Analysing the role of Income and Power Inequality on Groundwater levels

Rachita Saha (2019082)
Aayush Ranjan (2021003)
Aniket Kanojia (2021377)
Priyanshu Sehrawat (2021407)

Introduction

Groundwater is a vital natural resource, and its availability is crucial for sustaining agricultural and industrial activities. However, groundwater resources are under threat due to several factors, including climate change, population growth, and economic development. In this project, we aim to investigate the role of income and power inequality in influencing groundwater levels.

Income and power inequality are pressing issues in today's world, and their effects are felt across various domains, including economic growth, social development, and environmental sustainability.

The project aims to contribute to the existing literature by providing empirical evidence on the relationship between income and power inequality and groundwater levels.

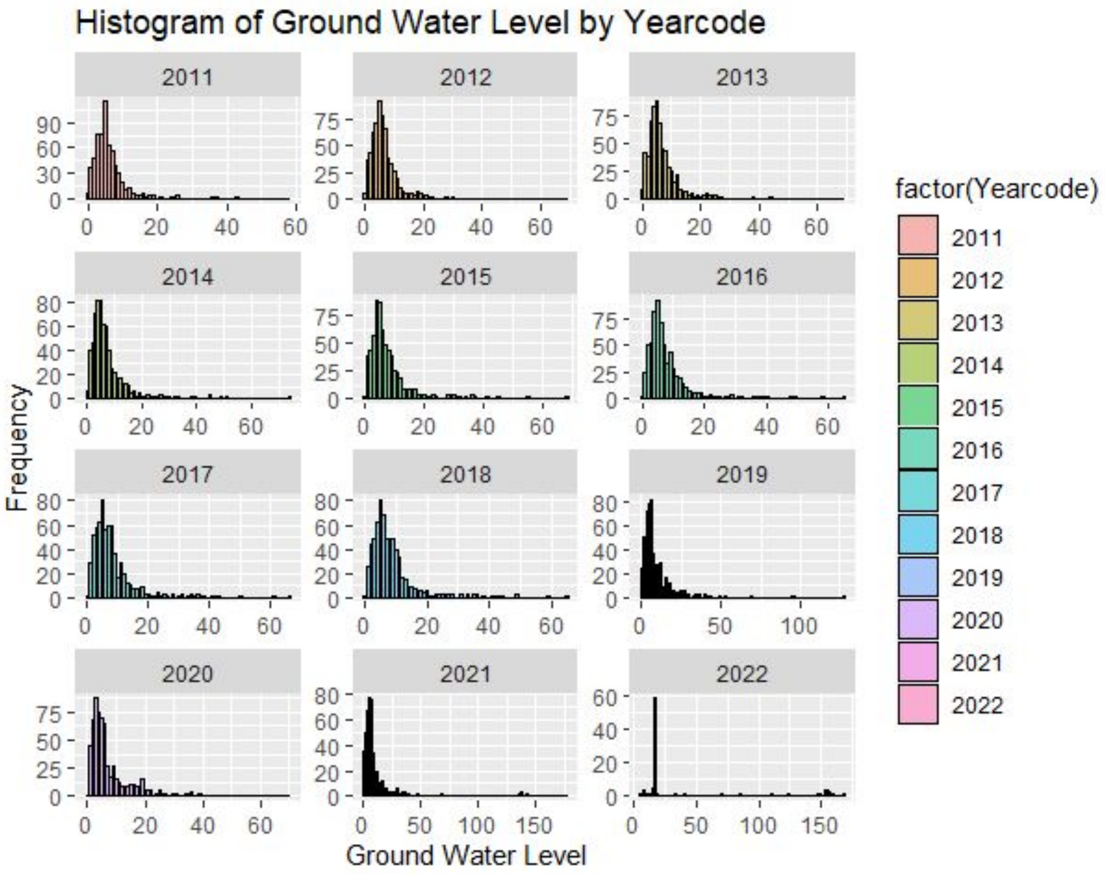
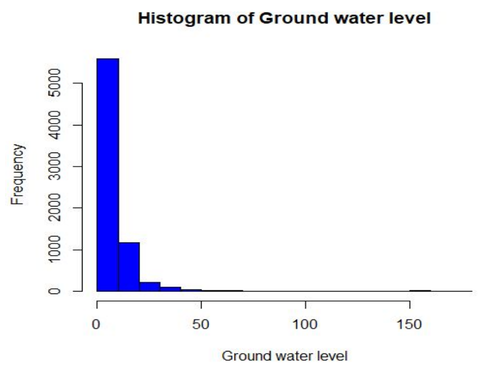
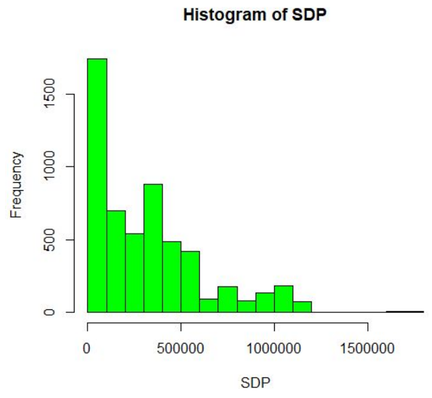


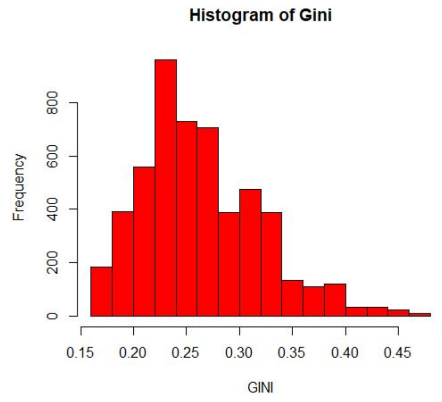
Fig: Groundwater levels across districts in India over the years 2011 to 2022

Variables

Variable	Acronym	Min	1st Qu	Median	Mean	3rd Qu	Max	St. Deviation	Description
Ground water level	Ground.water.level	0.01879	3.733	5.734	8.74	9.32339	177.618	13.002	It is measured as the vertical distance in meters from the ground to the water.
State Domestic Product	SDP	4405	16588	255739	313337	470328	1782903	3033215.7	Measured as per capita net state domestic product in rupees.
Gini Index	GINI	0.1600	0.2300	0.2600	0.268	0.31	0.48	0.057	It is ratio that represent the income distribution.





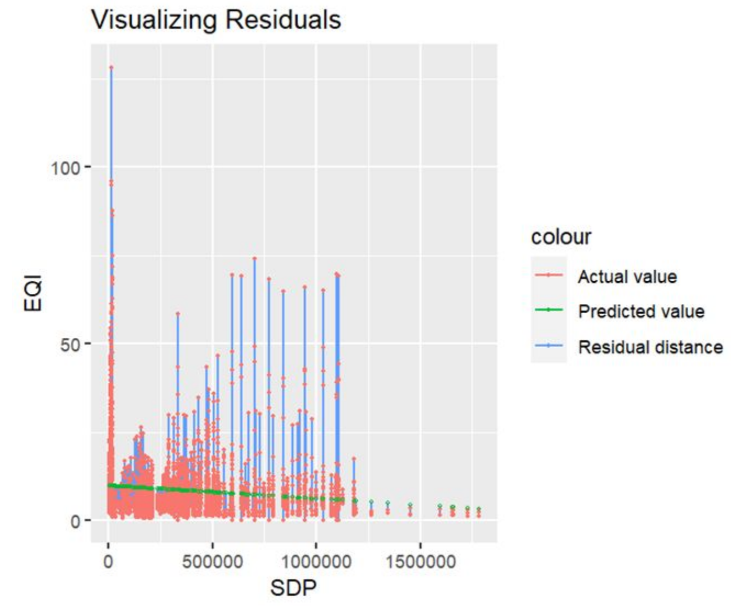


Baseline Model

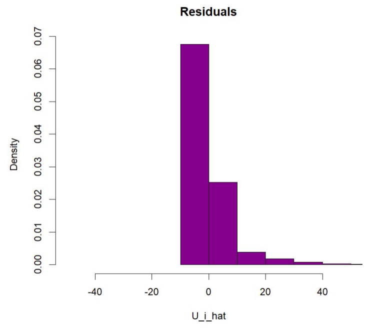
Baseline Model Equation:

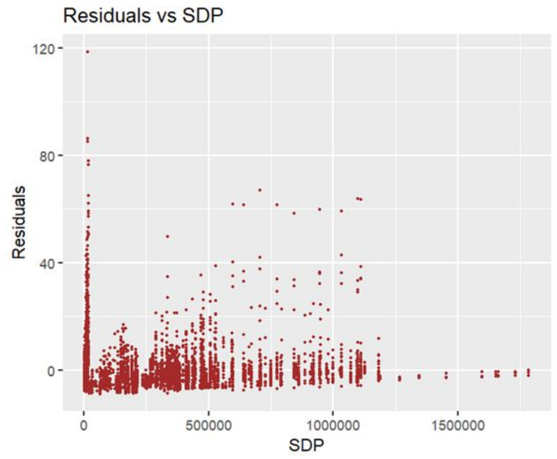
$$\text{Ground.water.level}_{i,t} = \beta_0 + \beta_1 SDP_{i,t} + \beta_2 SDP^2_{i,t} + \beta_3 SDP^3_{i,t} + \beta_4 Gini_{i,t}$$

Explanatory Variables	**Coefficient**
SDPscaled	-2.830*10^-2
SDP2	4.528*10^-5
SDP3	-1.976*10^-8
Gini	2.276*10^1
Intercept	6.376



Residual Standard Error	8.453 with 4247 DF
Multiple R-squared	0.07782
Adjusted R-squared	0.07695
F-statistic	89.6





Enhanced Model

Enhanced Model Equation:

$$year\_state\_GWL_{i,t} = \beta_0 + \beta_1 SDP_{i,t} + \beta_2 SDP^2_{i,t} + \beta_3 SDP^3_{i,t} + \beta_4 StateGini_{i,t} + \beta_5 party\_count_{i,t} + \beta_6 election\_margin_{i,t} + \beta_7 percent\_women\_elected_{i,t} + \beta_8 rainfall_{i,t} + \beta_9 population\_density_{i,t} + \gamma_{i,t}$$

Explanatory Variables	Estimate	Standard Error
SDPscaled	6.837e-04	4.515e-03
$SDPscaled^2$	-2.060e-06	7.851e-06
$SDPscaled^3$	2.859e-10	3.421e-09
StateGini	-3.113e+01 ***	6.096e+00
election_margin	2.203e-01 ***	3.036e-02
Party_count	3.455e-01	2.666e-01
percent_women_elected	-1.241e-01 **	4.729e-02
population_density	-3.883e-03 *	1.704e-03
rainfall	-3.827e-03	1.479e-03
Intercept	1.586e+01 ***	1.492e+00

Residual Standard Error	4.061 with 198 DF
Multiple R-squared	0.3132
Adjusted R-squared	0.282
F-statistic	10.03

New Regressors Descriptions

- **election_margin:** The difference in no. of votes won by the winning party and the party that came second in State Lok Sabha election
- **party_count:** The number of effective parties competing in the state.
- **percent_women_elected:** The percentage on seats won by women in the state out of all available seats.
- **rainfall:** The total rainfall (in mm) received by the state in a given year.
- **population_density:** The number of people per sq km in a state of India.

Variables	year_state_GWL	SDP	State Gini	Party_count	election_margin	percent_women_elected	Rainfall	Population_density
Min.	1.485	4405	0.1640	1.000	0.00	0.000	0.0	17
1st Qu	4.383	47696	0.2370	1.000	2.00	0.000	5.0	189.00
Median	5.968	283875	0.2700	2.000	5.00	7.692	46.9	350.00
Mean	8.418	377350	0.2775	2.648	12.38	8.485	123.5	826.2
3rd Qu	9.761	553009	0.3400	4.000	19.00	14.286	154.1	555.0
Max.	135.088	1782903	0.4010	7.000	66.00	28.571	1612.0	11297.0
St.Deviation	11.326	388914.3	0.069	1.663	15.244	7.580	196.577	2,123.897

Monte Carlo Simulations

On performing Monte Carlo simulations using 500 iterations, we found out that on gradually increasing the number of samples, the mean value of the estimates came closer to the actual OLS estimates and their variances kept on decreasing.

Conclusion: Our OLS estimates are consistent.

Statistic	N	Mean	St. Dev.	Min	Max
beta_1	500	-0.002	0.011	-0.031	0.032
beta_0	500	8.457	2.841	0.223	16.949

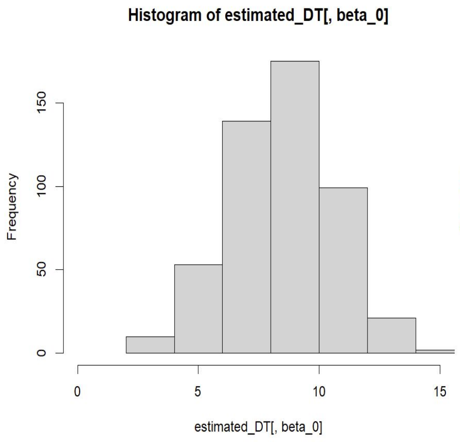
N=10

Statistic	N	Mean	St. Dev.	Min	Max
beta_1	500	-0.002	0.002	-0.011	0.005
beta_0	500	8.528	1.064	4.881	11.770

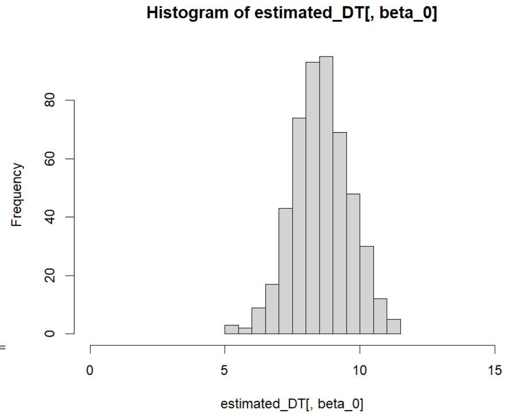
N=30

Statistic	N	Mean	St. Dev.	Min	Max
beta_1	500	-0.003	0.001	-0.006	0.001
beta_0	500	8.608	0.677	6.563	10.381

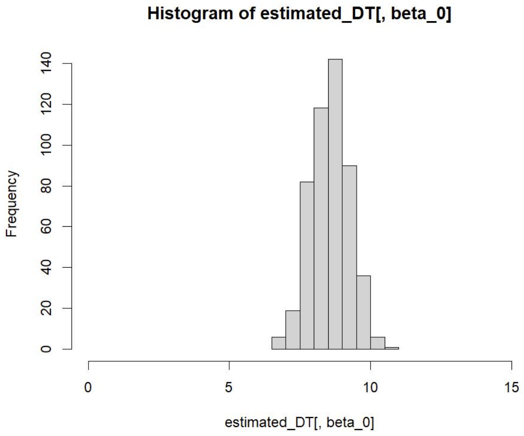
N=80



N=10



N=30



N=80

Breusch-Pagan Test

Null Hypothesis: The variance of the residuals across the groups is similar.

Alternate Hypothesis: The variance of the residuals is a function of the explanatory variables.

On applying OLS estimation to our enhanced model, the p-value is 1.131e-14 << 0.05. Hence, we can reject the null hypothesis. The assumption of homoskedasticity in our OLS model is therefore violated. This can cause bias in the model.

To account for heteroskedasticity, we use Weighted Least Squares method. This method assigns each data point i with some positive weight of 1/var(e_i). We would give more priority to the data points that show lesser variance.

	Dependent variable:	
	year_state_GWL	
	OLS S.E	FGLS
	(1)	(2)
SDPscaled	0.001	-0.0001
	(0.005)	(0.003)
SDPscaled2	-0.00000	0.00000
	(0.00001)	(0.00000)
SDPscaled3	0.000	-0.000
	(0.000)	(0.000)
StateGini	-31.125***	-30.518***
	(6.096)	(3.778)
election_margin	0.220***	0.308***
	(0.030)	(0.029)
party_count	0.345	0.548***
	(0.267)	(0.195)
population_density	-0.004**	-0.005***
	(0.002)	(0.001)
percent_women_elected	-0.124***	-0.200***
	(0.047)	(0.024)
rainfall	-0.004**	-0.001
	(0.001)	(0.0004)
Constant	15.862***	14.578***
	(1.492)	(1.084)
Observations	208	208
R2	0.313	0.582
Note:	*p<0.1; **p<0.05; ***p<0.01	

Fig: BP Test Before FGLS

```
data:  model
BP = 85.797, df = 9, p-value = 1.131e-14
```

Fig: BP Test After FGLS

```
data:  model_fgls
BP = 9.3097, df = 9, p-value = 0.4092
```

Maximum Likelihood Estimation

We have estimated the parameter coefficients using the maximum likelihood estimation technique assuming gamma distribution.

Since, groundwater level was coming out to be right-skewed, we chose gamma distribution to model it.

Since, MLE is a more flexible technique compared to the OLS method, it does not require the homoscedasticity and the normal distribution assumption and hence estimates the parameters in a model having other types of distribution of errors i.e Gamma distribution in our model.

Explanatory Variables	Estimate
SDPscaled	6.837e-04
SDPscaled^2	-2.060e-06
SDPscaled^3	2.859e-10
StateGini	-3.113e+01
election_margin	2.203e-01
Party_count	3.455e-01
percent_women_ele cted	-1.241e-01
population_density	-3.883e-03
rainfall	-3.827e-03
Intercept	1.586e+01

Table: OLS Estimates

Explanatory Variables	Estimate
SDPscaled	6.836700e-04
SDPscaled^2	-2.059890e-06
SDPscaled^3	2.859135e-10
StateGini	-3.112547e+01
election_margin	2.202858e-01
Party_count	3.454568e-01
percent_women_ele cted	-1.2410974e-01
population_density	-3.882949e-03
rainfall	1.04757e+00
Intercept	1.586e+01

Table: MLE Estimates

Structural Breaks in the Model

Both t-test and the Chow test are used to check for structural breaks in the given model.

However, the t-test is used for comparing the means of 2 groups of data, we can also use it to conclude if there are differences in the coefficients of the models of each group. Chow-test is more suited to checking for breakpoints in the main model across different groups. We introduced these breakpoints as a dummy variable and checked differences where this dummy variable changed its value.

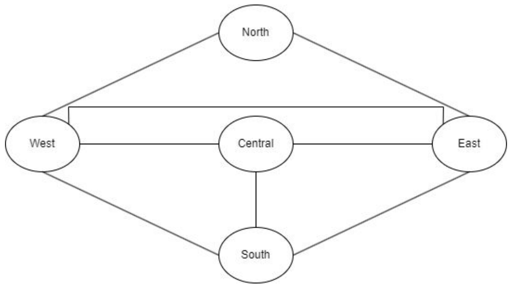


Fig: Chow Test

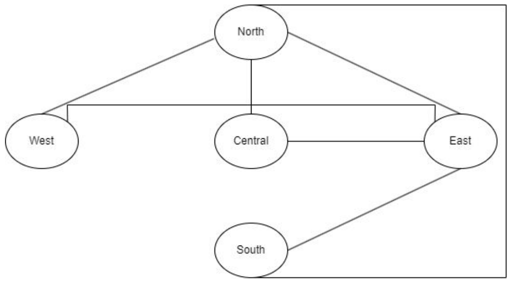


Fig: t-Test

The links show the structural breaks across the groups

Intersection	Chow-Test (p-value)	T-Test (p-value)
South-North	0.3829785	0.002461
South-East	2.761309e-17	2.357e-12
South-West	7.861446e-04	0.3051
East-North	4.830868e-04	2.773e-06
West-North	7.969244e-03	.001097
South-Central	1.407954e-05	0.2088
Central-North	0.07365334	0.0007976
East-West	2.820927e-18	4.134e-13
West-Central	2.125114e-04	0.57
East-Central	1.202069e-08	3.742e-05

Conclusion

- **Income inequality:** Gini index is statistically significant and negatively correlated with the GWL. This suggests that higher the gini (income inequality), lower is the GWL.
- **Power Inequality**: Election_margin and party_count are positively correlated with GWL, whereas percent_elected_women is negatively correlated.
- **Control Variables**: Population_density is negatively correlated with groundwater level which means that the groundwater level reduces if the population of people per unit area in a state increases. Rainfall on the other hand is negatively correlated to the groundwater level which is contrary to what we expected.
- Using BP test we got our p-value to be **1.31e-14** which means that our data was **heteroskedastic.** We fixed it using the FGLS method of weighted least squares after which we got our p value to be **0.4092**.
- Since the distribution of Ground water level is highly right-skewed, we perform MLE estimation assuming a gamma distribution for the population. The coefficient estimates observed after this were similar to the OLS estimates with the exception of rainfall variable.
- We encountered structural break points across the state groups in the model. This suggests that the coefficients of estimations across the state groups vary significantly.

Data sources

1) https://eci.gov.in/files/file/2851-party-wise-state-wise-seat-won-and-votes-polled-lok-sabha-2009/
2) https://datasource.kapsarc.org/explore/dataset/statewise-census-data-in-india-1901-2011/export/?disjunctive.location_name&disjunctive.variable_name&sort=date&q.timerange.date=date:%5B2001-01-01+TO+2011-01-01%5D&refine.variable_name=Population+Density
3) https://mausam.imd.gov.in/imd_latest/contents/index_rainfall_state_new.php

Thank You