With Great Growth, Comes Greater Consumption: Energy Consumption of Countries Around the world in Relation to Their annual GDP growth*

An analysis of economic growth and energy consumption globally between 1992 to 2022

Rayan Awad Alim

April 22, 2024

This study explores the relationship between energy consumption and GDP growth across various income levels from 1992 to 2022 using a linear regression model. By integrating global data on energy usage in terawatt-hours and economic performance, preliminary results suggest a positive correlation between increased energy consumption and GDP growth, particularly in lower-income countries driven by industrialization. In contrast, high-income regions exhibit a nuanced relationship, possibly due to advanced energy efficiencies and shifts towards service-oriented economies. These findings emphasize the need for energy-aware economic policies that align with sustainable development goals, highlighting differentiated strategies for income groups in the face of global environmental challenges

Table of contents

*Code and data are available at: RayanAlim/Energy-GDP-Analysis

1 Introduction

1.1 Background And Literature Review

Economic growth and energy consumption are deeply interlinked processes that influence each other in both direct and indirect ways. Historically, the economic output of a country has been closely tied to its energy usage, with periods of high growth accompanied by increases in energy consumption (Stern 2000) This relationship is often conceptualized through the lens of the Energy-Economic Growth Nexus which posits that energy is a critical input in economic production and, conversely, economic growth can lead to increased energy consumption due to higher industrial activity, transportation needs, and residential and commercial usage (Sorrell, Dimitropoulos, and Sommerville 2009).

Several theoretical frameworks have been proposed to explain this nexus. The most prominent among them is the Jevons Paradox, which suggests that as technological advancements increase energy efficiency, the rate of consumption of that energy may actually increase as its effective cost decreases (Alcott 2005). Conversely, the Khazzoom-Brookes postulate extends this paradox to the macroeconomic level, arguing that increased energy efficiency may lead to faster economic growth, which in turn increases total energy demand (Saunders 1992).

Empirical studies have shown diverse results depending on the economic structure and developmental stage of the country in question. For developed nations, research has indicated a decoupling of energy consumption and economic growth, attributed to advances in energy efficiency and a shift towards service-oriented economies which are less energy-intensive (Anthony D. Owen 2006). In contrast, developing countries tend to exhibit a strong correlation between energy consumption and GDP growth, driven by industrialization and infrastructure expansion which are highly energy-dependent (Bhattacharyya 2019).

1.2 Motivation

The policy implications of these findings are significant, as they influence national energy strategies and their alignment with economic objectives. The transition towards sustainable energy sources is also a critical factor in this equation, with renewable energy adoption seen as a key element in sustaining long-term economic growth without the environmental degradation associated with fossil fuels (Wei, Jiandong, and Saleem 2023).

Understanding the specific dynamics of energy consumption and economic growth within the context of a particular country requires a detailed examination of its economic policies, energy resources, and technological advancements. This paper aims to explore these themes within the context of countries globally, providing insights that may help in formulating strategies for sustainable economic and energy development.

2 Data

2.1 Dataset

The dataset used in this study encompasses annual data from 1992 to 2022, sourced from reputable international databases including the World Bank, OECD National Accounts, U.S. Energy Information Administration, and the Energy Institute's Statistical Review of World Energy. It comprises information on GDP growth rates and total energy consumption, measured in terawatt-hours (TWh), for a diverse set of countries globally. The dataset is organized by country-year pairs, enabling analyses of the relationship between economic growth and energy usage across different income levels. Quality assurance measures were implemented to handle missing data and ensure consistency across variables, all data handling, explained in datasheet in appendix A.

The analysis was carried out using the statistical programming language `R` (R Core Team 2023), using the `tidyverse`(Wickham et al. 2019), `here`(Müller 2020), `dplyr`(Wickham et al. 2022), and `readr`(Wickham, Hester, and Bryan 2022) packages. The figures and tables in the paper are generated using the, respectively, `ggplot2`(Wickham 2016) and `knitr`(Xie 2023) packages. Other packages used include `kableextra`(Zhu 2021),

2.2 Variables

The first dataset has the following variables shown in Table 1: **Entity Year Total__Energy__Consumption__TWh Average__GDP__Growth**

Table 1: First ten rows of a dataset countries energy consumption and gdp growth

Country	Year	Energy Consumption (TWH)	GDP Growth %
Afghanistan	2003	5.2	8.8
Afghanistan	2004	4.8	1.4
Afghanistan	2005	6.2	11.2
Afghanistan	2006	7.7	5.4
Afghanistan	2007	9.2	13.8
Afghanistan	2008	16.1	3.9
Afghanistan	2009	26.7	21.4
Afghanistan	2010	33.3	14.4
Afghanistan	2011	42.0	0.4
Afghanistan	2012	40.3	12.8

From the first dataset, we create a secoind one that is segmented by income level, it has has the following variables, a sample of 10 enteries is shown in Table 2: **Standardized__Entity Year Total__Energy__Consumption__TWh Average__GDP__Growth**

Table 2: First ten rows of a dataset countries grouped by income level energy consumption and gdp growth

Income level group	Year	Energy Consumption (TWH)	GDP Growth %
High Income	1992	57434.3	2.2
High Income	1993	58304.0	1.4
High Income	1994	59454.3	3.3
High Income	1995	61122.4	2.9
High Income	1996	63053.9	3.0
High Income	1997	63730.5	3.5
High Income	1998	64000.3	2.9
High Income	1999	64988.3	3.5
High Income	2000	66263.9	4.1

High Income	2001	66166.6	1.5

3 Model

Our models goal is to to understand the dynamics between energy consumption and economic growth across various nations and years. Recognizing the critical role of energy in economic development, this analysis seeks to quantify the extent to which changes in energy consumption influence GDP growth.

3.1 Model set-up

We make use of a linear regression model to investigate this relationship, formalized as follows:

$$GDP\_growth_{it} = \beta_0 + \beta_1 \times Energy\_consumption_{it} + \epsilon_{it}$$

Where:

$$GDP\_growth_{it}$$

- The GDP growth rate (%) for country (i) in year (t).

$$Energy\_consumption_{it}$$

- The primary energy consumption (in terawatt-hours, TWh) for country (i) in year (t).

$$\beta_0$$

- The intercept term, representing the baseline GDP growth when energy consumption is zero.

$$\beta_1$$

- The coefficient of energy consumption, reflecting the change in GDP growth rate associated with a one-unit increase in energy consumption.

$$\epsilon_{it}$$

- The error term, capturing all other factors affecting GDP growth not included in the model.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

- **Linear Assumption**: Preliminary data exploration suggested a potential linear relationship between the logarithm of energy consumption and GDP growth rates. This linear model provides a clear and interpretable framework to estimate the sensitivity of economic growth to changes in energy usage.
- **Economic Rationale**: Theoretically, energy is a fundamental input in production processes; thus, variations in energy supply are expected to directly impact economic output and growth. The linear regression model allows us to quantify this impact and provide insights valuable for policy formulation.
- **Statistical Considerations**: Linear regression is robust, widely understood, and provides comprehensive diagnostic tools to check for model assumptions. It serves as an excellent preliminary analytical tool, especially when the relationship does not involve complex interactions or non-linear dynamics.

3.1.2 Model justification

- **Normality**: We assume the residuals of the regression, (_{it}), are normally distributed, facilitating the use of classical statistical tests for hypothesis testing (e.g., t-tests for coefficients).
- **Independence**: Observations across years and countries are assumed to be independent of each other.
- **Homoscedasticity**: The error terms are assumed to have constant variance across different levels of energy consumption.
- **Linearity**: The relationship between GDP growth and energy consumption is presumed to be linear after appropriate transformations if required (e.g., logarithmic).

4 Results

4.1 Data Results

These two graphs depict the trends in global economic activity and energy usage. The first graph shows how average GDP growth has changed over time, offering insights into economic trends. The second graph illustrates the total energy consumption across different income levels from 1992 to 2022, showing how economic status influences energy usage patterns.

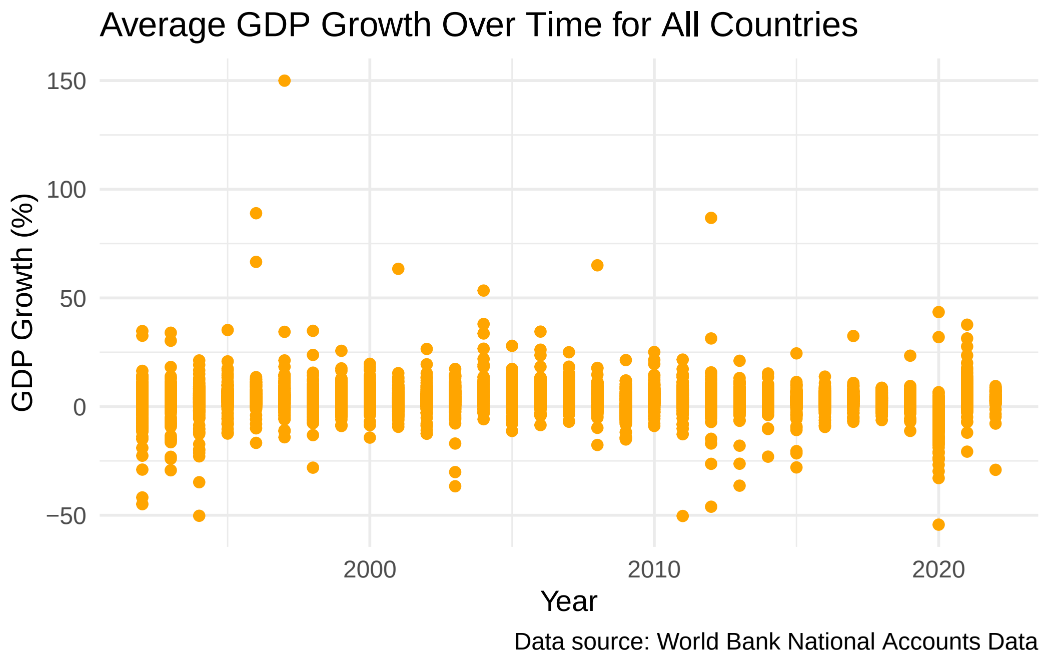

Figure 1: Total GDP Growth of all countries Over Time 1992-2022

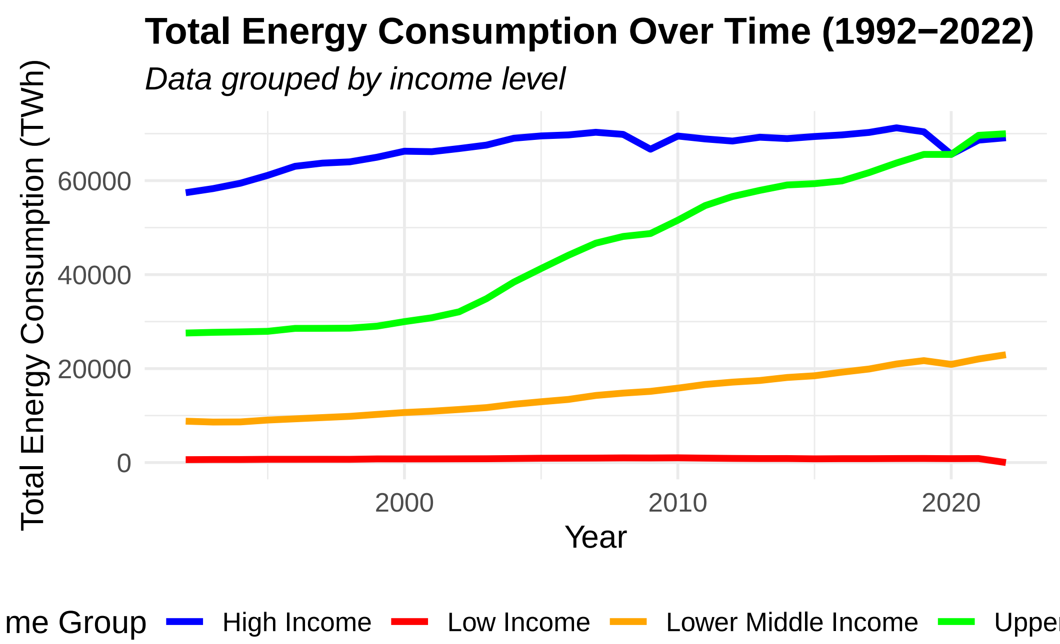

Figure 2: Total Energy Consumption Over Time 1992-2022

4.2 Model Results and Discussion

4.2.0.1 Model Estimation and Diagnostics

The statistical model employed in this analysis was constructed to ascertain the association between Average Gross Domestic Product (GDP) Growth and logarithmic transformation of Total Energy Consumption (TWh). Bayesian inference, implemented through Markov Chain Monte Carlo (MCMC) sampling, was utilized to estimate the posterior distributions of the model parameters. The stan_glm function in R, with a Gaussian family and identity link function, was used to facilitate the analysis, with a total posterior sample size of 4000.

Comparison of Observed Data and Model Predictions: Posterior Predictive Check

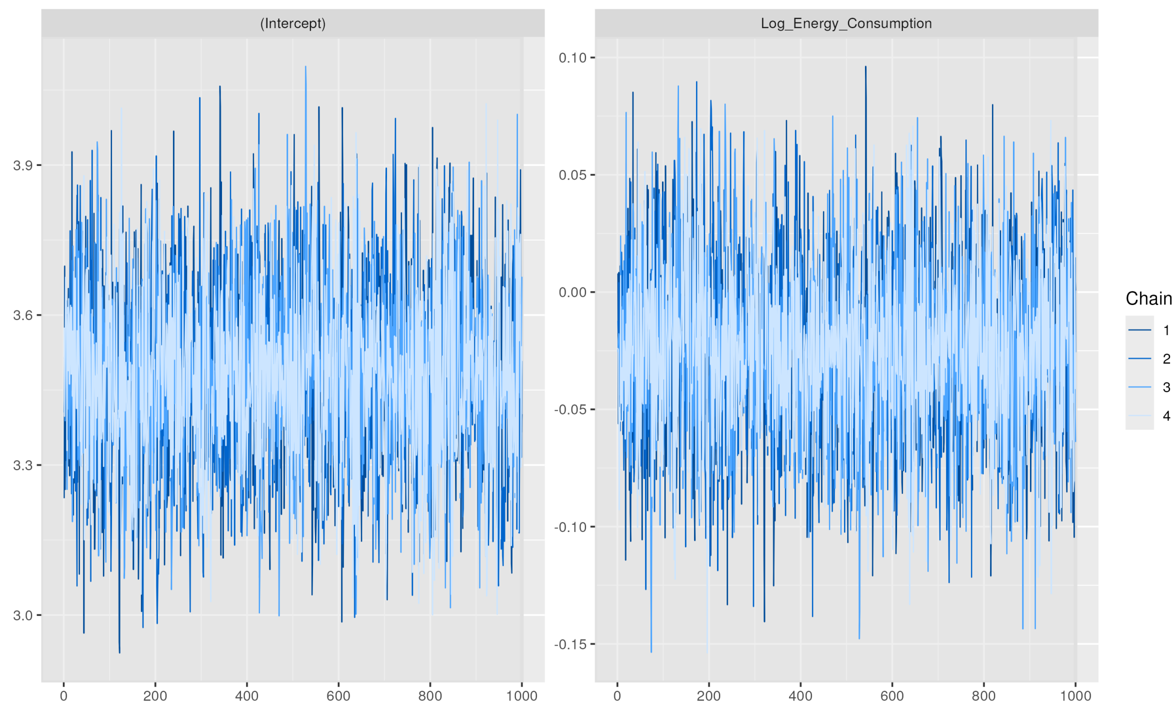

Figure *Markov Chain Monte Carlo (MCMC) trace plots for the Bayesian linear regression model parameters. The left panel displays the trace for the model's intercept, and the right panel shows the trace for the coefficient of Log_Energy_Consumption. Each plot contains four overlaid chains (Chain 1 to Chain 4), illustrated in different shades of blue. The trace plots illustrate the sampling path of the chains through the parameter space over 1000 iterations post warm-up. The plots exhibit stable convergence and adequate mixing of the chains, suggesting that the sampling process has reached equilibrium and that the posterior distributions are well-estimated.*

The primary variables included in the model were the intercept, representing the baseline level of GDP growth when energy consumption is at its logarithmic mean, and the coefficient for Log_Energy_Consumption, which indicates the degree of variability in GDP growth attributable to changes in energy consumption.

The results from the Bayesian regression analysis yielded the following parameter estimates (summarized by their posterior mean and standard deviation, with 10%, 50% (median), and 90% quantiles): - The Intercept was estimated at 3.5, with a standard deviation of 0.2. - The coefficient for Log_Energy_Consumption was centered around 0, with an effectively negligible standard deviation, indicating little to no linear effect of energy consumption on GDP growth in the model.

The residual error of the model (sigma) was estimated at 6.3, with a relatively small standard deviation, reflecting a consistent level of variability in GDP growth that is unexplained by the model.

4.2.0.2 Diagnostic Measures and Model Fit

The Mean Posterior Predictive Distribution (mean_PPD) of the outcome variable was estimated at 3.4, providing an aggregated measure of the model's predictions across all observations and serving as a diagnostic for the model's overall fit.

Comparison of Observed Data and Model Predictions: Posterior Predictive Check

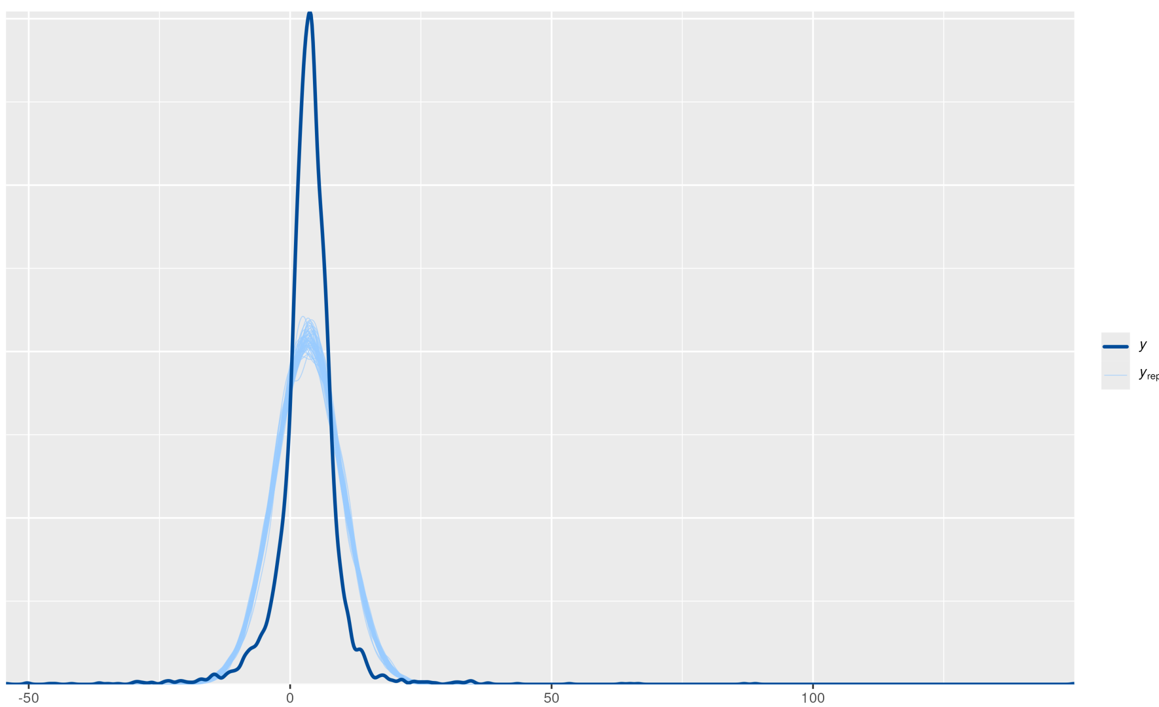

Figure *Posterior predictive check plot comparing the observed values of the outcome variable (y, in dark blue) to the replicated data (yrep, in light blue) generated from the model's posterior distribution. The sharp peak and significant overlap between the observed and replicated distributions around the zero mark demonstrate that the model's predictions closely align with the observed data, indicating an adequate fit for the central tendency of the data. The tails of the distribution should be examined for potential discrepancies between observed and predicted values.*

Monte Carlo standard errors were found to be virtually zero for all parameters, indicating precision in the MCMC estimates. Effective sample sizes (n_eff) for the Intercept and Log_Energy_Consumption parameters were both over 4000, suggesting a sufficient number of independent samples were drawn to estimate the posterior distributions accurately. The potential scale reduction factor (Rhat) was at the optimal value of 1 for all parameters, indicating that convergence was achieved, and the chains mixed well.

4.2.0.3 Trace and Posterior Distribution Plots

The trace plots for both the Intercept and Log_Energy_Consumption display the MCMC sampling paths over iterations, showing dense and well-mixed chains, which are indicative of good convergence properties and reliable posterior estimates. These graphical diagnostics support the numerical evidence from the Rhat statistics.

The posterior density plots for the Intercept and Log_Energy_Consumption exhibit peaked distributions centered around their respective mean estimates, confirming the point estimates presented earlier.

Posterior Distributions and Parameter Correlation for Bayesian Model of Energy Consumption and GDP Growth

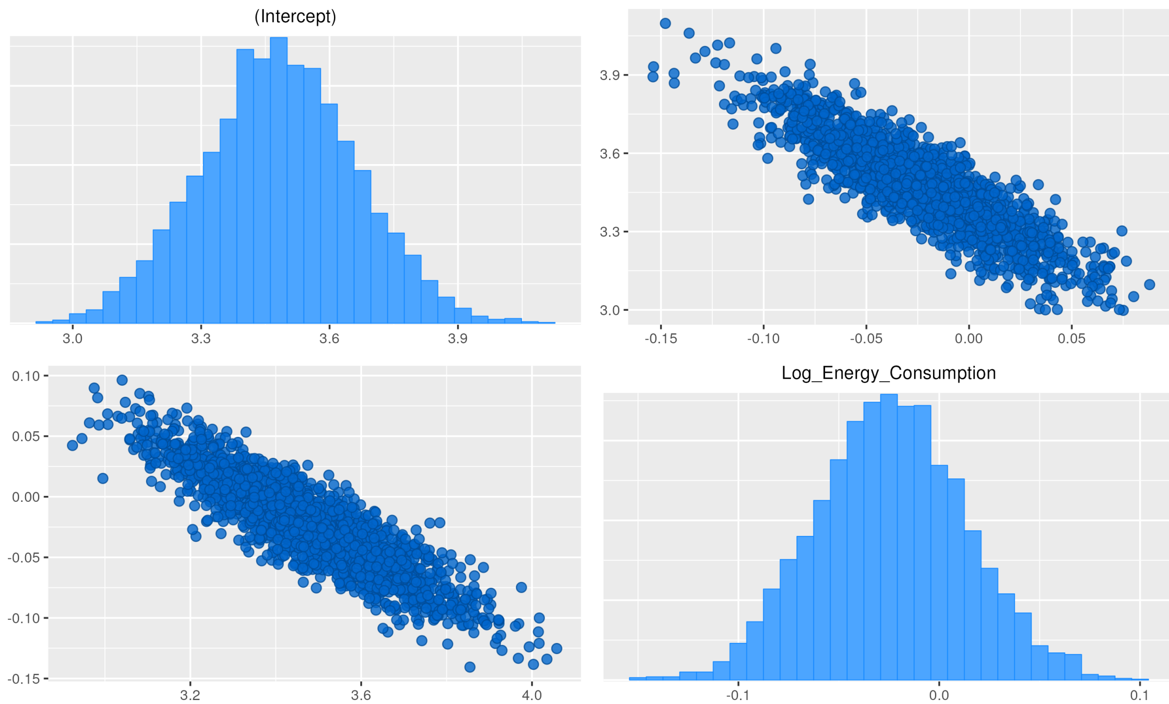

Figure *Pairs plot depicting the posterior distributions and bivariate relationships of the Bayesian linear regression model parameters. The diagonal panels show the density plots of the intercept (top left) and Log_Energy_Consumption (bottom right), both illustrating the posterior distribution of these parameters, which are unimodal and symmetric, suggesting precise estimates. The off-diagonal panels depict the scatterplot between the intercept and Log_Energy_Consumption with a fitted regression line, indicating a negligible correlation between the parameters, as shown by the horizontal trend of the fitted line and the even spread of points around it.*

4.2.0.4 Residuals Analysis

The residuals plot displays the spread of residuals against the predicted values, indicating the differences between observed and predicted GDP growth. The plot reveals a concentration of residuals around zero, suggesting a good model fit for most observations. However, a number of outliers can be observed, indicating instances where the model's predictions deviate substantially from the actual values. This warrants further investigation into potential sources of these discrepancies, such as non-linearity, omitted variable bias, or influential observations.

Residual Analysis of GDP Growth Predictions from Energy Consumption Model

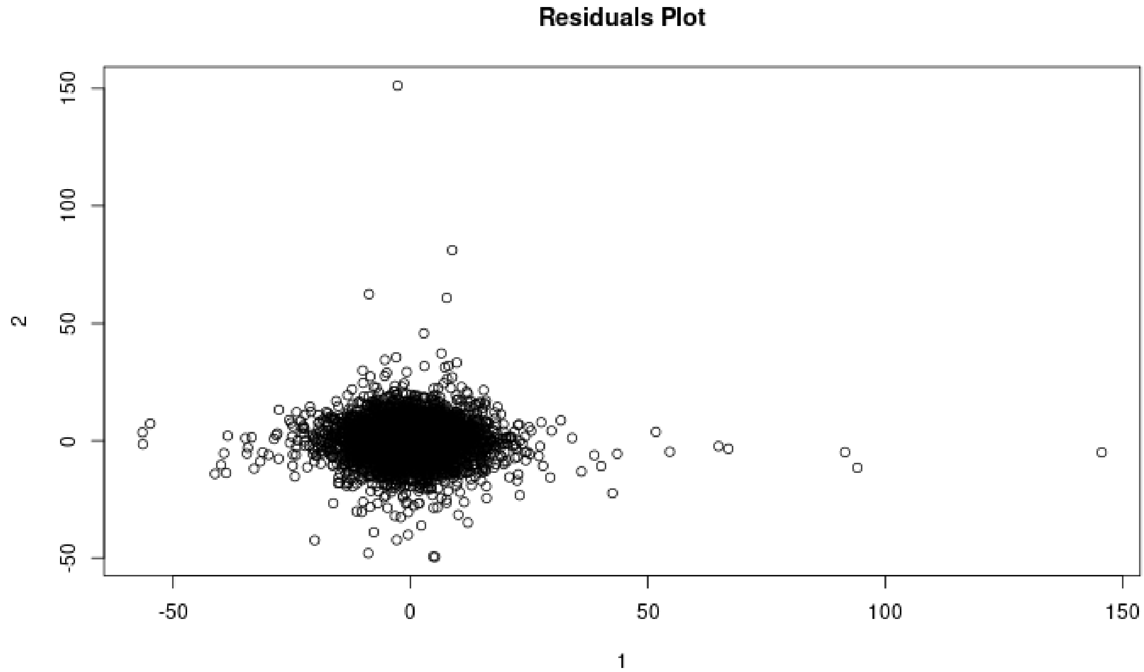

Figure *Scatter plot of the residuals from the Bayesian linear regression model. The residuals (the differences between observed and predicted values of Average GDP Growth) are plotted against the predicted values. The concentration of points around the horizontal axis (residuals of zero) indicates that the model generally fits the data well. However, the presence of points distant from the horizontal axis reflects outliers or instances where the model has significant predictive errors, suggesting areas where model assumptions may not hold or where further model refinement is necessary.*

4.3 Weaknesses and next steps

5 Model Complexity

The linear regression model employed in this analysis assumes a linear relationship between energy consumption and GDP growth. However, the actual relationship may be more complex, involving non-linear dynamics or interaction effects. Future studies could explore better modeling technique to capture these complexities.

6 Causality and Endogeneity

While the analysis identifies correlations between energy consumption and GDP growth, establishing causality requires careful consideration of potential endogeneity issues. Future research could employ instrumental variable approaches or quasi-experimental designs to mitigate endogeneity bias and provide more robust causal inference.

7 Temporal Dynamics

The analysis spans from 1992 to 2022, capturing trends over time. However, economic and energy policies, technological advancements, and global events may influence the relationship between energy consumption and GDP growth differently across different time periods. Future research could explore temporal dynamics and structural breaks to understand how the relationship evolves over time.

8 Policy Implications

The study highlights the need for energy-aware economic policies aligned with sustainable development goals. However, translating research findings into actionable policy recommendations requires careful consideration of political, social, and economic contexts. Future research could involve policy simulations or scenario analyses to evaluate the potential impacts of different policy interventions on energy consumption and economic growth.

9 Environmental Considerations

While economic growth is important, it is essential to consider environmental sustainability and climate change implications associated with increased energy consumption. Future research could integrate environmental indicators and assess trade-offs between economic development and environmental conservation to inform more holistic policymaking.

10 Cross-Country Variations

The analysis aggregates data across countries, masking a very big amount of heterogeneity in the relationship between energy consumption and GDP growth across different regions and income levels. Future research could conduct subgroup analyses or country-specific studies to better understand variations in this relationship and tailor policy recommendations accordingly.

A Appendix {A}

A.1 Datasheet

Extract of the questions from Gebru et al. (2021).

Dataset Abstract

This datasheet presents a dataset combining annual GDP growth data from the World Bank and OECD National Accounts with per capita primary energy consumption data from the U.S. Energy Information Administration and the Energy Institute's Statistical Review of World Energy. The dataset spans from 1965 to 2022 and includes information for 266 countries, enabling analyses of the relationships between economic growth and energy usage.

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

   - The dataset was specifically created to examine the relationship between economic growth, measured through GDP growth rates, and energy consumption on a per capita basis across various countries and years. The intent was to analyse the link in economic metrics with energy usage statistics in an effort to understand how economic trends are related to energy consumption pattern.

2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*

   - There are two datasets merged into one in this dataset. |The first dataset is from the World Bank national accounts data, and OECD National Accounts data files and the second datset is from U.S. Energy Information Administration (2023), Energy Institute - Statistical Review of World Energy (2023), Population based on various sources (2023) – with major processing by Our World in Data. They were merged by Rayan Awad Alim, they were merged for research purposes and the above outlined motivation.

3. *Who funded the creation of the dat9aset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*

   - The collection of the data was funded by World Bank and Organisation for Economic Co-operation and Development (OCED).

4. *Any other comments?*

   - This dataset coule be particularly valuable for policymakers and researchers interested in the implications of economic policies on energy consumption and environmental impact.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

    - Each instance in the dataset represents a country-year pair, with metrics for GDP growth and energy consumption per capita for that particular year and country

2. *How many instances are there in total (of each type, if appropriate)?*

    - `The dataset includes data from 1965 to 2022 for x countries, as well as by region. I`

3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

    - The dataset is not a sample it includes all countrries recognized by the UN aqre accounted for.

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

    - Each instance includes the following data points; Country name, Year, Annual GDP growth rate (percentage), Energy consumption per capita (kilowatt-hours per person)

5. *Is there a label or target associated with each instance? If so, please provide a description.*

    - There are no explicit labels or target variables as this dataset is designed for exploratory and inferential statistical analyses rather than predictive modeling.

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

    - Some instances have missing data due to non-reporting countries or years where the data collection was not feasible. Missing data points are noted and are a dropped in the analysis to avoid biased interpretations.

7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*

- The dataset primarily represents independent instances without explicit relational data between them

8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*

   - N/A

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

   - Some country names werew spelled differenltly, but they were standardized using their ISO3 names.

10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

    - The dataset is self-contained with regards to the information it provides. External links to the data sources are maintained for verification and detailed reference but are not necessary for the primary analysis tasks the dataset is intended to do

11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*

    - No

12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*

    - No.

13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*

    - yes, it identifies regions for country clusters, eg africa, north africa, middle east etc

14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*

- No.

15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*

    - No

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

    - The data for this dataset was sourced from authoritative international databases including the World Bank, OECD National Accounts data files, the U.S. Energy Information Administration (EIA), and the Energy Institute's Statistical Review of World Energy. These sources provide quantifiable, directly observable data compiled from governmental and energy sector reporting.

2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*

    - Data was downloaded from the websites: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD. and https://ourworldindata.org/energy#explore-data-on-energy

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*

    - Not a sample, the entire dataset.

4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*

    - The data compilation and preliminary analysis were conducted by a team of data scientists and economists employed by the respective data-providing organizations. These individuals are experts in their fields and are compensated as full-time employees by their institutions.

5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old*

news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

- The dataset includes data collected annually from 1965 to 2022. Each data point is recorded after the end of the reporting year.

6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*

    - No information provided.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

    - The data is publicly available in these 2 sites
    - https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2022&start=2022&view=bar
    - https://ourworldindata.org/energy#explore-data-on-energy

8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

    - The websites state that usage is made available to the public and require no consent to use the data as long as they are cited (U.S. Energy Information Administration, Energy Institute, and Our World in Data 2023)

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

    - The dataset initially contained 7,209 entries. Through data cleaning, entries with missing values for energy consumption or GDP growth, totaling 1,041, were removed, leaving 6,168. Subsequently, entries reporting zero energy consumption were deemed inaccurate and eliminated, amounting to 501 entries. The resulting cleaned dataset comprised 5,667 entries, suitable for analysis.

2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*

    - Yes both raw and clean files are saved as wel;l as the scripts made to clean them.

3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

- The programming software R (R Core Team 2023).

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
    - Only for the analysis this datasheet is accompanying.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
    - No. This dataset will exist in the following github repo which has an MIT license https://github.com/RayanAlim/Energy-Analysis/

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
    - The authors do not intend to maintain this dataset, but it will exist in the following github repo which has an MIT license for others to use https://github.com/RayanAlim/Energy-Analysis/
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
    - rayan.alim@mail.utoronto.ca
3. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
    - - The authors do not intend to maintain this dataset, but it will exist in the following github repo which has an MIT license for others to use https://github.com/RayanAlim/Energy-Analysis/.

References

Alcott, Blake. 2005. "Jevons' Paradox." *Ecological Economics* 54 (1): 9–21. https://doi.org/https://doi.org/10.1016/j.ecolecon.2005.03.020.

Anthony D. Owen. 2006. "Renewable Energy: Externality Costs as Market Barriers." *Energy Policy* 34 (5): 632–42. https://doi.org/https://doi.org/10.1016/j.enpol.2005.11.017.

Bhattacharyya, Subhes C. 2019. *Energy Economics: Concepts, Issues, Markets and Governance*. 2nd ed. London, UK: Springer London. https://doi.org/10.1007/978-1-4471-7468-4.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Saunders, Harry D. 1992. "The Khazzoom-Brookes Postulate and Neoclassical Growth." *The Energy Journal* 13 (4): 131–48. http://www.jstor.org/stable/41322471.

Sorrell, Steve, John Dimitropoulos, and Matt Sommerville. 2009. "Empirical Estimates of the Direct Rebound Effect: A Review." *Energy Policy* 37 (4): 1356–71. https://doi.org/https://doi.org/10.1016/j.enpol.2008.11.026.

Stern, David I. 2000. "A Multivariate Cointegration Analysis of the Role of Energy in the US Macroeconomy." *Energy Economics* 22 (2): 267–83. https://doi.org/https://doi.org/10.1016/S0140-9883(99)00028-6.

U.S. Energy Information Administration, Energy Institute, and Our World in Data. 2023. "Primary Energy Consumption Per Capita." U.S. Energy Information Administration, "International Energy Data"; Energy Institute, "Statistical Review of World Energy"; Various sources, "Population". https://www.eia.gov/international/data/world; https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html.

Wei, Shanxiang, Wen Jiandong, and Hummera Saleem. 2023. "The Impact of Renewable Energy Transition, Green Growth, Green Trade and Green Innovation on Environmental Quality: Evidence from Top 10 Green Future Countries." *Frontiers in Environmental Science* 10. https://doi.org/10.3389/fenvs.2022.1076859.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Gram-*

mar of Data Manipulation. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.