xinhang Liu, Vibhor Janey, LiChai Epperson, Matthew Hui

Abstract

In 2015, the United Nations proposed 17 SDG goals to promote economic prosperity while protecting the earth. In this project, we explore the progress being made by countries towards SDG 7 (access to clean and affordable energy) from the perspective of academic research impact, instead of a policy impact. Academic research impact is defined in 2 ways: how bulk (global) knowledge contributes to SDG 7 progress and how countries' own knowledge contributes to their SDG 7 progress. SDG 7 progress has been measured through relative change and the distance countries have to travel to achieve numerical targets set by the OECD. Using quantile regression we found that OECD countries are approaching the SDG 7 energy efficiency goal as research impact increases. This means that as the spread and influence of research related to SDG 7 increases, energy efficiency in OECD countries is improving, leading to less greenhouse emissions and more sustainable energy infrastructures.

1 Introduction

Sustainable development is a framework for meeting human improvements and needs while maintaining the ability and function of natural systems on our planet. This framework has been adopted and popularized by the United Nations' Sustainable Development Goals (SDGs).The SDGs consist of 17 goals, complemented with 169 targets, which together specify a 2030 development agenda to reach sustainable societies [1]. All 193 UN member countries agreed to work towards the 17 SDGs and fulfill the UN 2030 agenda.

At the same time, sustainability related research has become an important focus in academia [2], with the goal of exploring and improving technologies, or policies that improve sustainability. Specifically, there has been a growth in academic research on clean and affordable energy encompassing topics like renewable energies, energy efficiency, electric vehicles, and grid modernization, to name a few. As such, we are interested in seeing if there is a correlation between progress being made on SDGs and impactful academic research related to SDGs.

We focused on SDG 7, which the UN defines as ensuring 'access to affordable, reliable, sustainable and modern energy for all'. SDG 7 has three targets defined by the UN, which can be broadly categorized as improving access to energy, increasing the use of renewable energy and improving energy efficiency [3]. These are measured by a total of 27 indicators (like proportion of population with access to electricity, the full list is available in the appendix). As we were interested in change/progress in SDG 7 indicators over time, instead of using the raw SDG 7 indicator data provided by the World Bank [4], we calculated the relative change in SDG 7 indicators from 1991 till 2020 (based on data availability), described in our literature review and methods sections.

Moreover, Bréchet et al. (2016) [5] discussed the relation between the success of international agreements and individual domestic policies in each country. International agreements like the SDGs and UN's 2030 Agenda are only possible if actionable domestic policies are implemented. Impactful domestic research in areas of SDG 7 related sustainable growth can play an important role towards international SDG 7 progress, by being translated into domestic policies that lead to innovation and improvement in a country's system and technologies. Jaffe et al. (2020) [6] also discussed the relationship between research and economic progress.

As such, we are focusing on the impact of academic research related to SDG 7 by the UN member countries, as a means of leading to SDG 7 progress. Reed et al. (2021) [7] defines research impact as 'the process of assessing the significance and reach of both positive and negative effects of research'. It is common practice [8] to measure impact using citation counts i.e. citation impact. Citation impact is a measure of how many times a research paper is cited by other research papers. There are multiple ways of calculating citation impact [9] but we focus on 2 ways based on our literature review: AR-index [10] and log of total citations in 2 years of a publication (log(2 years)) [11]. The calculations, benefits, and drawbacks of these 2 metrics are discussed in our literature review section.

Thus, our predictor variables were AR-index and log of 2 year citations while relative change in the SDG 7 indicators were our response variables.

Based on this background our research question was: **As the impact of academic research related to the three SDG 7 targets increases, do we see positive global progress towards the UN 2030 SDG 7 targets?**

There are two main contributions of this research. First, we used 3 different methods to calculate relative change in SDG 7 indicators over time and assessed the relation between all 3 types of calculations with the 2 citation impact metrics (AR-index and log of citation in 2 years). These three methods are the percent change, logarithmic difference, and distance measure, and are further described in the literature review and methods sections. Second, and most importantly, we focused on isolating the relationship between SDG 7 progress through an academic research impact lens, instead of the domestic or international policy impact lenses. Two perspectives on academic research were used and we have named these bulk knowledge and country-specific. Bulk knowledge measures the citation impact metrics on a global (bulk) level such that all research, regardless of the country it is coming from, is assumed to contribute towards SDG 7 progress for all countries. Country-specific measures the citation impact metrics produced by a country on its own SDG 7 progress. It does not assume that the research impact of papers from one country helps all countries towards their SDG 7 progress.

We computed linear and quantile regressions for each type of SDG 7 indicator calculation (percent change, log difference, distance measure) against AR-index and log of citation in 2 years, individually. We found a few strong signals for both our bulk knowledge and country-specific citation impacts which showed that as either AR-index or log(2 years) increases, the distance measure for an indicator used to measure energy efficiency decreases. This meant that as research impact increased, the energy efficiency target was met but this was the case for only a group of 38 countries part of the Organization for Economic Co-operation and Development (OECD).

2 Literature Review

2.1 Calculating Citation Impact

The number of citations a research paper gets is dependent on the age of the publication i.e. how old a paper is. An older paper may, therefore, appear more impactful when it may have high citation counts only due to age. Gonzalez et al. (2013) [12] and Gonzalez et al.(2016) [13] discussed impact maturity times and citation time windows. While impact factors are calculated for journals, usually, over a 2 year time window, the 2 year time window can be used to evaluate the impact of an individual research publication by using citation counts in 2 years of a paper being published. In our approach we decided to take the log of sum of citation in 2 years, instead of averaging the citations over 2 years, as the log was a more appropriate transformation based on the distribution of values.

A popular approach to measure citation impact is the h-index [14]. As the h-index was author specific instead of research paper specific, using the h-index our dataset was not possible. Additionally h-index does not increase or decrease over time and does not take into account age of a publication. Due to these issues, the h-index was used to develop other indices by Jin et al. (2007) [10] and we focus on one of them: AR-index. The AR-index takes the age of publications into account and can decrease over time. As can be seen in figure 1, it is defined as the square root of the sum of the average number of citations per year of articles included in the h-core. The h-core is defined as the highest rank such that the first h publications in a year received each at least h citations.

$$AR = \sqrt{\sum_{j=1}^{h} \frac{cit_j}{a_j}}$$

Figure 1: AR-Index formula

2.2 Calculating SDG 7 relative change

As we wanted to see the relationship between SDG 7 progress and research impact, it was necessary to calculate the change in SDG 7 indicators over time, rather than use the raw numbers for every year. Two common statistical methods used for relative change are percentage change [15] and the logarithmic difference between two values [16]. The benefits of using these methods is that we can see how the SDG 7 indicators are changing every year. The logarithmic difference has an advantage over percent change as the calculations are symmetrical going forward and backward. But the main issue with both these methods is that they do not take into account that different countries have different starting points and, therefore, different distances to progress towards the SDG 7 targets set by the UN.

Miola and Sculz (2019) [17] did a meta-analysis on methods to measure SDG performance or progress. One of these methods is called the distance measure and is a metric that the OECD created to measure the SDG 7 progress of 38 OECD countries (more information about OECD available in appendix A.4). It can be calculated at a larger goal level or at the level of individual indicators or targets in the goal. The OECD discussed this method, among others in its paper on measuring distance of countries to SDG targets [18].The other methods discussed by OECD were the ratio scale and time distance method. The paper concluded that the distance measure method is preferred to the other 2 methods due to its robustness. Figure 2 shows the calculation for this method. In essence, distance measure tells us how many standard deviations a country (x) is from the target (T).

The advantages of this method is that it normalizes for a country's starting point by having country/region specific targets and it lowers the scores when all countries are far from the target and perform similarly badly. The major disadvantage with this method is that the numerical target end-values (T) have to be calculated for individual countries or countries with similar economies and resources. While OECD proposed a framework to calculate these target end-values for multiple groups of countries, it only calculated the values for OECD countries. Based on World Bank income classifications[19], these OECD countries are only high-income and upper-middle income countries and so the target values for these are not applicable for non-OECD countries.

$$\max\left(\frac{T - x}{\sigma}, 0\right)$$

Figure 2: Distance Measure formula

3 Data and Methods

3.1 Data set 1: Citation Impact Data

This data comes from the Web of Science (WoS) [20] which has multiple databases that provide reference and citation data from academic journals, conferences etc. Sustainable development research specific data from WoS was collected by the Sunter Lab at Tufts University. This data set only includes data on academic research in the English language. The data had 87792 rows and 191 columns. There are 70 of these columns contain the metadata on each publication in the row, while the other 121 columns contain the number of citations a paper has from 1901 till 2021. Thee earliest publication from the data set was from 1981, and the latest paper was from early 2022.

For our purposes, the relevant columns were the title (TI), abstract (AB), keywords about the paper provided by the authors (DE), Publication Year (PY), all columns containing citation counts and authors' unique nationalities.

3.1.1 Data Cleaning

As our focus was research related to SDG 7, we filtered the citation impact data such that it only contained research papers pertaining to SDG 7 targets. Data set was filtered to only contain publications where the title, abstract or keywords contained the words 'energy' or 'electricity'. After this the publication years ranged from 1991 till 2022.

This data has 7538 rows i.e. data on 7538 papers related to SDG 7.

3.1.2 Feature Engineering

Using the sum of citation counts over a publication's lifetime and the sum of citation counts in 2 years of publication, we calculated the AR-index and log of citation in 2 years. Figure 1 shows the formula used to calculate the AR-index.

As mentioned in our introduction, we looked at the academic research from two perspectives that we have named bulk knowledge and country-specific. Our assumptions for these led to AR-index and log(2 years) being calculated differently. Details on how these calculations were done can be found in the appendix (A.1).

For the bulk knowledge perspective, the 2 citation impact metrics (AR-index and log(2 years)) were calculated without considering the authors' nationality. As a reminder, the goal of this approach is to see if global (bulk) research impact has an impact on SDG 7 progress for all.

For the country specific approach, these metrics were calculated on a yearly basis for each country. The goal of this approach is to measure the impact of each country's research on its own SDG 7 progress. As country-specific used the unique nationalities column, papers where authors had more than one unique nationality (i.e. the research paper was a collaboration between authors from different countries) were dropped. This was 27% of the SDG 7 target specific data set. Furthermore, to be able to join this country-specific citation impact data with the SDG 7 indicator data, the country names in unique nationalities column were renamed to match the country names in the SDG 7 indicator data (described in section 3.2)

We ended up with 2 data frames, one for each academic research perspective: bulk knowledge data frame and country-specific data frame. Figures 3 and 4 show snippets of these data sets.

Note: For the two cleaned data frames, the years range from 1991 till 2020, not till 2022. This is because we needed a paper published in 2020 to accumulate citation counts in 2 years (till 2022) for the log(2 years) calculation.

	PY	2 years window	ar index bulk
0	1991	6.0	0.875595
1	1992	2.0	0.454859
2	1993	4.0	0.823754
3	1994	6.0	1.122167
4	1995	3.0	1.850156

Figure 3: Bulk Knowledge Citation Impact Data

	PY	Country Name	2 years window	ar index country
0	1991	india	1.0	0.707107
1	1991	united kingdom	5.0	0.516398
2	1992	brazil	2.0	0.454859
3	1994	china	1.0	0.192450
4	1994	france	1.0	0.471405

Figure 4: Country-Specific Citation Impact Data

3.2 Data set 2: SDG 7 Indicator Data

The data is downloaded from the World Development Indicators (WDI) website [4], which contains data for all countries, and indicators for all SDG targets from 1960 till 2021. Based on the metadata on the WDI website, all SDG 7 indicators have an indicator code starting with 'EG.' and by applying that criteria on the indicator code column, we got only SDG 7 indicator data. This data had 383572 rows and 66 columns. There were 2 non-year columns containing the country names and indicators names, and 62 year columns which contained the indicator value of each country at the specific year (each row represented a unique country and indicator combination).

3.2.1 Data Cleaning

The data was then converted to a tidy data format such that the columns were: Country Name, Year (1966 till 2021) and 27 Indicator columns. A list of all 27 SDG 7 indicators is available in Appendix A.2. Each row was a unique combination of country and year for 27 indicator values.

As the citation impact data contained data only from 1991 till 2020, only data from these years was used. This data then had 6650 rows.

Moreover, countries report data at different frequencies. For example, country A may report data on the percentage of population with access to electricity on a yearly basis, but country B reports it every two years. To deal with this missing data where countries do not typically report data, we used linear interpolation [21]. Linear interpolation was concluded to be a robust way to estimate missing values compared to mean substitution, which leads to large errors in the correlation matrix, degrading the performance of statistical modeling.[22]. Data that was outside the range of interpolation i.e. missing data before the first instance of data reported by a country and missing data outside of the last instance of data reported by a country were dropped. This was 20% of the 6650 rows.

3.2.2 Feature Engineering

We then calculated the relative change metrics for SDG 7 indicators. The year-on-year percentage change and year-on-year logarithmic differences were calculated as per the literature [15] [16] for all 27 indicators.

For the distance measure described in section 2.2, only 5 numerical target values were calculated by OECD for 5 indicators. We estimated 7 more based on our literature review on OECD target estimates [18] [23]. Using these 12 values (1), we calculated 12 distance measures for only OECD countries and all countries.

The reason for these 2 distance measure calculations are that figure 2 shows the formula utilizes standard deviation (σ). σ is the standard deviation across indicator values for all countries for whom, the target (T) value is available.

When countries for whom the T value is not available (in our case non-OECD countries), but instead OECD targets were used, the standard deviation was calculated on indicator values for all countries. This distorts the distance measure value. With the awareness that distance measure for all countries would possibly not give significant regressions [18], we calculated distance measure in the 2 ways to see if a signal could be found using distance measure for all countries, in addition to distance measure for only OECD countries.

SDG 7 Indicator	Unit	Numerical target value	Source
Proportion of population with access to electricity	%	100	OECD
Proportion of population with access to electricity, rural	%	100	Estimated
Proportion of population with access to electricity, urban	%	100	Estimated
Proportion of population with primary reliance on clean fuels and technology for cooking	%	97.50	OECD
Proportion of population with primary reliance on clean fuels and technology for cooking, rural	%	100	Estimated
Proportion of population with primary reliance on clean fuels and technology for cooking, urban	%	100	Estimated
Renewable energy share in the total final energy consumption	%	25.18	OECD
Renewable electricity share in total electricity generation	%	38.65	OECD
Energy Use (Energy Productivity/Efficiency)	kg per capita	18949.33	OECD
Fossil fuel energy consumption (% of total)	%	0	Estimated
Electricity production from coal sources (% of total)	%	0	Estimated
Electricity production from oil, gas and coal sources (% of total)	%	0	Estimated

Table 1: Estimated and OECD target values used in the distance measure method

3.3 Data set 3: Income-classification Data

This data is downloaded from the World Bank (WB) website [19]. It contains income levels for all 217 countries and regions defined by the World Bank. There are 217 rows and 2 columns (income level, country name). Income level is calculated using Gross National Income (GNI) per Capita. There are 4 classifications: low-income, lower-middle-income, upper-middle-income, and high-income. Table 3 in Appendix A.3 contains the 2020 income classifications thresholds from the World Bank. The World Bank website for this data says that income level is a measure of the economy level of each country.

We used 2020 income classifications as our data range for citation impact and SDG 7 indicator data ended at 2020.

The purpose of this data was to color data points in our regressions by income level to see whether countries with higher income levels have better SDG 7 progress. This idea of research impact and a country's wealth being connected came from Jaffe et al. (2020) [6]

3.4 Data Merging

Recall that citation impact metrics were calculated by bulk knowledge and country-specific assumptions. Bulk knowledge citation impact data was merged with SDG 7 indicator data and WB income classification data, which is shown in figure 5. Similarly, country-specific citation impact data was merged with the other 2 data sets and a merged country-specific data set was created. Therefore, we had 2 main clean, merged data sets for our regression analyses.

Each data set had the following columns:

- Country Name, Year
- Predictor variables: AR-index, log of citation in 2 years

- Response variables (SDG 7 Indicators): percent change, log difference, distance measure for OECD countries, distance measure for all countries

Year	2 years window	ar_index_bulk	Country Name	cooking	rural_cooking	urban_cooking	electricy	rural_electricy
2001	100.0	7.981228	argentina	0.631579	2.824859	0.203459	-0.284209	-4.463882
2001	100.0	7.981228	belgium	0.000000	0.000000	0.000000	0.000000	0.000000
2001	100.0	7.981228	brazil	0.786517	2.177858	0.103093	1.721347	10.048946
2001	100.0	7.981228	canada	0.000000	0.000000	0.000000	0.000000	0.000000
2001	100.0	7.981228	china	0.714286	-0.425532	-0.073153	0.258696	0.452353

Figure 5: Snippet of the merged bulk-knowledge data frame

3.4.1 Simple Linear Regression

To look for correlations between change in SDG 7 indicators and the citation impact metrics, we plotted and computed simple linear regressions. Table 2 summarizes the 156 simple linear regression conducted on each of the bulk knowledge and country-specific data sets.

The main assumptions we tested were for linearity and residual normality. First, for all linear regressions, the relationship between the response and predictors was linear. This was tested using the residuals vs fitted values plots.

Second, for all significant regressions, we checked if the residuals were normal using the Shapiro-Wilks test [24]. None of the significant regressions has normally distributed residuals.

Therefore, as none of the significant regressions met the residual normality assumption, we moved to quantile regression.

3.4.2 Quantile Regression

Quantile regression is an extension of linear regression that is used when the conditions of linear regression are not met [25]. This method, instead of estimating the mean of the response like in linear regression, estimates the median of the response when we set the quantile (τ) to 0.5.

All linear regressions in table 2 were repeated using quantile regressions at $\tau = 0.5$.

# of indicators = # of regressions	Response (Y)	Predictor (X)
27	% change of indicator	log(citations in 2 years)
27	% change of indicator	AR-index
27	log difference	log(citations in 2 years)
27	log difference	AR-index
12	distance measure for OECD countries	log(citations in 2 years)
12	distance measure for OECD countries	AR-index
12	distance measure for all countries	log(citations in 2 years)
12	distance measure for all countries	AR-index

Table 2: All simple linear regressions done: 156 regression per data frame i.e. a total of 312 regressions.

We have discussed the 4 strongest, interpretable signals found using this approach in the next section. These signals were for the SDG 7 target pertaining to energy efficiency.

4 Results and Discussion

4.1 Energy Efficiency and Significant Quantile Regressions

Energy efficiency is the third target of SDG 7, which 'calls for global progress on energy efficiency by doubling the rate of improvement in energy efficiency globally by 2030' [26].

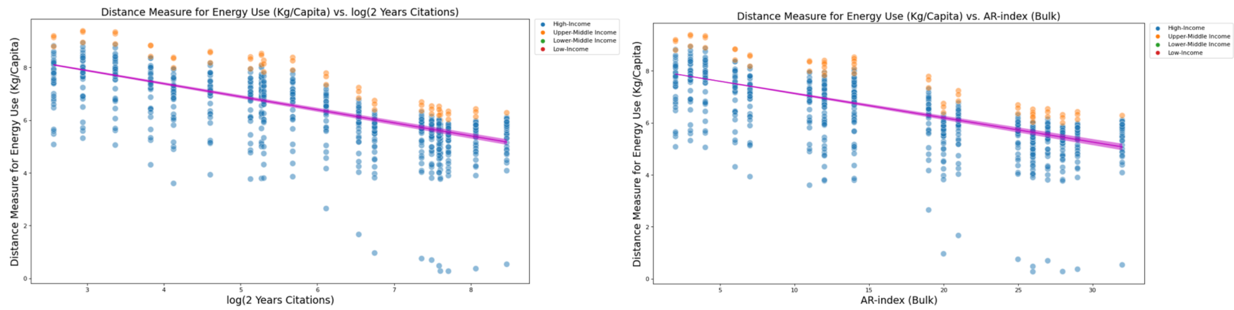
Our significant quantile regressions pertain to an indicator that measures energy efficiency. It is the energy use (kg per capita) indicator defined as "the use of primary energy before transformation to other end-use fuels." [27].

4.1.1 Regression results with Bulk Knowledge data

In the bulk knowledge data set, only the distance measurement of energy use for OECD countries was significant (plots and regression summaries in figure 6). The distance measure of energy use is decreasing with the increase of log(2 years) and AR-index. The average test MSE across the 5-fold cross-validation for the model with log(2 years) is 1.0297, and model with the AR-index is 1.0932. A 95% confident interval is visible around the regression line, and we are 95% confident that the true regression line is between the interval.

4.1.2 Regression results with Country-specific data

Same as the bulk knowledge data results, for the country-specific data set, only the distance measure of energy use for OECD was significant (plots and regression summaries in figure 7). We saw the same trend as before: distance measure of energy use decreases as the two




```
                         QuantReg Regression Results
==============================================================================
Dep. Variable:     energy_use_kg_capita   Pseudo R-squared:             0.3342
Model:                         QuantReg   Bandwidth:                    0.4041
Method:                   Least Squares   Sparsity:                      1.761
Date:                  Sat, 17 Dec 2022   No. Observations:                756
Time:                          20:22:40   Df Residuals:                    754
                                          Df Model:                        1
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      9.3604      0.110     84.988      0.000       9.144       9.577
log_2_Years   -0.4952      0.018    -27.313      0.000      -0.531      -0.460
==============================================================================
```

```
                         QuantReg Regression Results
==============================================================================
Dep. Variable:     energy_use_kg_capita   Pseudo R-squared:             0.3359
Model:                         QuantReg   Bandwidth:                    0.3996
Method:                   Least Squares   Sparsity:                      1.742
Date:                  Sat, 17 Dec 2022   No. Observations:                756
Time:                          20:23:08   Df Residuals:                    754
                                          Df Model:                        1
================================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept       8.0579      0.065    123.456      0.000       7.930       8.186
ar_index_bulk  -0.0934      0.003    -27.540      0.000      -0.100      -0.087
================================================================================
```

Figure 6: Bulk knowledge regression plots and summaries for distance measure of energy use (kg/capita) vs log(2 Years Citations) and vs AR-index

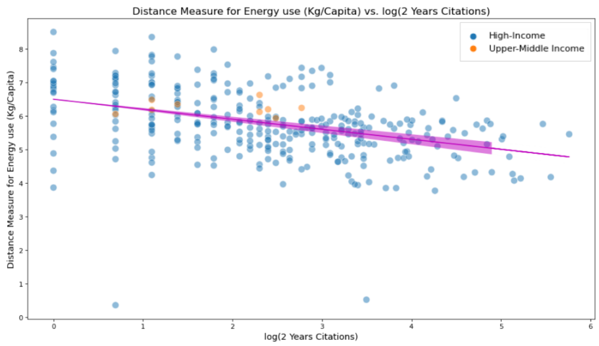
predictors increase. The average test MSE across the 5-fold cross-validation for the model with log(2 years) is 0.9510, and the AR-index is 0.9938. A 95% confident interval is visible around the regression line, and we are 95% confident that the true regression line is between the interval.

4.1.3 Discussion

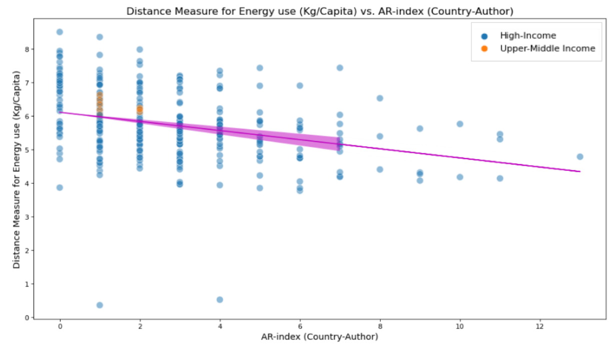
The distance measure tells us how far a country is from reaching the target. Therefore, the distance measure for energy use tells us how far countries are from reaching the energy use target. For both, bulk knowledge and country-specific, as citation impact increases, distance measure decreases. For bulk knowledge, this means that as the spread and influence of research related to SDG 7 on a global level increases, OECD countries are closer to reaching the energy use target. For country-specific, we observed the same trend and interpret that as research related to SDG 7 spreads in OECD countries, these countries are closer to reaching the target.

OECD countries are mainly high-income and upper-middle income countries (as can be seen in the plots in figures 6 and 7. As high-income countries use almost 5 times as much energy on a per capita basis, than lower-middle and low-income countries [27], improvements towards energy use (and thus improvement in energy efficiency) are very important. An improvement in energy efficiency is often the most economic and readily available means of improving energy security and reducing greenhouse gas emissions.[27]

Moreover, in terms of answering our research question, we observed a positive correlation between increase in SDG 7 related research impact and progress towards the energy efficiency




```
                         QuantReg Regression Results
==============================================================================
Dep. Variable:     energy_use_kg_capita   Pseudo R-squared:            0.07756
Model:                         QuantReg   Bandwidth:                    0.5756
Method:                   Least Squares   Sparsity:                      2.112
Date:                  Sat, 17 Dec 2022   No. Observations:                355
Time:                          21:55:57   Df Residuals:                    353
                                          Df Model:                        1
=============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------
Intercept      6.5000      0.114     56.808      0.000       6.275       6.725
log_2_Years   -0.2983      0.041     -7.193      0.000      -0.380      -0.217
```




```
                         QuantReg Regression Results
==============================================================================
Dep. Variable:     energy_use_kg_capita   Pseudo R-squared:            0.04355
Model:                         QuantReg   Bandwidth:                    0.6371
Method:                   Least Squares   Sparsity:                      2.235
Date:                  Sat, 17 Dec 2022   No. Observations:                355
Time:                          21:56:37   Df Residuals:                    353
                                          Df Model:                        1
=============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------
Intercept      6.1115      0.091     67.441      0.000       5.933       6.290
ar_index_ca   -0.1361      0.025     -5.411      0.000      -0.186      -0.087
```

Figure 7: Country-Specific regression plots and summaries for distance measure of energy use (kg/capita) vs log(2 Years Citations) and vs AR-index

target for only OECD countries. We did not see correlations for other targets and other countries.

Specifically regarding the quantile regression results for the bulk knowledge data, most of the high research participation countries that make up the majority of the citation impact data (except China) are in the OECD countries. It is difficult to conclude whether non-OECD countries' research impact helps OECD countries' energy efficiency progress. Importantly, China has the highest research participation in our data and is a non-OECD. It is possible that Chinese SDG 7 research may help energy efficiency progress for OECD countries but more research is needed to determine that.

5 Conclusion

5.1 Limitations

Our current approach has lacks target values for each indicator of SDG 7. The UN-STAT defines goals for each SDG target but does not set individual numerical targets on a country-wise basis[28]. The distance measure that we proposed requires a target value for a country or group of countries at the same economic, educational, and SDG level. Including other countries in the distance measure calculation using OECD-specific targets did not lead to any significant quantile regressions, as was expected. Targets need to be calculated for all countries to be able to measure the global progress towards SDGs in general.

Another limitation in our findings was that quantile regressions for the distance measure indicators were significant only for OECD countries, which are all high-income and upper-middle income countries. It is unclear whether global research impact helps lower-middle and low-income countries to improve their SDG 7 indicators. There was also an inability to generalize these results for all countries and SDGs. They may only be generalizable to SDG 7 progress in countries similar in economic stature to OECD countries, but not to overall SDG progress (our research only pertains to SDG 7).

5.2 Biases

Our citation impact dataset contained research done primarily by upper-middle and high-income countries, leading to selection bias towards these countries.

Additionally, the citation impact metrics were only calculated for academic research published in the English language. We are potentially missing data where academic research is published in non-English languages. Lastly, only academic research was included, which means potentially beneficial and impactful research conducted by private companies or other non-academic organizations was excluded. All this skews data towards high-income and upper-middle income countries, which generally have more academic institutions and higher research participation than lower-middle or low-income countries.

5.3 Broader Impacts

Sustainable development academic research has an important role to play towards SDG 7 progress. Our research helps measure SDG 7 progress, relating it directly to academic research related to SDG 7 and finding strong signals for further analyses. Academic research in sustainable development should be prioritized as it can help towards SDG progress.

Progress being made on an SDG indicator can now be measured using citation impact metrics, comparing it to the distance measure to the SDG indicator target model we used. As most sustainability research on SDGs are on a higher goal level, this research methodology can be used to assess SDG performance at more granular levels for other SDGs as well.

Countries with lower research participation rates may also benefit from collective academic research related to SDG 7. Further research should be conducted to understand the relation between the global (bulk) research pool and SDG 7 progress for low research participation countries.

References

[1] Owsianiak M. Dong Y. Kravchenko M. Molin C. Hauschild M. Z. Laurent, A. Assessing the sustainability implications of research projects against the 17 un sustainable

development goals. *Procedia CIRP*, 90, 2020.

[2] Jurgilevich A. Pietikäinen J. Korhonen-Kurki K Soini, K. Universities responding to the call for sustainability: A typology of sustainability centres. *Journal of Cleaner Production*, 170, 2018.

[3] Goal 7 — department of economic and social affairs. `https://sdgs.un.org/goals/goal7`.

[4] Wdi: Sdg goals and targets. `https://datatopics.worldbank.org/sdgs/sdg-goals-targets.html`.

[5] Hritonenko N. Yatsenko Y. Bréchet, T. Domestic environmental policy and international cooperation for global commons. *Resource and Energy Economics*, 44, 2016.

[6] ter Horst E. Gunn L. H. Zambrano J. D. Molina G. Jaffe, K. A network analysis of research productivity by country, discipline, and wealth. *PLOS ONE*, 15, 2020.

[7] Ferré M. Martin-Ortega J. Blanche R. Lawford-Rolfe R. Dallimer M. Holden J. Reed, M. S. Evaluating impact from research: A methodological framework. *Research Policy*, 50(4), 2021.

[8] Measuring your impact: Impact factor, citation analysis, and other metrics: Citation analysis. `https://researchguides.uic.edu/c.php?g=252299&p=1683205`.

[9] Ludo Waltman. A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2):365–391, 2016.

[10] Liang L. M.-Rousseau R. Egghe L. Jin, B. H. The r- and ar-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 2007.

[11] Pablo González and I. Dorta. Impact maturity times and citation time windows: The 2-year maximum journal impact factor. *Journal of Informetrics*, 7:593–602, 04 2013.

[12] Dorta I. González, P. Impact maturity times and citation time windows: The 2-year maximum journal impact factor. *Journal of Informetrics*, 7:593–602, 04 2013.

[13] Dorta I. González, P. Do fixed citation windows affect the impact maturation rates of scientific journals? *Journal of Informetrics*, 30, 2016.

[14] L. Bornmann. Measuring the societal impact of research. *EMBO Reports*, 13(8), 2012.

[15] Williams C Curran-Everett, D. Explorations in statistics: The analysis of change. *Advances in Physiology Education*, 39(2), 2015.

[16] Vartia P. Vartia Y. O. Tornqvist, L. How should relative changes be measured? the american statistician. *The American Statistician*, 39(1), 1985.

[17] A. Miola and F. Schiltz. Measuring sustainable development goals performance: How to monitor policy action in the 2030 agenda implementation? *Ecological Economics*, 164:106373, 2019.

[18] OECD. Measuring distance to the sdg targets. *OECD*, 2017.

[19] How does the world bank classify countries? `https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries`.

[20] Caroline Birkle, David A. Pendlebury, Joshua Schnell, and Jonathan Adams. Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1):363–376, 02 2020.

[21] Alexandre M. Bayen and Timmy Siauw. Chapter 14 - interpolation. In Alexandre M. Bayen and Timmy Siauw, editors, *An Introduction to MATLAB® Programming and Numerical Methods for Engineers*, pages 211–223. Academic Press, Boston, 2015.

[22] A.M.M. Al-Bakir. Filling missing data using interpolation methods: Study on the effect of fitting distribution. *Key Engineering Materials*, 594-595:889–895, 01 2014.

[23] M. Shinwell Cohen, G. How far are oecd countries from achieving sdg targets for women and girls?: Applying a gender lens to measuring distance to sdg targets. *OECD Statistics Working Papers*, 2020.

[24] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[25] Robert N. Rodriguez. Five things you should know about quantile regression. 2017.

[26] Sdg 7.2 - renewable energy. `https://www.seforall.org/goal-7-targets/renewable-energy`, Nov 2022.

[27] `https://databank.worldbank.org/metadataglossary/world-development-indicators/series/EG.USE.PCAP.KG.OE`.

[28] Sdg indicators (2019). `https://unstats.un.org/sdgs/indicators/indicators-list/`.

[29] "the oecd creates better policies for better lives. read the oecd's main figures to find out more about their work.". `https://www.oecd.org/about/`.

A Appendix

A.1 Details on Citation Impact Metrics

The citation impacts were calculated two separate ways, one that considers the entirety of academic research as a collective pool for all countries' access and development (Bulk Knowledge), and the other that takes a more granular approach by attributing the country's research prowess to its own SDG 7 progress (Country-Specific).

In the case of Bulk knowledge, log(2 years) was calculated by summing all citations in a 2 year window for all papers published in a particular year. We then took the log of these sums to normalize the dataset on a log-scale. This also accounted for any particularly aggressive outliers in our dataset. The AR Index was calculated by summing the total citations a paper received in its entire lifespan, then aggregating the dataset by year in a descending order of total citations of a paper. We then considered h papers' citation counts such that the first h articles in that year received at least h citations, and calculated the AR-Index using the formula described in section 2.1.

In our Country-specific data, log(2 years) was calculated by summing all citations in a 2 year window for all papers published by a country in a particular year. We then took the log of these sums to normalize the dataset on a log-scale. The AR Index was calculated by summing the total citations a paper received in its entire lifespan, aggregating the dataset by country and subsequently by year in a descending order of total citations of a paper. We then considered h papers' citation counts such that the first h articles published by a country in that year received at least h citations, and calculated the AR-Index using the formula described in section 2.1.

A.2 SDG 7 Indicators

The 27 SDG 7 indicators measuring three main SDG 7 targets are:

1. Access to clean fuels and technologies for cooking (% of population)
2. Access to clean fuels and technologies for cooking, rural (% of rural population)
3. Access to clean fuels and technologies for cooking, urban (% of urban population)
4. Access to electricity (% of population)
5. Access to electricity, rural (% of rural population)
6. Access to electricity, urban (% of urban population)
7. Alternative and nuclear energy (% of total energy use)
8. Combustible renewables and waste (% of total energy)

Income Group	GNI per Capita Threshold
High-Income	greater than $12,535
Upper-Middle Income	$4,046 -$12,535
Lower-Middle Income	$1,035 – $4,045
Low-Income	less than $1,035

Table 3: World Bank Income Classifications for 2020 [19]

9. Electric power consumption (kWh per capita)
10. Electric power transmission and distribution losses (% of output)
11. Electricity production from coal sources (% of total)
12. Electricity production from hydroelectric sources (% of total)
13. Electricity production from natural gas sources (% of total)
14. Electricity production from nuclear sources (% of total)
15. Electricity production from oil sources (% of total)
16. Electricity production from oil, gas and coal sources (% of total)
17. Electricity production from renewable sources, excluding hydroelectric (% of total)
18. Electricity production from renewable sources, excluding hydroelectric (kWh)
19. Energy imports, net (% of energy use)
20. Energy intensity level of primary energy (MJ/$2017 PPP GDP)
21. Energy use (kg of oil equivalent per capita)
22. Energy use (kg of oil equivalent) per $1,000 GDP (constant 2017 PPP)
23. Fossil fuel energy consumption (% of total)
24. GDP per unit of energy use (constant 2017 PPP $ per kg of oil equivalent)
25. GDP per unit of energy use (PPP $ per kg of oil equivalent)
26. Renewable electricity output (% of total electricity output)
27. Renewable energy consumption (% of total final energy consumption)

A.3 World Bank Income Classification Table

See table 3.

A.4 OECD Countries

OECD desribes itself as 'an international organisation that works to build better policies for better lives. Our goal is to shape policies that foster prosperity, equality, opportunity and well-being for all. We draw on 60 years of experience and insights to better prepare the world of tomorrow. Together with governments, policy makers and citizens, we work on establishing evidence-based international standards and finding solutions to a range of social, economic and environmental challenges. From improving economic performance and creating jobs to fostering strong education and fighting international tax evasion, we provide a unique forum and knowledge hub for data and analysis, exchange of experiences, best-practice sharing, and advice on public policies and international standard-setting.' [29]

38 countries are part of the OECD and these are: Australia, Austria, Belgium, canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, USA, UK, Colombia, Costa Rica, South Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, and Poland.