Assessing the Relationship Between Sustainability Research and Sustainable Development Goal 7

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

Content

Introduction

- In 2015, United Nations (UN) set 17 interlinked Sustainable Development Goals (SDGs) to be achieved by 2030. These goals were adopted by all 193 UN member countries.
- UN describes them as a "shared blueprint for peace and prosperity for people and the planet, now and into the future"

Introduction

- We focus on SDG 7 which is described by the UN as **'ensuring access to affordable, reliable, sustainable and modern energy for all'**
- UN defines 3 main targets for SDG 7:
  - **Access to energy**: 'By 2030, ensure universal access to affordable, reliable and modern energy services'
  - **Increased use of renewable energy**: 'By 2030, increase substantially the share of renewable energy in the global energy mix'
  - **Energy efficiency**: 'By 2030, double the global rate of improvement in energy efficiency'
- The 3 targets have sub-targets. SDG 7 is measured by a total of 27 indicators by the World Bank. These include numerical measures like:
  - Proportion of population with access to electricity
  - Renewable energy share in the total final energy consumption

Introduction

- While countries are working to reach these SDG targets, academia is simultaneously producing more sustainable development related research
- Some of the major research is happening in the clean and affordable energy space
- For example: research on renewable energies, energy efficiency, electric vehicles, grid modernization

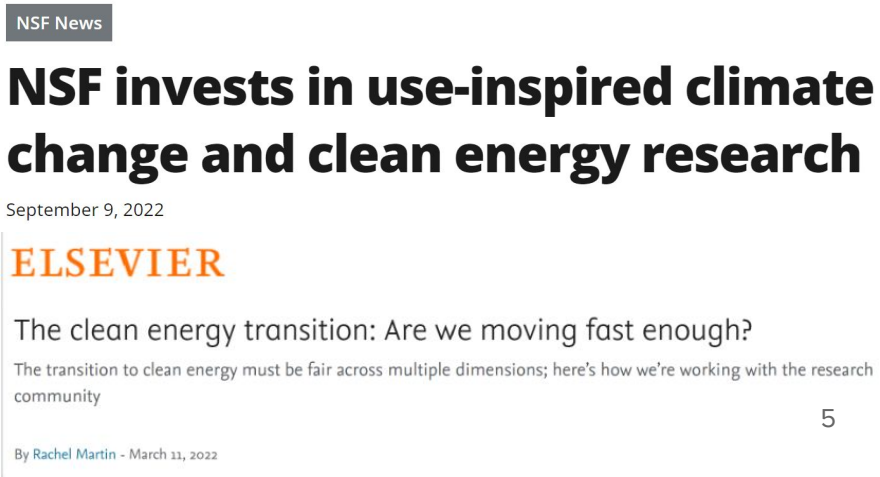

SCIENCE|BUSINESS®

Academic sector makes moves to going green

12 May 2022 | News

As the EU transitions towards a net zero economy, universities are stepping up to play their part. But funding is needed for initiatives such as greening ancient buildings

By Goda Naujokaitytė


NSF News

NSF invests in use-inspired climate change and clean energy research

September 9, 2022

ELSEVIER

The clean energy transition: Are we moving fast enough?

The transition to clean energy must be fair across multiple dimensions; here's how we're working with the research community

By Rachel Martin - March 11, 2022

Introduction

- Our project focuses on understanding if there is a correlation between the progress being made on SDG 7 compared to impactful academic research being done on SDG 7 related targets.
  - Does academic research influence SDG 7 progress?
- Research question:
  - **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?**
- Important parameters to note:
  - **We are looking at academic research and SDG 7 indicator data on a global scale**
    - Not focusing on only one country's progress towards SDG 7
  - **Data ranges from 1991 to 2020**
    - Majority of our data is prior to the official definition of SDGs established by the UN (2015)

Background: Research Impact

- How is academic research impact quantified?
- Impact is measured using citation counts, also known as Citation Impact
  - **Citation impact**: A measure of how many times an academic journal article, book, or author is cited by other articles, books, and authors.
- Citation impact can be calculated in many ways. We are using two following calculations, based on our literature review:
  - **AR-index**
    - Research impact of paper scaled by age of publication
  - **Log of citations in 2 years on publication year**
    - Total citation counts in 2 years of paper's publication
  - We will go into more details about these in the feature engineering slides.

3 Main Contributions

1. **We are focusing on SDG 7 and measuring progress for each individual SDG 7 target**
    - Prior research on SDG progress primarily focused on the SDGs as a whole, not at individual SDG targets
2. **We are using and comparing 3 different methods to calculate the relative change in SDG 7 indicators**
    - Generally, it is difficult to measure SDG progress for each country using the raw data because different countries progress at different rates and have different starting points.
    - The methods we use to track relative changes are:
        - **Year-on-year percentage change**
        - **Year-on-year logarithmic difference**
        - **Year-on-year distance measure of indicator from 2030 end-value**

3 Main Contributions

3. We're looking at understanding SDG progress from an academic research lens

- Most research looks at SDG progress through a policy lens.
- Our project looks at academic research impact from two perspectives:
  - **'Bulk Knowledge'**
    - Measuring citation impact of research produced globally (in bulk) towards SDG 7 progress for individual countries.
    - Research done (mostly produced in high-income and upper-middle income countries) has a global impact towards SDG 7, even in countries with lower participation in research (which are often lower-middle and low income countries)
  - **'Country-Specific'**
    - Measuring citation impact of research produced by a country on its own SDG 7 progress.
    - Not assuming that citation impacts of research from a country contribute to SDG 7 progress for all countries.

3 Data Sets

- **Citation Impact data**
  - Comes from the Web of Science (WoS) 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.
- **SDG 7 indicators data**
  - Collected from the World Bank website (World Development Indicator dataset) for all 193 UN member countries, for all 27 indicators measuring SDG 7
- **World Bank income level data**
  - WB classifies countries by income level using GNI per capita in US Dollars
  - 4 classifications: High-income, upper-middle income, lower-middle income, low-income

Data Set 1: Citation Impact Data

- Relevant columns: Title, Abstract, Keywords describing paper, Publication Year, Citation counts for every year since publication, and Author Nationalities
- Filtered to get research related to SDG 7 by only getting papers where the title, keywords, or abstract contained the words **'energy'** or **'electricity'**
  - The earliest published paper from this criteria was in 1991.
  - Data available for 1991-2020

	AU	DE	TI	AB	1901.0	1902.0	1903.0	...	2017.0	2018.0	2019.0	2020.0	2021.0	Nationalities	Unique Nationalities
0	Sztumski, W	sustainable development; ideology of sustainab...	The Mythology of Sustainable Development	In this article I present a proposition that a...	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	[[ ]]	[[ ]]
1	Degai, TS; Petrov, AN	sustainable development; Arctic; Indigenous pe...	Rethinking Arctic sustainable development agen...	The year 2020 has urged Humanity to rethink th...	0.0	0.0	0	...	0	0	0	0	1	[Russia, Russia, USA, Russia]	[Russia, USA]
2	Kanapathy, S; Lee, KE; Mokhtar, M; Sivapalan, ...	Education for sustainable development; Chemist...	Enculturing Sustainable Development Concept Th...	The enculturation of sustainable development c...	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	[Malaysia, Malaysia, Malaysia, Malaysia]	[Malaysia]
3	Selvi, M; Selvi, M	environmental education; sustainable developme...	Environmental education for sustainable develo...	Sustainable development has been a key concept...	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	[Turkey]	[Turkey]

Citation Impact Data: Feature Engineering

‘Bulk Knowledge’

- Calculate 2 research impact metrics without considering the authors’ country of origin.
- The goal of this approach is to see if global (bulk) research impact has an impact on SDG 7 progress for all.

	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

‘Country-specific’

- Calculate 2 research impact metrics 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.

	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

Citation Impact Data: Feature Engineering

- **Log(Number of Citations in 2 Years)**
  - **Calculation**
    - Log of the sum of citation counts in the 2 years after an article is published.
  - **Advantages**
    - Removes the issue of older publications having more citations.
    - Adjusts for outliers by taking the log of each sum.
  - **Disadvantages**
    - Does not accommodate for how the article's relevance changes with progress in research.
- **AR-index**
  - **Calculation**
    - The root sum of citations of the first h articles in a year, divided by its age.
    - h is defined as the the highest rank such that the first h publications in a year received each at least h citations.
  - **Advantages**
    - AR takes the age of the publications into account.
    - Allows for an index that can actually decrease over time.

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

Data Set 2: SDG 7 Indicators

- 27 indicators measuring the 3 SDG 7 targets
  - Access to energy, increased use of renewable energy, energy efficiency
  - Data is available for each year from 1960 to 2021, for all UN member countries
  - Countries report data at different frequencies
    - **Forward Linear Interpolation** was done to fill gaps in years where countries don't typically report data
  - Data outside the range of interpolation was dropped (~20% of 6650 rows)

	Country Name	Country Code	Indicator Name	Indicator Code	1996	1997	1998	...	2014	2015	2016	2017	2018	2019	2020
72100	albania	ALB	Access to clean fuels and technologies for coo...	EG.CFT.ACCS.ZS	NaN	NaN	NaN	...	74.600000	76.200000	77.800000	78.700000	79.8	80.70	81.3
72101	albania	ALB	Access to clean fuels and technologies for coo...	EG.CFT.ACCS.RU.ZS	NaN	NaN	NaN	...	55.700000	57.600000	59.050000	60.500000	62.5	63.25	64.6
72102	albania	ALB	Access to clean fuels and technologies for coo...	EG.CFT.ACCS.UR.ZS	NaN	NaN	NaN	...	89.900000	90.700000	91.100000	91.600000	92.2	92.30	92.9
72103	albania	ALB	Access to electricity (% of population)	EG.ELC.ACCS.ZS	100.0	100.0	100.0	...	99.949997	99.980003	99.889999	99.889999	100.0	100.00	100.0

SDG 7 Indicators: Feature Engineering

Calculate relative change of indicators over time

- **% change**
  - Common statistical method to calculate relative change
  - Issue: does not take into account that different countries have different starting points
    - For example: if country A has 15% electrification in 2010 and 20% electrification in 2011; and country B has 90% electrification in 2010 and 95% electrification in 2011. Both have a +5% change for 2011, but country B is ahead of country A in electrification; countries are in different phases of development.

$$\text{Percent Change} = \frac{\text{New Value} - \text{Old Value}}{\text{Old Value}} \times 100\%$$

- **Log difference**
  - Common statistical method to calculate relative change
  - Values are symmetrical going forward and backward
  - Same issue as above

$$Log_a\left(\frac{x}{y}\right) = Log_a(x) - Log_a(y)$$

SDG 7 Indicators: Feature Engineering

Calculate relative change of indicators over time

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

- **Distance measure (aka 'Modified z-score')**
  - Tells us that Country X is Y standard deviations short of target T
    - T: target value that countries have to meet by 2030 for an indicator
    - Standard deviation: measured across indicator value for all countries in a specific year
    - X: the current indicator value for country X in a specific year
  - Resolves the issue of countries starting at different points because T is country-specific
  - Issues
    - End-values (T) for SDG 7 targets are not available for all countries
      - Need to be calculated on a country-by-country basis (based on economy, national goals, resources, starting point)
      - Cannot generalize one target value for all countries
      - *For example: if T for access to electricity is 100%. This might not be a viable target for a low-income country but 50% may be*
    - 12 indicator end-values are available for 38 countries part of the Organization for Economic Cooperation and Development (OECD). These end-values were calculated by OECD.
      - Mix of high-income and upper-middle income countries including US, UK, Canada
      - **We are using these values to calculate distance measure for only OECD countries and all countries**

https://www.oecd.org/wise/measuring-distance-to-the-sdgs-targets.htm

Main data frames: Merging the 3 data sets

- **Citation Impact Metrics + SDG 7 Indicators + Income-Level Classification**
- Two main data frames
  - **Bulk knowledge data frame and Country-specific data frame**
  - Have the same columns, just different ways of calculating citation impact
    - Year, country, income-level
    - **2 Predictor variables:** *log(2 year citations), AR-index*
    - **4 Response variables:**
      - % change of all indicators
      - log difference of all indicators
      - distance measure calculated for OECD countries
      - distance measure calculated for all countries using OECD target end-values.

Simple Linear Regression

# 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

156 regressions were done for each of the 2 main data frames

Total: 156

Simple Linear Regression: Assumptions

- **Assumption 1: Linearity**
  - Checked using residual vs fitted value plots
    - No heteroskedasticity or patterns indicating non-linearity were observed for all regressions
- Assumption 2: Normality of residuals
  - Checked using Shapiro-Wilk test (if p-value<0.05, not normally distributed)
    - For regressions where predictor was significant (p-value < 0.05), residual normality was tested
    - **None of the significant regressions had normal residuals**

Quantile Regression

- Quantile regression is an extension of linear regression that is used when the conditions of linear regression are not met (normality of residuals)
- Instead of estimating the mean of the response like OLS, QR estimates the **median of the response** (quantile = 0.5)
- Found some **significant regressions** showing **strong signals** using QR

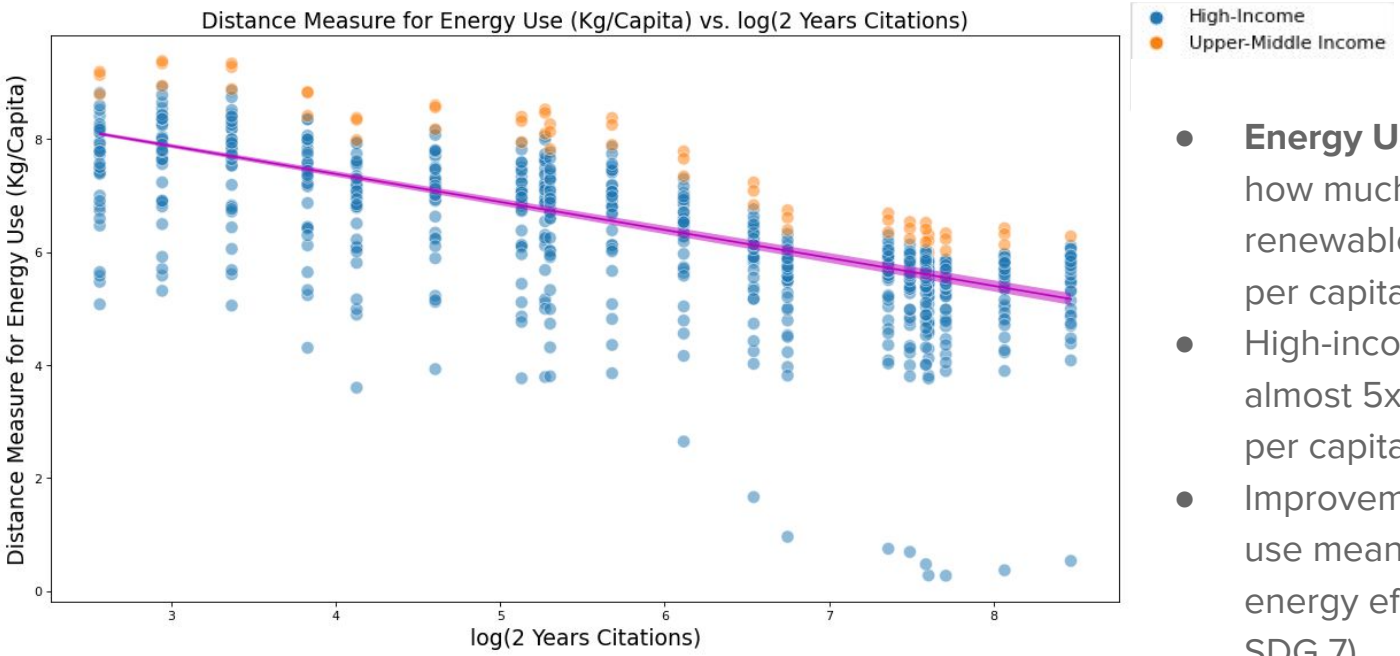
https://towardsdatascience.com/quantile-regression-ff2343c4a03

QR analysis on Bulk Knowledge Dataframe

(only distance measure regressions for OECD were significant)

- **X (Predictor)**
  - **log(citation in 2 years)**
  - **AR-Index**
- **Y (Response)**
  - **Distance measurement on SDG 7 indicator for OECD**

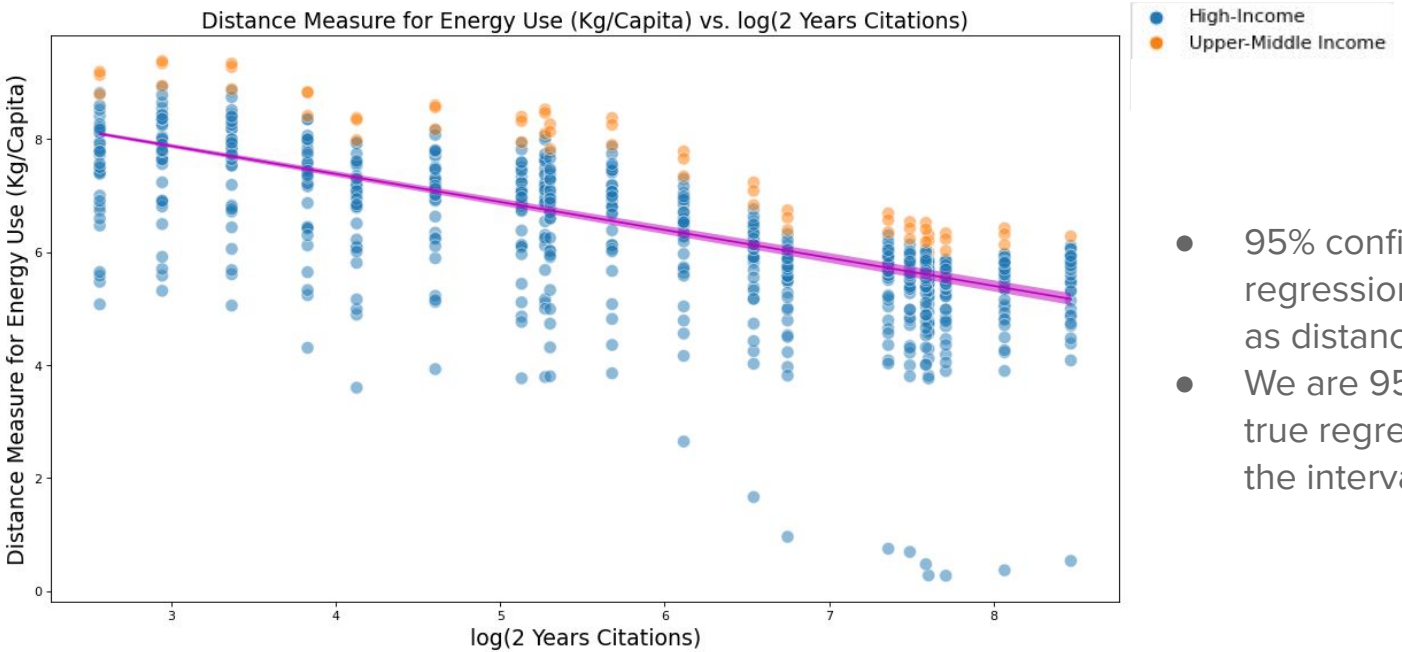
Distance measure of Energy Use (Kg/Capita) vs log(citations in 2 years) [only OECD]



- **Energy Use** is a measure of how much energy from both renewables fossil fuels is used per capita.
- High-income countries use almost 5x as much energy on a per capita basis.
- Improvement towards energy use means improvements in energy efficiency (Target# 3 of SDG 7)

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

Distance measure of Energy Use (Kg/Capita) vs log(citations in 2 years) [only OECD]



- 95% confidence band around regression line becomes wider, as distance measure decreases.
- We are 95% confident that the true regression line is between the interval.

Distance measure of Energy Use (Kg/Capita) vs log(citations in 2 years) [only OECD]

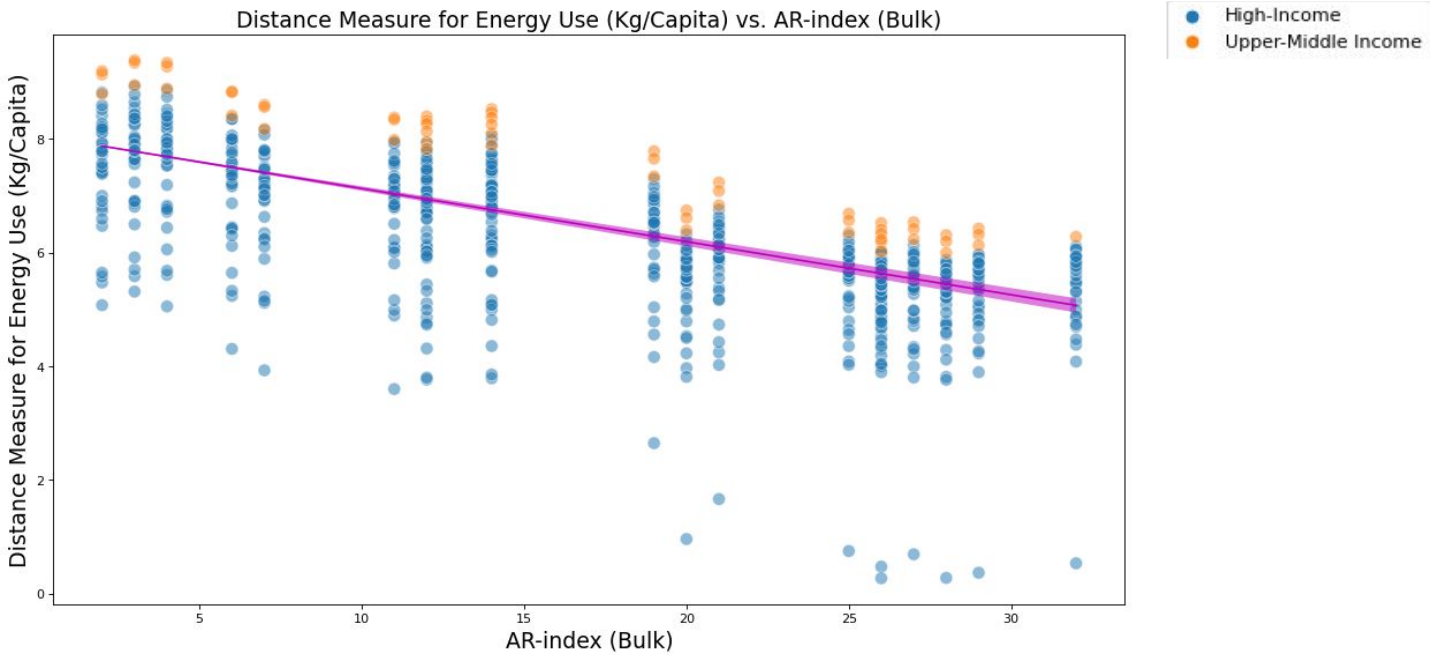
QuantReg Regression Results

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

- log(citation in 2 years) is **statistically significant**
- Average 5-fold cross validation test **MSE: 1.0927**

Distance measure of Energy Use (Kg/Capita) vs AR-Index [only OECD]



Distance measure of Energy Use (Kg/Capita) vs AR-Index [only OECD]

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

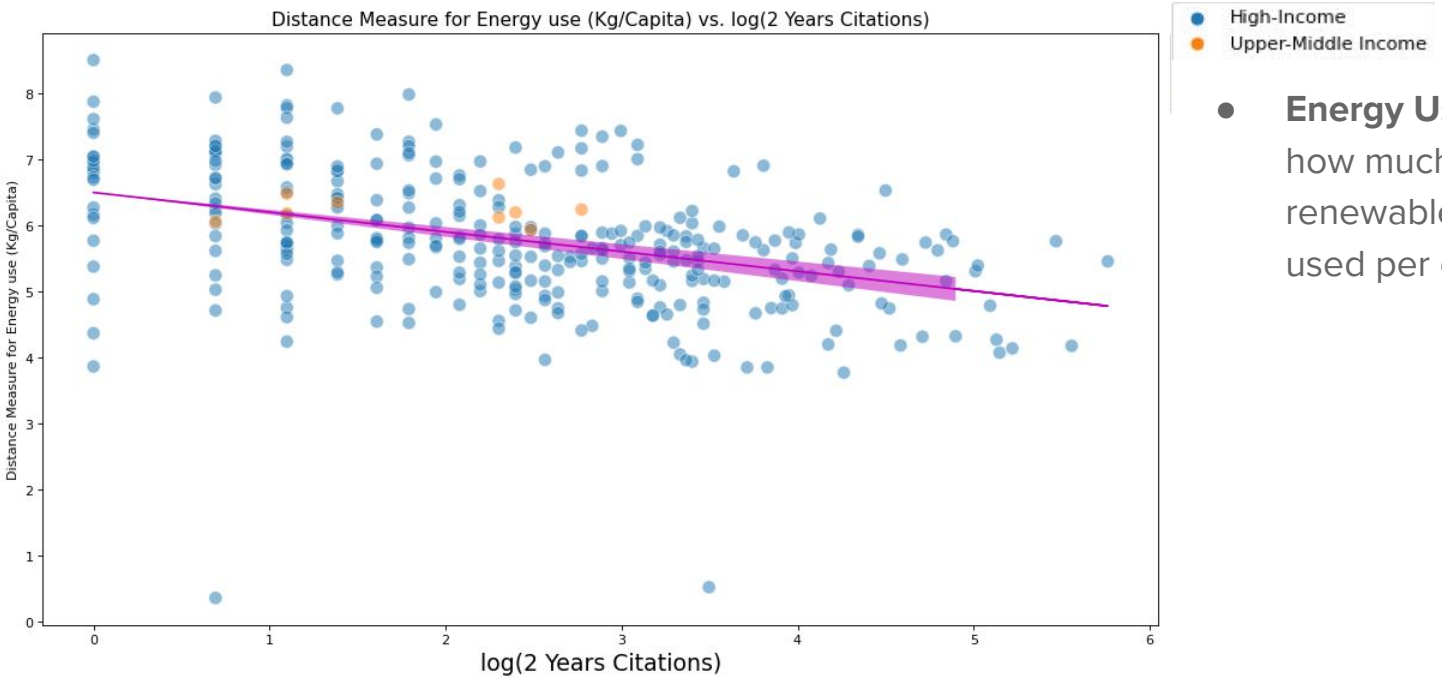
- AR-index is **statistically significant**
- Average 5-fold cross validation test **MSE: 1.0932**

QR analysis on Country Specific Dataframe

(only distance measure regressions for OECD were significant)

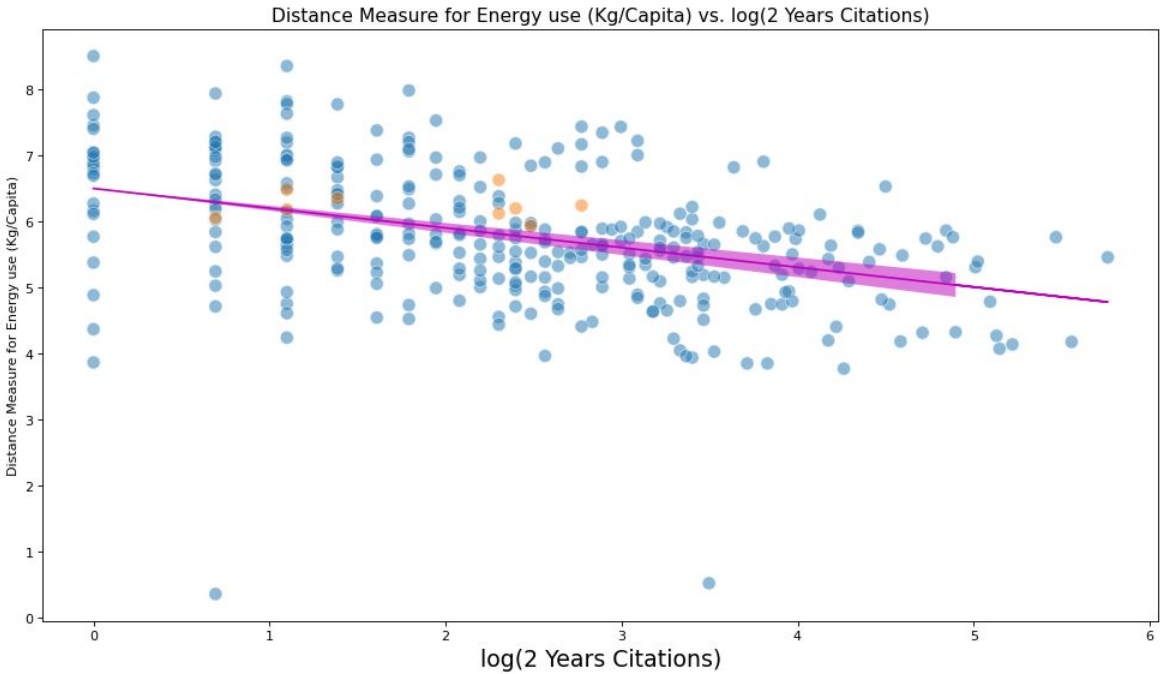
- **X (Predictor)**
  - **log(citation in 2 years)**
  - **AR-Index**
- **Y (Response)**
  - **Distance measurement on SDG 7 indicator for OECD**

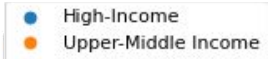
Distance measure for Energy use (Kg/Capita) vs log(2 Years Citations) [only OECD]



- **Energy Use** is a measure of how much energy from both renewables fossil fuels is used per capita.

Distance measure for Energy use (Kg/Capita) vs log(2 Years Citations) [only OECD]





- 95% confidence band around regression line becomes wider with decrease of distance measure and becomes fuzzy around the end with minimal data points.
- We are 95% confident that the true regression line is between the interval.

Distance measure for Energy use (Kg/Capita) vs log(2 Years Citations) [only OECD]

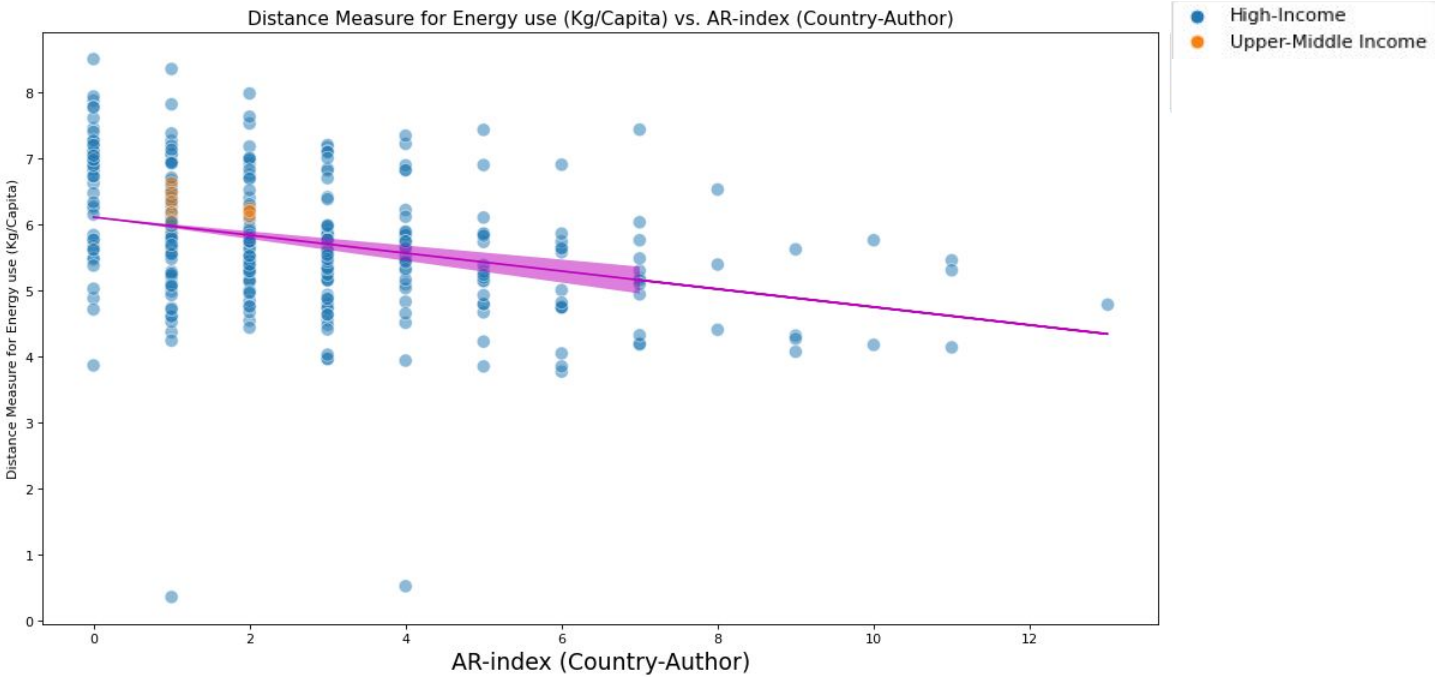
QuantReg Regression Results

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

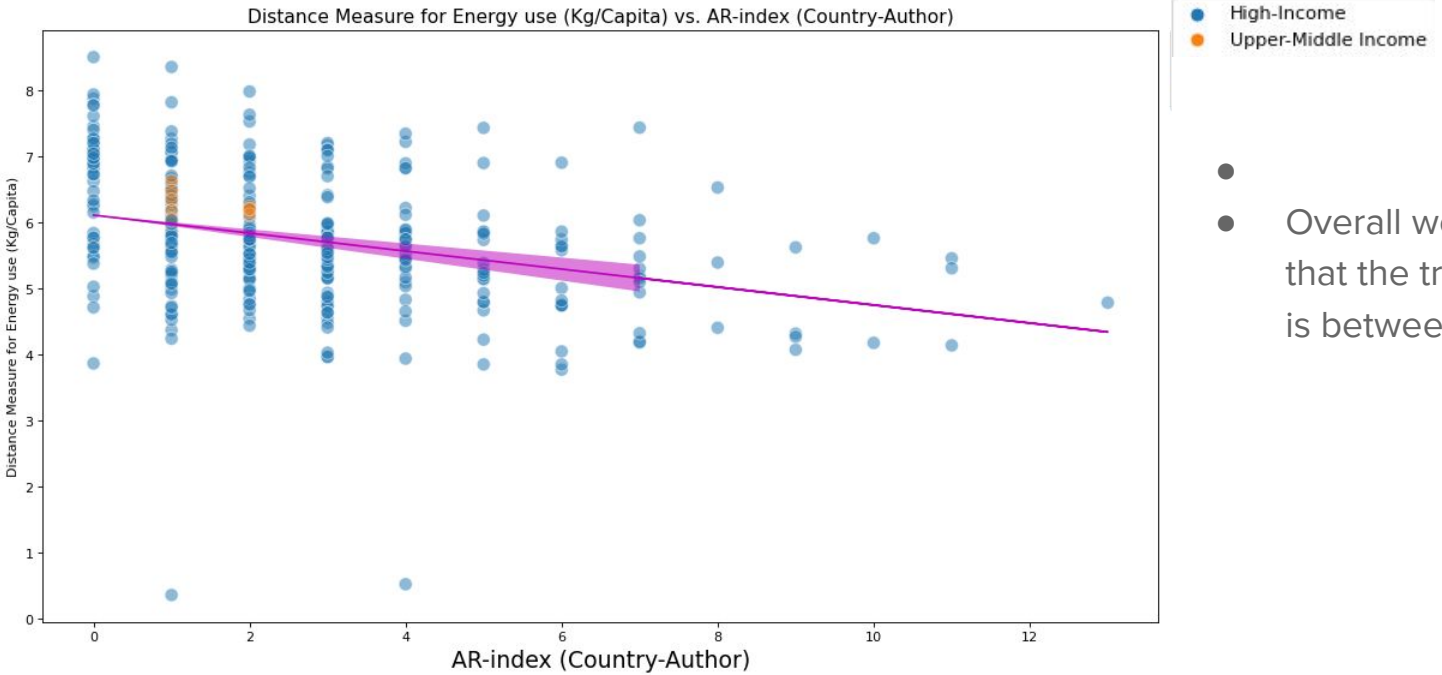
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.5000	0.114	56.808	0.000	6.275	6.725
log_2_Years	-0.2983	0.041	-7.193	0.000	-0.380	-0.217

- log(citation in 2 yrs) is **statistically significant**
- Average 5-fold cross-validation test **MSE: 0.9510**

Distance measure for Energy use (Kg/Capita) vs AR-Index [only OECD]



Distance measure for Energy use (Kg/Capita) vs AR-Index [only OECD]



-
- Overall we are 95% confident that the true regression line is between the interval.

Distance measure for Energy use (Kg/Capita) vs AR-Index [only OECD]

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

- AR-Index is **statistically significant**
- Average 5-fold cross-validation test **MSE: 0.9938**

Interpretation: Quantile Regressions

- **For Bulk and Country-specific:**
  - **All 4 quantile regressions were significant** on distance measure of Energy Use (Kg/Capita) vs:
    - log(citations in 2 years)
    - AR-index
- The distance measure tells us how far a country is from reaching the target
  - ***As citation impact metrics increase, i.e. as research related to SDG 7 targets increases and has more influence and spread, distance measure decreases which means OECD countries are closer to the energy use target.***
    - Energy Use measures energy efficiency (SDG 7 target # 3)

Interpretation: Answering Our Research Question

- ***"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?"***
  - Positive correlation between increase in SDG 7 related research impact and progress towards the energy efficiency target for only OECD countries

Interpretation: Drawing Conclusions

- **Bulk Knowledge**:
  - Most of the high research participation countries that make up the majority of the citation impact data (except China) are in the OECD countries
    - Hard to say if non-OECD countries' research impact helps OECD countries' energy efficiency progress
    - China has the highest research participation in our data and is non-OECD.

Limitations

- The distance measure requires a target value for a country or group of countries at the same economic, educational, and SDG level.
- The quantile regressions for the distance measure indicators were only for OECD countries, which are all high-income and upper-middle income countries
- It is difficult to generalize this result for all countries and SDGs

Biases

- High and upper-middle income countries have higher research participation
  - The citation impact data we used has more papers from these wealthier countries
- The citation impacts are only calculated for English-language academic research published.
- Only academic research is included. Impactful research conducted by private companies and non-profit organizations is excluded.
  - This can also bias data towards high-income and upper-middle income countries which have more academic institutions and so higher research participation.

Broader Impacts

- Sustainable development academic research has an important role to play towards SDG 7 progress for certain countries.
- Progress being made on an SDG indicator target can be measured using distance measure and citation impact metrics.
- Countries with lower research participation rates may also benefit from collective academic research related to SDG 7.

References

- World Bank. (n.d.). World development indicators. DataBank. https://databank.worldbank.org/source/world-development-indicators
- Miola, A., & Schiltz, F. (2019). Measuring sustainable development goals performance: How to monitor policy action in the 2030 Agenda implementation? Ecological Economics, 164, 106373. doi:10.1016/j.ecolecon.2019.106373
- Jin, B., Liang, L., Rousseau, R. et al. The R- and AR-indices: Complementing the h-index. CHINESE SCI BULL 52, 855–863 (2007). https://doi.org/10.1007/s11434-007-0145-9
- Tornqvist, L., Vartia, P., & Vartia, Y. O. (1985). How Should Relative Changes Be Measured? The American Statistician, 39(1), 43–46. https://doi.org/10.2307/2683905
- OECD (2018), Measuring Distance to the SDG Targets 2017: An Assessment of Where OECD Countries Stand, OECD Publishing, Paris, https://doi.org/10.1787/9789264308183-en.
- Manpower Research and Statistics Department (MRSD). (n.d.). Absolute vs relative change concepts and definitions. https://stats.mom.gov.sg/SL/Pages/Absolute-vs-Relative-Change-Concepts-and-Definitions.aspx

Thank you for listening!