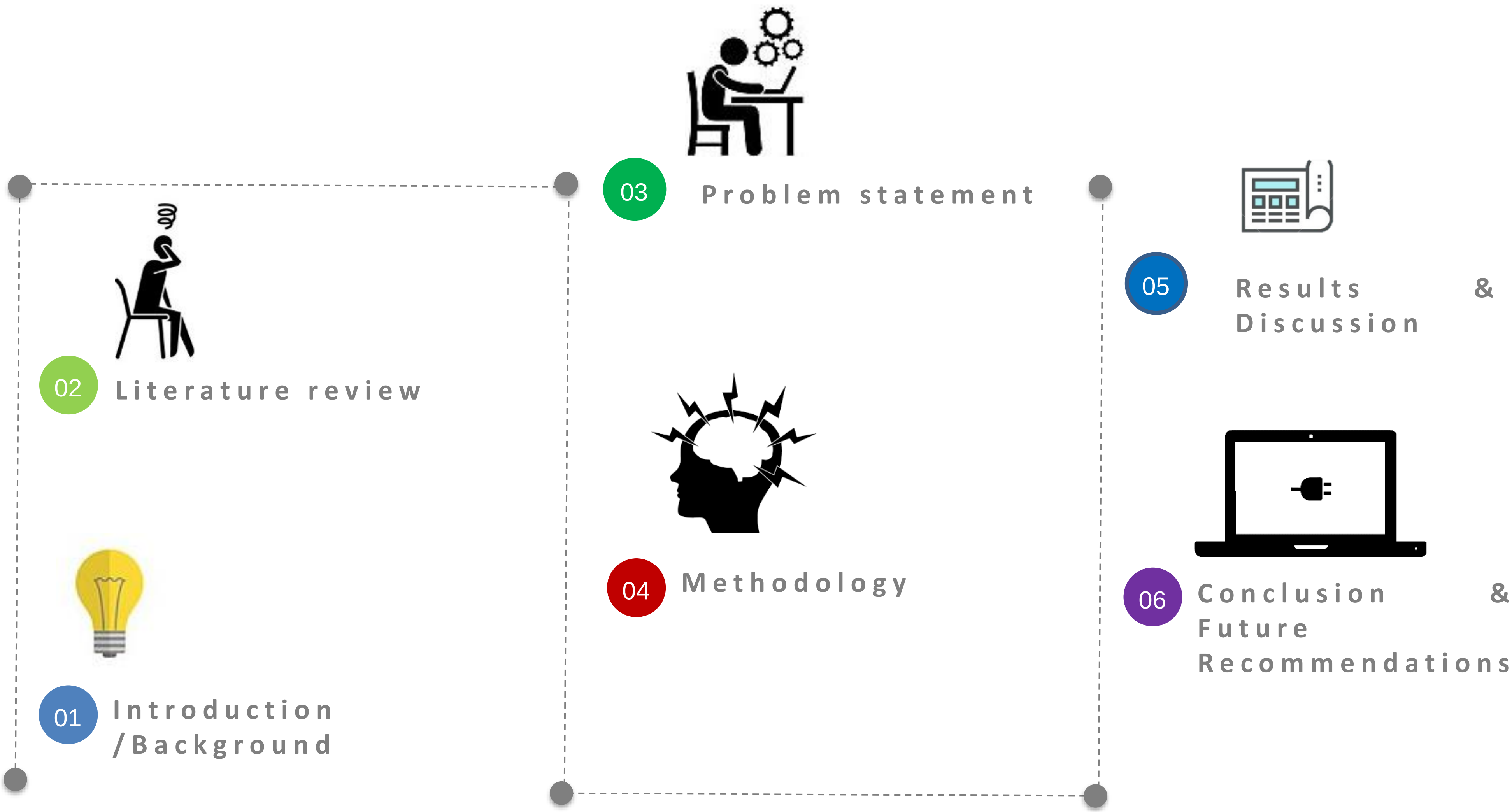
A STUDY OF THE RELATIONSHIP BETWEEN GDP, POPULATION & CO2 EMISSIONS BASED ON TIME SERIES

CHETAN TIPPA

DEC 2023

LIVERPOOL JOHN MOORES UNIVERSITY


01
Introduction /Background
02
Literature review
03
Problem statement
04
Methodology
05
Results & Discussion
06
Conclusion & Future Recommendations

INTRODUCTION/BACKGROUND

- Human-caused greenhouse gas (GHG) emissions drive climate change.
- About 60% of greenhouse gas (GHG) emissions come from 10 countries, while the 100 countries with the lowest emissions account for less than 3%.
- China was the biggest CO2 emitter in 2020, accounting for 30% of the global total. The US followed closely with 12% of global CO2 emissions, while India contributed 6.2% (Figure 1 & 2).
- The Prime Minister of India, Narendra Modi, has expressed his dedication to attaining carbon neutrality by the year 2070 during the 26th Conference of the Parties (COP 26) convened in Glasgow.
- The study is to examine the relationship between population growth, economic development, and CO2 emissions.

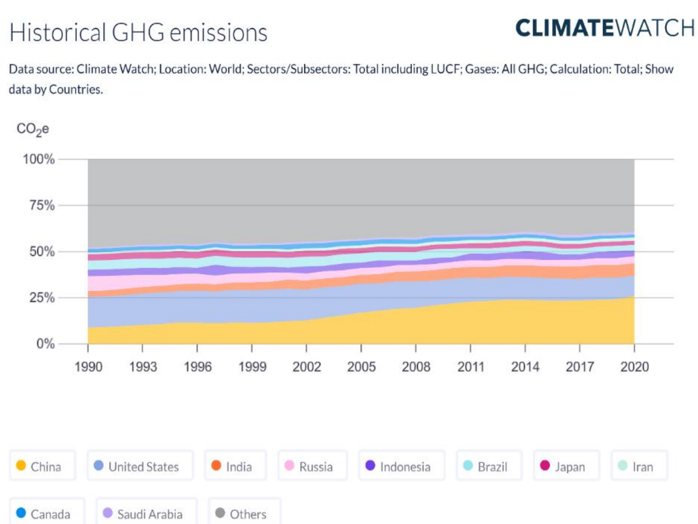


Figure 1 - GHG emissions country-wise percentage from 1990 – 2020 (Washington, DC: World Resources Institute, 2022)

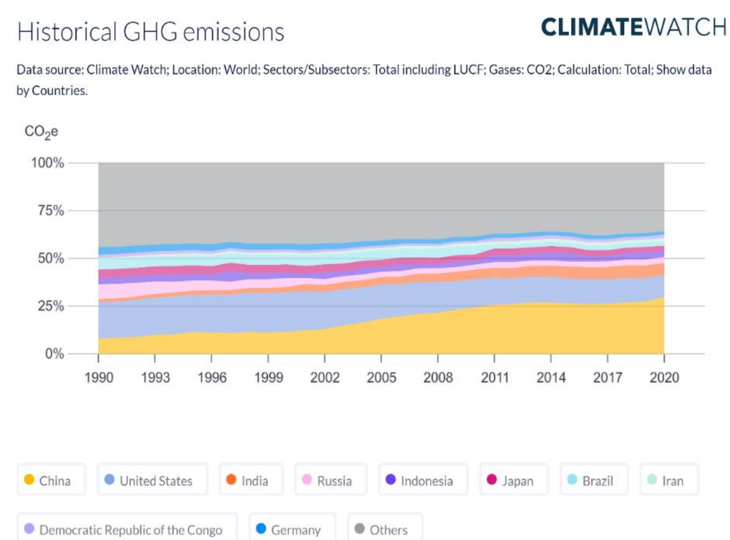


Figure 2 - CO2 emissions country-wise percentage from 1990 – 2020 (Washington, DC: World Resources Institute, 2022)

AIM

- The main aim of this study is to examine the relationship between population growth, economic development, and CO_2 emissions from 1850 to 2021. This study aims to investigate the link between economic development and environmental deterioration.

OBJECTIVE

The research objectives are formulated based on the aim of this study which is as follows:

- To analyze and identify the relationship between economic growth and environmental degradation.
- To explore the causality between economic growth and environmental degradation.
- To assess the VAR and multiple Linear regression models and identify the most precise one to find the relationship between economic growth and environmental degradation.
- To use the Vector Autoregressive Model of Timeseries Modeling to forecast CO_2 levels of actual values vs predicted values from 2011 to 2021 based on historical data.

LITERATURE REVIEW

Sl. No	Year	Title	Author(s)	Dataset	Algorithms	Findings
1	2023	Revisiting economic growth and CO2 emissions nexus in Taiwan using a mixed-frequency VAR model	Tsangyao Chang, Chen-Min Hsu, Sheng-Tung Chen, Mei-Chih Wang, Cheng-Feng Wu	1. GDP – Taiwan Economic Journal (TEJ) database. 2. Annual primary energy consumption and CO2 emissions data from the BP Statistic Review of World Energy, 2020 database with time periods spanning from 1970 to 2019. for Taiwan.	MF-VAR model to investigate the causal association between economic progress and CO2 emissions.	The MF-VAR model shows evidence that feedback phenomena exist between economic progress and CO2 emissions.
2	2020	Economic performance of India amidst high CO2 emissions	Edmund Ntom Udemba, Hasan Güngör,Festus Victor Bekun,Dervis Kirikkaleli	The data and variables are retrieved from the World Development Indicators of the World Bank (WDI, 2020) and the British Petroleum database (2020) for the case of carbon dioxide emission (CO2) with all series later transformed into logarithm values.	ARDL technique, Granger's causality	1.A mix (positive and negative) of relationships between the explanatory variables and the explained variable (GDP) except for the case of trade openness, which is insignificant in the long run. 2. Granger causality displayed give insight and credence to the link among the selected variables, namely CO2 emission, energy utilization, liberalization (openness), and population.
3	2020	The relationship between energy consumption, economic growth and carbon dioxide emissions in Pakistan	Muhammad Kamran Khan, Muhammad Imran Khan and Muhammad Rehan	World Development indicator (World Bank): https://data.worldbank.org/products/wdi	ARDL model. To investigate the nexus between energy consumption, economic growth and CO2 emission in Pakistan by from 1965 to 2015	1. ADF & Philips Peron Unit root tests indicate the attributes are stationary at first order difference. 2. Long-run ARDL results pointed out that energy consumption has a positive effect on CO2 emissions in Pakistan.

LITERATURE REVIEW

Sl. No	Year	Title	Author(s)	Dataset	Algorithms	Findings
4	2018	Effect of energy consumption and economic growth on environmental degradation in India: A time series modelling	Krishan K.Pandey, Harshil Rastogi	Central Electricity Authority (CEA), Central Pollution Control Board (CPCB), World Bank, Government of India, BP Statistical Review Reports, BP Energy Outlook, Emissions Database for Global Atmospheric Research (EDGAR).	ADF test, Cointegration test , and Granger Causality test to examine the interrelationship among Electricity consumption, Economic growth and CO2 emissions.	Results: Cointegration 1. Economic growth and CO2 emissions 2. CO2 emissions and electricity consumption for different sectors 3. Economic growth and electricity consumption for different sectors 4. Total electricity consumption and CO2 emissions for different sectors Results: Granger Causality 1. Bidirectional Causality : GDP & CO2 emissions, electricity consumption in agricultural, commercial, and industrial sectors, economic growth, and electricity consumption in agricultural, traction, railways, and industrial sectors, Total electricity consumption and CO2 emissions. 2. Unidirectional Causality : Domestic electricity consumption and economic growth, Commercial electricity consumption and economic growth,

PROBLEM STATEMENT

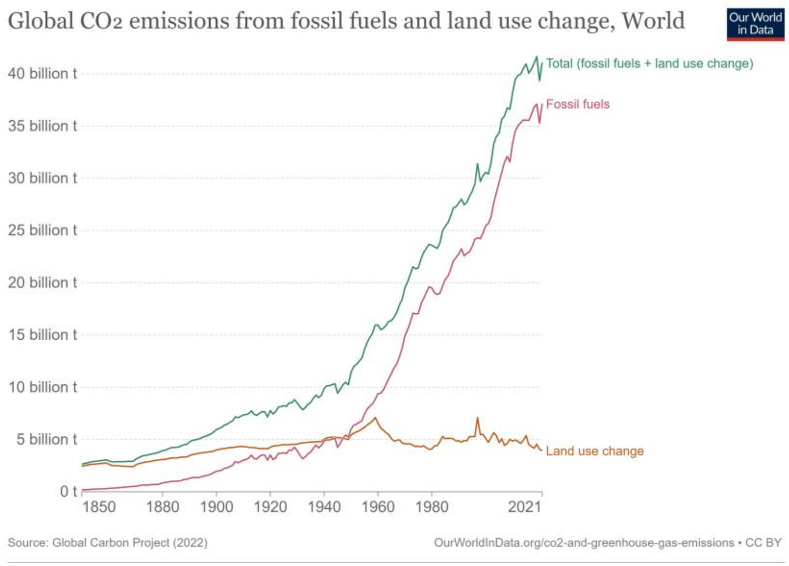


Figure 3 - Global CO2 emissions from fossil fuels and land use change (Ritchie, 2020)

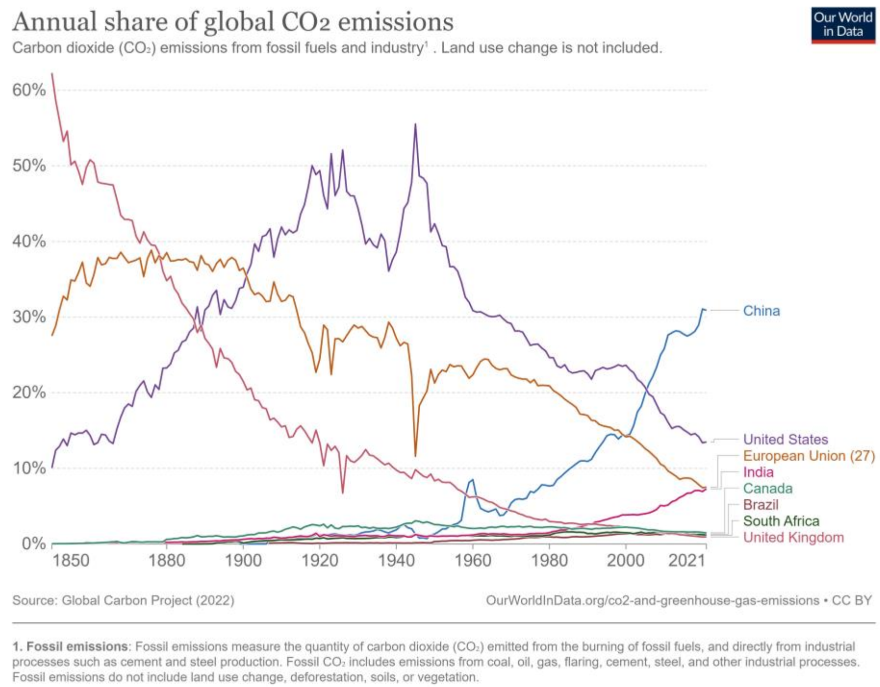


Figure 4 - Annual share of Global CO2 emissions from fossil fuels and Industry (Ritchie, 2020)

- Fossil fuel and land use CO2 emissions (Figure 3).
- CO2 emissions through industrialization (Figure 4).
- Greenhouse gas emissions contribute to global warming.
- Understanding how changes in GDP and population impact CO2 emissions and vice versa requires sophisticated modelling and data analysis techniques, making it a multifaceted research challenge.

METHODOLOGY

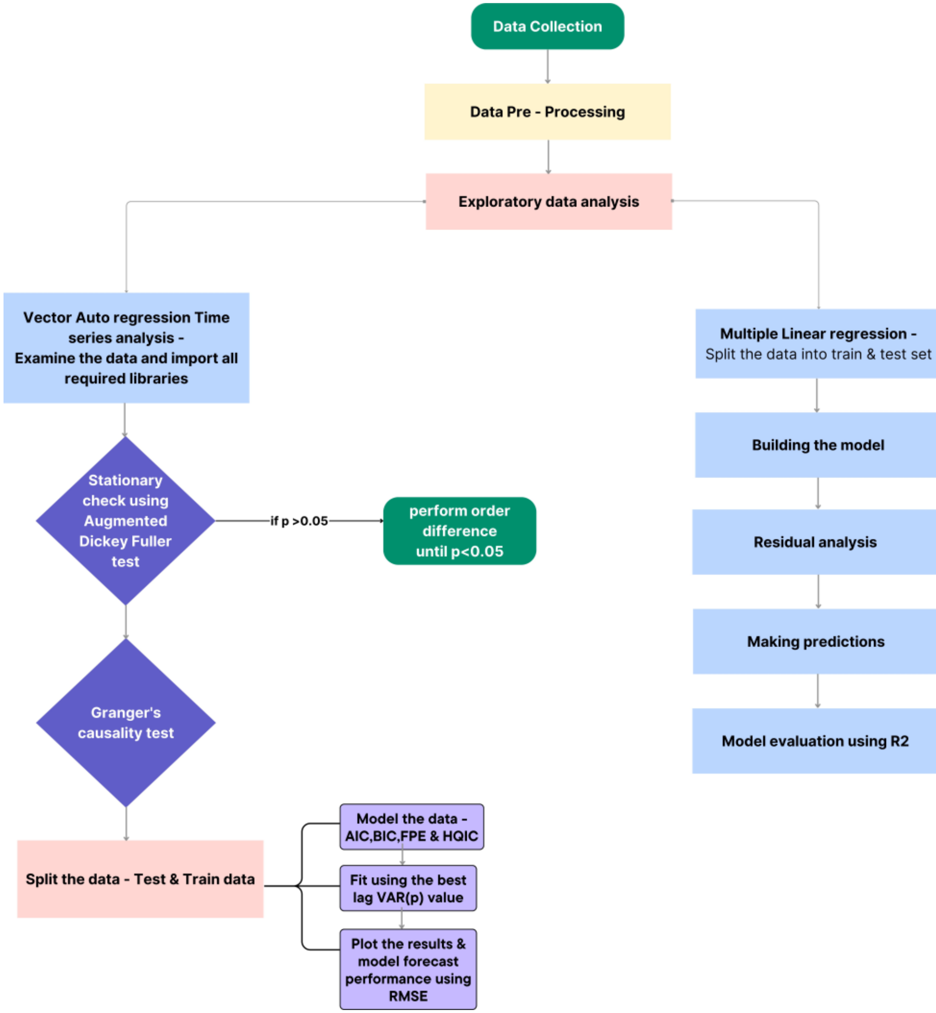


Figure 5 - VAR & Multiple linear regression Framework illustrating workflow.

- Identification of dataset.
- Data preprocessing and exploratory data analysis.
- VAR & MLR model – Train & test
- Evaluation using RMSE & R^2 values

RESULTS AND DISCUSSION – MISSING VALUES, HEAT MAP, BOX PLOT & DENSITY PLOT

```
country              0.000000
population           0.000000
gdp                 19.186047
co2_including_luc   11.046512
dtype: float64
```

Figure 6 - Missing values (percentage) in the dataset

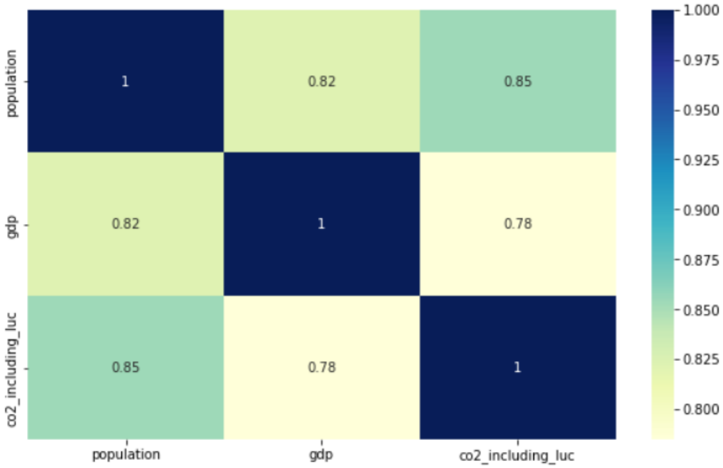


Figure 7 – Heat Map plot

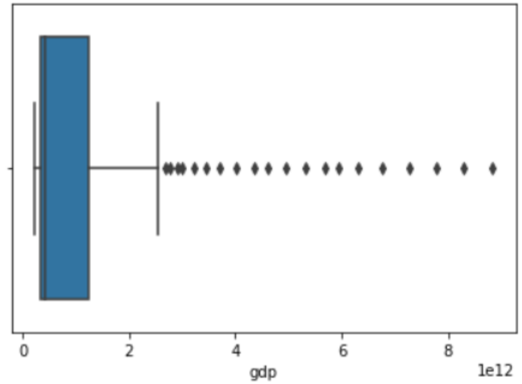


Figure 8 – Box plot of gdp

<AxesSubplot:xlabel='gdp', ylabel='Density'>

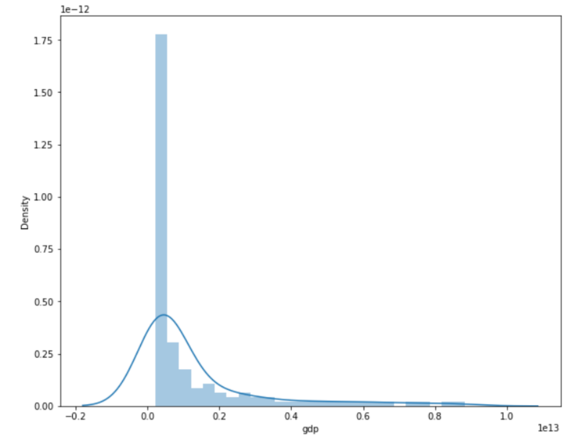


Figure 9 – Density plot of gdp

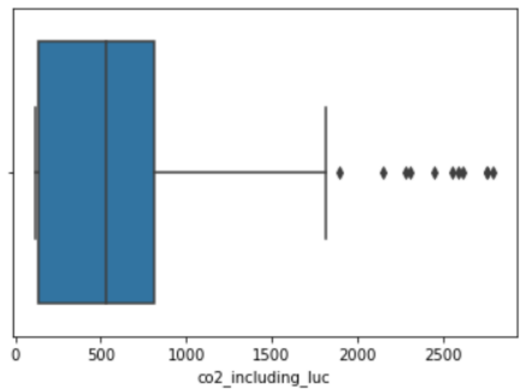


Figure 10 – Box plot of co2_including_luc

<AxesSubplot:xlabel='co2_including_luc', ylabel='Density'>

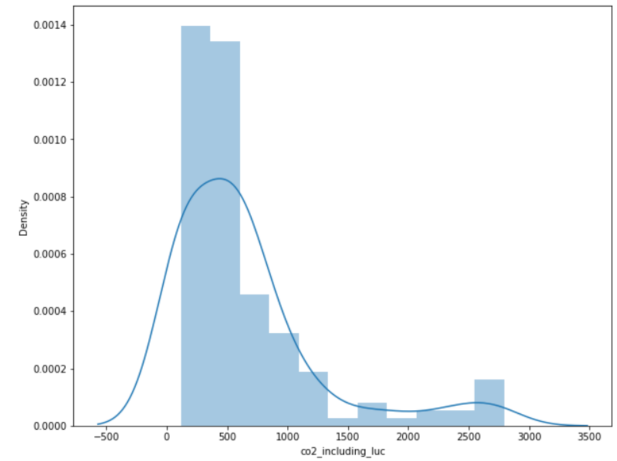


Figure 11 – Density plot of co2_including_luc

- Figure 6 – Missing values
- Figure 7 – Correlation between variables
- Figures 8,9,10 & 11 – Impute of missing values.

RESULTS AND DISCUSSION – SCALING, OLS ,VIF & ERROR TERMS

	population	gdp	co2_including_luc	India
year				
1945	0.128148	0.026188	0.177763	1
1901	0.049663	0.006702	0.003425	1
1996	0.644075	0.260271	0.339043	1
1942	0.135644	0.025568	0.174988	1
1977	0.358378	0.089860	0.210667	1

Figure 12 – Variable values after applying Min-Max scaler

	Features	VIF
0	population	3.09
1	gdp	3.09
2	India	1.73

Figure 13 – VIF results of variables

- Figure 12 – Min max scaler
- Figure 13 – VIF values
- Figure 14 – OLS regression results
- Figure 15 – Distribution of error terms

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      co2_including_luc   R-squared:                       0.732
Model:                            OLS   Adj. R-squared:                  0.727
Method:                 Least Squares   F-statistic:                     159.7
Date:                Sat, 28 Oct 2023   Prob (F-statistic):           3.60e-34
Time:                        11:02:18   Log-Likelihood:                 93.714
No. Observations:                 120   AIC:                            -181.4
Df Residuals:                     117   BIC:                            -173.1
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
population     0.5001      0.065      7.701      0.000       0.371       0.629
gdp            0.2666      0.094      2.848      0.005       0.081       0.452
India          0.0565      0.013      4.199      0.000       0.030       0.083
==============================================================================
Omnibus:                       17.261   Durbin-Watson:                   2.375
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               20.259
Skew:                           0.865   Prob(JB):                     3.99e-05
Kurtosis:                       4.029   Cond. No.                         11.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 14 – OLS Regression results

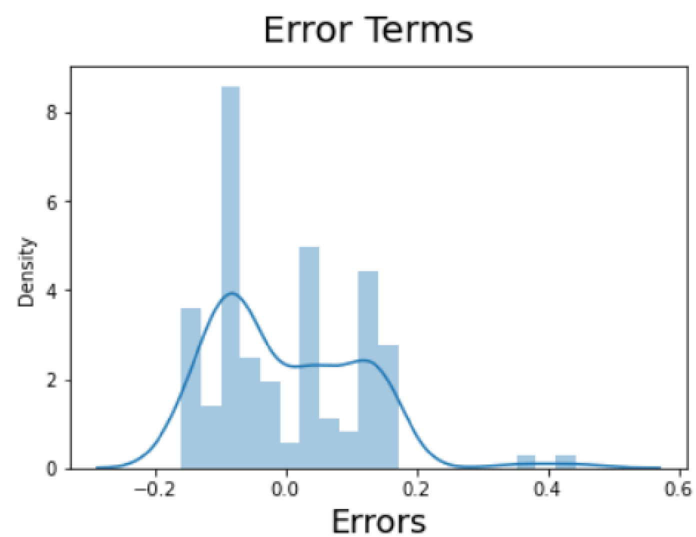


Figure 15 – Distribution of error terms

RESULTS AND DISCUSSION – SCATTER PLOT & VAR LINE PLOT

year	Actual Value	Predicted Value	Difference
1857	0.200176	0.066026	0.134149
1974	0.158461	0.236625	-0.078165
1861	0.001981	0.060991	-0.059010
1865	-0.000170	0.069416	-0.069586
1981	0.164757	0.290430	-0.125673
1909	0.003553	0.091505	-0.087952
1968	0.128986	0.199487	-0.070502
1947	0.180389	0.122034	0.058355
1891	0.005110	0.078227	-0.073117
2000	0.404335	0.495753	-0.091418
1876	0.200176	0.072992	0.127183
1989	0.234201	0.369572	-0.135371
1851	0.200176	0.056959	0.143217
2007	0.513874	0.604398	-0.090524
1951	0.250517	0.116760	0.133757
2015	0.883846	0.757732	0.126114
2002	0.383517	0.519697	-0.136181
1984	0.173489	0.317609	-0.144120
1986	0.200160	0.335991	-0.135831
1935	0.160534	0.120864	0.039670
1925	0.134924	0.101868	0.033056

Figure 17 – Difference values of Actual value and predicted value

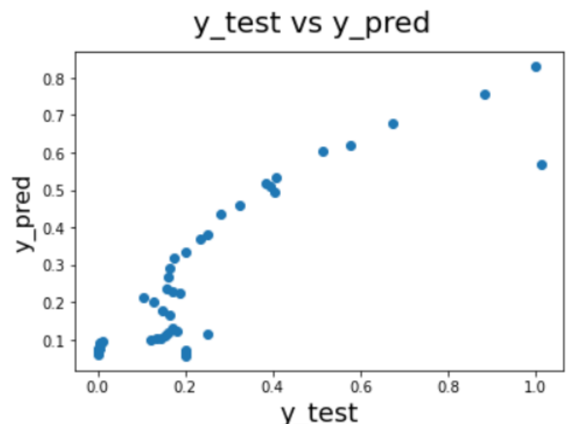


Figure 16 – Scatter plot of actual vs predicted values

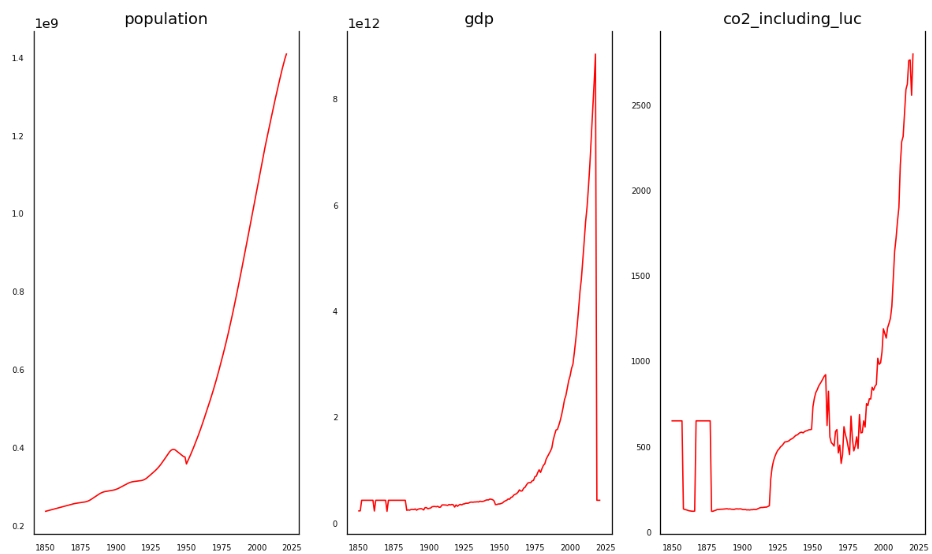


Figure 18 – Line plot of gdp, population and co2_including_luc

- Figure 16 indicates the scatter plot of y_test vs y_predicted.
- Figure 17 indicates the difference value of the actual and predicted values.
- Figure 18 shows the trend and seasonality of the variables.

RESULTS AND DISCUSSION – ADF & GRANGER CAUSALITY TEST

```
Augmented Dickey-Fuller Test:
ADF test statistic         -0.203267
p-value                     0.938131
# lags used                 3.000000
# observations            168.000000
critical value (1%)        -3.469886
critical value (5%)        -2.878903
critical value (10%)       -2.576027
Weak evidence against the null hypothesis
Fail to reject the null hypothesis
Data has a unit root and is non-stationary
```

Figure 19 - ADF test of population

```
Augmented Dickey-Fuller Test:
ADF test statistic          2.319969
p-value                     0.998968
# lags used                 1.000000
# observations            170.000000
critical value (1%)        -3.469413
critical value (5%)        -2.878696
critical value (10%)       -2.575917
Weak evidence against the null hypothesis
Fail to reject the null hypothesis
Data has a unit root and is non-stationary
```

Figure 20 - ADF test of gdp

```
Augmented Dickey-Fuller Test:
ADF test statistic      -1.307838e+01
p-value                  1.893166e-24
# lags used              1.000000e+00
# observations           1.680000e+02
critical value (1%)     -3.469886e+00
critical value (5%)     -2.878903e+00
critical value (10%)    -2.576027e+00
Strong evidence against the null hypothesis
Reject the null hypothesis
Data has no unit root and is stationary
```

Figure 22 – Differenced ADF test of population

```
Augmented Dickey-Fuller Test:
ADF test statistic         -5.460848
p-value                     0.000003
# lags used                 5.000000
# observations            162.000000
critical value (1%)        -3.471374
critical value (5%)        -2.879552
critical value (10%)       -2.576373
Strong evidence against the null hypothesis
Reject the null hypothesis
Data has no unit root and is stationary
```

Figure 23 – Differenced ADF test of gdp

```
Augmented Dickey-Fuller Test:
ADF test statistic          2.319969
p-value                     0.998968
# lags used                 1.000000
# observations            170.000000
critical value (1%)        -3.469413
critical value (5%)        -2.878696
critical value (10%)       -2.575917
Weak evidence against the null hypothesis
Fail to reject the null hypothesis
Data has a unit root and is non-stationary
```

Figure 21 - ADF test of co2_including_luc

```
Augmented Dickey-Fuller Test:
ADF test statistic      -6.122516e+00
p-value                  8.785796e-08
# lags used              2.000000e+00
# observations           1.680000e+02
critical value (1%)     -3.469886e+00
critical value (5%)     -2.878903e+00
critical value (10%)    -2.576027e+00
Strong evidence against the null hypothesis
Reject the null hypothesis
Data has no unit root and is stationary
```

Figure 24 – Differenced ADF test of co2_including_luc

	population_1d_x	gdp_1d_x	co2_including_luc_1d_x
population_1d_y	1.0000	0.3957	0.3301
gdp_1d_y	0.9046	1.0000	0.0474
co2_including_luc_1d_y	0.0009	0.0000	1.0000

Figure 25 – Granger Causality Matrix

- Figures 19,20 and 21 – ADF test
- Figures 22,23 and 24 – Differenced ADF test
- Figure 25 – Granger causality matrix

RESULTS AND DISCUSSION – VAR FORECAST

VAR Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	88.05	88.11	1.729e+38	88.07
1	87.47	87.72*	9.707e+37	87.57*
2	87.40*	87.84	9.050e+37*	87.58
3	87.45	88.08	9.544e+37	87.71
4	87.48	88.30	9.863e+37	87.81
5	87.57	88.57	1.077e+38	87.98
6	87.47	88.67	9.811e+37	87.96
7	87.52	88.90	1.025e+38	88.08
8	87.59	89.16	1.113e+38	88.23
9	87.52	89.28	1.042e+38	88.24
10	87.58	89.52	1.107e+38	88.37
11	87.50	89.63	1.027e+38	88.36
12	87.50	89.82	1.044e+38	88.45
13	87.56	90.07	1.112e+38	88.58
14	87.63	90.32	1.207e+38	88.72
15	87.57	90.45	1.152e+38	88.74

Figure 26 – VAR order selection

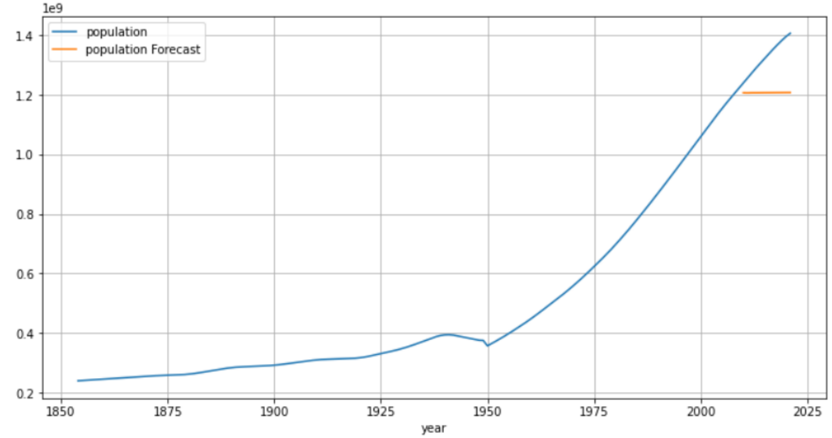


Figure 27 – Actual vs Forecast data of population

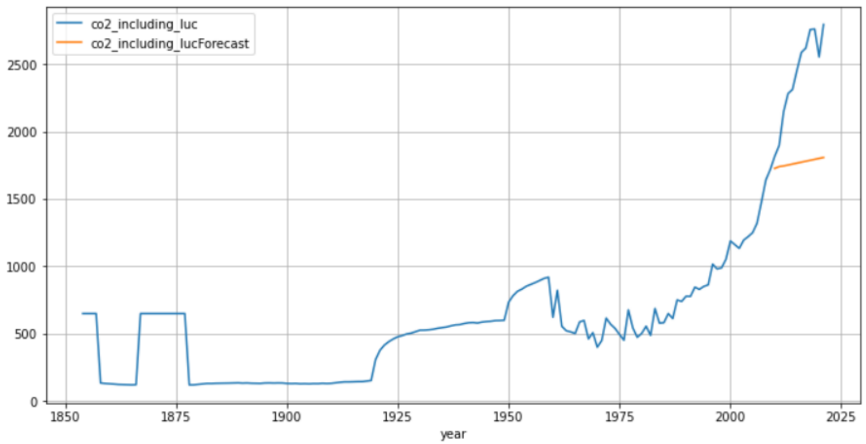


Figure 28 – Actual vs Forecast data of co2_including_luc

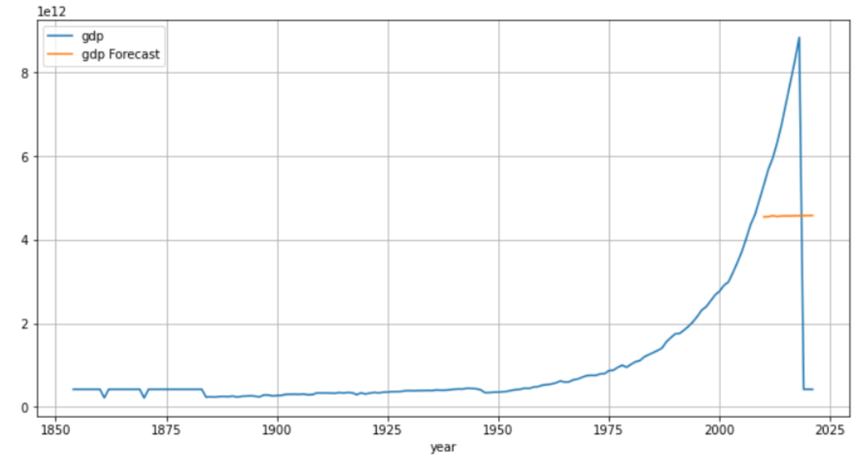


Figure 29 – Actual vs Forecast data of gdp

- Figure 26 – VAR order selection.
- Figures 27, 28, and 29 – Actual data vs Forecasted data of population,co2_including_luc & gdp respectively.

RESULTS AND DISCUSSION - EVALUATION

- The variable ‘gdp’ ranges from 2.15E+11 to 8.84E+12. The variable ‘population’ ranges from 0.23 billion to 1.41 billion. Finally, ‘co2_including_luc’ ranges from 118 to 2795 million tonnes. gdp, population, and co2_including_luc have RMSEs of 5.80E+12, 1.2 billion, and 1683 million tonnes, respectively. Findings suggest forecast accuracy is unreliable.
- Vector autoregression (VAR) is less accurate than multiple linear regression (MLR). MLR has a 78% R^2 value. RMSE values for gdp, population, and co2_including_luc are 5.80E+12, 1.2 billion, and 1683 million tonnes. The RMSE value for the variable ‘co2_including_luc’ falls within the median range, while the RMSE values for ‘population’ and ‘gdp’ are located towards the higher end of the range that encompasses the minimum and maximum values for each respective variable.

CONCLUSION

Let's revisit the research questions and try to address the initial queries:

- **How carbon dioxide (CO_2) emissions, population size, and economic growth are correlated?**
  - The ‘Heat map’ plot compares ‘gdp’, ‘population’, and ‘co2_including_luc’. ‘co2_including_luc’ and ‘population’ are highly correlated compared to ‘gdp’ and ‘co2_including_luc’.
- **Which of the factors, population, or economic growth, exhibits a causal relationship with CO_2 emissions?**
  - Based on the obtained value of 0.0474 establish that the variable co2_including_luc_ Granger causes the variable gdp. Hence, there exists a significant relationship between gdp and co2_including_luc.
- **Which model MLR or VAR gives the highest performance of accuracy in terms of evaluation?**
  - The MLR model has the highest accuracy (78% R^2) compared to the VAR model. The RMSE value for the variable ‘co2_including_luc’ falls within the median range, while the RMSE values for ‘population’ and ‘gdp’ are located towards the higher end of the range that encompasses the minimum and maximum values for each respective variable.

CONCLUSION

Let us reexamine the study's objectives to determine if they have been successfully achieved.

- **To analyze and identify the relationship between economic growth and environmental degradation.**

  - ‘co2_including_luc’ and ‘population’ are highly correlated compared to ‘gdp’ and ‘co2_including_luc’.

- **To explore the causality between economic growth and environmental degradation.**

  - Based on the obtained value of 0.0474 establish that the variable co2_including_luc_ Granger causes the variable gdp. Hence, there exists a significant relationship between gdp and co2_including_luc.

- **To assess the VAR and multiple Linear regression models and identify the most precise one to find the relationship between economic growth and environmental degradation.**

  - The MLR model has superior accuracy, as evidenced by its high R2 value of 78%. The RMSE value for the variable ‘co2_including_luc’ falls within the median range, while the RMSE values for ‘population’ and ‘gdp’ are located towards the higher end of the range that encompasses the minimum and maximum values for each respective The accuracy percentage of R2 value is comparatively lower because of a smaller number of attributes chosen for the study.

- **To use the Vector Autoregressive Model of Timeseries Modeling to forecast CO2 levels of actual values vs predicted values from 2011 to 2021 based on historical data.**

  - The forecasted values for the variables gdp', 'population', and 'co2_including_luc' demonstrate a slight similarity to the observed data.

FUTURE RECOMMENDATIONS

- Elimination of outliers from gdp.
- Study the historical data of 50 years to reduce redundancy.
- The incorporation of supplementary variables.
- Study of other greenhouse gases.
- The identification of locations or cities that serve as significant sources of carbon dioxide (CO2) and other greenhouse gases (GHGs).