Econometrics Final Project

An Analysis of Sociopolitical Factors Affecting Hydrogencarbonate Quantity in Groundwater in India

Introduction

- India faces several environmental degradation challenges including water pollution. Many of India's rivers and lakes are heavily contaminated with industrial and agricultural waste.

- While Hydrogencarbonate is a naturally occurring anion in groundwater, it is possible for high quantities of Hydrogencarbonate to be present due to pollution.
  For example, the use of fertilizers and pesticides can lead to increased levels of Hydrogencarbonate in groundwater.

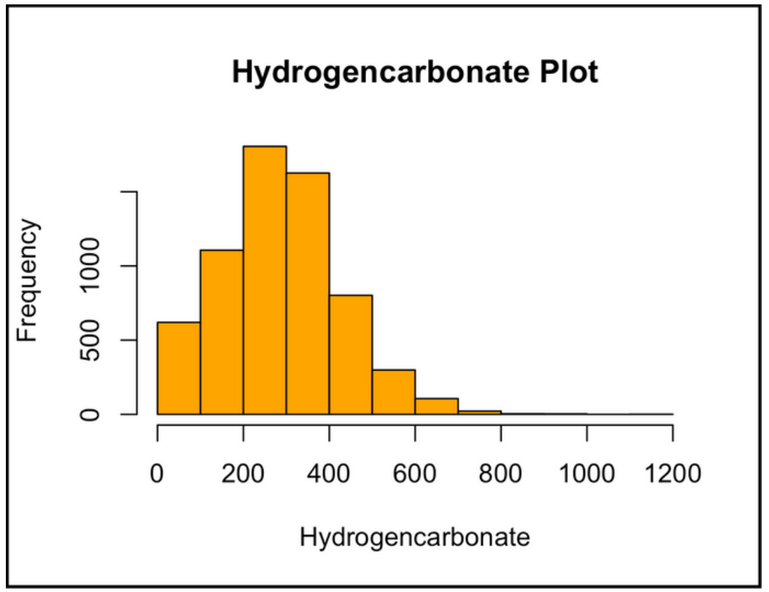
- Thus, higher levels of Hydrogencarbonate in groundwater may indicate that the water is contaminated.

- In our previous analysis of levels of Hydrogencarbonate in groundwater across districts, we mainly looked at economic factors.

- Looking at economic factors along with other social and political factors can help us understand environment degradation better.

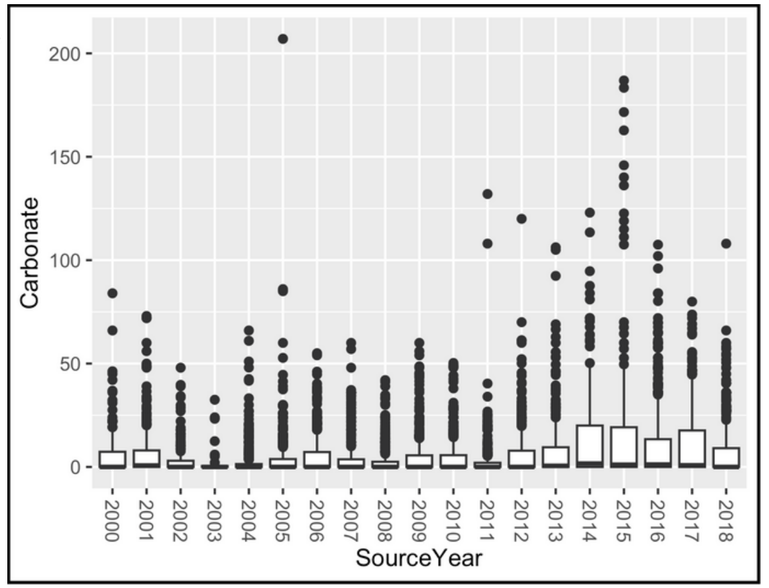
Data Visualisation

Statistic	N	Mean	St. Dev.	Min	Max
Hydrogencarbonate	6,395	286.566	139.174	0.000	1,118.333



Skew: 0.4290402

Shape of Distribution: Since the coefficient of skewness is slightly greater than 0, the graph is said to be positively skewed, with the majority of data values less than mean.
Although the data is distributed fairly normally, majority of the values are on the left side of the graph.



Outliers: There are a few outliers visible in the box plot.
For example, the data point around value 200 in 2005.
Units: Not mentioned (assume milligrams per liter)

Base Model

$$EQI(i,t) = \alpha\_0 + \alpha\_1 SDP(i,t) + \alpha\_2 SDP2(i,t) + \alpha\_3 SDP3(i,t) + \alpha\_4 GINI(i) + \gamma(i,t)$$

Variable	Description	Acronym
State Domestic Product	Total value of goods and services produced during any financial year within the geographical boundaries of a state (in Rs Crore)	*SDP*
Square of State Domestic Product	Square of the State Domestic Product in (Rs Crore)^2	*SDP*2
Cube of State Domestic Product	Cube of the State Domestic Product in (Rs Crore)^3	*SDP*3
Gini Coefficient	Measure of economic inequality in the district	GINI

Coefficients	Estimate	Std. Error	Signif. Code
α0	2.201e+02	9.119e+00	***
α1	4.986e-04	3.114e-05	***
α2	-6.987e-10	5.338e-11	***
α3	2.661e-16	2.338e-17	***
α4	-1.663e+00	3.034e+01	
R-squared	0.04918	Standard Error:	136

Signif. Codes:

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

- In a given state, all districts are assigned the state's SDP (and hence SDP2 and SDP3).
- **α0** tells us that in a district with no income inequality and whose state has an SDP of Rs. 0 Cr, the expected level of Hydrogencarbonate in the groundwater is **220.1 mg/L.**
- Since **α1, α2** and **α3** all depend on various powers of SDP, the effect of a unit increase in SDP will thus be dependent on the combined effect of α1, α2 and α3.

- **α4** tells us that if a district changes from having no economic inequality (Gini = 0) to complete economic inequality (Gini = 1), then we expect a decrease in Hydrogencarbonate quantity by **1.663 mg/L.**
- Note that change in the Gini Coefficient for a district level can only be equal to 1 under special circumstances (in the case described above) since the value of Gini Coefficient is between 0 and 1. So generally, the actual change in EQI level in most cases will be some number between 0 and 1 multiplied by α4.
- The negative sign of **α4** seems counterintuitive, since we would expect that with decreasing inequality, environmental conditions should improve. However, Gini is not a significant indicator.
  Perhaps we would fare better using a state level model.

State Level Model

$$EQI(I,T) = \alpha\_0' + \alpha\_1' SDP(I,T) + \alpha\_2' SDP2(I,T) + \alpha\_3' SDP3(I,T) + \alpha\_4' GINI\_STATE(I) + \gamma(I,T)$$

Variable	Description	Acronym
State Domestic Product	Total value of goods and services produced during any financial year within the geographical boundaries of a state (in Rs Crore)	*SDP*
Square of State Domestic Product	Square of the State Domestic Product in (Rs Crore)^2	*SDP*2
Cube of State Domestic Product	Cube of the State Domestic Product in (Rs Crore)^3	*SDP*3
Gini Coefficient	Measure of economic inequality in the state	*GINI*_STATE

Coefficients	Estimate	Std. Error	Signif. Code
$\alpha 0'$	1.122e+02	2.524e+01	***
$\alpha 1'$	7.445e-04	1.027e-04	***
$\alpha 2'$	-1.070e-09	1.948e-10	***
$\alpha 3'$	4.221e-16	8.993e-17	***
$\alpha 4'$	1.302e+02	6.646e+01	.
R-squared	0.1745	Standard Error:	113.6

Signif. Codes:

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

- In the above model, we have used data on the state level.
- The coefficeints of SDP, SDP2 and SDP3 are highly significant in both, the district model as well as the state model.

- However, it is evident that the state level model is better than the district level model, for the following reasons:

1. R - squared has increased from 0.04918 to 0.1745.
2. The standard error of a regression represents the average distance that the observed values fall from the regression line. The residual standard error has decreased from 136 to 113.6, indicating a better fit.
3. The coefficient of the Gini variable now has a positive sign, which is now in accordance with our expectations, that a more unequal society is results in higher pollution levels. Furthermore, coefficient is now also significant at 0.95 level.

Perhaps economic inequality is not the only indicator of environmental degradation. Maybe investigating power inequality or social inequality could yield some useful results.

Enhanced Model

$$EQI(i,t) = \alpha\_0 + \alpha\_1 SDP(i,t) + \alpha\_2 GINI\_STATE(i) + \alpha\_3 wm(i,t) + \alpha\_4 lr(i,t) + \alpha\_5 enpp(i,t) + \gamma(i,t)$$

Variable	Description	Acronym
Win Margin	Win margin refers to the percentage difference in the number of seats won by the winning party compared to the runner-up party in state assembly elections	wm
Effective Number of Political Parties	Measures the effective number of political parties in a state based on distribution of seats won by all parties	enpp
Literacy Rate	Literacy rate measures the percentage of a population that is able to read and write at a basic level	lr

Coefficients	Estimate	Std. Error	Signif. Code
$\alpha 0$	7.694e+02	7.955e+01	***
$\alpha 1$	1.650e-04	1.994e-05	***
$\alpha 2$	-2.038e+02	8.365e+01	*
$\alpha 3$	-3.469e-01	3.386e-01	
$\alpha 4$	-5.977+00	7.710e-01	***
$\alpha 5$	-2.283e+01	5.472e+00	***
R-squared	0.2343	Standard Error:	1.747

Signif. Codes:

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

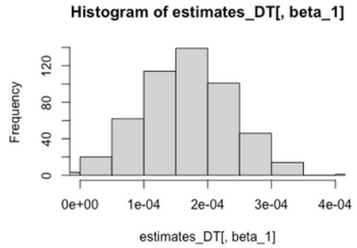
Number of Data Points = 387

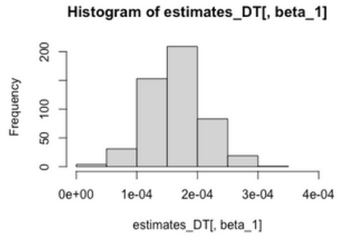
- Initially, bptest() gave us a p-value of 0.01127, indicating heteroskedasticity in our data.
- After transforming the variables, bptest() gave us a p-value of almost 1. Thus, we have rectified heteroskedasticity and we now proceed with running the regression and drawing inferences.
- Literacy rate and effective number of political parts are found to be significant.

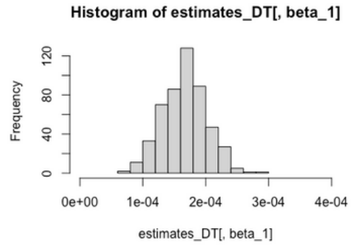
Methods of Testing

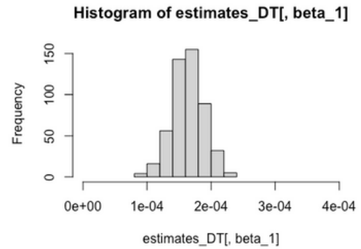
Monte-Carlo Simulations

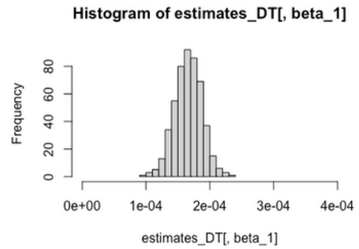
- We ran Monte-Carlo simulations for sample sizes of 40, 80, 160, 260 and 360.
- For all variables, standard deviation decreased and the average value came closer to the true assumed value.











- Hence, our estimates are consistent.

Maximum Likelihood Estimate

Regressors	SDP	gini_state	win_margin	literacy rate	ENPP
Coefficient Estimates	1.6636e-4	-2.2691e+2	-2.5524e-1	-6.1053e+0	-2.2152e+1

- We minimized the negative log-likelihood function for estimating the coefficients of the given linear regression model.
- We improved the estimation of the enhanced model by finding the set of coefficients that minimizes the negative log-likelihood of observing the given data. In other words, MLE finds the parameters that make the model most likely to produce the observed data.

Testing For Break

$$EQI(i,t) = \alpha\_0 + \alpha\_1 SDP(i,t) + \alpha\_2 GINI\_STATE(i) + \alpha\_3 wm(i,t) + \alpha\_4 lr(i,t) + \alpha\_5 enpp(i,t) + \alpha\_6 North(i) + \alpha\_7 South(i) + \alpha\_8 East(i) + \alpha\_9 West(i) + \alpha\_10 (North*enpp)(i,t) + \alpha\_11 (South*enpp)(i,t) + \alpha\_12 (East*enpp)(i,t) + \alpha\_13 (West*enpp)(i,t) + \gamma(i,t)$$

Coefficients	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_10	α_11	α_12	α_13
Estimate	4.085e+02	1.071e-04	1.590e+02	9.965e-01	-4.084e+00	-8.354e+00	1.567e+02	1.771e+02	-1.483e+02	6.997e+01	-1.841e+01	-3.864e+01	2.285e+01	-2.901e+00
Signif. Code	***	***	*	***	***		***	***	***			*	*	

R-squared	0.5695

Signif. Codes:

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

- We divide state categories into North, South, East, West and Central regions.
- If the interaction term for any of the state groups turns out to be significant then it indicates that an increase in enpp has a different effect in that state group compared to the others (i.e., a structural break across state groups).
- t-test indicates a structural break for Southern and Eastern states relative to the other state groups.

What's New

- When studying levels of pollutants in the environment, some of the first causes that come to mind are proximity to industries, levels of fertilizer usage, presence/absence of strict regulations, etc.

- Sociopolitical factors like income inequality (represented by the Gini coefficient), political power (represented by win margin), political diversity (represented by ENPP), and literacy rates are rarely studied in relation to pollutants and environmental quality.

- In fact, data on such unique indices isn't widely available. We created some of the indices from scratch.

- Our study analyses the effect of such sociopolitical factors, and has found that they do in fact influence the level of groundwater pollution to an extent.

Conclusion

- In the enhanced model, we found that literacy rate and the effective number of political parties have a highly significant effect on the level of hydrogencarbonate (HCO3) in groundwater, with a unit increase in literacy rate leading to a **5.988** unit decrease in HCO3 levels and a unit increase in ENPP leading to a **22.83** unit decrease in HCO3 levels.
- Furthermore, when testing for a structural break we found that a structural break does exist and that ENPP has a differential impact on the level of hydrogencarbonate in groundwater, based on whether the state lies in the North, South, Centre, West or East.
- While Central, Northern, and Western states did not experience a significant change in HCO3 levels, for a unit increase in ENPP, Southern states experienced a decrease of **38.64** units of HCO3. On the other hand, Eastern states experienced an increase of **22.85** units of HCO3 for a unit increase in ENPP.

- Thus, our study finds that not only do sociopolitcal factors have an effect on the environmental quality, in some areas they have a differential effect on environmental quality. This finding could be the basis for further interesting research to uncover why sociopolitical factors are more effective in reducing pollution in some states and less effective in others.

END