Applied Microeconometrics

Computer Assignment - 1 on PANEL DATA

Analysis on the Impact of Forest Cover by Socio-Economic Indicators

Dilipkumaar (EE/2023-25/008)

Harish Chander P (EE/2023-25/011)

Priyadharshini R (EE/2023-25/020)

Vasanth D (EE/2023-25/026)

Supraja A (AE/2023-25/036)

References:
DOI: 10.20659/jfp.18.2_141

https://www.princeton.edu/~otorres/Panel101R.pdf

INTRODUCTION

Ever rising population, fierce competition between countries to become the largest economy, rapidly increasing developmental activities has altered the course of earth in the past century. Economic growth is always connected with an increase in pollution, deforestation and climate change. With the world starting to realize that there is only one earth, countries have started in taking oaths to restore earth, moving towards sustainable growth. To analyze the impact caused by various initiatives that were started by nations to move a step closer to a sustainable world, this study analyzes a common notion that the Forest cover has been declining rapidly over the past few decades, making it imperative to study the effects of socio-economic factors on this trend. This analysis examines 82 countries from Europe, Asia and Oceania over the period from 2000 to 2020, with data collected at five-year intervals, sourced from the World Development Indicators (WDI) database.

SECTION 1

1. Objective:

This paper seeks to explore the complex interplay between socio- economic factors that contribute to changes in forest cover. The ongoing reduction in forested areas has emerged as a critical issue on the global stage, with numerous countries implementing strategies to combat deforestation and mitigate its impact. Given the multidimensional challenges, understanding the drivers behind forest cover changes is essential for developing effective policies.

This study employs panel data, which is particularly well-suited for this type of analysis. Panel data allows for the examination of variations both across different countries and over time, providing a more comprehensive understanding of how environmental and economic factors influence forest cover in diverse contexts.

2. Analysis in Relation to the Reference Paper:

The present analysis differs from the reference paper in both the selection of explanatory variables and the dataset utilized. Total greenhouse gas emissions (measured in kilotons of CO2 equivalent) have been employed in place of adult literacy, with the focus on examining how greenhouse gas emissions impact forest area. This approach allows for a technical understanding of the relationship between economic productivity, as reflected in CO2 emissions, and changes in forest cover.

Additionally, population density has been used instead of total population, and land area has been incorporated as a time-invariant variable. Furthermore, while the reference paper treats regional clusters of countries as a single entity, this study considers individual countries as distinct entities for analysis.

3. Variables Used in the Model:

The dependent variable in this analysis is forest area (in square kilometres), which serves as an indicator of forest degradation. The study evaluates the effect of total greenhouse gas emissions (measured in kilotons of CO2 equivalent) as a key explanatory variable. Other explanatory variables include population density (people per square kilometre of land area), value added by the agricultural sector (in constant 2015 US$), gross domestic product (in constant 2015 US$), rural population, and life expectancy at birth. These independent variables exhibit variation across both time and entities. Additionally, land area (in square kilometers) is treated as a time-invariant explanatory variable within the model.

For the ease of empirical analysis, logarithmic transformations have been applied to the forest area, land area, gross domestic product, value added by the agricultural sector, and total greenhouse gas emissions.

4. Data Source and Variable Description:

The data utilized in this analysis has been obtained from the World Development Indicators (WDI) database. The table below provides a detailed description of the variables used, as defined in WDI.

Table-1: Variable description

Variable	Notation	Unit	Description
Forest area	*forest_ area*	Sq. km	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems & trees in urban parks and gardens.
Land area	*land_ area*	Sq. km	Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones.
Population density	*pop_ dens*	People per sq. km of land area	Population density is midyear population divided by land area in square kilometers. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship--except for refugees not permanently settled in the country of asylum.
Rural Population	*rural_ pop*	Percentage of total population	Rural population refers to people living in rural areas as defined by national statistical offices. It is calculated as the difference between total population and urban population.

Life expectancy at birth	*life_ exp*	Total (years)	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
Gross Domestic Product (GDP)	*gdp_ constant*	Constant 2015 US$	GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2015 prices, expressed in U.S. dollars.
Value added by Agricultural sector	*Agriculture_va*	Constant 2015 US$	Agriculture, forestry, and fishing corresponds to ISIC divisions 01-03 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. Data are in constant 2015 prices, expressed in U.S. dollars.
Total greenhouse gas emissions	*total_CO2*	kt of CO2 equivalent	Total greenhouse gas emissions in kt of CO2 equivalent are composed of CO2 totals excluding short-cycle biomass burning (such as agricultural waste burning and savanna burning) but including other biomass burning (such as forest fires, post-burn decay, peat fires and decay of drained peatlands), all anthropogenic CH4 sources, N2O sources and F-gases (HFCs, PFCs and SF6).
Forest cover %	*fl*	%	$\frac{forest_area}{land_area}$x100

The econometric equation used for the analysis:

$$\log(forest\_area)_{it} = \beta_0 + \beta_1 \log(total\_CO_2)_{1it} + \beta_2 \log(land\_area)_{2it} + \beta_3 \log(popn\_dens)_{3it} + \beta_4 (life\_exp)_{4it} + \beta_5 (rural\_popn)_{5it} + \beta_6 \log(GDP)_{6it} + \beta_7 \log(agri\_\ dollars)_{7it} + U_{it}$$

Where,

$\log(total\_CO_2)_{1it}$ = log of total greenhouse gas emission of country i for time t.

$\log(land\_area)_{2it}$ = log of land area of country i for time t.

$\log(popn\_dens)_{3it}$ = log of population density of country i for time t.

$(life\_exp)_{4it}$ = life expectancy at birth of country i for time t.

$(rural\_popn)_{5it}$ = rural population of country i for time t.

$\log(GDP)_{6it}$ = log of Gross domestic Product (GDP) of country i for time t.

$\log(agri\_dollars)_{7it}$ = log of Value added for agriculture sector of country i for time t.

5. Data Features:

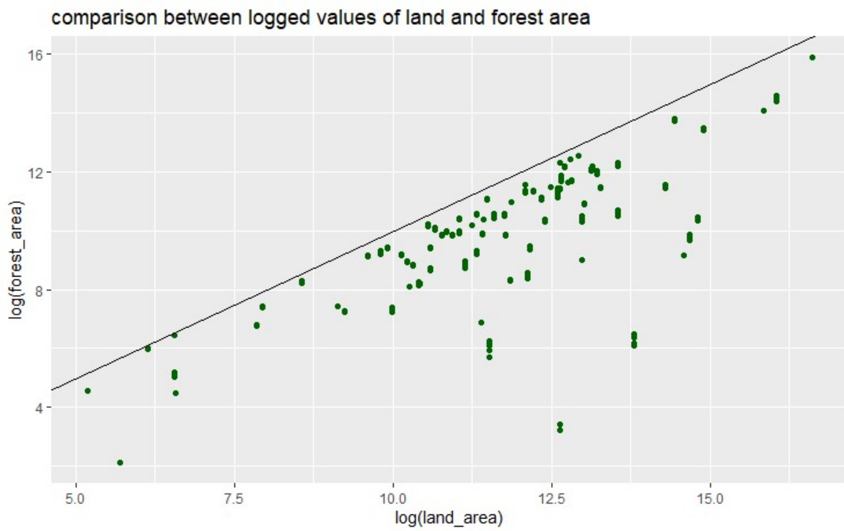
The analysis is conducted over five time periods (**T = 5**), spanning from 2000 to 2020, with observations taken at five-year intervals. The sample includes 82 countries (**N = 82**) from Asia, Europe, and Oceania, representing a mix of both developed and developing nations.

The dataset is structured as a strongly balanced panel, ensuring consistent observations across all countries and time periods.

Consequently, **the total number of observations = N*T= 82*5 = 410.**

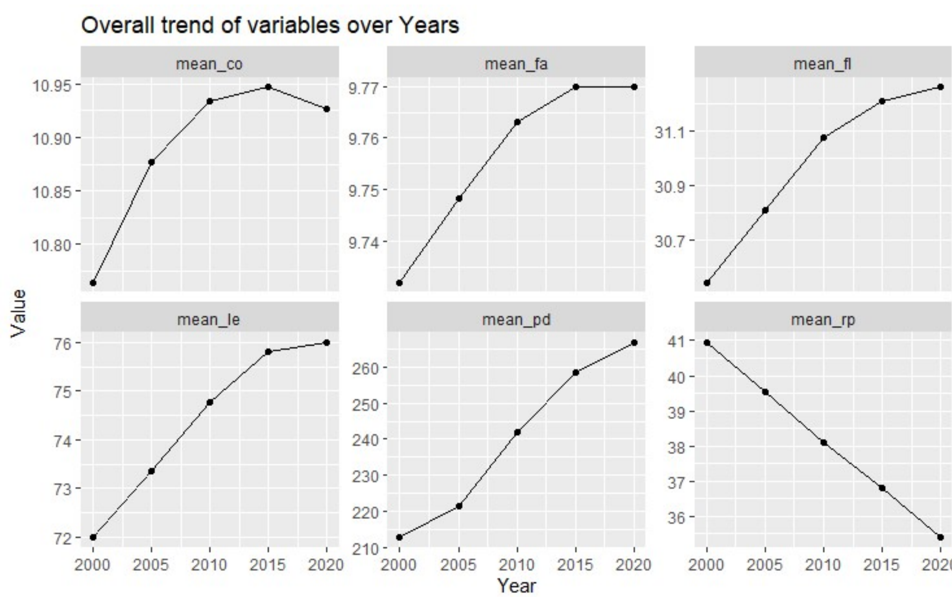
SECTION 2

(6)



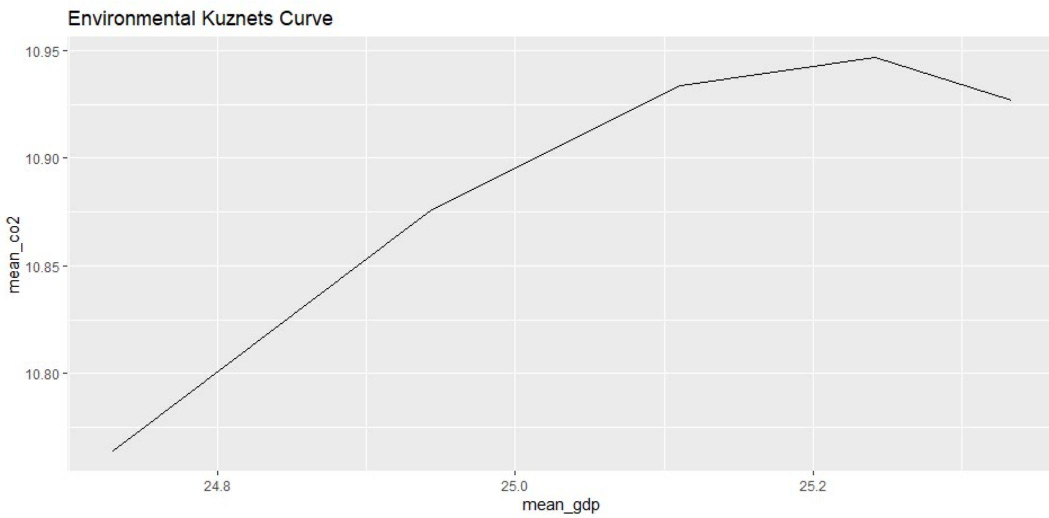
Graph1: Comparison between logged values of land and forest area

The line represents y=x, meaning forest_area = land_area. With countries getting near this line represents that they are having higher proportions of forest cover. Few outliers visible in the graph has occurred due to taking countries from the middle east as part of the Asian continent where they are fully or mostly deserted in nature.



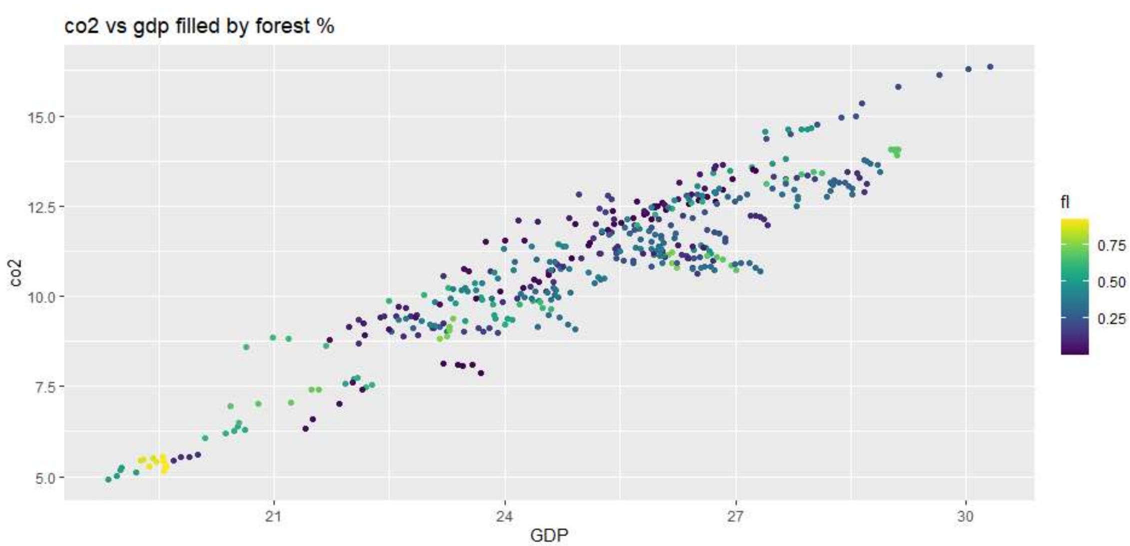
Graph 2: Mean of variables plotted over years

A constant growth in the average of different variables like forest area/percentage, co2, life expectancy at birth and population density are clearly visible. A constant decrease is found in rural population which may be a result of migrations, rapid urbanization, etc.



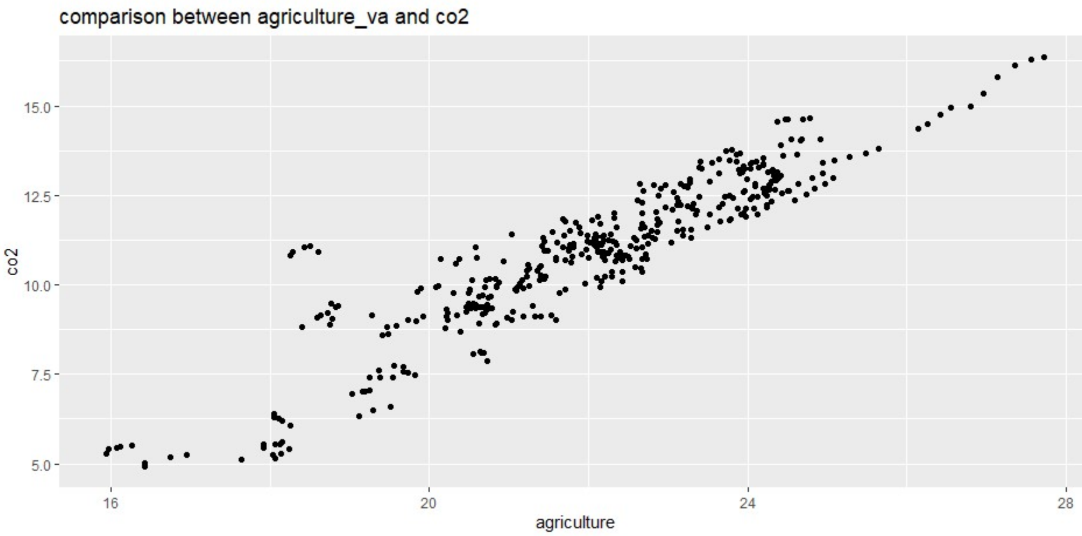
Graph 3: Environmental Kuznets Curve

Plotting mean GDP (income) with mean CO2 (pollution) variable gives us the Environmental Kuznets Curve and as per the EKC Hypothesis we find it true that after a period of time the pollution has started to reduce as GDP increases. We can see the curve has started to take a downward path.



Graph 4: Comparison between GDP, co2 and percentage of forest cover

GDP over the x axis brings a cluster in the centre representing a huge group of middle-income countries, low- and high-income countries on either corner. The forest percentage is filled by colours showing that middle income countries are having a large mix indicating some countries are focusing on sustainable practises and/or economic development. This also shows that the 2 variables correlate highly with each other.



Graph 5: Comparison between Agriculture and CO2

The relationship between the value added by the agricultural sector (x-axis) and total greenhouse gas emissions (measured in CO2 equivalents on the y-axis) suggests that an increase in the agricultural sector's value added is associated with a rise in total greenhouse gas emissions. This correlation may be attributed to factors such as methane emissions from paddy cultivation and other agricultural activities.

Additionally, the variance inflation factor (VIF) for the explanatory variables—total greenhouse gas emissions (measured in CO2 equivalents), value added by the agricultural sector, and GDP—exceeds 10, indicating the presence of multicollinearity among these variables. To address this issue, the value added by the agricultural sector(agriculture_va) and GDP(gdp_constant) will be excluded from further analysis.

(7) Summary Statistics:

Table-2: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
forest_area	205840	927564	8.2	8153116
land_area	737853.3	2236156	180	16378668
pop_dens	240.21	788.34	2.48	7918.95
rural_pop	38.15	21.45	0	86.6
life_exp	74.39	5.48	58.57	84.56
total_co2	332298.1	1139816	134.76	12942868

We have taken countries from Asia, Europe and Oceania consisting of different countries varying in geographical, climatic, political and economic conditions. Large range of values are found in almost all the variables taken in this study. The set of countries selected vary largely either side of extremes, from deserted to island nation, from dry to tropical weather, hot and cold climate throughout the year. Many factors like economic well-being, geographical positioning of the country, climatic conditions, vegetation affect the variables. Singapore is having 0% of rural population, island nations and deserted countries having lesser forest cover, economically well-developed countries experiencing higher life expectancy, developing countries focusing on economic development releasing higher concentrations of pollution are some of the key takeaways from this statistical summary

SECTION 3.1

(8)

(a) Pooled Model:

Without time dummy:

$$log(forest\_area)_{it} = \beta_0 + \beta_1\, log(total\_CO_2)_{1it} + \beta_2\, log(land\_area)_{2it} + \beta_3\, log(popn\_dens)_{3it}$$
$$+ \beta_4\, (life\_exp)_{4it} + \beta_5\, (rural\_popn)_{5it} + U_{it}$$

Estimated Equation:

$$\widehat{log(forest\_area)}_{it} = \underset{(1.5634)}{-3.3361} \underset{(0.0815)}{-0.32810\, log(total\_CO2)} \underset{(0.0802)}{+0.54033\, log(land\_area)} \underset{(0.0001)}{-0.0003\, (popn\_dens)}$$
$$\underset{(0.0186)}{+0.0375\, (life\_exp)} \underset{(0.0046)}{+0.0153\, (rural\_popn)}$$

N = 410, R^2 = 0.62539

With time dummy:

$$log(forest\_area)_{it} = \beta_0 + \delta_0\, d(2010)_t + \delta_1 d(2015)_t + \delta_2\, d(2020)_t + \beta_1\, log(total\_CO_2)_{1it}$$
$$+ \beta_2\, log(land\_area)_{2it} + \beta_3\, log(popn\_dens)_{3it} + \beta_4\, (life\_exp)_{4it}$$
$$+ \beta_5\, (rural\_popn)_{5it} + U_{it}$$

Estimated Equation:

$$\widehat{log(forest\_area)}_{it} = \underset{(1.6089)}{-\ 3.4543} \underset{(0.2409)}{-\ 0.0815\, d(2010)} \underset{(0.2447)}{-0.0944\, d(2015)} \underset{(0.2451)}{-0.0705\, d(2020)}$$
$$\underset{(0.0822)}{+\ 0.3261\, log(total\_CO2)} \underset{(0.0808)}{+\ 0.5422\, log(land\_area)} \underset{(0.0001)}{-\ 0.0003\, (popn\_dens)} \underset{(0.0196)}{+0.0398(life\_exp)} \underset{(0.0046)}{+\ 0.0156(rural\_popn)}$$

N= 410, R^2 = 0.62555

The pooled OLS model which doesn't account for the unobserved individual heterogeneity, α_i which is in our case the *land_area* that doesn't vary across periods. Thus, the U_{it} considered in the model only accounts for the time varying effects that is correlated the X_{jit} affecting the *forest_area,* say, the *total_CO₂* which not only explains for the variations in the *forest_area* but also the *Climate change* (U_{it}) which is caused from the *total_CO₂* that also explains for the variations in the *forest_area.*

Thus, when ignoring the α_i in the model often results in heterogeneity bias which is the bias caused by omitting a time constant variable which results in the autocorrelation of U_{it} and U_{is}; $t \neq s$ and the problem of heteroscedasticity.

Therefore, the Pooled model when estimated using OLS often gives a biased and inconsistent estimator of β_j. This issue can be solved by including a composite error term into the model: $V_{it} = \alpha_i + U_{it}$, that will be investigated in the FE model.

The results for the pooled OLS model's with and without time dummy being almost the same for the coefficient estimates and the intercept estimate is because both the model's estimates are not controlled for countries and the time periods, thus showing results which aren't country-specific or time-specific and also to the ignorance of heterogeneity bias which could be the main factor that made the time dummies insignificant in the Pooled model.

(b) <u>FE Model:</u>

<u>FE Model (within)Transformation:</u>

FE Model:

$$log(forest\_area)_{it} = \alpha_i + \beta_1 \, log(total\_CO_2)_{1it} + \beta_2 \, log(land\_area)_{2it} + \beta_3 \, log(popn\_dens)_{3it} + \beta_4 \, (life\_exp)_{4it} + \beta_5 \, (rural\_popn)_{5it} + U_{it}$$

averaging over time:

applying $\frac{1}{T} \sum_{t=1}^{T} \{all\ variables\}$, we get:

Time Demeaned Model:

$$\overline{log(forest\_area)}_{it} = \alpha_i + \beta_1 \, \overline{log(total\_CO_2)}_{1it} + \beta_2 \, \overline{log(land\_area)}_{2it} + \beta_3 \, \overline{log(popn\_dens)}_{3it} + \beta_4 \, (\overline{life\_exp})_{4it} + \beta_5 \, (\overline{rural\_popn})_{5it} + \overline{U_{it}}$$

subtracting the FE model from the time demeaned model, we get:

FE Estimator Model:

$$\log(\widehat{forest\_area})_{it}^{*} = \beta_1 \log(total\_CO_2)_{1it}^{*} + \beta_3 \log(popn\_dens)_{3it} + \beta_4 (life\_exp)_{4it}^{*} + \beta_5 (rural\_popn)_{5it}^{*} + U_{it}^{*}$$

Without Time Dummy:

$$\log(forest\_area)_{it} = \beta_1 \log(total\_CO_2)_{1it} + \beta_3 (popn\_dens)_{3it} + \beta_4 (life\_exp)_{4it} + \beta_5 (rural\_popn)_{5it} + V_{it}\ (\alpha_i + U_{it})$$

Estimated Equation:

$$\log(\widehat{forest\_area})_{it} = \underset{(0.0126)}{-0.0522 \log(total\_CO2)} \underset{(0.0016)}{+0.0051\ (life\_exp)} \underset{(0.00002)}{-0.000046\ (popn\_dens)} \underset{(0.0010)}{-0.0036\ (rural\_popn)}$$

N = 410, R^2 = 0.13135

With Time Dummy:

$$\log(forest\_area)_{it} = \delta_0\, d(2010)_t + \delta_1\, d(2015)_t + \delta_2\, d(2020)_t + \beta_1 \log(total\_CO_2)_{it} + \beta_3 \log(popn\_dens)_{it} + \beta_4 (life\_exp)_{it} + \beta_5 (rural\_popn)_{it} + V_{it}\ (\alpha_i + U_{it})$$

Estimated Equation:

$$\log(\widehat{forest\_area})_{it} = \underset{(0.0099)}{0.0370\, d(2010)} \underset{(0.0120)}{+\ 0.0430\, d(2015)} \underset{(0.129)}{+\ 0.0391\, d(2020)} \underset{(0.0129)}{-\ 0.04699 \log(total\_CO2)} \underset{(0.0024)}{-\ 0.0013\ (life\_exp)} \underset{(0.0002)}{-0.000047\ (popn\_dens)} \underset{(0.0010)}{-\ 0.0025\ (rural\_popn)} \underset{(0.0010)}{-\ 0.0025\ (rural\_popn)}$$

N =410, R^2 = 0.17109

Notice that, when the FE model allows for any arbitrary correlation between α_i and the X_{jit}, i.e. $cov(\alpha_i , X_{jit}) \neq 0$, the unobserved time constant effects term, α_i and *land_area* getting time demeaned in the within transformation FE model because of the inclusion of the composite error term: $V_{it} = \alpha_i + U_{it}$ into the model unlike the Pooled model.

The FE model now contains the time demeaned data and nevertheless the coefficient estimates and Std. errors are estimated from the time demeaned data but interpreted using the initial FE model.

One such improvement that can be made from the pooled model which gives a joint intercept for all the countries across time, whereas, now each country's intercept is estimated showing the base depletion rate of forest irrespective of the socio-economic indicators considered in the model by assigning a dummy variable to each explanatory variables of the FE model to get the Least Square Dummy Variable (LSDV) model which yields the same result as the FD model.

The strict exogeneity and homoscedasticity of independent variables along with serial uncorrelation of the time varying unobserved effects is to be satisfied for the FE model, since we can no longer allow for correlation of lagged dependent X_{jit} on $Y_{i,t-1}$ dependent variable for the above reason of $cov(\alpha_i , X_{jit}) \neq 0$, where suppose, the previous *law enforcements* (U_{is}) in certain countries for the reduction in $total\_CO_2$, can contribute to the control over depletion of *forest_area.*

RE Model:

Without Time Dummy:

$$log(forest\_area)_{it} = \beta_0 + \beta_1 \, log(total\_CO_2)_{1it} + \beta_2 \, log(land\_area)_{2it} + \beta_3 \, log(popn\_dens)_{3it}$$
$$+ \beta_4 \, (life\_exp)_{4it} + \beta_5 \, (rural\_popn)_{5it} + V_{it} \, (\alpha_i + U_{it})$$

Estimated Equation:

$$\widehat{log(forest\_area)}_{it} = \underset{(0.9008)}{0.0793} - \underset{(0.0126)}{0.0502} \, log(total\_CO2) + \underset{(0.0764)}{0.8667} \, log(land\_area) - \underset{(0.00002)}{0.000047} \, (popn\_dens)$$
$$+ \underset{(0.0016)}{0.0052} \, (life\_exp) - \underset{(0.0010)}{0.0034} \, (rural\_popn)$$

$N= 410, R^2 = 0.29415$

With Time Dummy:

$$log(forest\_area)_{it} = \beta_0 + \delta_0 \, d(2010)_t + \delta_1 \, d(2015)_t + \delta_2 \, d(2020)_t + \beta_1 \, log(total\_CO_2)_{it}$$
$$+ \beta_2 \, log(land\_area)_{it} + \beta_3 \, log(popn\_dens)_{it} + \beta_4 \, (life\_exp)_{it}$$
$$+ \beta_5 \, (rural\_popn)_{it} + V_{it} \, (\alpha_i + U_{it}$$

Estimated Equation:

$$\widehat{log(forest\_area)}_{it} = \underset{(0.9073)}{0.4853} + \underset{(0.0099)}{0.0431\, d(2010)} + \underset{(0.0120)}{0.0335\, d(2015)} + \underset{(0.0129)}{0.0394\, d(2020)}$$
$$- \underset{(0.0129)}{0.0449\, log(total\_CO2)} + \underset{(0.0764)}{0.8627\, log(land\_area)} - \underset{(0.00002)}{0.000048\, (popn\_dens)} - \underset{(0.0024)}{0.0012\, (life\_exp)} - \underset{(0.0010)}{0.0024\, (rural\_popn)}$$

N = 410, R^2 = 0.31662

The RE model explicitly includes intercept into the model by the assumption of unobserved effect, α_i has a zero mean such that there will be $cov(\alpha_i , X_{jit}) = 0$, existing in the model making the prior model's estimators of FE and Pooled OLS to be inefficient if the above assumption is met.

The method of having the time-invariant explanatory variable into the model by the assumption of no heterogeneity bias for the unobserved time invariant variable α_i to be correlated with $total\_CO_2$. In this case, the RE model's α_i is now concerned with the unmeasured fraction (or) intensity of the $total\_CO_2$ that contributes directly to the forest depletion.

Note: The $total\_CO_2$ considered here is the CO_2 emissions equivalent of all ghg emissions, not only the CO_2 emissions.

Table- 3: Estimates of pooled, FE and RE model

Variables	Model 1		Model 2		Model 3	
	Pooled Model		Fixed Effects Model		Random Effects Model	
	Without Time Dummy	With time dummy	Without Time Dummy	With time dummy	Without Time Dummy	With time dummy
Constant	-3.3361 (1.5634) **	-3.4543 (1.6089) **			0.0793 (0.9008)	0.4853 (0.9073)
log (land_area)	0.5403 (0.0802) ***	0.5422 (0.0808) ***			0.8667 (0.0764) ***	0.8627 (0.0764) ***
pop_dens	-0.0003 (0.0001) ***	-0.0003 (0.0001) ***	-0.000046 (0.00002) *	-0.000047 (0.00002) *	-0.000047 (0.00002) *	-0.000048 (0.00002) *
rural_pop	0.0153 (0.0046) ***	0.0156 (0.0046) ***	-0.0036 (0.0010) ***	-0.0026 (0.0010) **	-0.0034 (0.0010) ***	-0.0024 (0.0010) *
life_exp	0.0375 (0.0186) **	0.0398 (0.0196) **	0.0051 (0.0016) ***	-0.0013 (0.0024)	0.0052 (0.0016) ***	-0.0013 (0.0024)
log (total_co2)	0.3281 (0.0815) ***	0.3262 (0.0822) ***	-0.0522 (0.0126) ***	-0.0470 (0.0129) ***	-0.0503 (0.0126) ***	-0.0450 (0.0129) ***
year2005		-0.0509 (0.2376)		0.0202 (0.0079) **		0.0201 (0.0079) **
year2010		-0.0816 (0.2409)		0.0370 (0.0099) ***		0.0432 (0.0099) ***
year2015		-0.0945 (0.2447)		0.0431 (0.0120) ***		0.0335 (0.0120) ***
year2020		-0.0706 (0.2451)		0.0391 (0.0129) ***		0.0395 (0.0129) ***

Values in Parenthesis are Std. Error and the base year is 2000
Level of significance (p- value): <0.01(***), <0.05 (**), <0.1 (*)

(d)

Table-4: F-Statistic Summary

F- STATISTIC	Model 1		Model 2		Model 3	
	Pooled Model		Fixed Effects Model		Random Effects Model	
	Without time dummy	With time dummy	Without time dummy	With time dummy	Without time dummy	With time dummy
VALUES	134.889	74.2494	12.2485	8.25612	168.357	185.324
P-VALUES	0.000	0.000	0.000	0.000	0.000	0.000

By looking at the table, it can be analyzed that each model irrespective of the time dummy or without time dummy, provides large F-stat value and small p-values showing joint significance of the variables considered in the model.

(e)

The total greenhouse gas (GHG) emissions, measured in CO2 equivalent, serve as the key variable in this analysis, demonstrating significant relevance at 5% across all three econometric models—pooled, fixed effects, and random effects—both with and without the inclusion of a time dummy. In the pooled model, CO2 emissions and rural population exhibit a positive effect, which shifts to a negative effect in the fixed and random effects models. This change underscores the importance of accounting for unobserved heterogeneity and omitted variable bias in pooled model.

Land area emerges as a highly significant variable, displaying a positive relationship in both the pooled and random effects models, regardless of the inclusion of the time dummy. However, in the fixed effects model, land area is demeaned due to its time-invariant nature. Population density consistently shows an overall negative effect across all models, with or without the time dummy.

The introduction of time dummies results in slight adjustments to the coefficients, yet the general trends remain stable across the models. While the time dummy is insignificant in the pooled model, it is significant in both the fixed and random effects models, contributing to an increase in forest cover. The explanatory variables, life expectancy and rural population, are significant in models without the time dummy, but become insignificant when the time dummy is included.

(f)

Since total greenhouse gas (GHG) emissions, measured in CO2 equivalent is the key variable.

Without time dummy

A one percent increase in kilotons of CO2 equivalent is associated with a 0.328% increase in forest cover (measured in square kilometers) with respect to pooled model.

A one percent increase in kilotons of CO2 equivalent is associated with a 0.052% decrease in forest cover (measured in square kilometers) with respect to FE model

A one percent increase in kilotons of CO2 equivalent is associated with a 0.051% decrease in forest cover (measured in square kilometers) with respect to RE model.

With time dummy

A one percent increase in kilotons of CO2 equivalent is associated with a 0.326% increase in forest cover (measured in square kilometers) with respect to pooled model.

A one percent increase in kilotons of CO2 equivalent is associated with a 0.047% decrease in forest cover (measured in square kilometers) with respect to FE model.

A one percent increase in kilotons of CO2 equivalent is associated with a 0.045% decrease in forest cover (measured in square kilometers) with respect to RE model.

SECTION 3.2

(9)

Table-5: Test of hypothesis for the best fitting model		
Models tested	**Null hypothesis**	**Alternate hypothesis**
Pooled with FE	Pooled Model is preferred ($\alpha i = 0$)	FE model is preferred (At least one of the $\alpha i \neq 0$)
Pooled with RE	Pooled Model is preferred ($Var\{\alpha i\} = 0$)	RE model is preferred ($Var\{\alpha i\} \neq 0$)
FE with RE	RE model is preferred	FE model is preferred

Table-6: Test Statistic for Models without Time dummy					
Models tested	**Test**	**Test Statistic**	**Degrees of Freedom**	**Value**	**p- value**
Pooled with FE	F- test	F Test	324	5239.2	0.000
Pooled with RE	Breusch Pagan LM Test	Chi square	1	810.45	0.000
FE with RE	Hausman Test	Chi square	4	1.7691	0.7781

Table-7: Test Statistic for Models with Time dummy					
Models tested	**Test**	**Test Statistic**	**Degrees of Freedom**	**Value**	**p- value**
Pooled with FE	F- test	F Test	320	5420.3	0.000
Pooled with RE	Breusch Pagan LM Test	Chi square	1	811.11	0.000
FE with RE	Hausman Test	Chi square	8	5.4523	0.7083

(a) Pooled with FE:

Null hypothesis (H_0): $\alpha_i = 0$

Alternate hypothesis (H_1): At least one of the time invariant individual heterogeneities, $\alpha_i \neq 0$

<u>Without time dummy</u>

F- Statistic value is 5239.2 which is very large with degree of freedom 324. P- Value is 0.000.

<u>Decision</u>: We reject null hypothesis at 5% level of significance.

<u>Without time dummy</u>

F- Statistic value is 5420.3 which is very large with degree of freedom 320. P- Value is 0.000.

<u>Decision</u>: We reject null hypothesis at 5% level of significance.

Therefore, FE model is better fit than Pooled model with and without time dummy.

(b) Pooled with RE:

Null hypothesis (H_0) : $\sigma^2_\alpha = 0$

Alternate hypothesis (H_1) : $\sigma^2_\alpha \neq 0$

$$\boxed{Corr(V_{it}, V_{is}) = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_u}}$$

<u>Without time dummy</u>

BP LM test Statistic value is 810.45 which is very large with degree of freedom 1. P- Value is 0.000.

<u>Decision</u>: We reject null hypothesis at 5% level of significance.

<u>Without time dummy</u>

BP LM test Statistic value is 811.11 which is very large with degree of freedom 1. P- Value is 0.000.

<u>Decision</u>: We reject null hypothesis at 5% level of significance.

Therefore, RE model is better fit than Pooled model with and without time dummy.

(c) FE with RE:

Null hypothesis (H_0) : RE model is preferred

Alternate hypothesis (H_1) : FE model is preferred

H test:

$$H_w = (\hat{\beta}_{RE} - \hat{\beta}_{FE})' [V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})]^{-1} (\hat{\beta}_{RE} - \hat{\beta}_{FE})$$

Without time dummy

Hausman test Statistic value is 1.7691 which is very large with degree of freedom 4. P-Value is 0.7781.

Decision: We fail reject null hypothesis at 5% level of significance.

Therefore, RE model is better fit than FE model.

Without time dummy

Hausman test Statistic value is 5.4523 which is very large with degree of freedom 8. P-Value is 0.7083.

Decision: We fail reject null hypothesis at 5% level of significance.

Therefore, RE model is better fit than Pooled model with and without time dummy.

(d) The analysis concludes that the Random Effects (RE) model provides a better fit for the data. Based on the RE model, it is found that land area, rural population, and total greenhouse gas emissions significantly impact overall forest cover.

Conclusion

The study's findings indicate that while economic growth and industrial activities contribute to environmental degradation, particularly through CO2 emissions, the impact on forest cover is multifaceted, involving a trade-off between development and environmental conservation.

The analysis confirms that higher economic activities, particularly in middle-income countries, are correlated with both positive and negative environmental outcomes. The Environmental Kuznets Curve observed in the study suggests that while pollution levels initially rise with GDP growth, they tend to decrease as countries become wealthier and adopt more sustainable practices.

Unobserved variables, often referred to as omitted variables, are factors that are not included in the model but could influence the dependent variable. These variables can introduce bias into the estimated relationships between the included independent variables and the dependent variable. In the context of our study on the relationship between socio-economic factors (like greenhouse gas emissions, population density, etc.) and forest area, possible unobserved variables might include:

Unobserved variables that might influence the relationship between socio-economic factors and forest area include governance quality, corruption, environmental regulation, global commodity prices, foreign direct investment, and climate change. These factors could affect forest cover but are not captured by the model.

End of Assignment