ECO 221

ECONOMETRICS PROJECT

Analysis of Role of Power Inequality using Ground Water Level

project presentation

GROUP 21

GUNAR SIDHWANI

MEDHA HIRA

KRITIKA GUPTA

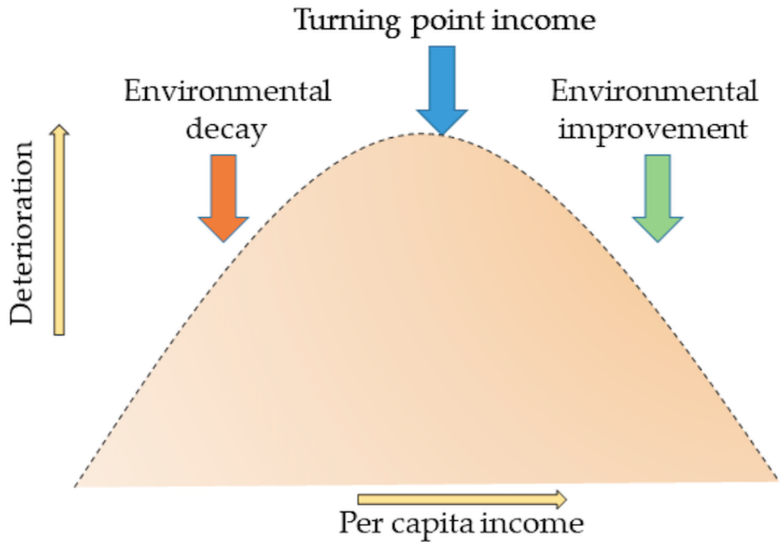
SUMEET MEHRA

ECO 221

Describing Key Concepts

Environmental Kuznet's Curve

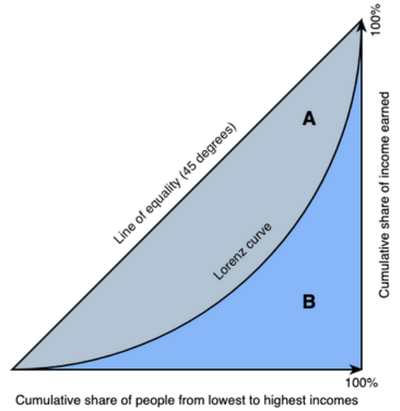
Modified by Gene Grossman and Alan Krueger



Pre-Industrial Era: Economic indicators deteriorate as the economy industrializes.
Then there is a **Turning Point**.
Post-Industrial Era: The indicators improve again or degrade less when new tech is introduced, and more money is funneled back to society.

Understanding Inequality

There isn't a single society without it.
Inequality generally calculated using a GINI Index.



GINI Index:
Gives us no information about the distribution of wealth between demographics.
It doesn't tell us how easy or difficult it is to escape poverty.
No insight into how society arrived at that level of inequality.

Today, India is considered a **mixed economy**, the private and public sectors co-exist, and the country leverages international trade.

ECO 221

Exploring the chosen environmental variable

Environmental Variable

Ground Water Level

Affect of it on Environment

The level of groundwater can significantly impact the environment, as it can affect the availability of water for plants and animals, as well as the overall health of ecosystems.

As an indicator of climate change

In addition to providing insights into the health of ecosystems, groundwater level can also be used as an indicator of the impacts of climate change.

Implication of High Ground Water

High groundwater levels can also help to recharge surface water sources such as lakes, rivers, and wetlands, which can help to maintain healthy water levels in these areas.

Implication of Low Ground Water

Low groundwater levels can also increase the risk of wildfires, as dry vegetation is more susceptible to burning.

Total Number of Outliers for the Ground Water Level Data = 351

ECO 221

Results derived by running regression models

Intercept, also known as β_0 is being estimated here.
We have used the OLS estimate because of the property of unbiasedness, this means that $E(\beta^{hat}_{0,OLS}) = \beta_0$

β_1 is being estimated here. It represents the change in EQI if the SDP has been increased by value of 1.
It is $\Delta(EQI)/\Delta(SDP)$ when $\Delta u = 0$.

$$(EQI)_{i.t} = \beta_0 + \beta_1 SDP_{i.t} + u_{i.t}$$

Ground.water.level

Predictors	*Estimates*	*CI*	*p*
(Intercept)	7.74	7.34 – 8.14	**<0.001**
VALUE	0.00	0.00 – 0.00	**<0.001**
Observations	4628		
R^2 / R^2 adjusted	0.009 / 0.008		

$$(EQI)_{i,t} = \alpha_0 + \alpha_1(SDP)_{i,t} + \alpha_2(SDP)^2_{i,t} + \alpha_3(SDP)^3_{i,t} + \alpha_4(GINI) + \gamma_{i,t}$$

Ground.water.level

Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.25	-1.52 – 1.01	0.694
VALUE	0.00	0.00 – 0.00	**<0.001**
VALUE^2	-0.00	-0.00 – -0.00	**<0.001**
VALUE^3	0.00	0.00 – 0.00	**0.002**
ginivalue	16.87	12.92 – 20.82	**<0.001**
Observations	4628		
R^2 / R^2 adjusted	0.061 / 0.060		

The Regression Model estimates the relationship between the independent variable VALUE (sdpValue), value^2,value^3, ginivalue measured as the sdpValue and the dependent variable Ground.water.value.
(**α**0) : This is the intercept term, which represents the value of the dependent variable (EQI) when all of the predictor variables (VALUE, VALUE^2, VALUE^3, and ginivalue) are equal to zero. In other words, a represents the y-intercept of the regression line. This is basically the Ground Water Level Value whenSDP and Ginivalue are null and zero.

(**α**1) :This is the coefficient for the first predictor variable VALUE), which represents the linear relationship between VALUE and ginivalue. It indicates the change in Ground.water.value for a one-unit increase in VALUE, holding all other variables constant.**

(**α**2) : This is the coefficient for the second predictor variable (VALUE^2), which represents the quadratic relationship between VALUE and Ground.water.value. It indicates the change in Ground.water.value for a one-unit increase in VALUE^2, holding all other variables constant. **

(**α**3): This is the coefficient for the third predictor variable (VALUE^3), which represents the cubic relationship between VALUE and Ground.water.value . It indicates the change in Ground.water.value for a one-unit increase in VALUE^3, holding all other variables constant.**

(**α**4) : This is the coefficient for the fourth predictor variable (ginivalue),which represents the linear relationship between ginivalue and Ground.water.value. It indicates the change in Ground.water.value for a one unit increase in ginivalue, holding all other variables constant.

** - In practice, it is often difficult to hold some variables constant while manipulating others.It is mathematically not possible to perform an increase in x^2 or x^3 without increasing x or vice versa, but the interpretations of individual coefficients can be only found out when this is performed.
The Variables are not independent of each other,

The OLS estimator is:
Unbiased, Efficient and Consistent with the estimation of the true value
Therefore it is a good starting point in the regression analysis.

Link to Summary Sheet:
https://docs.google.com/spreadsheets/d/1hPy6dREqwi8ugUcx0eFFjvpFID57bKVEkjej29ElfkA/edit#gid=0

Introduction of Political Variables and Regression Results

1. The Margin of Election Victory
2. Index for Effective Number of Political Parties in a State Assembly
3. Literacy Rate
4. Women's Representation in Politics
5. Average Monthly Salary

```
Residuals:
    Min      1Q  Median      3Q     Max
-13.932  -3.949  -1.239   1.751 117.494

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              8.308e+01  5.729e+00  14.503  < 2e-16 ***
as.numeric(VALUE)        2.374e-05  3.685e-06   6.442 1.32e-10 ***
stateginivalue          -4.932e+01  4.314e+00 -11.432  < 2e-16 ***
vote_share              -1.704e-01  3.093e-02  -5.509 3.84e-08 ***
enop                    -1.543e+00  3.319e-01  -4.650 3.43e-06 ***
avgmonthlysalary        -1.514e+00  2.775e-01  -5.457 5.13e-08 ***
women_rep                5.197e-01  5.731e-02   9.067  < 2e-16 ***
litrate                 -4.149e-01  3.682e-02 -11.269  < 2e-16 ***
I(as.numeric(VALUE)^2) -1.297e-11  4.510e-12  -2.877  0.00404 **
I(as.numeric(VALUE)^3)  1.853e-18  1.646e-18   1.126  0.26020
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.53 on 4018 degrees of freedom
Multiple R-squared:  0.1829,    Adjusted R-squared:  0.1811
F-statistic: 99.95 on 9 and 4018 DF,  p-value: < 2.2e-16
```

Granularity : State

```
Residuals:
    Min      1Q  Median      3Q     Max
-15.488  -3.869  -1.265   1.666 117.928

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              8.043e+01  5.827e+00  13.802  < 2e-16 ***
as.numeric(VALUE)        4.132e-05  3.336e-06  12.389  < 2e-16 ***
ginivalue                9.837e+00  2.318e+00   4.244 2.25e-05 ***
vote_share              -1.759e-01  3.150e-02  -5.584 2.50e-08 ***
enop                    -1.440e+00  3.364e-01  -4.280 1.91e-05 ***
avgmonthlysalary        -3.632e+00  2.033e-01 -17.862  < 2e-16 ***
women_rep                6.439e-01  5.725e-02  11.248  < 2e-16 ***
litrate                 -1.508e-01  2.823e-02  -5.342 9.72e-08 ***
I(as.numeric(VALUE)^2) -2.731e-11  4.372e-12  -6.247 4.61e-10 ***
I(as.numeric(VALUE)^3)  5.551e-18  1.635e-18   3.395 0.000694 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.635 on 4018 degrees of freedom
Multiple R-squared:  0.1601,    Adjusted R-squared:  0.1582
F-statistic:  85.1 on 9 and 4018 DF,  p-value: < 2.2e-16
```

Granularity : District

Introduction of Political Variables and Regression Results

1. The Margin of Election Victory
2. Index for Effective Number of Political Parties in a State Assembly
3. Literacy Rate
4. Women's Representation in Politics
5. Average Monthly Salary

```
Residuals:
    Min      1Q  Median      3Q     Max
-13.932  -3.949  -1.239   1.751 117.494

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           8.308e+01  5.729e+00  14.503  < 2e-16 ***
as.numeric(VALUE)     2.374e-05  3.685e-06   6.442 1.32e-10 ***
stateginivalue       -4.932e+01  4.314e+00 -11.432  < 2e-16 ***
vote_share           -1.704e-01  3.093e-02  -5.509 3.84e-08 ***
enop                 -1.543e+00  3.319e-01  -4.650 3.43e-06 ***
avgmonthlysalary     -1.514e+00  2.775e-01  -5.457 5.13e-08 ***
women_rep             5.197e-01  5.731e-02   9.067  < 2e-16 ***
litrate              -4.149e-01  3.682e-02 -11.269  < 2e-16 ***
I(as.numeric(VALUE)^2) -1.297e-11  4.510e-12  -2.877  0.00404 **
I(as.numeric(VALUE)^3)  1.853e-18  1.646e-18   1.126  0.26020
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.53 on 4018 degrees of freedom
Multiple R-squared:  0.1829,    Adjusted R-squared:  0.1811
F-statistic: 99.95 on 9 and 4018 DF,  p-value: < 2.2e-16
```

Granularity : State

```
Residuals:
    Min      1Q  Median      3Q     Max
-15.488  -3.869  -1.265   1.666 117.928

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           8.043e+01  5.827e+00  13.802  < 2e-16 ***
as.numeric(VALUE)     4.132e-05  3.336e-06  12.389  < 2e-16 ***
ginivalue             9.837e+00  2.318e+00   4.244 2.25e-05 ***
vote_share           -1.759e-01  3.150e-02  -5.584 2.50e-08 ***
enop                 -1.440e+00  3.364e-01  -4.280 1.91e-05 ***
avgmonthlysalary     -3.632e+00  2.033e-01 -17.862  < 2e-16 ***
women_rep             6.439e-01  5.725e-02  11.248  < 2e-16 ***
litrate              -1.508e-01  2.823e-02  -5.342 9.72e-08 ***
I(as.numeric(VALUE)^2) -2.731e-11  4.372e-12  -6.247 4.61e-10 ***
I(as.numeric(VALUE)^3)  5.551e-18  1.635e-18   3.395 0.000694 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.635 on 4018 degrees of freedom
Multiple R-squared:  0.1601,    Adjusted R-squared:  0.1582
F-statistic:  85.1 on 9 and 4018 DF,  p-value: < 2.2e-16
```

Granularity : District

Discussing the Results

Role of Standard Errors in Regression Analysis

1. Standard Error is the estimated standard deviation of the sampling distribution of the statistic, and it measures the degree of uncertainty in the estimate of the true population coefficient due to sampling variation.
2. Smaller standard errors indicate greater precision in the estimate of the population parameter or statistic.
3. Standard errors can be used to carry out Hypothesis Testing, particularly with the t-test.
4. t-values far away from confidence interval indicate more significance of the predictor variable.
5. In this regression analysis, all of the predictor variables are significant at the 0.001 level, which indicates a very high level of significance.(Refer p-value)

* One asterisk (*) indicates significance at the 0.05 level (p < 0.05)
* Two asterisks (**) indicate significance at the 0.01 level (p < 0.01) * Three asterisks (*) indicate significance at the 0.001 level (p < 0.001)

Anova Table

term	df	sumsq	meansq	statistic	p.value
as.numeric(VALUE)	1	1.931762e+03	1931.762454	38.9858640	0.0000000
I(as.numeric(VALUE)^2)	1	1.165798e+04	11657.975339	235.2754297	0.0000000
I(as.numeric(VALUE)^3)	1	2.202179e+03	2202.178811	44.4432718	0.0000000
ginivalue	1	1.801675e+03	1801.675052	36.3605051	0.0000000
women_rep	1	5.685231e+03	5685.231335	114.7364964	0.0000000
enop	1	8.252062e+02	825.206170	16.6538983	0.0000457
vote_share	1	1.164317e+03	1164.317042	23.4976644	0.0000013
avgmonthlysalary	1	1.771315e+04	17713.146712	357.4778709	0.0000000
litrate	1	1.663157e+03	1663.156978	33.5650026	0.0000000
north	1	2.732872e+04	27328.723601	551.5346362	0.0000000
south	1	7.097009e+00	7.097009	0.1432283	0.7051125
west	1	7.952112e+03	7952.111823	160.4855450	0.0000000
center	1	1.539021e+01	15.390212	0.3105976	0.5773448
Residuals	4014	1.988950e+05	49.550331	NA	NA

ANOVA (Analysis of Variance) table is a statistical table that displays the results of an ANOVA analysis. The F-value is the ratio of the between-group mean square to the within-group mean square. The p-value is associated with the F-value, which indicates the result's significance level.

Results of Test

Results of Chow Test Suggests that structural breaks are present.
Splitting the regression into multiple regression for different region provides a better fit for data.
Having a singular regression model for the entire data is not a great fit

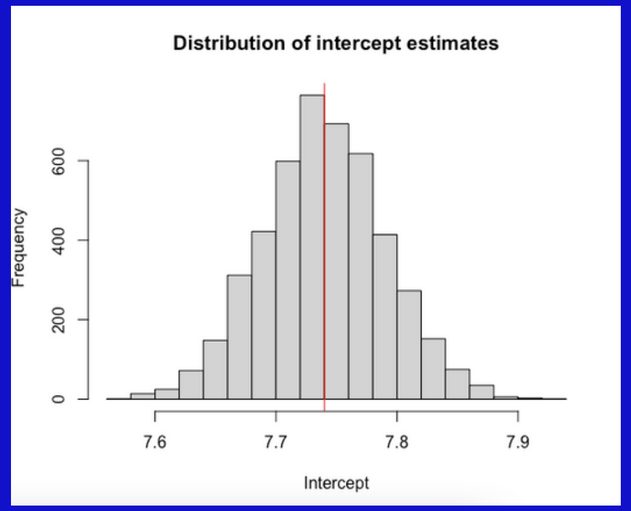
ECO 221

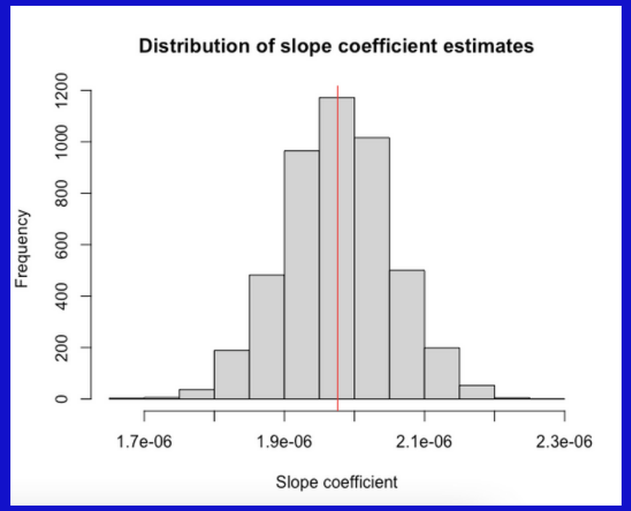
MonteCarlo Simulations & MLE

	Ground.water.level		
Predictors	*Incidence Rate Ratios*	*CI*	*p*
(Intercept)	1542.70	908.37 – 2619.06	**<0.001**
VALUE	1.00	1.00 – 1.00	**<0.001**
VALUE^2	1.00	1.00 – 1.00	**<0.001**
VALUE^3	1.00	1.00 – 1.00	**<0.001**
ginivalue	2.39	1.96 – 2.93	**<0.001**
women rep	1.04	1.03 – 1.04	**<0.001**
enop	0.75	0.72 – 0.77	**<0.001**
vote share	0.98	0.98 – 0.98	**<0.001**
avgmonthlysalary	0.80	0.78 – 0.81	**<0.001**
litrate	0.99	0.99 – 0.99	**<0.001**
north	1.97	1.91 – 2.02	**<0.001**
south	1.25	1.18 – 1.31	**<0.001**
west	1.45	1.39 – 1.50	**<0.001**
center	1.03	0.99 – 1.07	0.150
Observations	4028		
R^2 Nagelkerke	0.875		

MLE is widely used in statistical modeling and inference because it has desirable properties such as consistency, efficiency, and asymptotic normality.
Till now, we have only explored the normal distribution of the errors and Ground.water.level parameter. MLE gave us the opportunity to consider other possible distributions and get the optimal solution and best estimates.

The diagram on the left shows the poison distribution.





Value of alpha_0 used: 7.740e+00
Value of alpha_1 used: 1.976e-06
Number of iterations: length of df_8f

ECO 221

Testing for Heteroscedasticity

Breusch Pagan Test

Our assumptions while running the OLS tests had been the presence of a homoskedastic system. But this is not true in the real world. On running this test on the model, we certify the heteroscedastic nature.

```
> # Perform the Breusch-Pagan test
> bptest(model5)

        studentized Breusch-Pagan test

data:  model5
BP = 237.63, df = 13, p-value < 2.2e-16
```

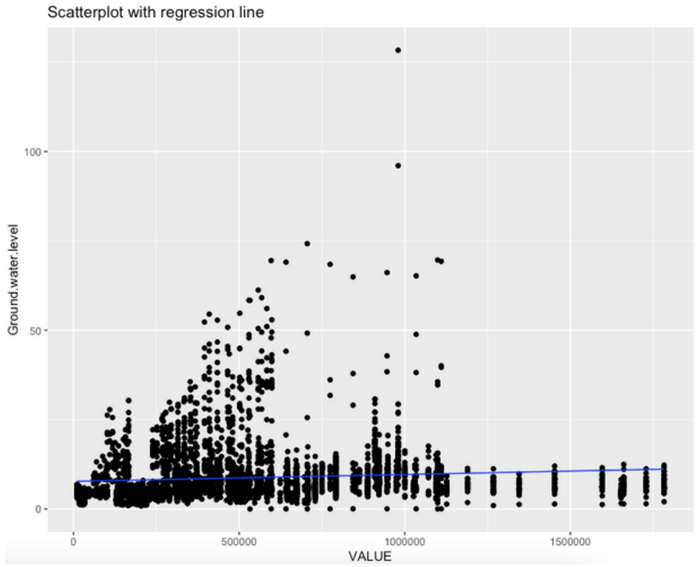
```
> print(variances)
  North States   South States    East States    West States Central States
     58.998713      11.214879       3.220341     186.165998       7.508700
```

Here we see a variation in data across each of the five groupings of states in India.

Then up until now, why did we use the OLS estimator? Would it have been wise to check for heteroskedasticity early on?

It violates the property of constant variance or homoskedasticity. But the heteroscedastic nature still holds onto the unbiasedness property of the OLS estimator, this helps us make an educated guess in the right direction.

To resolve the issue of heteroscedasticity, we can apply a transformation on x and y such that x(new) = Tx and y(new) = Ty. Here T is a matrix such that: T'T = omega^(-1) where omega = V(y|x). After this transformation, we can apply OLS estimates on x(new) and y(new) to obtain the results.



ECO 221

What's new in our project?

Objective: To understand the relationship between women's representation in politics and groundwater level (EQI).

Null hypothesis: After controlling for other variables in the model, there is no significant relationship between women's representation in politics and groundwater level.
Alternative hypothesis: After controlling for other variables in the model, there is a significant positive relationship between women's representation in politics and groundwater level.

The p-value obtained while running the hypothesis testing is:
1.759677e-07
While the significance level is set to 0.05 (alpha).
Indicating with 95% confidence and 5% significance level that we reject the null hypothesis.
There is a positive relationship between women's representation in politics and groundwater level after controlling for other variables.

ECO 221

Conclusions and Inferences and References

For the Regression Analysis :

- The R^2 Value obtained is 0.18 for State Level Granularity and 0.1582 for District Level Granularity.
- From the p-value and the hypothesis testing, we infer that there is a positive correlation between women's representation in politics and groundwater level.
- We see that when we introduce the dummy variables of north, south, west, and center, the R^2 value increases. We avoid including the east dummy variable as this would lead us into a dummy variable trap.
- We studied the inequality of income via the gini value in assignment one and power inequality by considering political factors, voting statistics, and even considering a person's monthly salary.
- We notice that variables of vote_share, avgmonthlysalary, enop, litrate affect the ground water level in a positive way.
- This study made us understand the significance on OLS estimators and later the draw back in case of a heteroscedastic system. In the case of which we choose to use the FGLS method.

References

https://www.indiacensus.net/literacy-rate.php
https://adrindia.org/sites/default/files/Women_representation_among_all_MPs_and_MLAs_English.pdf
https://www.statista.com/statistics/1305089/india-average-monthly-salary-by-state/
https://www.devdatalab.org/shrug

THANK YOU

MR. GAURAV ARORA
MS. SUMEDHA SHUKLA