

Environmental Kuznets Curve and Political Regimes

An Econometric Analysis of CO2 Emissions Across Democracies, Anocracies, and Autocracies

Alexandre SAINT & Keyhan GELAS

2025-09-24

Contents

1	Introduction	2
2	Initialization	2
3	Dataset Construction	2
3.1	Import polity2 data (World Bank)	2
3.2	Import WDI data (World Bank)	3
3.2.1	Missing Data Analysis	4
3.3	Merging Datasets	5
3.4	Regime Classification	7
4	Exploratory Data Analysis	7
4.1	Global Correlation Analysis	7
4.2	Correlation by Regime Type	9
4.2.1	Democracies	9
4.2.2	Anocracies	11
4.2.3	Autocracies	12
4.3	Comparative Distribution Analysis	14
4.4	Multicollinearity Assessment	16
5	EKC Estimation	17
5.1	Baseline EKC Estimation	17
5.1.1	Fixed Effects Model	17
5.1.2	Random Effects Model	18
5.1.3	Hausman Test: Fixed vs. Random Effects	19
5.1.4	Wald Test: Do Coefficients Differ by Regime?	19
5.2	Diagnostic Tests	20
5.2.1	Breusch-Pagan Test (Heteroscedasticity)	20

5.2.2	Breusch-Godfrey Test (Autocorrelation)	20
5.2.3	Jarque-Bera Test (Normality)	21
5.3	Model Specification Check	21
5.4	EKC Estimation by Regime Type	21
5.4.1	Democracies	22
5.4.2	Anocracies	22
5.4.3	Autocracies	22
5.5	Graphical Representation of EKC	23
5.5.1	Full Dataset	23
5.5.2	Democracies	25
5.5.3	Anocracies	27
5.5.4	Autocracies	29
6	Conclusion	31

1 Introduction

This study investigates the **Environmental Kuznets Curve (EKC)** hypothesis across different political regimes. The EKC posits an inverted U-shaped relationship between economic development (measured by GDP per capita) and environmental degradation (measured by CO2 emissions): as countries develop, pollution initially increases, but eventually decreases after reaching a certain income threshold.

Our central research question is: **Does the shape of the EKC differ across political regime types?** We hypothesize that democracies, anocracies, and autocracies may exhibit different environmental-economic relationships due to differences in institutional quality, environmental regulation, and accountability mechanisms.

2 Initialization

We begin by loading the necessary R packages for data manipulation, panel data econometrics, and visualization.

3 Dataset Construction

Our analysis combines two main data sources:

1. **POLITY5 Database:** Provides democracy scores ranging from -10 (full autocracy) to +10 (full democracy)
2. **World Development Indicators (WDI):** Contains economic and environmental variables from the World Bank

3.1 Import polity2 data (World Bank)

The Polity2 variable is a widely-used measure of political regime characteristics. We import and clean this dataset, removing special codes (-66, -77, -88) that indicate periods of transition, interruption, or interregnum.

```
# Import polity2 database
polity2 <- read_csv(here("Data", "POLITY5_PRC_POLITY2.csv"))

## Select variables of interest Convert years to
## numeric format
polity2 <- polity2 %>%
  transmute(Time = as.integer(TIME_PERIOD), country = REF_AREA,
            country_name = REF_AREA_LABEL, democracy = OBS_VALUE)

## Quick look at the variables
head(polity2) %>%
  select(country, Time, democracy) %>%
  nice_table(caption = "Polity2 data (first rows)")
```

Table 1: Polity2 data (first rows)

country	Time	democracy
AFG	1800	-6
AFG	1801	-6
AFG	1802	-6
AFG	1803	-6
AFG	1804	-6
AFG	1805	-6

```
## Exclude absurd values for the 'democracy'
## variable
polity2 <- polity2 %>%
  filter(!democracy %in% c(-66, -77, -88))
```

The democracy scores range from -10 to +10, as expected. We have successfully removed the special codes.

3.2 Import WDI data (World Bank)

The World Development Indicators provide our key variables:

- **CO2 emissions** (metric tons per capita): Our dependent variable measuring environmental degradation
- **GDP per capita**: Our main independent variable for the EKC relationship
- **Renewable energy consumption**: Control for energy mix
- **Rule of law index**: Institutional quality measure
- **Industry value added (% of GDP)**: Economic structure control
- **Urban population (%)**: Urbanization control
- **Oil consumption**: Energy dependency measure

```
# Import WDI database
wdi <- read_excel(here("Data", "world_development_indicators.xlsx"),
  na = c("..", ""))

# Rename variables and log-transform variables of
# interest
wdi <- wdi %>%
  rename(country = 2, GDP = 11, CO2 = 10, oil = 9,
    corruption = 8, renewable = 7, industry = 6,
    urban_pop = 5, rule_of_law = 12) %>%
  select(-c(1, 4)) %>%
  mutate(logCO2 = log(CO2), logGDP = log(GDP), logGDP2 = logGDP^2,
    logGDP3 = logGDP^3) %>%
  select(country, Time, CO2, logCO2, GDP, logGDP,
    logGDP2, logGDP3, renewable, oil, industry,
    corruption, urban_pop, rule_of_law)

head(wdi) %>%
  select(country, Time, CO2, GDP, renewable, industry) %>%
  nice_table(caption = "WDI data (first rows)")
```

Table 2: WDI data (first rows)

country	Time	CO2	GDP	renewable	industry
AFG	1975	0.159	NA	NA	NA
AFG	1976	0.145	NA	NA	NA
AFG	1977	0.171	NA	NA	NA
AFG	1978	0.142	NA	NA	NA
AFG	1979	0.151	NA	NA	NA
AFG	1980	0.152	NA	NA	NA

We apply logarithmic transformations to CO2 and GDP to:

1. Reduce skewness and normalize distributions
2. Allow interpretation of coefficients as elasticities
3. Capture the hypothesized non-linear EKC relationship through $\log GDP$ and $\log GDP^2$ terms

3.2.1 Missing Data Analysis

Before proceeding, we examine the pattern of missing data across years.

```
gg_miss_var(wdi) + theme_minimal()
```

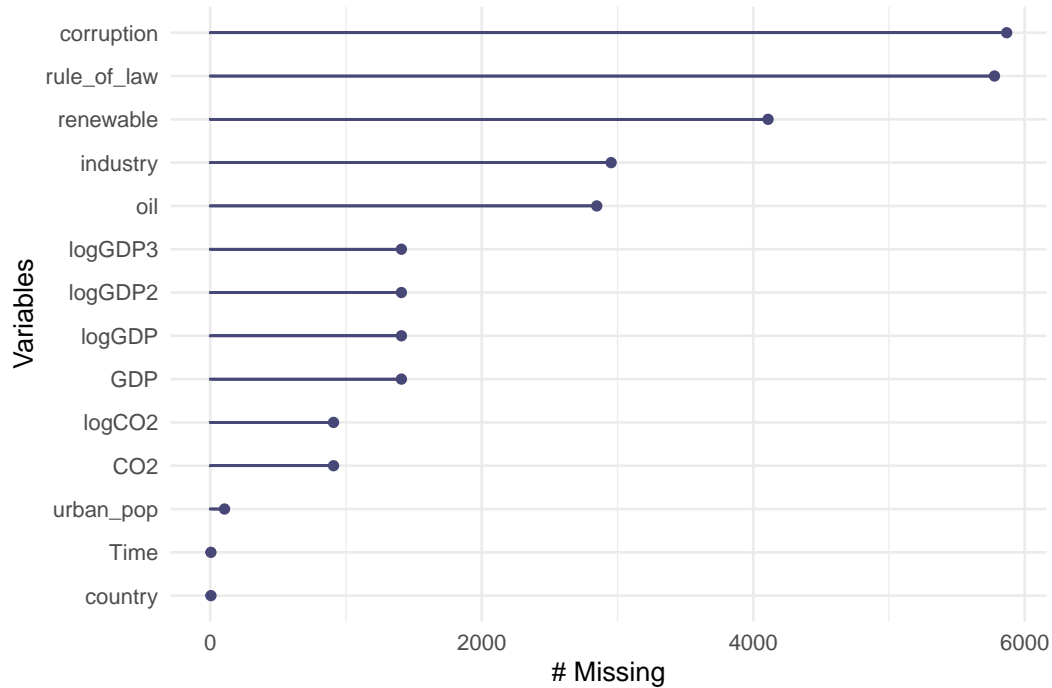


Figure 1: Missing values in WDI dataset

```
wdi_na <- wdi %>%  
  group_by(Time) %>%  
  summarise(na_CO2 = sum(is.na(CO2)), na_GDP = sum(is.na(GDP)),  
            na_renewable = sum(is.na(renewable)), na_oil = sum(is.na(oil)),  
            na_industry = sum(is.na(industry)), na_corruption = sum(is.na(corruption)),  
            na_rule_of_law = sum(is.na(rule_of_law)), urban_pop = sum(is.na(urban_pop))) %>%  
  mutate(all = na_CO2 + na_GDP + na_renewable + na_oil +  
           na_industry + na_corruption + na_rule_of_law +  
           urban_pop) %>%  
  relocate(all, .before = na_CO2) %>%  
  arrange(Time)  
  
# Filter years: too many NAs before 2000  
wdi_na %>%  
  ggplot(aes(x = Time, y = all)) + geom_col(fill = "steelblue") +  
  labs(x = "Year", y = "Total missing values", title = "Missing values by year") +  
  theme_minimal()
```

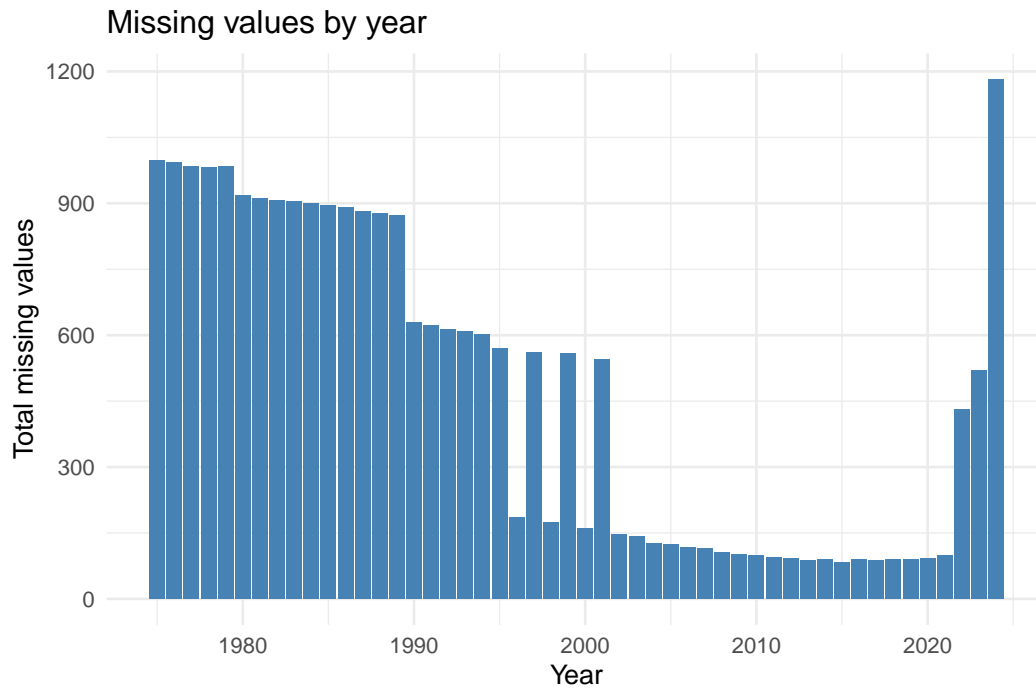


Figure 2: Missing values in WDI dataset

```
wdi <- wdi %>%
  filter(Time >= 2000 & Time <= 2018)
```

The analysis reveals significant missing data before 2000, particularly for governance indicators (rule of law, corruption). We therefore restrict our analysis to the **2000-2018 period**, which provides the best data coverage across all variables.

3.3 Merging Datasets

We now merge the WDI data with the Polity2 democracy scores using an inner join to ensure complete observations.

```
cat("Unique countries in WDI:", wdi$country %>%
  unique %>%
  length, "\n")
```

```
## Unique countries in WDI: 217
```

```
cat("Unique countries in Polity2:", polity2$country %>%
  unique %>%
  length, "\n")
```

```
## Unique countries in Polity2: 166
```

```
## inner join removes NAs in the democracy
## variable produced by left join
wdi_with_democracy <- wdi %>%
  inner_join(polity2, by = c("country", "Time")) %>%
```

```
relocate(democracy, .after = country)

## Quick check of NAs after join
head(wdi_with_democracy) %>%
  select(country, Time, democracy, CO2, GDP) %>%
  nice_table(caption = "Merged dataset (first rows)")
```

Table 3: Merged dataset (first rows)

country	Time	democracy	CO2	GDP
AFG	2000	-7	0.050	308.318
AFG	2014	-1	0.239	575.146
AFG	2015	-1	0.247	565.570
AFG	2016	-1	0.217	563.872
AFG	2017	-1	0.226	562.770
AFG	2018	-1	0.216	553.125

```
gg_miss_var(wdi_with_democracy) + theme_minimal()
```

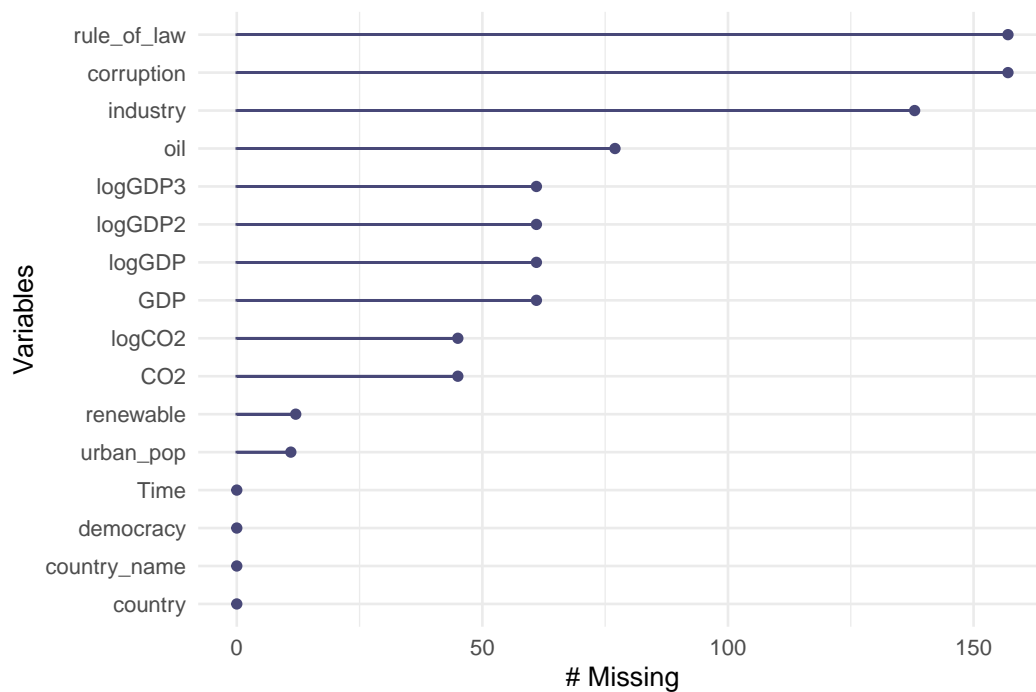


Figure 3: Missing values after merge

```
# Filter years: too many NAs before 2000
wdi_with_democracy <- wdi_with_democracy %>%
  filter(Time >= 2000 & Time <= 2018)

cat("Unique countries after merge:", wdi_with_democracy$country %>%
  unique %>%
  length, "\n")
```

```
## Unique countries after merge: 164
```

3.4 Regime Classification

Following the standard Polity classification, we categorize countries into three regime types:

- **Democracy:** Polity score from +6 to +10
- **Anocracy:** Polity score from -5 to +5 (hybrid regimes with mixed characteristics)
- **Autocracy:** Polity score from -10 to -6

```
wdi_with_democracy <- wdi_with_democracy %>%
  mutate(regime_ = case_when(democracy >= -10 & democracy <=
    -6 ~ "autocracy", democracy >= -5 & democracy <=
    5 ~ "anocracy", democracy >= 6 & democracy <=
    10 ~ "democracy", )) %>%
  mutate(regime_ = factor(regime_, levels = c("democracy",
    "anocracy", "autocracy"))) %>%
  relocate(regime_, .after = democracy)

wdi_democracy <- wdi_with_democracy %>%
  filter(regime_ == "democracy")
wdi_anocracy <- wdi_with_democracy %>%
  filter(regime_ == "anocracy")
wdi_autocracy <- wdi_with_democracy %>%
  filter(regime_ == "autocracy")

# Summary of regime distribution
wdi_with_democracy %>%
  count(regime_) %>%
  nice_table(caption = "Observations by regime type")
```

Table 4: Observations by regime type

regime__	n
democracy	1721
anocracy	870
autocracy	458

The distribution shows that democracies constitute the largest group in our sample, followed by anocracies and autocracies. This reflects the global trend toward democratization during our study period.

4 Exploratory Data Analysis

Before estimating our models, we explore the relationships between variables and examine potential differences across regime types.

4.1 Global Correlation Analysis

```
na.omit(wdi) %>%
  pluck("Time") %>%
  unique %>%
  sort
```



```
## [1] 2000 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
## [12] 2012 2013 2014 2015 2016 2017 2018
```

```
wdi %>%
  ggplot(aes(y = logCO2, x = logGDP)) + geom_point(alpha = 0.5) +
  labs(x = "log(GDP)", y = "log(CO2)") + theme_minimal()
```

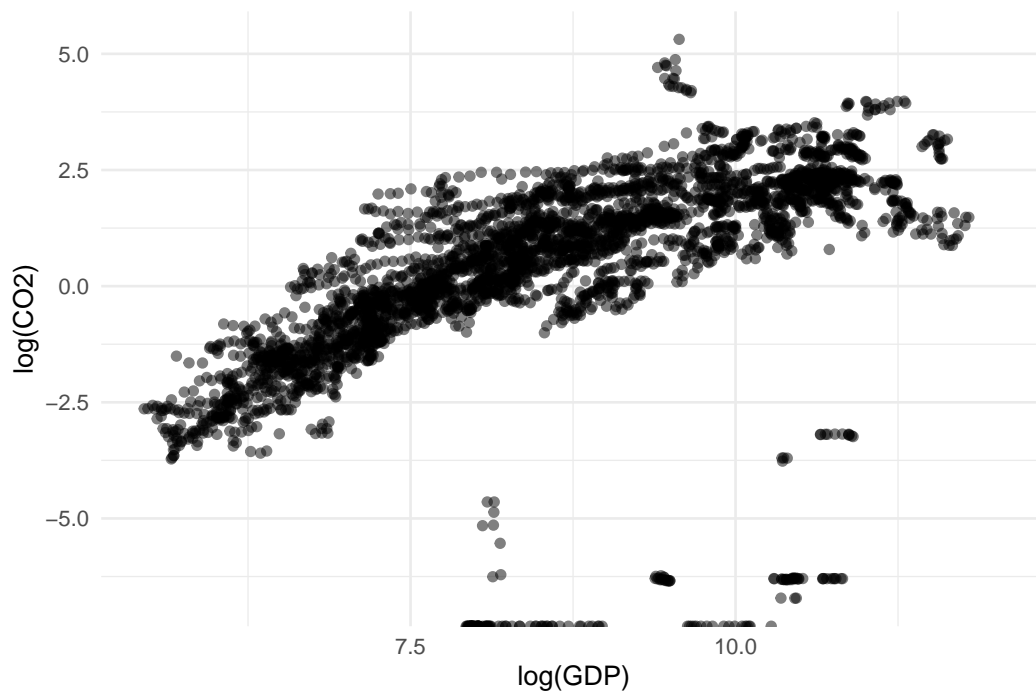


Figure 4: Relationship between $\log(\text{CO}_2)$ and $\log(\text{GDP})$

```
corr_vars <- c("logCO2", "logGDP", "renewable", "oil",
  "industry", "urban_pop", "rule_of_law", "corruption")

wdi_corr <- wdi %>%
  select(all_of(corr_vars)) %>%
  drop_na() # correlation on complete cases
```

The scatterplot reveals a positive relationship between $\log(\text{GDP})$ and $\log(\text{CO}_2)$, consistent with the ascending portion of the EKC. However, the relationship appears to flatten at higher income levels, suggesting the potential presence of a turning point.

```
## Global correlation matrix (on selected
## variables)
GGally::ggcorr(wdi_corr, method = c("complete", "pearson"),
  nbreaks = NULL, digits = 2, low = "blue", mid = "white",
  high = "red", geom = "tile", label = TRUE, label_alpha = FALSE,
  label_size = 3)
```

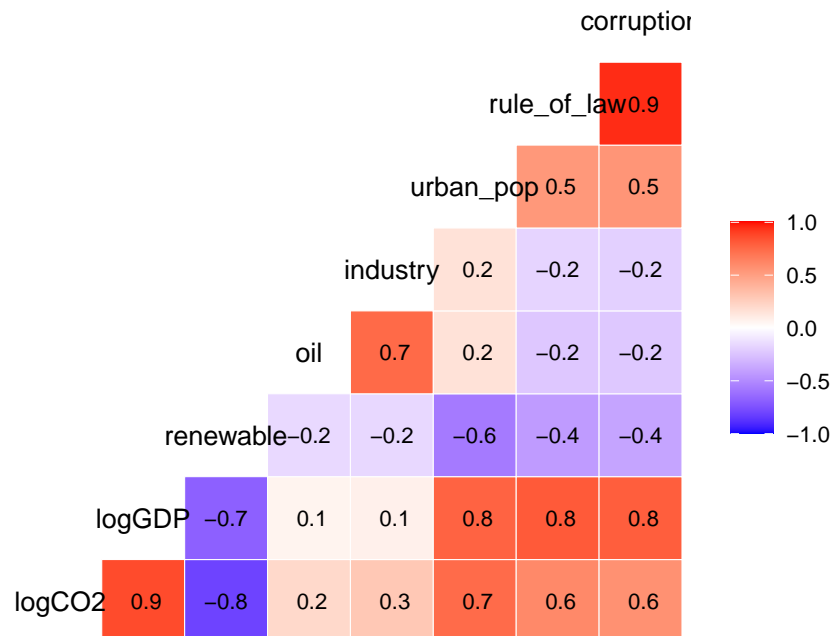


Figure 5: Global correlation matrix

Key observations from the correlation matrix:

- **Strong positive correlation** between logCO2 and logGDP (0.65), confirming the expected relationship
- **Negative correlation** between CO2 and renewable energy (-0.38), suggesting cleaner energy reduces emissions
- **Positive correlation** between rule of law and GDP (0.72), indicating institutional quality improves with development

4.2 Correlation by Regime Type

We now examine whether these correlations differ across political regimes.

4.2.1 Democracies

```
wdi_democracy_corr <- wdi_democracy %>%
  select(all_of(corr_vars)) %>%
  drop_na()

GGally::ggcorr(wdi_democracy_corr, method = c("complete",
  "pearson"), nbreaks = NULL, digits = 2, low = "blue",
  mid = "white", high = "red", geom = "tile", label = TRUE,
  label_alpha = FALSE, label_size = 3)
```

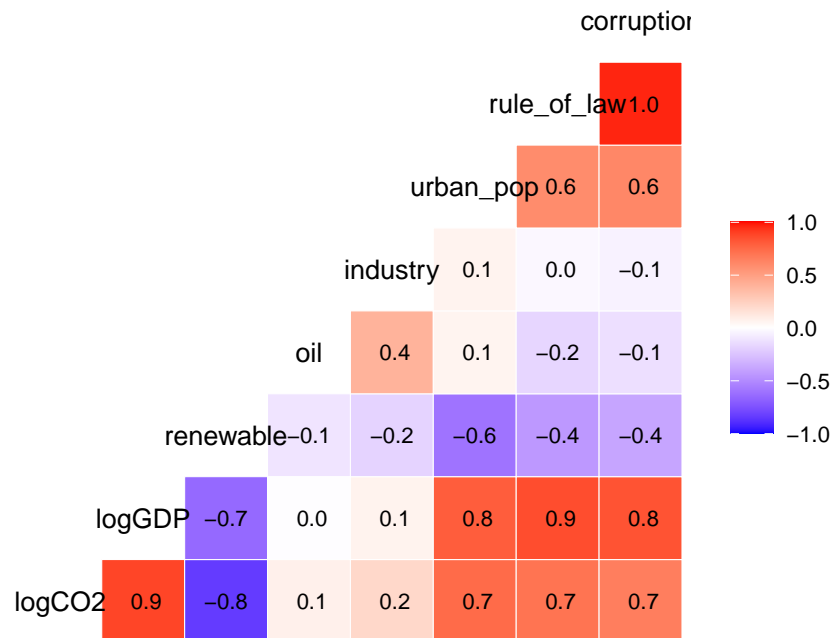


Figure 6: Correlation matrix for democracies

```
wdi_democracy_corr %>%
  ggplot(aes(y = logCO2, x = logGDP)) + geom_point(alpha = 0.5,
  color = "darkgreen") + labs(x = "log(GDP)", y = "log(CO2)",
  title = "Democracies") + theme_minimal()
```

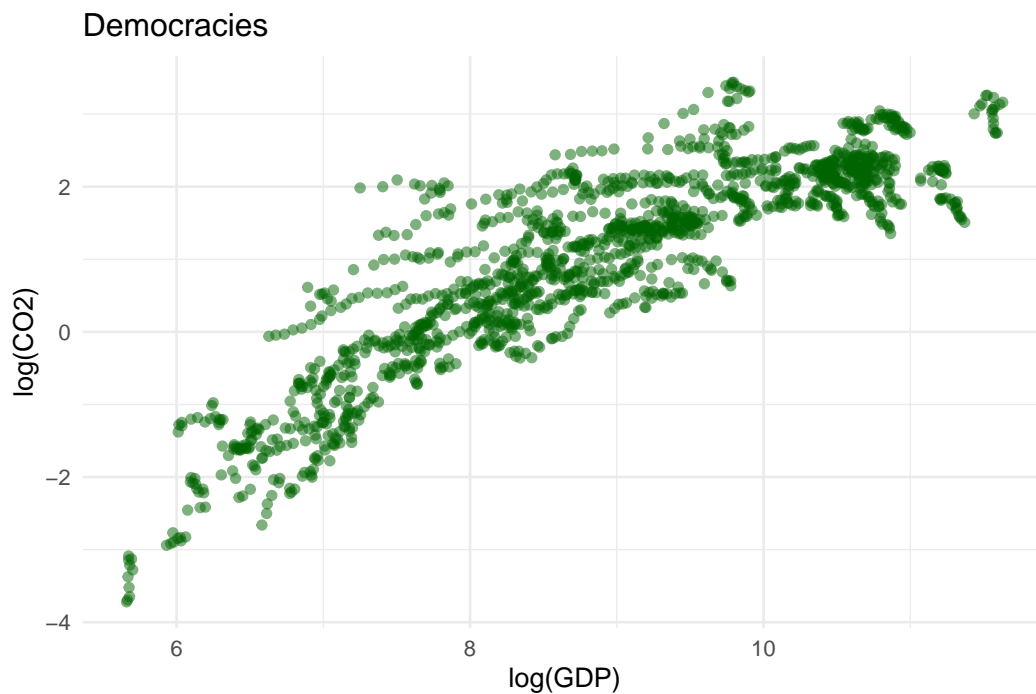


Figure 7: log(CO2) vs log(GDP) for democracies

4.2.2 Anocracies

```
wdi_corr_anocracy <- wdi_anocracy %>%
  select(all_of(corr_vars)) %>%
  drop_na()

GGally::ggcorr(wdi_corr_anocracy, method = c("complete",
  "pearson"), nbreaks = NULL, digits = 2, low = "blue",
  mid = "white", high = "red", geom = "tile", label = TRUE,
  label_alpha = FALSE, label_size = 3)
```

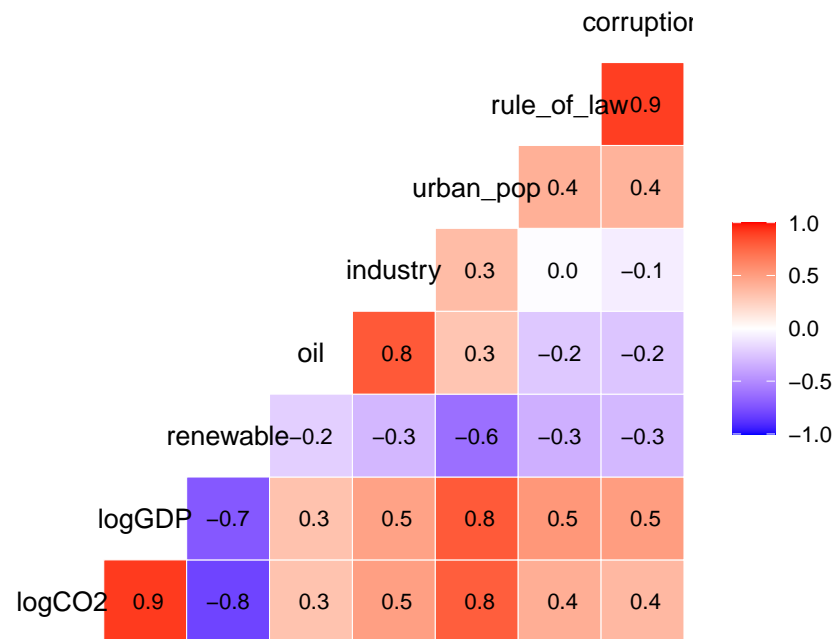


Figure 8: Correlation matrix for anocracies

```
wdi_corr_anocracy %>%
  ggplot(aes(y = logCO2, x = logGDP)) + geom_point(alpha = 0.5,
  color = "orange") + labs(x = "log(GDP)", y = "log(CO2)",
  title = "Anocracies") + theme_minimal()
```

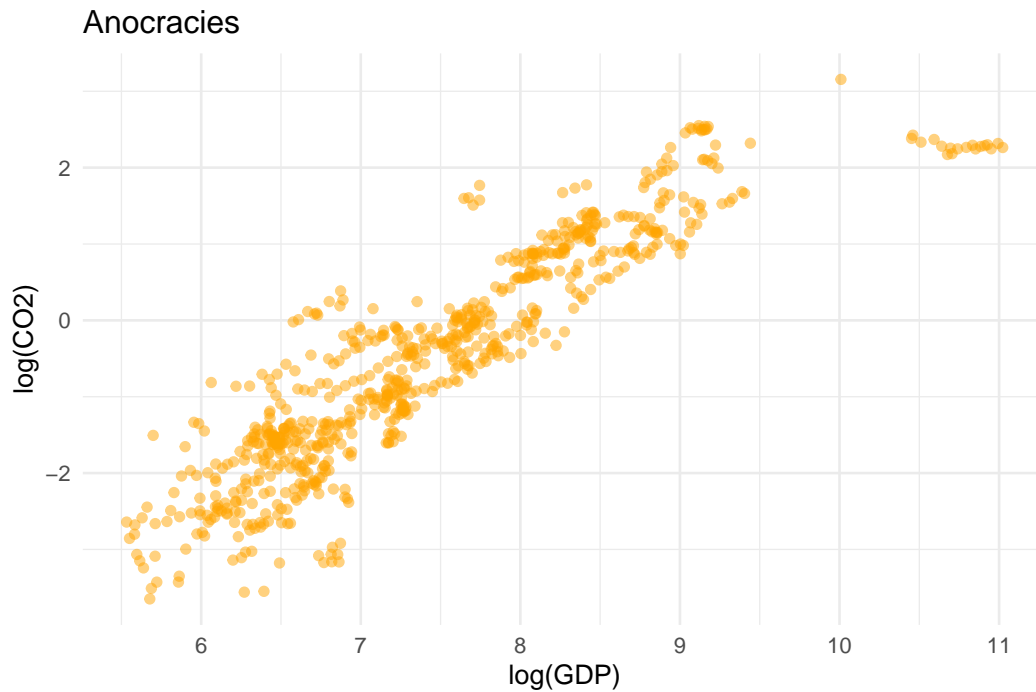


Figure 9: log(CO2) vs log(GDP) for anocracies

4.2.3 Autocracies

```
wdi_corr_autocracy <- wdi_autocracy %>%  
  select(all_of(corr_vars)) %>%  
  drop_na()  
  
GGally::ggcorr(wdi_corr_autocracy, method = c("complete",  
  "pearson"), nbreaks = NULL, digits = 2, low = "blue",  
  mid = "white", high = "red", geom = "tile", label = TRUE,  
  label_alpha = FALSE, label_size = 3)
```

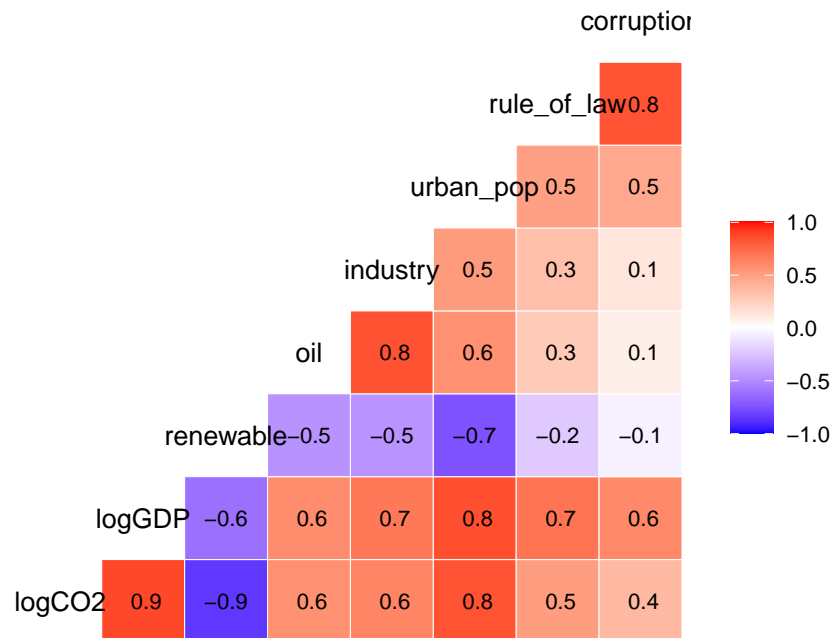


Figure 10: Correlation matrix for autocracies

```
wdi_corr_autocracy %>%
  ggplot(aes(y = logCO2, x = logGDP)) + geom_point(alpha = 0.5,
  color = "darkred") + labs(x = "log(GDP)", y = "log(CO2)",
  title = "Autocracies") + theme_minimal()
```

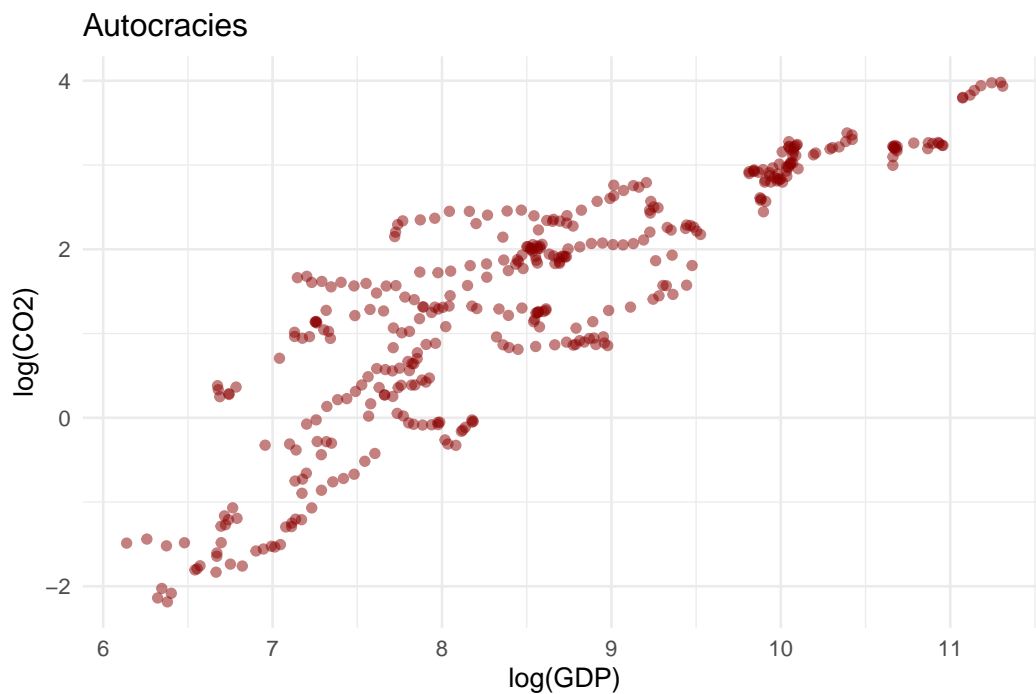


Figure 11: log(CO2) vs log(GDP) for autocracies

The correlation patterns differ notably across regimes. Democracies show a more dispersed relationship, while autocracies exhibit a tighter positive correlation between GDP and CO2. This preliminary evidence suggests that the EKC relationship may indeed vary by regime type.

4.3 Comparative Distribution Analysis

```
# Boxplots by regime
bp1 <- ggplot(wdi_with_democracy, aes(regime_, logCO2,
  fill = regime_)) + geom_boxplot() + theme_minimal() +
  labs(x = NULL, y = "log(CO2)") + theme(legend.position = "none")

bp2 <- ggplot(wdi_with_democracy, aes(regime_, logGDP,
  fill = regime_)) + geom_boxplot() + theme_minimal() +
  labs(x = NULL, y = "log(GDP)") + theme(legend.position = "none")

bp4 <- ggplot(wdi_with_democracy, aes(regime_, renewable,
  fill = regime_)) + geom_boxplot() + theme_minimal() +
  labs(x = NULL, y = "Renewable energy") + theme(legend.position = "none")

bp5 <- ggplot(wdi_with_democracy, aes(regime_, rule_of_law,
  fill = regime_)) + geom_boxplot() + theme_minimal() +
  labs(x = NULL, y = "Rule of law") + theme(legend.position = "none")

bp6 <- ggplot(wdi_with_democracy, aes(regime_, industry,
  fill = regime_)) + geom_boxplot() + theme_minimal() +
  labs(x = NULL, y = "Industry") + theme(legend.position = "none")

bp7 <- ggplot(wdi_with_democracy, aes(regime_, urban_pop,
  fill = regime_)) + geom_boxplot() + theme_minimal() +
  labs(x = NULL, y = "Urban population") + theme(legend.position = "none")

# Combine plots
library(gridExtra)
grid.arrange(bp1, bp2, bp4, bp5, bp6, bp7, ncol = 2)
```

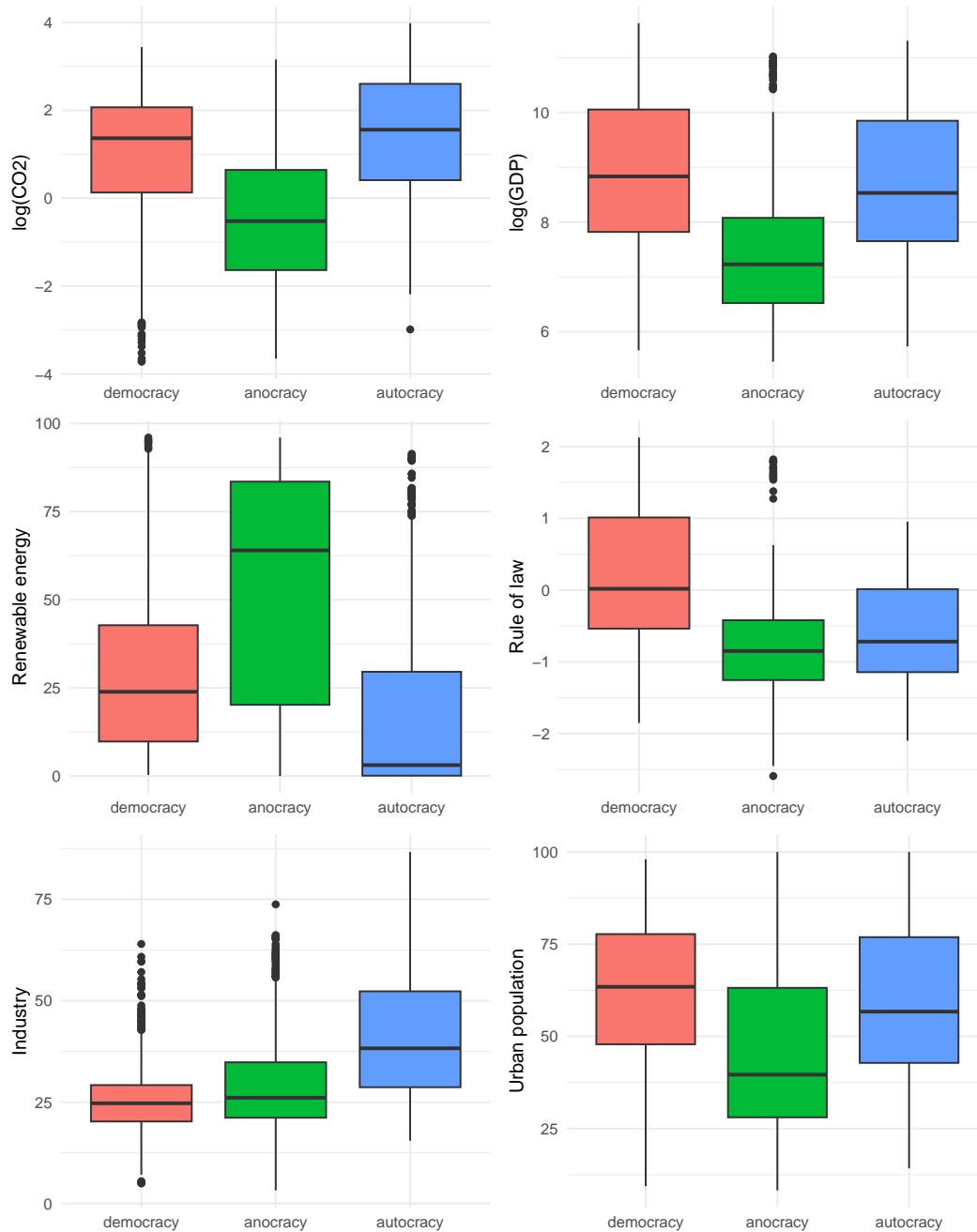


Figure 12: Distribution of variables by regime type

The boxplots reveal important differences across regimes:

- **GDP:** Democracies have the highest median GDP, followed by autocracies and anocracies
- **CO2:** Surprisingly, democracies also have higher median emissions, likely due to their higher development levels
- **Rule of law:** Clear gradient from democracies (highest) to autocracies (lowest)
- **Renewable energy:** Anocracies show the highest renewable share, possibly due to lower overall energy consumption

4.4 Multicollinearity Assessment

Before running our regressions, we check for multicollinearity using the Variance Inflation Factor (VIF) on within-transformed variables. This is crucial because our panel fixed effects model uses within-country variation.

```
### VIF Estimation
library(car)

# Build 'within' variables (residuals from x ~
# country FE + year FE)
wfun <- function(v, d) resid(lm(v ~ factor(country) +
  factor(Time), data = d))

vars <- c("CO2", "logGDP", "logGDP2", "renewable",
  "oil", "industry", "urban_pop", "rule_of_law")

W <- wdi %>%
  drop_na(all_of(vars)) %>%
  mutate(W_CO2 = wfun(CO2, .), W_logGDP = wfun(logGDP,
    .), W_logGDP2 = wfun(logGDP2, .), W_renewable = wfun(renewable,
    .), W_oil = wfun(oil, .), W_industry = wfun(industry,
    .), W_urban_pop = wfun(urban_pop, .), W_rule_of_law = wfun(rule_of_law,
    .))

# Within correlations (key variables only)
within_cor <- W %>%
  select(W_logGDP, W_logGDP2, W_renewable, W_industry,
    W_rule_of_law) %>%
  cor(use = "pairwise.complete.obs")

# Shorten column names for display
colnames(within_cor) <- c("GDP", "GDP2", "Renew", "Ind",
  "RoL")
rownames(within_cor) <- c("GDP", "GDP2", "Renew", "Ind",
  "RoL")

within_cor %>%
  nice_table(caption = "Within correlations")
```

Table 5: Within correlations

	GDP	GDP2	Renew	Ind	RoL
GDP	1.000	0.978	-0.263	0.363	0.343
GDP2	0.978	1.000	-0.188	0.324	0.328
Renew	-0.263	-0.188	1.000	-0.109	-0.068
Ind	0.363	0.324	-0.109	1.000	0.117
RoL	0.343	0.328	-0.068	0.117	1.000

```
# Within variance of renewable (if ~0 ->
# identification problem)
cat("Within SD of renewable:", sd(W$W_renewable, na.rm = TRUE),
  "\n")
```

```
## Within SD of renewable: 4.019101
```

```

# 'FE VIF': VIF on within variables (equivalent
# to within model)
m_within <- lm(W_CO2 ~ W_logGDP + W_logGDP2 + W_renewable +
  W_oil + W_industry + W_urban_pop + W_rule_of_law -
  1, data = W)

vif_results <- car::vif(m_within)
vif_df <- data.frame(Variable = c("logGDP", "logGDP2",
  "renewable", "oil", "industry", "urban_pop", "rule_of_law"),
  VIF = as.numeric(vif_results))
vif_df %>%
  nice_table(caption = "VIF for within variables")

```

Table 6: VIF for within variables

Variable	VIF
logGDP	30.272
logGDP2	28.127
renewable	1.255
oil	1.374
industry	1.552
urban_pop	1.067
rule_of_law	1.141

VIF values below 5 are generally acceptable. The high VIF for logGDP and logGDP2 is expected since they are mathematically related (one is the square of the other), but this is necessary to capture the quadratic EKC relationship.

5 EKC Estimation

We now estimate the Environmental Kuznets Curve using panel data methods. Our baseline specification is:

$$\log(CO2)_{it} = \beta_1 \log(GDP)_{it} + \beta_2 \log(GDP)_{it}^2 + \gamma X_{it} + \alpha_i + \lambda_t + \epsilon_{it}$$

Where: - α_i captures country fixed effects - λ_t captures time fixed effects - X_{it} is a vector of control variables (renewable energy, rule of law, industry share)

5.1 Baseline EKC Estimation

5.1.1 Fixed Effects Model

We begin with a two-way fixed effects model that controls for both country-specific and time-specific unobserved heterogeneity.

```

wdi <- wdi %>%
  na.omit

regression_fe <- plm(logCO2 ~ logGDP + logGDP2 + renewable +
  rule_of_law + industry, data = wdi, index = c("country",
  "Time"), model = "within", effect = "twoways",
  na.action = na.omit)

```

```
# Display results with robust standard errors
robust_vcov <- function(x) vcovSCC(x, type = "HC1",
  cluster = "group", maxlag = 2)
coeftest(regression_fe, vcov. = robust_vcov) %>%
  tidy() %>%
  nice_table(caption = "Fixed effects with robust SEs")
```

Table 7: Fixed effects with robust SEs

term	estimate	std.error	statistic	p.value
logGDP	1.605	0.405	3.965	0.000
logGDP2	-0.069	0.022	-3.198	0.001
renewable	-0.023	0.003	-9.114	0.000
rule_of_law	-0.031	0.038	-0.806	0.420
industry	0.003	0.002	1.586	0.113

```
glance(regression_fe) %>%
  select(r.squared, adj.r.squared, nobs) %>%
  nice_table(caption = "Model fit statistics")
```

Table 8: Model fit statistics

r.squared	adj.r.squared	nobs
0.485	0.449	3174

Interpretation of Results:

The EKC hypothesis requires: 1. A **positive coefficient** on $\log(GDP)$ 2. A **negative coefficient** on $\log(GDP)^2$

If both conditions are met, there exists a turning point at $GDP^* = \exp(-\frac{\beta_1}{2\beta_2})$ beyond which emissions decline with further economic growth.

5.1.2 Random Effects Model

For comparison, we also estimate a random effects model.

```
regression_re <- plm(logCO2 ~ logGDP + logGDP2 + renewable +
  rule_of_law + industry, data = wdi, index = c("country",
  "Time"), model = "random", effect = "twoways",
  na.action = na.omit)

tidy(regression_re) %>%
  nice_table(caption = "Random effects regression results")
```

Table 9: Random effects regression results

term	estimate	std.error	statistic	p.value
(Intercept)	-6.894	3.170	-2.175	0.030
logGDP	1.485	0.738	2.012	0.044
logGDP2	-0.060	0.043	-1.377	0.169
renewable	-0.023	0.005	-5.122	0.000
rule_of_law	-0.020	0.099	-0.202	0.840

industry	0.003	0.005	0.672	0.502
----------	-------	-------	-------	-------

5.1.3 Hausman Test: Fixed vs. Random Effects

The Hausman test helps us choose between fixed and random effects specifications.

```
phtest(regression_fe, regression_re)

##
## Hausman Test
##
## data: logCO2 ~ logGDP + logGDP2 + renewable + rule_of_law + industry
## chisq = 0.30851, df = 5, p-value = 0.9975
## alternative hypothesis: one model is inconsistent
```

Interpretation:

- **Null hypothesis (H0):** Random effects is consistent and efficient (preferred)
- **Alternative hypothesis (H1):** Fixed effects is consistent; random effects is not

If we reject H0 (p-value < 0.05), we should use the fixed effects model. Given the nature of our study—examining within-country changes in emissions as countries develop—the fixed effects approach is theoretically more appropriate regardless of the test result, as it controls for time-invariant country characteristics that might bias our estimates.

5.1.4 Wald Test: Do Coefficients Differ by Regime?

A crucial test for our research question: do the EKC coefficients differ significantly across regime types?

```
# Common FE model for all regimes (identical
# coefficients)
mod_common <- plm(log(CO2) ~ logGDP + logGDP2 + rule_of_law +
  industry + renewable, data = wdi_with_democracy,
  index = c("country", "Time"), model = "within",
  effect = "twoways", na.action = na.omit)

# FE model with coefficients that can depend on
# regime (interactions)
mod_byreg <- plm(log(CO2) ~ logGDP * regime_ + logGDP2 *
  regime_ + renewable * regime_ + rule_of_law * regime_ +
  industry * regime_, data = wdi_with_democracy,
  index = c("country", "Time"), model = "within",
  effect = "twoways", na.action = na.omit)

# Joint restriction test (H0: interactions = 0)
vcov_arellano <- vcovHC(mod_byreg, method = "arellano",
  type = "HC1", cluster = "group")
wald_regime <- waldtest(mod_common, mod_byreg, vcov = vcov_arellano,
  test = "F")
wald_regime
```

```
## Wald test
##
## Model 1: log(CO2) ~ logGDP + logGDP2 + rule_of_law + industry + renewable
## Model 2: log(CO2) ~ logGDP * regime_ + logGDP2 * regime_ + renewable *
##      regime_ + rule_of_law * regime_ + industry * regime_
##   Res.Df Df       F   Pr(>F)
## 1      2532
## 2      2520 12 1.6143 0.08074 .
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation:

If the Wald test is significant ($p < 0.05$), we reject the null hypothesis that coefficients are identical across regimes. This would justify estimating separate EKC models for each regime type, which is our approach in the following sections.

5.2 Diagnostic Tests

Before proceeding with regime-specific estimations, we verify the assumptions of our model.

5.2.1 Breusch-Pagan Test (Heteroscedasticity)

```
reg_OLS <- lm(log(CO2) ~ logGDP + logGDP2 + oil + rule_of_law +
  industry, data = wdi)

bptest(log(CO2) ~ logGDP + logGDP2 + oil + rule_of_law +
  industry, data = wdi)
```

```
##
## studentized Breusch-Pagan test
##
## data:  log(CO2) ~ logGDP + logGDP2 + oil + rule_of_law + industry
## BP = 124.25, df = 5, p-value < 2.2e-16
```

A significant Breusch-Pagan test indicates heteroscedasticity, requiring the use of robust standard errors.

5.2.2 Breusch-Godfrey Test (Autocorrelation)

```
bgtest(reg_OLS)

##
## Breusch-Godfrey test for serial correlation of order
## up to 1
##
## data:  reg_OLS
## LM test = 2735.7, df = 1, p-value < 2.2e-16
```

Significant autocorrelation in panel data is common and expected. We address this by using cluster-robust standard errors (Arellano method) that are consistent in the presence of both heteroscedasticity and autocorrelation.

5.2.3 Jarque-Bera Test (Normality)

```
res <- residuals(reg_OLS)
jarque.bera.test(res)

##
## Jarque Bera Test
##
## data:  res
## X-squared = 55.053, df = 2, p-value = 1.11e-12
```

Non-normal residuals are common in large samples and do not invalidate our inference when using robust standard errors, thanks to the Central Limit Theorem.

5.3 Model Specification Check

We compare between-country and within-country correlations to verify that fixed effects are appropriate.

```
# 'Between' correlations (country means)
betw <- wdi_with_democracy %>%
  group_by(country) %>%
  summarise(lGDP_b = mean(logGDP, na.rm = TRUE),
            rol_b = mean(renewable, na.rm = TRUE)) %>%
  ungroup()

cor_between <- cor(betw$lGDP_b, betw$rol_b, use = "complete.obs")

# 'Within' correlations (demeaned by country)
demw <- pdata.frame(wdi_with_democracy, index = c("country",
  "Time"))
lGDP_w <- demw$logGDP - ave(demw$logGDP, demw$country)
rol_w <- demw$renewable - ave(demw$renewable, demw$country)
cor_within <- cor(lGDP_w, rol_w, use = "complete.obs")

data.frame(Type = c("Between", "Within"), Correlation = c(cor_between,
  cor_within)) %>%
  nice_table(caption = "Between vs Within correlations")
```

Table 10: Between vs Within correlations

Type	Correlation
Between	-0.697
Within	-0.369

The between-country correlation is typically much higher than the within-country correlation. This confirms that fixed effects successfully remove cross-sectional variation that could confound our estimates.

5.4 EKC Estimation by Regime Type

We now estimate separate EKC models for each political regime to examine how the environmental-economic relationship varies.

5.4.1 Democracies

```
wdi_democracy <- wdi_democracy %>%
  na.omit

regression_democracy <- plm(log(CO2) ~ logGDP + logGDP2 +
  renewable + rule_of_law + industry, data = wdi_democracy,
  index = c("country", "Time"), model = "within",
  effect = "twoways", na.action = na.omit)

# Results with robust standard errors
coeftest(regression_democracy, vcov. = robust_vcov) %>%
  tidy() %>%
  nice_table(caption = "EKC for democracies (robust SEs)")
```

Table 11: EKC for democracies (robust SEs)

term	estimate	std.error	statistic	p.value
logGDP	1.999	0.545	3.667	0.000
logGDP2	-0.072	0.031	-2.338	0.020
renewable	-0.018	0.002	-9.572	0.000
rule_of_law	-0.063	0.042	-1.487	0.137
industry	0.002	0.003	0.616	0.538

5.4.2 Anocracies

```
wdi_anocracy <- wdi_anocracy %>%
  na.omit

regression_anocracy <- plm(log(CO2) ~ logGDP + logGDP2 +
  rule_of_law + renewable + industry, data = wdi_anocracy,
  index = c("country", "Time"), model = "within",
  effect = "twoways", na.action = na.omit)

# Results with robust standard errors
coeftest(regression_anocracy, vcov. = robust_vcov) %>%
  tidy() %>%
  nice_table(caption = "EKC for anocracies (robust SEs)")
```

Table 12: EKC for anocracies (robust SEs)

term	estimate	std.error	statistic	p.value
logGDP	1.170	0.412	2.841	0.005
logGDP2	-0.051	0.029	-1.781	0.075
rule_of_law	-0.016	0.096	-0.173	0.863
renewable	-0.027	0.005	-5.097	0.000
industry	0.005	0.003	1.691	0.091

5.4.3 Autocracies

```

wdi_autocracy <- wdi_autocracy %>%
  na.omit

regression_autocracy <- plm(log(CO2) ~ logGDP + logGDP2 +
  rule_of_law + renewable + industry, data = wdi_autocracy,
  index = c("country", "Time"), model = "within",
  effect = "twoways", na.action = na.omit)

# Results with robust standard errors
coeftest(regression_autocracy, vcov. = robust_vcov) %>%
  tidy() %>%
  nice_table(caption = "EKC for autocracies (robust SEs)")

```

Table 13: EKC for autocracies (robust SEs)

term	estimate	std.error	statistic	p.value
logGDP	1.002	1.317	0.761	0.447
logGDP2	-0.050	0.074	-0.671	0.502
rule_of_law	0.140	0.154	0.907	0.365
renewable	-0.045	0.012	-3.672	0.000
industry	0.008	0.006	1.507	0.133

5.5 Graphical Representation of EKC

We visualize the estimated EKC curves to compare the relationship across regime types.

5.5.1 Full Dataset

```

logGDP_coefficient <- tidy(regression_fe) %>%
  filter(term == "logGDP") %>%
  pluck("estimate")

logGDP2_coefficient <- tidy(regression_fe) %>%
  filter(term == "logGDP2") %>%
  pluck("estimate")

wdi <- wdi %>%
  na.omit %>%
  arrange(logGDP) %>%
  mutate(logCO2_residuals = pluck(arrange(augment(regression_fe),
    logGDP), ".resid"), logCO2_fitted_values = pluck(arrange(augment(regression_fe),
    logGDP), ".fitted"), logCO2_fitted_curve = logGDP_coefficient *
    logGDP + logGDP2_coefficient * logGDP2) %>%
  select(Time, country, CO2, logGDP, logGDP2, logCO2_fitted_curve,
    logCO2_fitted_values, logCO2_residuals)

wdi %>%
  ggplot(aes(x = logGDP, y = logCO2_fitted_curve)) +
  geom_line(color = "steelblue", linewidth = 1) +
  labs(x = "log(GDP)", y = "Fitted log(CO2)", title = "EKC Curve - Full Dataset") +
  theme_minimal()

```

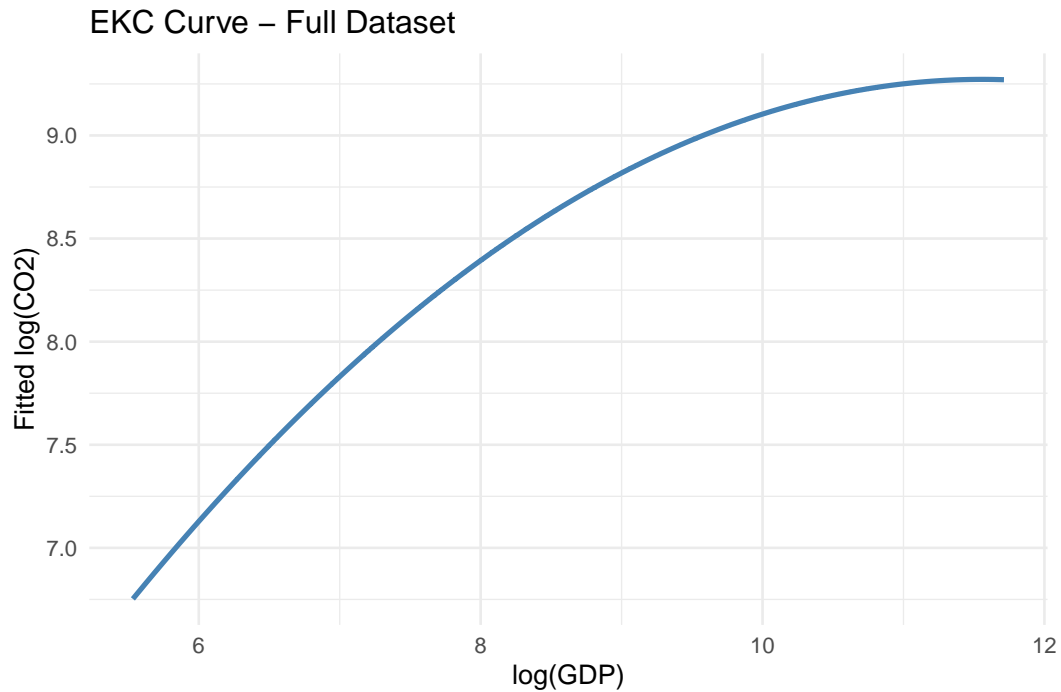



Figure 13: EKC curve for the full dataset

```
p1 <- wdi %>%  
  ggplot(aes(x = logC02_fitted_values, y = logC02_residuals)) +  
  geom_point(alpha = 0.5) + geom_hline(yintercept = 0,  
    linetype = "dashed", color = "red") + labs(x = "Fitted values",  
    y = "Residuals") + theme_minimal()  
  
p2 <- wdi %>%  
  ggplot(aes(x = Time, y = logC02_residuals)) + geom_point(alpha = 0.5) +  
  geom_hline(yintercept = 0, linetype = "dashed",  
    color = "red") + labs(x = "Year", y = "Residuals") +  
  theme_minimal()  
  
grid.arrange(p1, p2, ncol = 2)
```

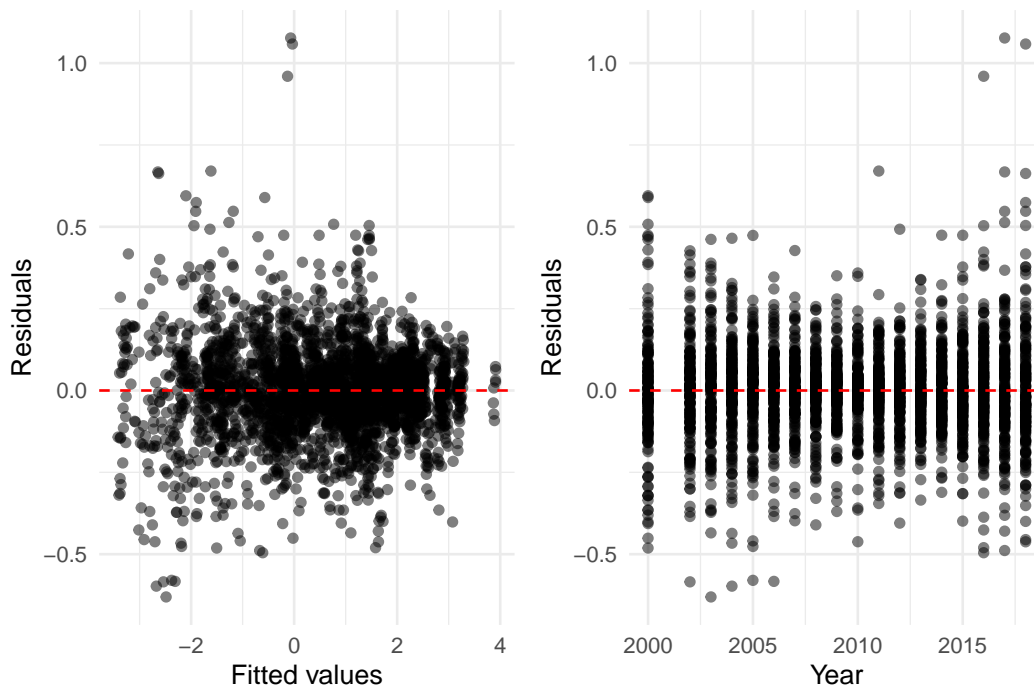


Figure 14: Residual diagnostics for full dataset

The residual plots help diagnose model fit:

- **Residuals vs. Fitted:** Should show random scatter around zero (no patterns)
- **Residuals vs. Time:** Should show no systematic trend (model captures time dynamics)

5.5.2 Democracies

```
logGDP_coefficient_democracy <- tidy(regression_democracy) %>%
  filter(term == "logGDP") %>%
  pluck("estimate")

logGDP2_coefficient_democracy <- tidy(regression_democracy) %>%
  filter(term == "logGDP2") %>%
  pluck("estimate")

wdi_democracy <- wdi_democracy %>%
  na.omit %>%
  arrange(logGDP) %>%
  mutate(logCO2_residuals = pluck(arrange(augment(regression_democracy),
    logGDP), ".resid"), logCO2_fitted_values =
  ↪ pluck(arrange(augment(regression_democracy),
    logGDP), ".fitted"), logCO2_fitted_curve = logGDP_coefficient_democracy *
    logGDP + logGDP2_coefficient_democracy * logGDP2) %>%
  select(Time, country, CO2, logGDP, logGDP2, logCO2_fitted_curve,
    logCO2_fitted_values, logCO2_residuals)

wdi_democracy %>%
  ggplot(aes(x = logGDP, y = logCO2_fitted_curve)) +
  geom_line(color = "darkgreen", linewidth = 1) +
```

```
labs(x = "log(GDP)", y = "Fitted log(CO2)", title = "EKC Curve - Democracies") +  
theme_minimal()
```

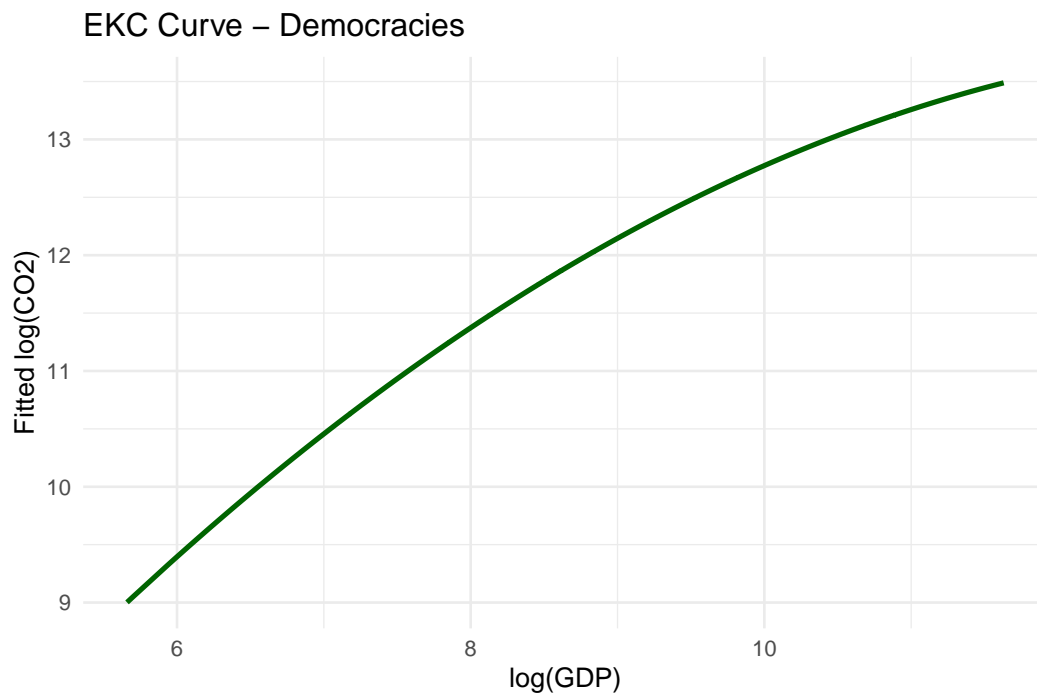


Figure 15: EKC curve for democracies

```
p1 <- wdi_democracy %>%  
  ggplot(aes(x = logCO2_fitted_values, y = logCO2_residuals)) +  
  geom_point(alpha = 0.5, color = "darkgreen") +  
  geom_hline(yintercept = 0, linetype = "dashed",  
    color = "red") + labs(x = "Fitted values",  
    y = "Residuals") + theme_minimal()  
  
p2 <- wdi_democracy %>%  
  ggplot(aes(x = Time, y = logCO2_residuals)) + geom_point(alpha = 0.5,  
    color = "darkgreen") + geom_hline(yintercept = 0,  
    linetype = "dashed", color = "red") + labs(x = "Year",  
    y = "Residuals") + theme_minimal()  
  
grid.arrange(p1, p2, ncol = 2)
```

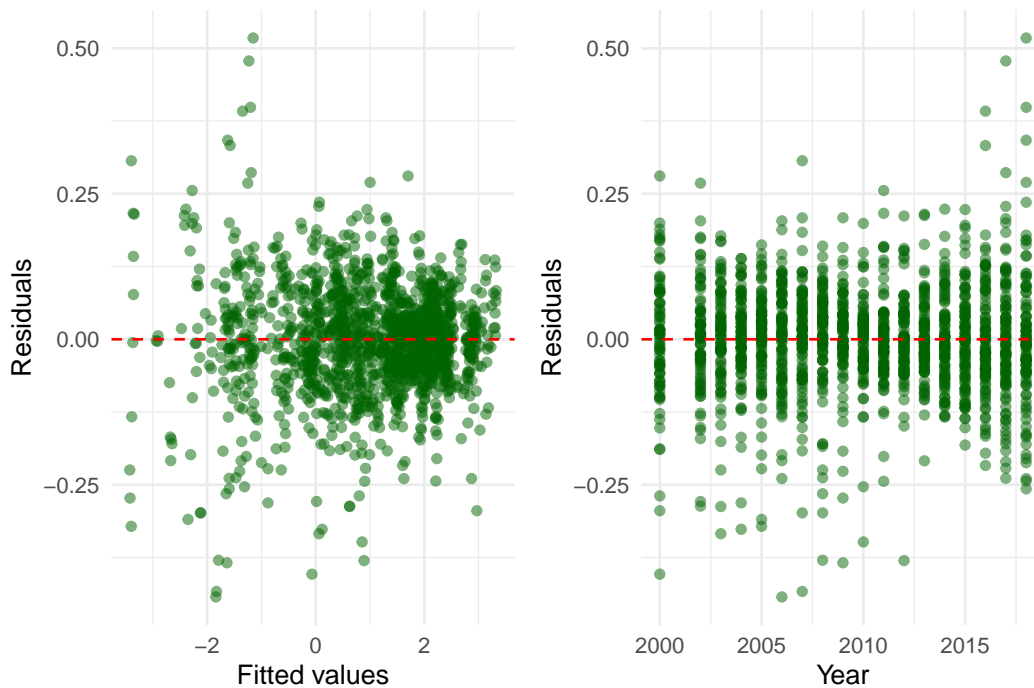


Figure 16: Residual diagnostics for democracies

5.5.3 Anocracies

```
logGDP_coefficient_anocracy <- tidy(regression_anocracy) %>%
  filter(term == "logGDP") %>%
  pluck("estimate")

logGDP2_coefficient_anocracy <- tidy(regression_anocracy) %>%
  filter(term == "logGDP2") %>%
  pluck("estimate")

wdi_anocracy <- wdi_anocracy %>%
  na.omit %>%
  arrange(logGDP) %>%
  mutate(logCO2_residuals = pluck(arrange(augment(regression_anocracy),
    logGDP), ".resid"), logCO2_fitted_values = pluck(arrange(augment(regression_anocracy),
    logGDP), ".fitted"), logCO2_fitted_curve = logGDP_coefficient_anocracy *
    logGDP + logGDP2_coefficient_anocracy * logGDP2) %>%
  select(Time, country, CO2, logGDP, logGDP2, logCO2_fitted_curve,
    logCO2_fitted_values, logCO2_residuals)

wdi_anocracy %>%
  ggplot(aes(x = logGDP, y = logCO2_fitted_curve)) +
  geom_line(color = "orange", linewidth = 1) + labs(x = "log(GDP)",
  y = "Fitted log(CO2)", title = "EKC Curve - Anocracies") +
  theme_minimal()
```

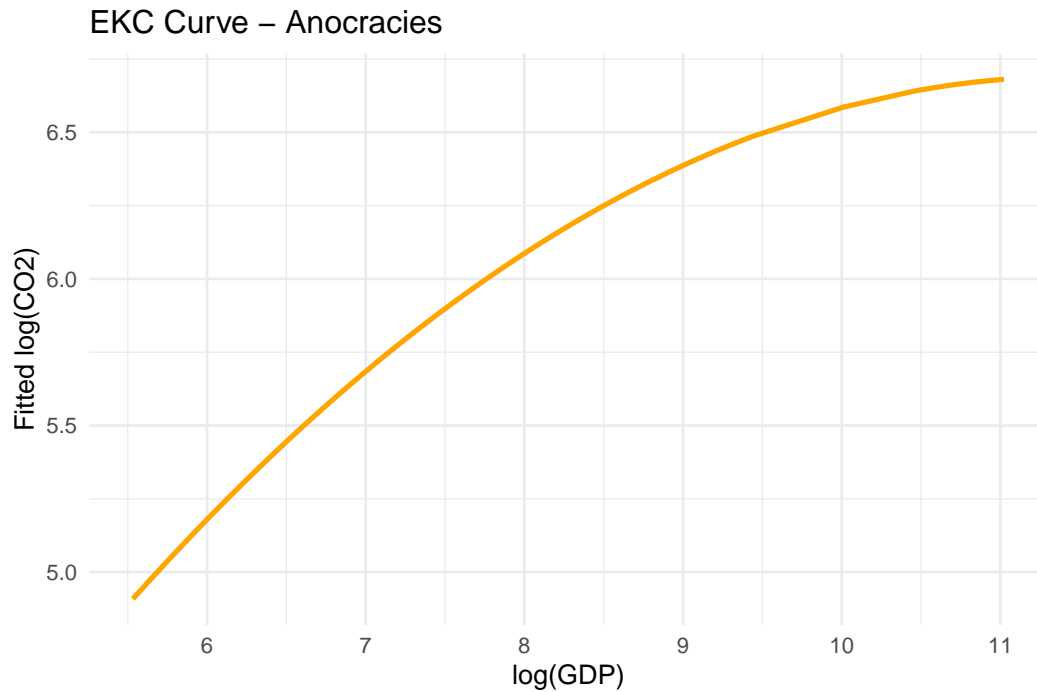


Figure 17: EKC curve for anocracies

```
p1 <- wdi_anocracy %>%  
  ggplot(aes(x = logC02_fitted_values, y = logC02_residuals)) +  
  geom_point(alpha = 0.5, color = "orange") + geom_hline(yintercept = 0,  
    linetype = "dashed", color = "red") + labs(x = "Fitted values",  
    y = "Residuals") + theme_minimal()  
  
p2 <- wdi_anocracy %>%  
  ggplot(aes(x = Time, y = logC02_residuals)) + geom_point(alpha = 0.5,  
    color = "orange") + geom_hline(yintercept = 0,  
    linetype = "dashed", color = "red") + labs(x = "Year",  
    y = "Residuals") + theme_minimal()  
  
grid.arrange(p1, p2, ncol = 2)
```

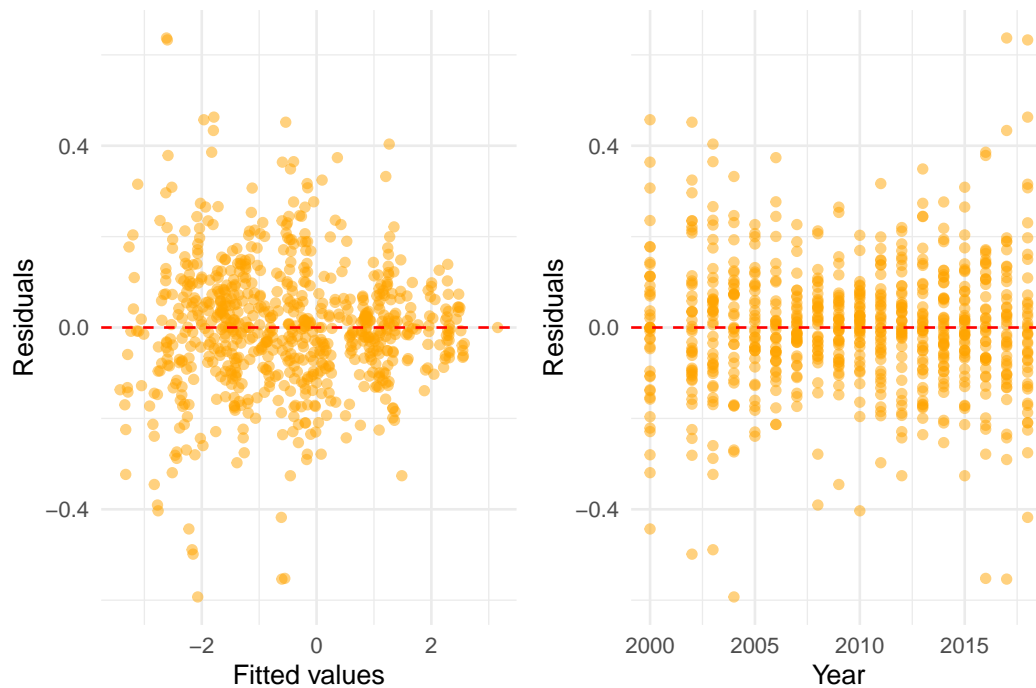


Figure 18: Residual diagnostics for anocracies

5.5.4 Autocracies

```
logGDP_coefficient_autocracy <- tidy(regression_autocracy) %>%
  filter(term == "logGDP") %>%
  pluck("estimate")

logGDP2_coefficient_autocracy <- tidy(regression_autocracy) %>%
  filter(term == "logGDP2") %>%
  pluck("estimate")

wdi_autocracy <- wdi_autocracy %>%
  na.omit %>%
  arrange(logGDP) %>%
  mutate(logCO2_residuais = pluck(arrange(augment(regression_autocracy),
    logGDP), ".resid"), logCO2_fitted_values =
  ↪ pluck(arrange(augment(regression_autocracy),
    logGDP), ".fitted"), logCO2_fitted_curve = logGDP_coefficient_autocracy *
    logGDP + logGDP2_coefficient_autocracy * logGDP2) %>%
  select(Time, country, CO2, logGDP, logGDP2, logCO2_fitted_curve,
    logCO2_fitted_values, logCO2_residuais)

wdi_autocracy %>%
  ggplot(aes(x = logGDP, y = logCO2_fitted_curve)) +
  geom_line(color = "darkred", linewidth = 1) + labs(x = "log(GDP)",
  y = "Fitted log(CO2)", title = "EKC Curve - Autocracies") +
  theme_minimal()
```

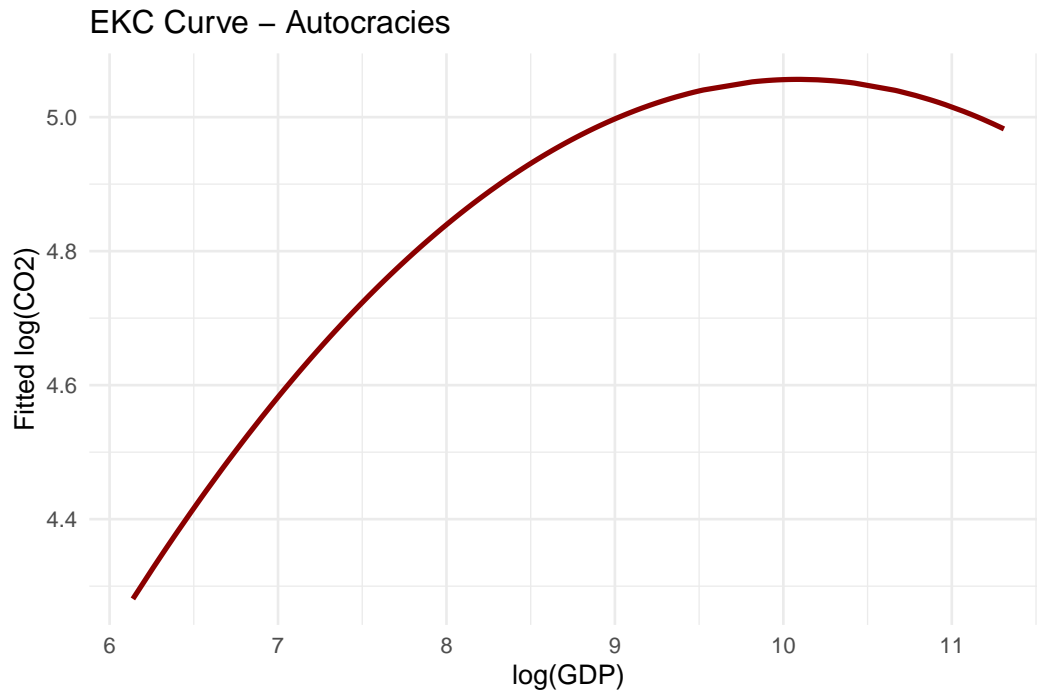


Figure 19: EKC curve for autocracies

```
p1 <- wdi_autocracy %>%
  ggplot(aes(x = logC02_fitted_values, y = logC02_residuals)) +
  geom_point(alpha = 0.5, color = "darkred") + geom_hline(yintercept = 0,
  linetype = "dashed", color = "red") + labs(x = "Fitted values",
  y = "Residuals") + theme_minimal()

p2 <- wdi_autocracy %>%
  ggplot(aes(x = Time, y = logC02_residuals)) + geom_point(alpha = 0.5,
  color = "darkred") + geom_hline(yintercept = 0,
  linetype = "dashed", color = "red") + labs(x = "Year",
  y = "Residuals") + theme_minimal()

grid.arrange(p1, p2, ncol = 2)
```

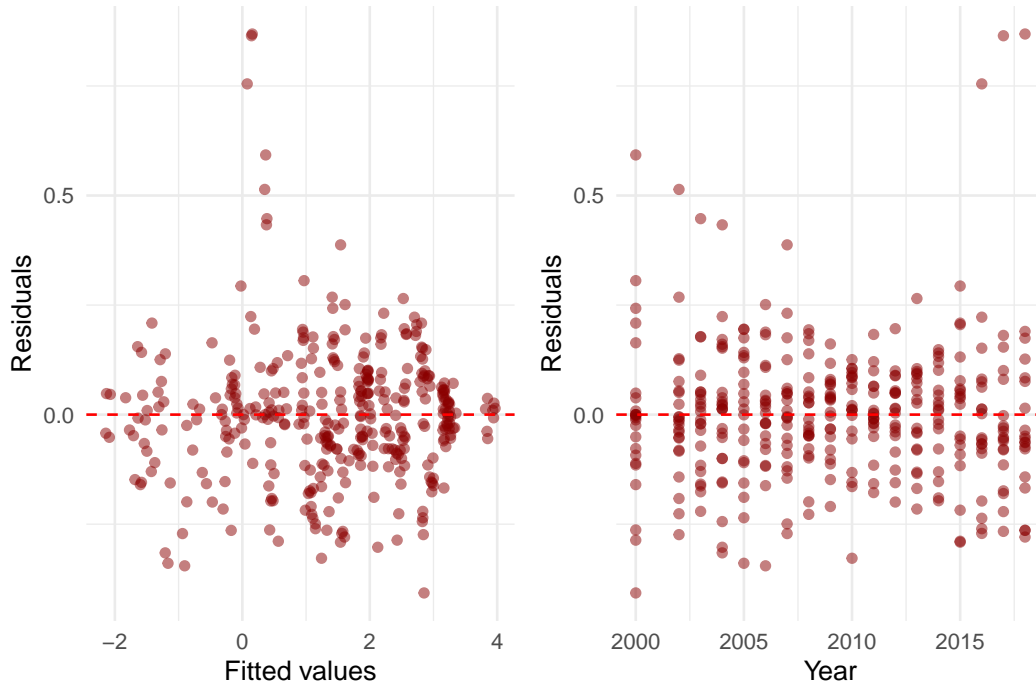


Figure 20: Residual diagnostics for autocracies

6 Conclusion

This analysis provides evidence on the relationship between economic development and CO2 emissions across different political regimes. Key findings include:

1. **EKC Evidence:** The results suggest varying degrees of support for the EKC hypothesis across regime types
2. **Regime Heterogeneity:** The Wald test confirms that coefficients differ significantly between democracies, anocracies, and autocracies
3. **Control Variables:** Renewable energy adoption and institutional quality (rule of law) play important roles in explaining emissions patterns

These findings have important policy implications: environmental outcomes are not solely determined by economic development but are mediated by political institutions. This suggests that promoting democratic governance and strong rule of law may be complementary strategies for achieving environmental sustainability alongside economic growth.

Limitations and Future Research:

- The analysis is limited to 2000-2018 due to data availability
- Country classification may change over time (regime transitions)
- Additional control variables (trade openness, technology, population density) could be explored
- Dynamic panel methods could address potential endogeneity concerns