Econometrics Final Project

Testing the Environmental Kuznets Curve (EKC) Hypothesis for River Water Pollution in India.

- **Aditya Rastogi (2018273)**
- **Krit Verma (2017348)**
- **Prarthana Kansal (2017357)**
- **Siddharth Sadhwani (2018313)**

Introduction

India is on its way to become a developed country yet social issues like Environmental Degradation and pollution still continue to be prevalent.

One of the factors that often complement development is economic growth and it is possible that various aspects of Environmental Degradation may get affected differently with changes in economy. As India grows and urbanizes, it’s water bodies degrade.

Apart from economic aspects, it is important to also include other dependencies that may affect Environment Degradation.

Introduction (contd.)

India is country of various religions. Hinduism is the religion followed by the majority of the Population (79%) [1] . Rivers hold a very important place in Hinduism. As a result, many rivers in India are subjected to huge level of pollution owing to these religious activities [2].

Thus, Hindu Population can be an important metric to judge water pollution of rivers in India.

Open defecation in India is still a big issue. Many people who still live in rural India do not have access to proper toilet facilities. Waste from these open *toilets* generally ends up in the rivers which again can constitute as a Major Pollution factor[3].

[1] Census of India Data .(2011)
[2] Hinduism Case Study-Climate Change 2018, Harvard Divinity Project, https://rlp.hds.harvard.edu/files/hds-rlp/files/climate_change_hinduism.pdf
[3] Open Defecation: A Prominent Source Of Pollution In Drinking Water In Villages - IJLBPR, A V Rajgire http://new.ijlbpr.com/jlbpradmin/upload/ijlbpr_50e55d24590f6.pdf

Kuznets Hypothesis

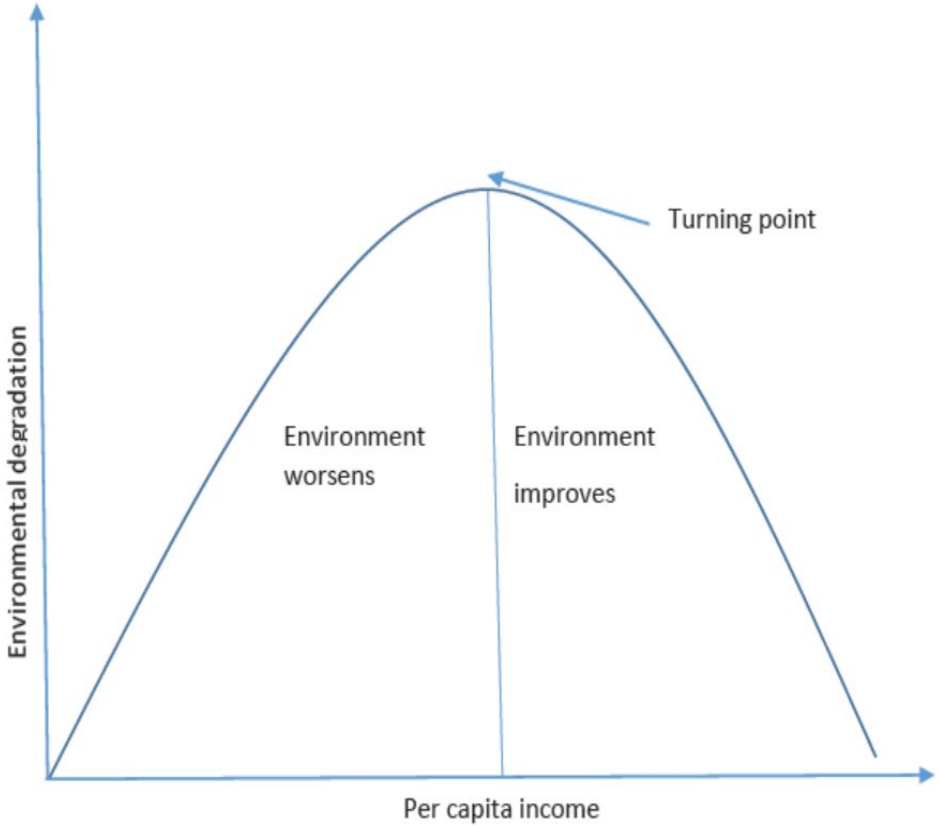
- Inverted *U*-curve hypothesis that posits the relationship between income and income inequality.

- EKC (Environmental Kuznets Curve) postulates the relationship between income and environmental deterioration.

Environmental Kuznets Curve

- At early stages of economic growth, environmental degradation rises

- After some threshold is reached, the co-movement tends to reverse at higher levels of economic growth



Power Inequality

- Kuznets suggested power inequality is a function of both income inequality and per capita income. [4]

- Technological change, education, the political process, and socioeconomic conditions may eventually lead to a decline in pollution. [5]

- Environmental degradation is affected due to mediation by social cohesion and cooperation to protect common resources. [6]

[4] M.Torras,J.K.Boyce:Ecological Economics25 (1998)
[5] Gene M. Grossman, Alan B. Krueger, Economic Growth and the Environment (1995)
[6] Cusing et al. The Haves, the Have-Nots, and the Health of Everyone (2015)

Power Inequality

- Ethnicity and different races contribute to increasing pollution.[7]
- Education and public transit that can help lessen a society's impact on the environment.[8]
- Unequal societies invest less in pro-environmental policies, monitoring, and research[9]

[7] Alcsina et al.(1999)
[8] Wisman JD. 2011. Inequality, social respectability, political power, and environmental devastation
[9] Nolen J et al 2012. Desigualdad y pol´ıtica ambiental enM´exico

Dataset Information

- The whole data has been taken from a span of 10 consecutive years i.e. 2004 to 2013.

- The main focus is 16 states of India which are home to major rivers of the country.

- Total Data points = 16*10

  = 160

Variable Description Dependent Variables

- **Fish Production:** An increase in fish production indicates an increase in overall river health (decrease in Pollution).
- **Conductivity:** Sudden change in its amount indicates the degradation of water quality (Affected by Chemical Affluents, Industrial Discharge etc.)
- **Coliform:** Increase in its amount indicates the degradation of water quality. (Coliform is a bacteria generally soil and faecal wastes of warm blooded animals)

All data taken for the Time Period: 2004-2013

Variable Description Dependent Variables

Variable	Description	Acronym
Fish Production	State-wise fish production (in '000 tonnes)	y_1
Conductivity	The Change in conductivity of a river from mean Conductivity (in µmhos/cm)	changeConductivity (y_2) Here y_2 takes the value '1' if the absolute value of change i.e. ($\lvert y_2 - y_{2\text{-avg}} \rvert$) > mean ($\lvert y_2 - y_{2\text{-avg}} \rvert$) otherwise y_2 takes '0'.
log of (1+Total Coliform)	log of Observed TC values in a river(in MPN/100ML)	log(1+totCol (y_3)) [5]

Variable Description Independent Variables

- To study the effect of economic variables on River Water Pollution we have taken SDP, State Domestic Product (per capita) .

Variable	Description	Acronym
SDP (Per Capita)	Per Capita value of goods & services produced in a state within a year (in Rupees)	sdp
SDP^2 (Per Capita)	Square of Per Capita value of goods & services produced in a state within a year (in Rupees)	sdp2

Variable Description Independent Variables

- For our **control** variable we have taken Urbanization.

Urbanization^2	Square of the distribution of rural and urban population state-wise.	urban2

- For our **income inequality** variable we have taken values of GINI index.

GINI Index	The difference in income between the extremes of the economic society. A higher value signifies a greater inequality. (in the range (0,1))	giniIndex

Variable Description Independent Variables

- For our **power inequality** variables we have taken values of

Variable	Description	Acronym
Hindu Population	Percentage of Hindus in the total population. (statewise)	hindus
Political Participation	Index to map out voter turnout to political awareness of each state.	politicpart
Infant Mortality Rate	Number of deaths of infants less than the age of 1, out of 1000 in each state.	imf
Literacy Rate	Literacy Rate Percentage state-wise	literacy

Data Summary

- We have taken the data for Literacy Rates (Person) annually for the years 2004-14 to model change in y_i (s) for the percentage of latrines inside household premises.
- We found the p-value of Pearson's Correlation for the two datasets for which we had the data, i.e. the year 2011(Literacy Rates-2011 & Standard of Living Data-2011)
- **Correlation:** 0.82
- A positive strong correlation value states that a change in 'x' results in a positive change in 'y'. Hence we can assume the percentage of closed latrines as an apt representation of trend in the literacy rates.

Variable Acronym	N	Mean	Median	SD	Min.	Max.
sdp	160	30,957	27,171	14,126.81	7,588	69,097
sdp2	160	1.16e+09	7.38e+08	1.03e+09	5.76e+07	4.77e+09
urban2	160	13.97	7.93	17.43	1.14	73.28
giniIndex	160	0.35	0.33	0.07	0.26	0.58
hindus	160	81.60	83.93	10.88	54.73	94.70
politicpart	160	0.58	0.60	0.06	0.40	0.68
imf	160	45.08	45.50	15.74	12.00	79.00
literacy	160	5,197	5,041	1,398.18	2,500	9,025

Environmental Kuznets Curve (EKC)

We estimate the empirical relation between Pollution variables (Y_i), Economic growth variables (x_i), Additional covariates (Z) and ε is the stochastic error term assumed to be normally distributed.:

$$\mathbf{Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + Z'\beta + \varepsilon}$$

In our model, $Z'\beta = \beta_3 \text{giniIndex} + \beta_4 \text{hindus} + \beta_5 \text{politicpart} + \beta_6 \text{imf} + \beta_7 \text{literacy} + \beta_8 \text{urban2}$

If $\boldsymbol{\beta_1 > 0}$ and $\boldsymbol{\beta_2 < 0}$ and both are statistically significant, then the pollution variables are said to display the inverted U-shaped relationship. [10][11]

[10] Rupasingha et al (2002), The environmental Kuznets Curve for US counties.
[11] Justin Tevie et al (2011), Testing the Environmental Kuznets Curve Hypothesis for Biodiversity Risk in the US: A Spatial Econometric Approach

Model, Hypothesis and Analysis

We initially frame our hypothesis for each regression model and the power inequalities.

We then proceed our study in two steps :

- We first run a regression for the basic EKC model i.e. Y_i=f (sdp,sdp2) and analysis of results for the same is done.
- Secondly, we introduce the power and control variables to the model and run a regression to capture the effect they might have on our initial model.

Lastly, we proceed to the final results and conclusion.

Fish Production (y_1)

EKC Hypothesis : y_1 displays an inverted-U relationship with the economic growth variables (sdp,sdp2).

i.e. for the model,

$$\mathbf{y_1\ (Fish\ Production) = \beta_0 + \beta_1 sdp + \beta_2 sdp2 + \varepsilon}$$

- $H_0 = \beta_1 > 0$ and $\beta_2 < 0$
- $H_a = \beta_1 < 0$ and $\beta_2 > 0$

Fish Production (y_1)

Fish Production (y_1)	Estimate (SE)
sdp	+3.99e-02 (1.06e-02)***
sdp2	-4.14e-07 (1.44e-07)**
N = 157 | **Multiple R-squared** = 0.16 | **Adjusted R-squared** = 0.15	

Signif. Codes:

1. 0 '***'
2. 0.001 '**'
3. 0.01 '*'
4. 0.05 '.'
5. 0.1 ' '
6. 1

As the values for both sdp and sdp2 are fairly significant, and according to the signs of β_1 & β_2 we **fail to reject the null Hypothesis (H_0)** i.e. our y_1 for the chosen fish production model follows the EKC Hypothesis and displays an inverted-U relationship with the economic growth variables.

We now test the effect of Power Inequality, Income Inequality and Control covariates on the model.

Fish Production (y_1)

EKC Hypothesis : y_1 displays an inverted-U relationship with the economic growth variable when additional covariates are also taken.

i.e. for the model,

$$\mathbf{y_1\ (Fish\ Production) = \beta_0 + \beta_1 sdp + \beta_2 sdp2 + \beta_3 giniIndex + \beta_4 hindus + \beta_5 politicpart + \beta_6\ imf + \beta_7\ literacy + \beta_8 urban2 + \varepsilon}$$

We now present the following Hypothesis for our power inequalities.

hindus	**politicalpart**	**imf**	**literacy**
$H_0 = \beta_4 < 0$	$H_0 = \beta_5 > 0$	$H_0 = \beta_6 < 0$	$H_0 = \beta_7 > 0$
$H_a = \beta_4 > 0$	$H_a = \beta_5 < 0$	$H_a = \beta_6 > 0$	$H_a = \beta_7 < 0$

Fish Production (y_1)

Fish Production (y_1)	Estimate (SE)
sdp	+7.02e-02 (1.44e-02)***
sdp2	-7.33e-07 (1.70e-07)***
urban2	+8.13e+00 (2.77e+00)**
giniIndex	+6.27e+02 (4.65e+02)
hindus	-8.84e+00 (4.15e+00)*
politicpart	+1.87e+03 (5.63e+02)**
imf	-7.68e+00 (2.97e+00)*
literacy	-2.08e+01 (5.61e+00)***
N = 151 | **Multiple R-squared** =0.33 | **Adjusted R-squared**=0.30	

Signif. Codes:

1. 0 '***'
2. 0.001 '**'
3. 0.01 '*'
4. 0.05 '.'
5. 0.1 ' '
6. 1

Conclusion

The signs of β_1 & β_2 are still the same and the additional covariates have only increased the significance of β_2 ,hence **we fail to reject the null Hypothesis(H_0)** i.e. it follows the EKC Hypothesis after considering the effects of additional covariates as well. For power inequality:

- **hindus :** We **fail to reject the null hypothesis** as $\beta_4 < 0$ and significant. Thus, Hindu Population Density is positively associated with water pollution.
- **politicalpart :** We **fail to reject the null hypothesis** as $\beta_5 > 0$ and significant.Thus, Political Participation Rate is inversely associated with water pollution.
- **imf :** We **fail to reject the null hypothesis** as $\beta_6 < 0$ and significant. Thus, Infant Mortality Rate is positively associated with water pollution.
- **literacy :** We **reject the null hypothesis** as $\beta_7 < 0$ and significant.This maybe because literacy rate is a very complex variable and might affect other aspects of the society too. For example it is possible that in a state where the literacy is high , the state might not necessarily be close to some major water body which hence would affect its fish production.

changeConductivity

EKC Hypothesis : changeConductivity displays an inverted-U relationship with the economic growth variables (sdp,sdp2).

i.e. for the model,

$$\textbf{changeconductivity} = \beta_0 + \beta_1 \text{sdp} + \beta_2 \text{sdp2} + \varepsilon$$

- $H_0 = \beta_1 > 0$ and $\beta_2 < 0$
- $H_a = \beta_1 < 0$ and $\beta_2 > 0$

changeConductivity

changeConductivity	Estimate (SE)
sdp	+2.27e-05 (1.25e-05).
sdp2	-3.11e-10 (1.71e-10).
N =157 | **Multiple R-squared** =0.02 | **Adjusted R-squared** =0.01	

Signif. Codes:

1. 0 '***'
2. 0.001 '**'
3. 0.01 '*'
4. 0.05 '.'
5. 0.1 ' '
6. 1

As the values for both sdp and sdp2 are significant, and according to the signs of β_1 & β_2 we **fail to reject the null Hypothesis (H_0)** i.e. our changeConductivity for the model follows the EKC Hypothesis and displays an inverted-U relationship with the economic growth variables.

We now test the effect of Power Inequality, Income Inequality and Control covariates on the model.

changeConductivity

EKC Hypothesis : changeConductivity displays an inverted-U relationship with the economic growth variable when additional covariates are also taken.

i.e. for the model,

$$\textbf{changeConductivity} = \beta_0 + \beta_1 \textbf{sdp} + \beta_2 \textbf{sdp2} + \beta_3 \textbf{giniIndex} + \beta_4 \textbf{hindus} + \beta_5 \textbf{politicpart} + \beta_6 \textbf{imf} + \beta_7 \textbf{literacy} + \beta_8 \textbf{urban2} + \varepsilon$$

We now present the following Hypothesis for our power inequalities.

hindus	**politicalpart**	**imf**	**literacy**
$H_0 = \beta_4 > 0$	$H_0 = \beta_5 < 0$	$H_0 = \beta_6 > 0$	$H_0 = \beta_7 < 0$
$H_a = \beta_4 < 0$	$H_a = \beta_5 > 0$	$H_a = \beta_6 < 0$	$H_a = \beta_7 > 0$

changeConductivity

changeConductivity	Estimate (SE)
sdp	+3.32e-05 (1.82e-05).
sdp2	-3.90e-10 (2.14e-10).
urban2	+7.29e-05 (3.48e-03)
giniIndex	+1.03e+00 (5.85e-01).
hindus	+3.95e-03 (5.22e-03)
politicpart	-1.21e+00 (7.08e-01).
imf	+1.01e-02 (3.73e-03)**
literacy	+2.37e-03 (7.05e-03)
N = 151 | **Multiple R-squared** =0.13 | **Adjusted R-squared**=0.08	

Signif. Codes:

1. 0 '***'
2. 0.001 '**'
3. 0.01 '*'
4. 0.05 '.'
5. 0.1 ' '
6. 1

Conclusion

The signs of β_1 & β_2 are still the same and the additional covariates have almost no effect on the significance of β_1 and β_2 ,hence **we fail to reject the null Hypothesis(H_0)** i.e. it follows the EKC Hypothesis after considering the effects of additional covariates as well.

For power inequality:

- **politicalpart** : We **fail to reject the null hypothesis** as $\beta_5 < 0$ and significant.Thus, Political Participation Rate is inversely associated with water pollution.
- **imf** : We **fail to reject the null hypothesis** as $\beta_6 > 0$ and significant. Thus, Infant Mortality Rate is positively associated with water pollution.
- The values for the remaining power inequalities were not significant enough to infer meaningful results.

Total Coliform (log (1+y_3))

EKC Hypothesis : log (1 + y_3) displays an inverted-U relationship with the economic growth variables (sdp,sdp2).

i.e. for the model,

$$\log (1 + y_3) = \beta_0 + \beta_1 sdp + \beta_2 sdp2 + \varepsilon$$

- $H_0 = \beta_1 > 0$ and $\beta_2 < 0$
- $H_a = \beta_1 < 0$ and $\beta_2 > 0$

Total Coliform (log (1+y_3))

Total Coliform (log (1 + y_3))	**Estimate (SE)**
sdp	-4.77e-06 (3.63e-05)
sdp2	+1.68e-10 (4.97e-10)
N = 157 | **Multiple R-squared** =0.01 | **Adjusted R-squared** = -0.01	

Signif. Codes:

1. 0 '***'
2. 0.001 '**'
3. 0.01 '*'
4. 0.05 '.'
5. 0.1 ' '
6. 1

As the values for both sdp and sdp2 are not significant, we **reject the Null Hypothesis (H_0)**, i.e. the EKC Hypothesis does not hold for our total Coliform Variable.

We now test the effect of Power Inequality, Income Inequality and Control covariates on the model.

Total Coliform (log (1+y_3))

EKC Hypothesis : log (1 + y_3) displays an inverted-U relationship with the economic growth variable when additional covariates are also taken.

i.e. for the model,

$$\log (1 + y_3) = \beta_0 + \beta_1 sdp + \beta_2 sdp2 + \beta_3 giniIndex + \beta_4 hindus + \beta_5 politicpart + \beta_6 imf + \beta_7 literacy + \beta_8 urban2 + \varepsilon$$

We now present the following Hypothesis for our power inequalities.

hindus	politicalpart	imf	literacy
$H_0 = \beta_4 > 0$	$H_0 = \beta_5 < 0$	$H_0 = \beta_6 > 0$	$H_0 = \beta_7 < 0$
$H_a = \beta_4 < 0$	$H_a = \beta_5 > 0$	$H_a = \beta_6 < 0$	$H_a = \beta_7 > 0$

Total Coliform (log (1+y_3))

Total Coliform (log (1 + y_3))	**Estimate (SE)**
sdp	-2.45e-05 (5.05e-05)
sdp2	+1.94e-10 (5.95e-10)
urban2	-9.76e-03 (9.67e-03)
giniIndex	-1.78e+00 (1.63e+00)
hindus	+2.58e-03 (1.45e-02)
politicpart	-8.45e+00 (1.97e+00)***
imf	-1.10e-02 (1.04e-02)
literacy	+2.45e-02 (1.96e-02)
N = 151 | **Multiple R-squared** =0.19 | **Adjusted R-squared**=0.14	

Signif. Codes:

1. 0 '***'
2. 0.001 '**'
3. 0.01 '*'
4. 0.05 '.'
5. 0.1 ' '
6. 1

Conclusion

As the values for both sdp and sdp2 are still not significant, we **reject the Null Hypothesis (H_0)**, i.e. the EKC Hypothesis does not hold for our total Coliform Variable even after the additional covariates have been added.

- **politicalpart** : We **fail to reject the null hypothesis** as $\beta_5 < 0$ and significant.Thus, Political Participation Rate is inversely associated with water pollution.
- The values for the remaining power inequalities were not significant enough to infer meaningful results.

Data Sources

1. **EPWRF :** EPW Research Foundation, India Time Series Data for Agricultural, Production and Banking sector of the GDP.
2. **CPCB :** Central Pollution Control Board of India for Conductivity, Faecal Coliform and Total Coliform ,data for pollutants in Indian Water bodies
3. **Census :** Census Population Data - 2001,2011**,** For Hindu Population
4. **NSSO :** National Sample Survey Office, For Literacy Rates
5. **MoEFW:** Environment Statistics of India