DEGGED - Dynamic Exploration of Greenhouse Gas Emissions and its Determinants using R and Shiny

CHOONG Shi Lian Selene[1], JIANG Weiling Angeline[1], WONG Wei Sheng Dylan[1]

[1] School of Computing and Information Systems, Singapore Management University

Introduction

Global warming is expected to result in a rise of the average global temperature between 1.1 to 6.4 degree celsius over the century, if no interventions are taken to reduce greenhouse gas emissions. Economic drivers such as electricity production, transportation and waste generation, contributes to greenhouse gas emissions. To mitigate these emissions, measures like emission taxes and use of renewable energy sources were implemented. Understanding the relationship between greenhouse gas emissions and its determinants will offer insights on the influence of determinants on emission levels.

Motivation

Findings from past researches to investigate the influence of determinants on greenhouse gas emission levels were generally static in nature. Often, most research papers only published the final model used without showing the intermediate steps performed. Our group developed an interactive analytical dashboard to enable the users to explore and analyse the determinants of greenhouse gas emissions, without the need for any programming knowledge.

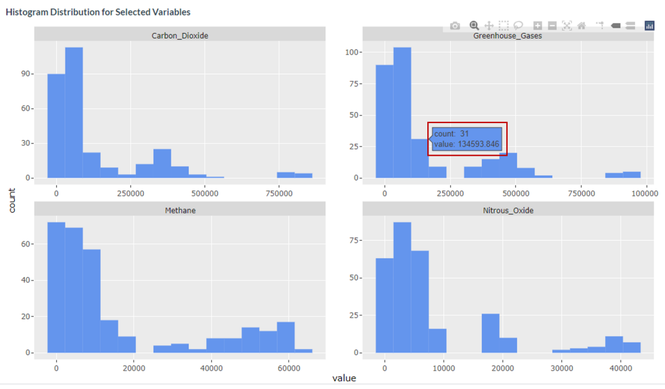
Approach

The application aims to provide an interactive user-interface design to allow users to: 1. Apply **exploratory data analysis** to understand the factors affecting emissions by country and time period; 2. Identify main factors contributing to emissions from cross-sectional view through **Ordinary Least Square (OLS) regression** with step-by-step guidance to detect any violation of OLS assumptions; 3. Identify main factors contributing to emissions from panel data view using **Panel data regression** and list of validation tests for panel analysis

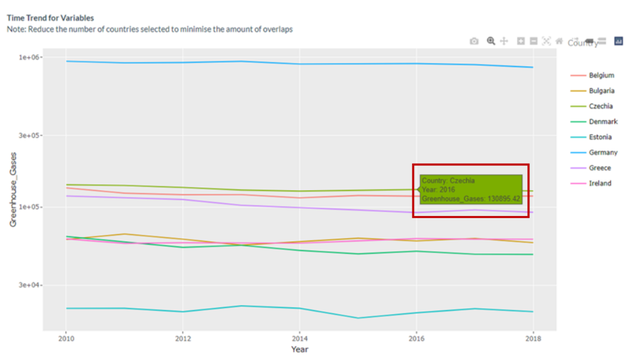
Tools used in building DEGGED include:

Exploratory Data Analysis

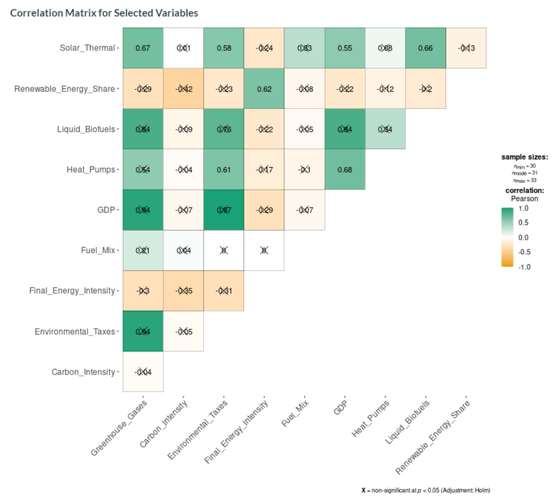
Descriptive Statistics Users can explore an interactive table containing the descriptive statistics of the selected variable is included. Histogram also displays the distribution of the values of selected variables, where each frame represent one variable. Users can hover over each bar to display the frequency count for that specific x value.



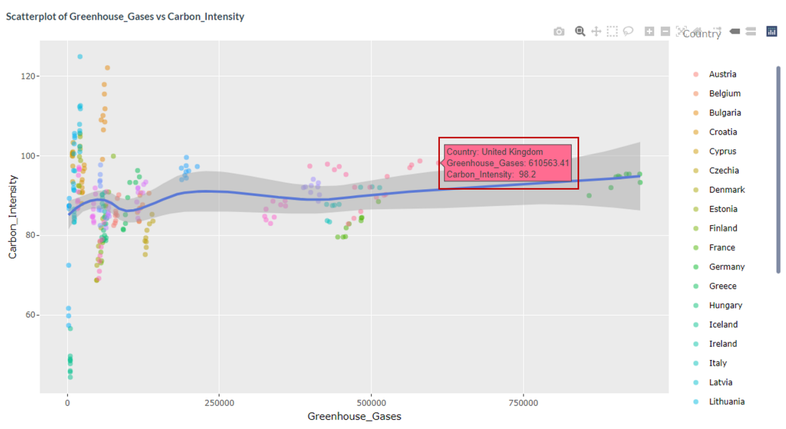
Time Trend Time trend lines allow users to visually view the changes in the variable over the years. The chart by individual countries offers a quick comparison across the countries, while the grouped chart offers a comparison across the variables. Hovering over the lines displays the relevant value for the hovered data point.



Bivariate Analysis The correlation matrix allows users to analyse the linear relationships between two variables, where the number in each box represents the correlation coefficient for that variable pair. The color of the boxes indicates if the pairs are positively (green), negatively (orange), or not (white) correlated at all. Crosses represent statistically insignificant correlation result.

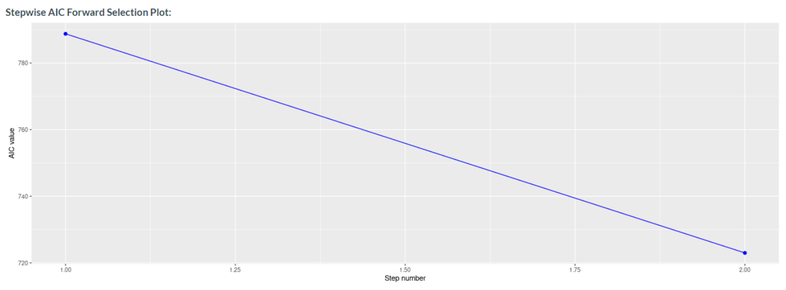


The scatterplot shows the relationship between the selected x and y variables. Each dot represents a data point along with loess (locally weighted smoothing) curve. Hovering over each element (dot/line/band) will show the respective x and y values.

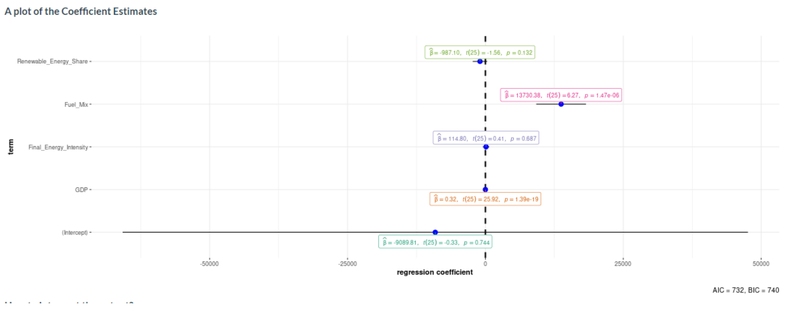


Ordinary Least Square (OLS) Regression

The OLS module consists of 2 sub-modules (Variable Selection and Model Selected), and provides a step-by-step guide in building OLS regression models. Users can first explore and identify the best subset of independent variables to include in the model using the Variable Selection sub-module. Apart from the ANOVA and regression summary of the steps taken at each search and the parameter estimates of the final model and the detailed output on the steps in the model ran, a plot of the criterion value and the step number is also provided to offer a visualisation of the change of the criterion value after each search step.

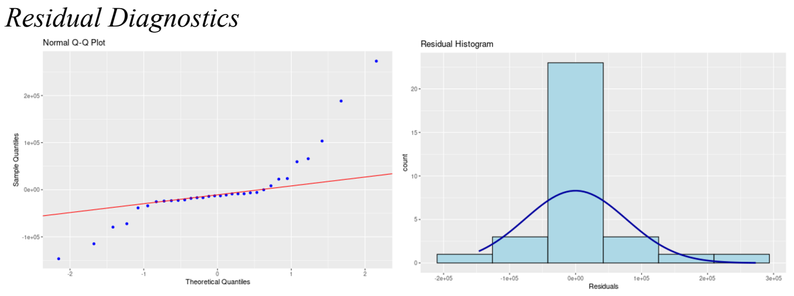


Using the best set of selected independent variables, users can input them under the Model Selected sub-module to obtain the parameter estimations and run the necessary OLS assumption tests. Parameter estimates will be presented as a dot plot for easy intepretation.

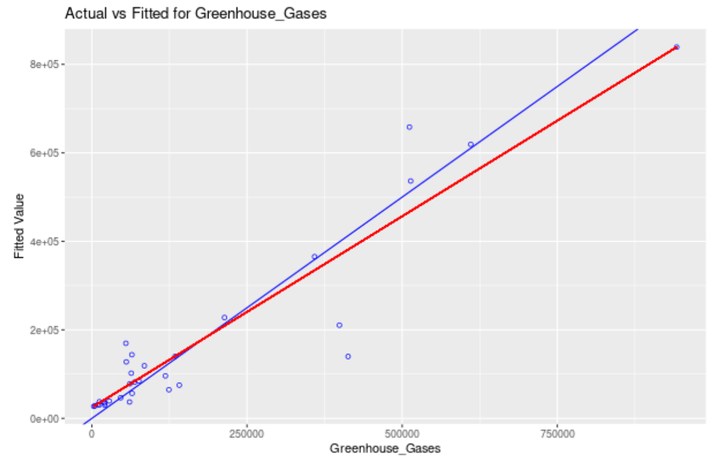


List of assumption tests performed include (i) Heteroskedasticity test, (ii) Residual diagnostics, (iii) Collinear diagnostics and (iv) Model fit assessment. Visualisation of results were implemented where appropriate.

Residual Diagnostics

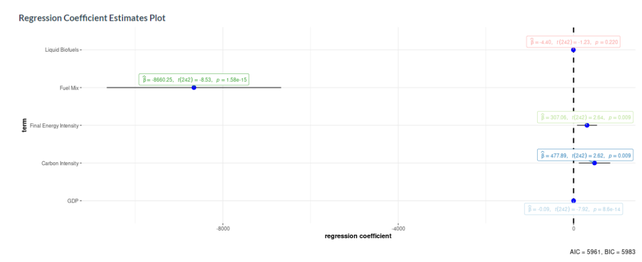


Model Fit Assessment



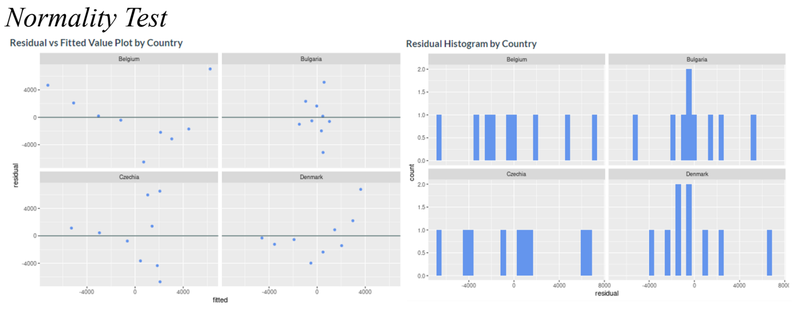
Panel Data Regression

The panel data regression module consists of 2 sub-modules (Model Building and Validation Tests), and provides a step-by-step guide in building panel regression models. Users can decide the type of panel data models (fixed or random effect) as well as the type of effect (individual, time and two-ways). Apart from the coefficient estimate table, the regression coefficient estimates are plotted as a dot-and-whisker plot for clearer visualisation. Model fit assessment tests (Hausman/ Chow / Lagrange Test) will also be performed and its outputs shown.

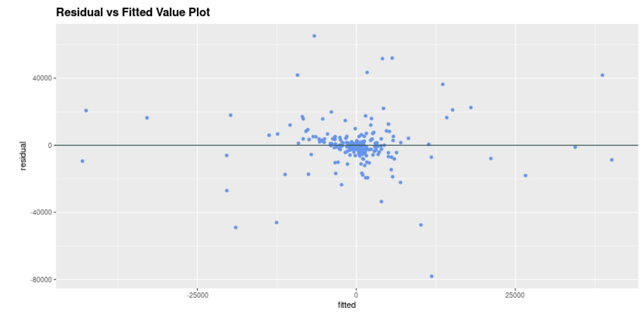


List of assumption tests performed include (i) Normality test, (ii) Serial Correlation Test and (iii) Heteroskedasticity Test. Visualisation of results were likewise implemented where appropriate.

Normality Test



Heteroskedasticity Test



Future Work

- Modify the olsrr package in R to incorporate reactive dataset and hence able to output a panel of fit criterion (R-square, Adjusted R-square, Mallow's Cp, AIC, SBIC and SBC)
- Extend scope to cover greenhouse gas emissions in other countries such as U.S. which also aims to cut greenhouse gas emissions by half by 2030 as part of its new commitment to the Paris climate agreement

The authors wish to thank Professor Kam Tin Seong for his support and guidance throughout this project