Econometrics

Econometric Analysis of Factors Affecting CO_2 Emissions

Data: World Bank (World Development Indicators)
Method: OLS regression + exploratory analysis
Unit of analysis: country

Lilit Gevorgyan, Monika Khachatryan, Maria Zakaryan

Goal

Understand which macro factors are associated with higher/lower CO_2 emissions across countries.

Main questions

- Does higher income (GDP per capita) relate to higher CO_2 emissions?
- Do renewables (REC%) reduce emissions?
- Does energy intensity (energy per unit of GDP) increase emissions?

Country groups

- Developed vs developing (proxied by World Bank income group / high-income dummy).
- Fossil-fuel exporters vs others (dummy variable).

Hypotheses (expected signs)

Hypothesis 1: The Impact of GDP on CO_2 Emissions
Hypothesis 2: Fossil Fuel Exporters vs. Non-Exporters (T-Test)

Source: World Bank (WDI). Analysis done in Python (statsmodels, pandas).

Dataset (country level)

What we used (summary)

- Country observations: 217
- Indicators from World Bank WDI
- Country-level averages across available years
- No deep dive on cleaning — focus is econometric relationships

Dependent variable

CO_2 emissions (log-transformed)

Why log?
- reduces skewness
- easier interpretation (elasticities)

Key explanatory variables

- GDP per capita (log)
- Renewable energy consumption share — REC(%)
- Energy intensity (log)
- Urban population share (%)
- Trade (%)
- Fossil-fuel exporter (1 = Yes)
- High-income dummy (World Bank classification)
  (included as a control; not significant in the final model)

Note: log variables were computed using log(1+x) to handle zero values.

Correlation structure (key variables)

Correlation matrix based on log-transformed variables and country averages.

What stands out

- log(CO_2) & log(GDPpc): strong + correlation (~0.75)
- log(CO_2) & REC(%): moderate − (~−0.56)
- Urbanization correlates with GDPpc (~0.65)

Why this matters

Correlations are not the final answer.

They motivate a multivariate model to separate the effect of GDP from energy structure (renewables, energy intensity) and from country type (fossil exporter).

Examples of “high GDP, low CO_2” countries

The most efficient countries (high GDP, low CO_2 emissions):

	Country_Name	GDP_per_capita	CO2_emissions	CO2_per_GDP
212	Virgin Islands (U.S.)	43040.785330	0.001891	4.392695e-08
65	Faroe Islands	69729.083799	0.039576	5.675631e-07
118	Macao SAR, China	99060.227050	2.900991	2.928512e-05
165	San Marino	64423.943223	2.538527	3.940347e-05
4	Andorra	62755.426890	2.538527	4.045112e-05
21	Bermuda	99742.291873	4.092796	4.103371e-05
188	Switzerland	82051.545759	4.135231	5.039797e-05
173	Sint Maarten (Dutch part)	45856.206211	2.538527	5.535842e-05
124	Malta	57136.842460	3.234078	5.660233e-05
187	Sweden	62150.682009	3.707578	5.965466e-05

Why include this slide?

Shows that income alone does not determine emissions.

Some high-income economies have low CO_2 because of:

- cleaner electricity mix
- services-oriented economies
- strong efficiency policies

Important caution

Many “top efficient” entries are small islands/territories or special cases.

A deeper analysis would control for population, industry structure, and energy exports.

Efficiency proxy here is CO_2 per unit of GDP; used only as an illustrative ranking.

OLS regression model

$$\log(CO_2) = \beta_0 + \beta_1 \log(GDPpc) + \beta_2 REC(\%) + \beta_3 \log(EnergyIntensity)$$
$$+ \beta_4 Urban(\%) + \beta_5 Trade(\%) + \beta_6 FossilExporter + \beta_7 Is\_High\_Income + \varepsilon$$

Interpretation (log model)

- If X is in logs: β is an elasticity (1% ↑ in X → β% ↑ in CO_2)
- If X is in % points (REC, Urban, Trade): β ≈ % change in CO_2 for +1 pp change
- Dummy variable (FossilExporter): exp(β)−1 ≈ % difference vs non-exporters

Why we used logs

- Emissions and GDP are highly skewed
- Logs make relationships closer to linear
- Coefficients become easy to interpret

We also checked a simple bivariate model (log CO_2 on log GDPpc) for intuition.

Method: OLS (statsmodels). Results shown are associations (not causal claims).

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Q("log_CO2_emissions")   R-squared:                       0.698
Model:                                OLS   Adj. R-squared:                  0.688
Method:                     Least Squares   F-statistic:                     69.01
Date:                    Thu, 25 Dec 2025   Prob (F-statistic):           6.58e-51
Time:                            06:39:49   Log-Likelihood:                -141.02
No. Observations:                     217   AIC:                             298.0
Df Residuals:                         209   BIC:                             325.1
Df Model:                               7
Covariance Type:                nonrobust
=================================================================================================
                                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
Intercept                      -5.3943      0.602     -8.965      0.000      -6.581      -4.208
Q("log_GDP_per_capita")         0.6087      0.066      9.291      0.000       0.480       0.738
Q("REC(%)")                    -0.0057      0.002     -3.671      0.000      -0.009      -0.003
Q("log_Energy_intensity")       0.6345      0.090      7.038      0.000       0.457       0.812
Q("Urban_population(%)")       -0.0007      0.002     -0.355      0.723      -0.004       0.003
Q("Trade(%)")                  -0.0007      0.001     -1.040      0.299      -0.002       0.001
Q("Is_High_Income")            -0.0172      0.120     -0.144      0.886      -0.253       0.219
Q("is_fossil_fuel_exporter")    0.3333      0.133      2.507      0.013       0.071       0.595
==============================================================================
Omnibus:                       60.952   Durbin-Watson:                   2.073
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              454.028
Skew:                          -0.833   Prob(JB):                     2.56e-99
Kurtosis:                       9.888   Cond. No.                     2.36e+03
==============================================================================
```

Key coefficients (dependent variable: log CO_2 emissions)

Model fit

$R^2 = 0.698$ | Adj. $R^2 = 0.689$
N = 217 countries

Controls: Urban(%), Trade(%), Fossil exporter dummy
(High-income dummy was not significant and omitted in final spec.)

Income (GDP per capita)

$\beta_1 = 0.602$ ($p < 0.001$)
+1% GDPpc ⟶ ~+0.60% CO_2

Energy structure: renewables

$\beta_2 = -0.0057$ ($p < 0.001$)
+1 pp REC ⟶ ~−0.57% CO_2

Energy efficiency: intensity

$\beta_3 = 0.633$ ($p < 0.001$)
+1% Energy Intensity ⟶ ~+0.63% CO_2

Additional findings

- Fossil-fuel exporter: $\beta = 0.333$ ($p = 0.013$)
  ⟶ exporters have ~40% higher CO_2 (exp(0.333)−1)
- Urbanization and trade: not statistically significant in this specification ($p > 0.1$)

Interpretation uses approximate percentage changes: 100·(e^β−1) for dummy variables and small β for %-point variables.

Hypothesis 1: The effect of GDP on CO_2 emissions

We consider the following hypothesis:

H0: log_ GDP_per_capita does not have a significant impact on log_ CO2_emissions-

H1:log_ GDP_per_capita has a significant impact on log_ CO2_emissions

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Q("log_CO2_emissions")   R-squared:                       0.570
Model:                                OLS   Adj. R-squared:                  0.568
Method:                     Least Squares   F-statistic:                     284.7
Date:                    Thu, 25 Dec 2025   Prob (F-statistic):           3.06e-41
Time:                            04:59:43   Log-Likelihood:                -179.44
No. Observations:                     217   AIC:                             362.9
Df Residuals:                         215   BIC:                             369.6
Df Model:                               1
Covariance Type:                nonrobust
===========================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
Intercept                  -4.3885      0.338    -12.971      0.000      -5.055      -3.722
Q("log_GDP_per_capita")     0.5865      0.035     16.872      0.000       0.518       0.655
==============================================================================
Omnibus:                       46.175   Durbin-Watson:                   2.078
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              264.686
Skew:                           0.633   Prob(JB):                     3.34e-58
Kurtosis:                       8.260   Cond. No.                         88.2
==============================================================================
```

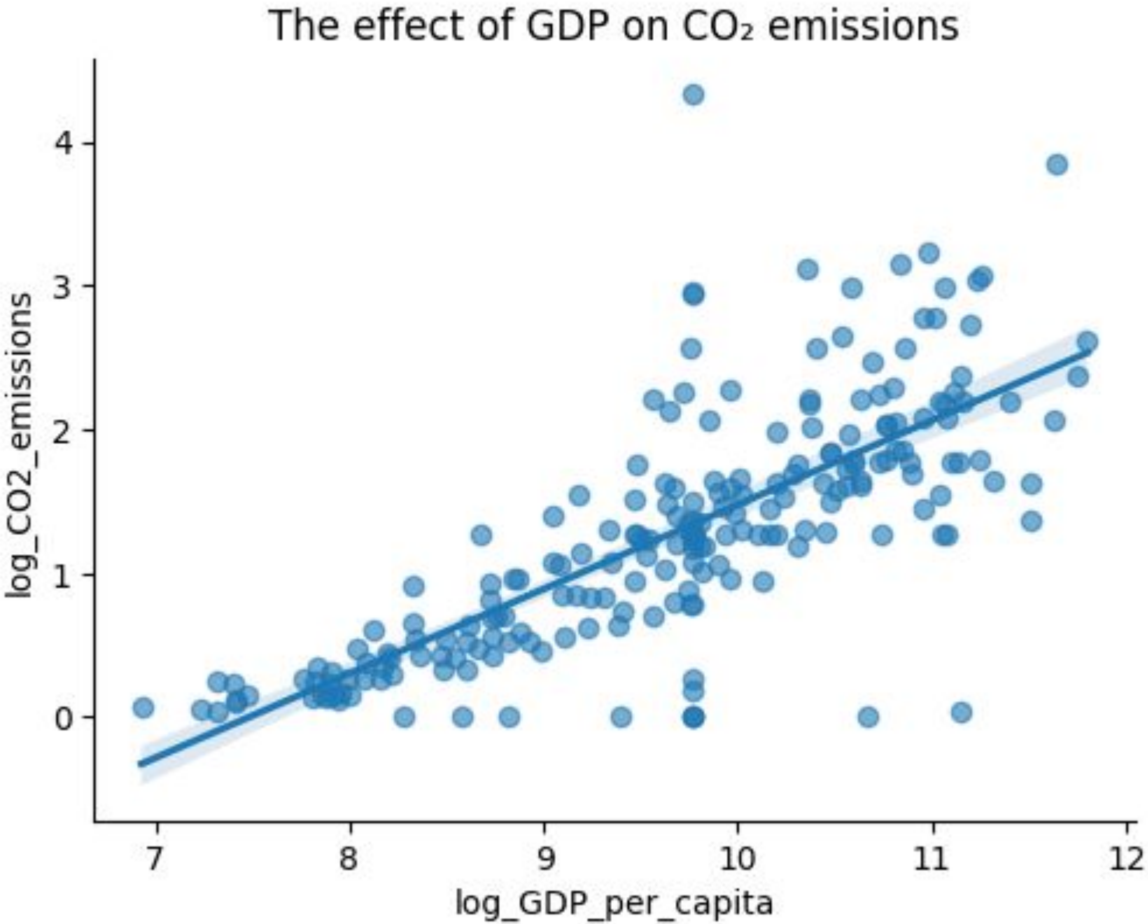
Regression (log-log)

$\log(CO_2) = \alpha + 0.586 \cdot \log(GDPpc)$

$R^2 \approx 0.57$ (N=217)

Interpretation:
+1% GDPpc ⟶ ~+0.59% CO_2

Pvalue=0<0.05=alpha, therefore, we have sufficient evidence to reject the null hypothesis, confirming a significant linear correlation between GDP and carbon emissions.



The line is an OLS fit. Points are countries (averaged values).

Hypothesis 2: Do average CO_2 emissions in fossil fuel exporting countries differ significantly from those in non-exporting countries?

H0: The mean CO_2 emissions of the **exporters** and **non-exporters** groups are equal.

H1: The mean CO_2 emissions of the **exporters** and **non-exporters** groups are not equal.

tcrit=1.97, tcalc=3.78>tcrit,and Pvalue<0.05, the results indicate a statistically significant difference between the two groups.

What we learned

Main takeaways

1) Economic development is strongly associated with higher CO_2.
2) Higher renewable share (REC%) is associated with lower CO_2.
3) Higher energy intensity is associated with higher CO_2.
4) Fossil-fuel exporters have higher CO_2 even after controls.

Implications (high level)

- Developing countries: grow with cleaner energy & efficiency upgrades.
- Developed countries: accelerate renewables and reduce energy intensity.
- Exporters: targeted policies for production and domestic consumption.

Limitations & next steps

Limitations:
- Cross-section associations (not causal)
- Possible omitted variables (industry, population, policy)

Next steps:
- Panel data with fixed effects
- Robust / heteroskedasticity-consistent standard errors
- Interaction terms: “developed × renewables” or “developing × intensity”

Thank you!