ECO Project

Group 19

MEMBERS

Ankur Bohra
Somay Jalan
Shaurya Bajaj
Abhinav Kumar Saxena

INDICATOR

hardnesstotal

Background

Goal: Understand the relationship between inequality, economic growth, and environmental pollution.

Proxies: Inequality: District-level Gini indices
Economic growth: Net state domestic products*
Environmental pollution: District-level <u>total hardness</u> readings

Total hardness:

- *Defined as the sum of the calcium and magnesium concentration both expressed as $CaCO_3$ in mg/L. [1]*
- *Requirement/acceptable limit:* ***200mg/L***
  Permissible limit in the absence of alternate source: ***600mg/L*** *[2]*

*At factor cost, at 2011-12 constant prices

[1] Central Ground Water Board, https://www.cgwb.gov.in/cgwbpnm/public/uploads/documents/17068003031756689658file.pdf

[2] Indian Standard IS 10500: 2012 (DRINKING WATER — SPECIFICATION)

Summary Statistics

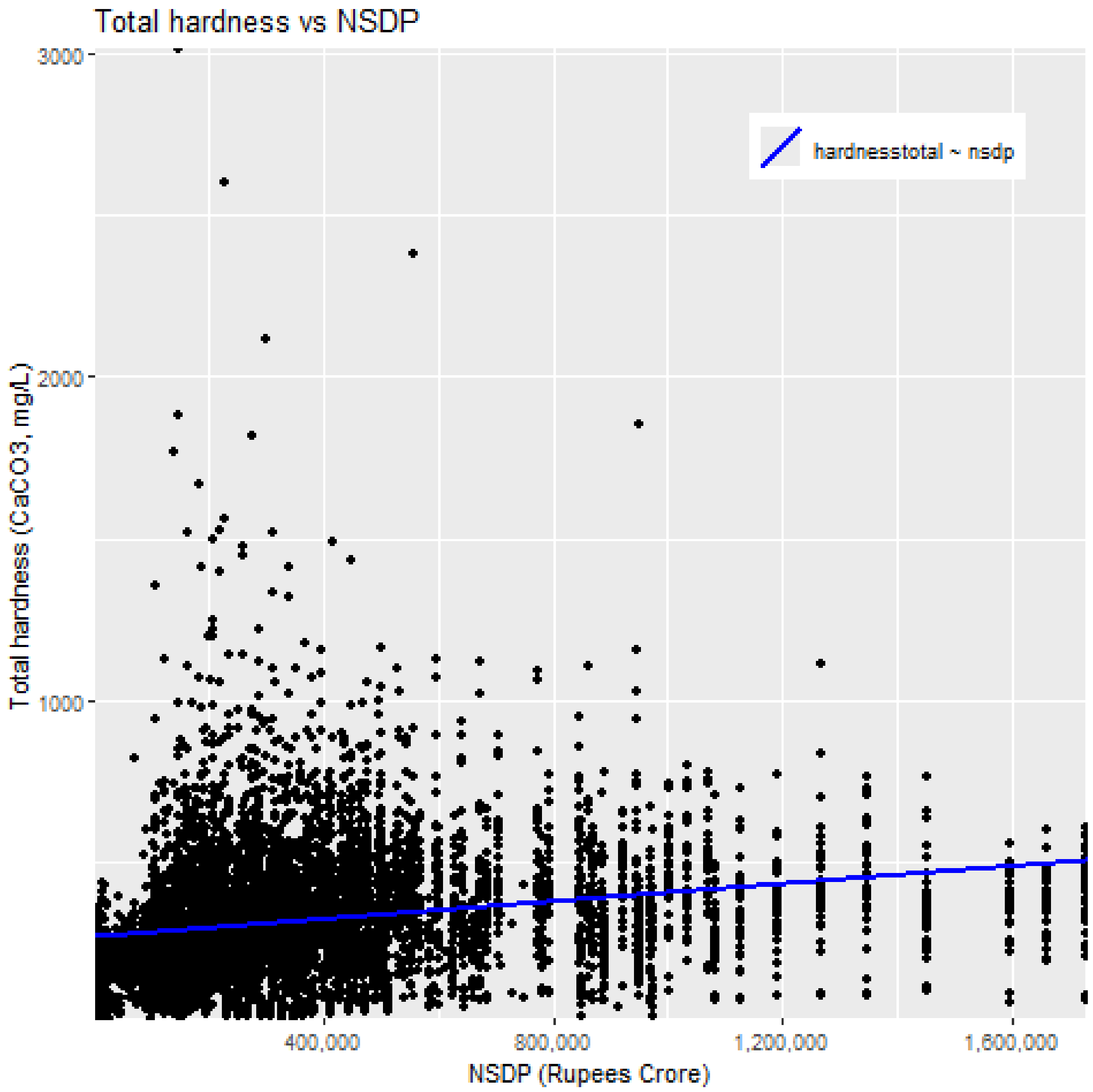
Variable Name	Mean	25th Percentile	50th Percentile	75th Percentile	95th Percentile
Total Hardness ($CaCO_3$, mg/L)	328.6217	195.9911	295.4653	417.2182	675.7037
Net State Domestic Product (Rupees Crore)	2,30,846.7	17,024.66	1,36,143.51	3,38,339.41	8,43,930.34
Gini Index	0.2668218	0.23	0.26	0.30	0.37

First model: GWQ ~ NSDP

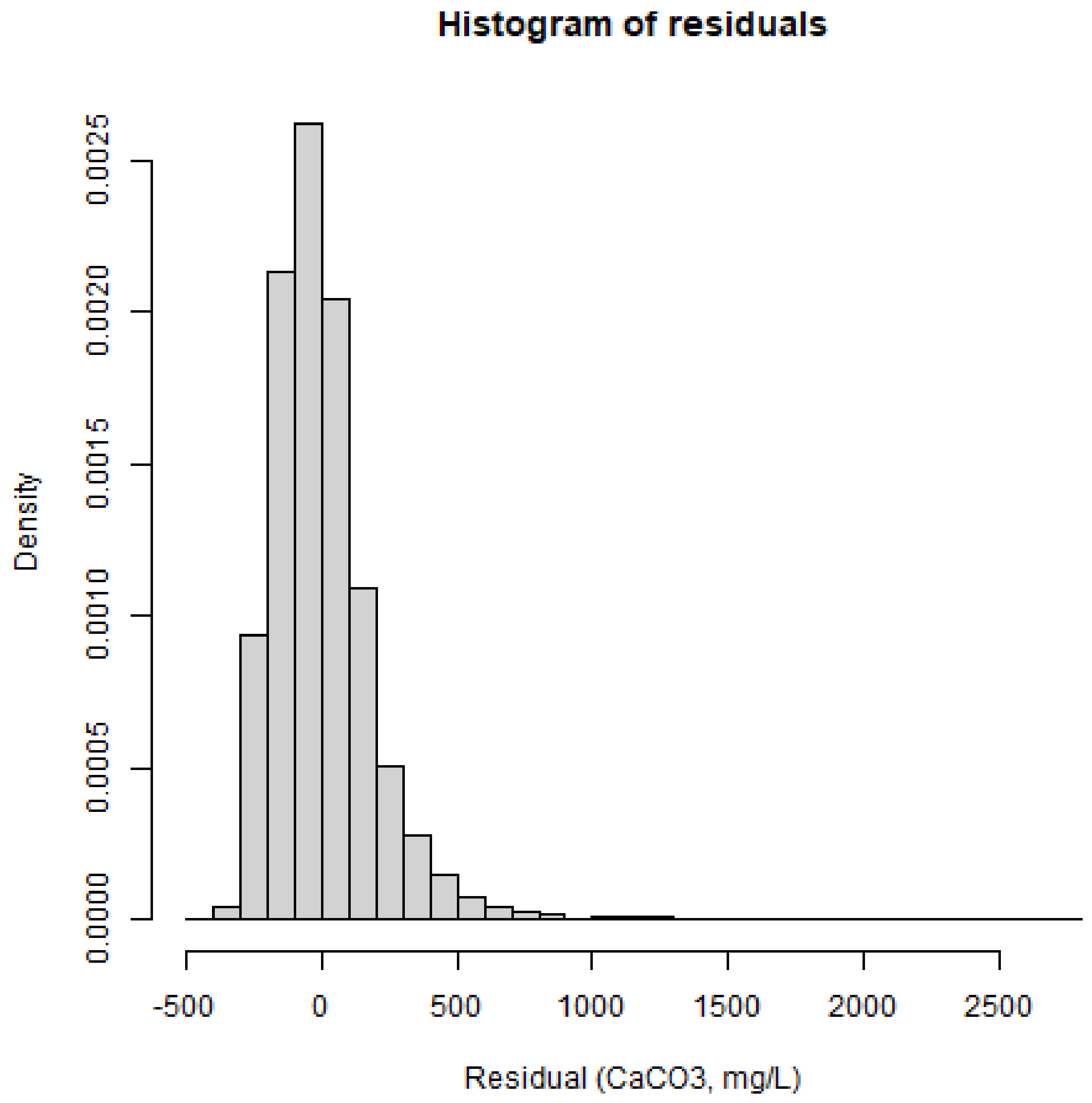
Model: $GWQ_{i,t} = \beta_0 + \beta_1 NSDP_{i,t} + u_{i,t}$

Variable	Coefficient (Standard Error)
Intercept	273.6 (3.647)
NSDP	0.00014 (0.000007346) (mg/L)/Crore 0.14 (0.007346) (μg/L)/Crore

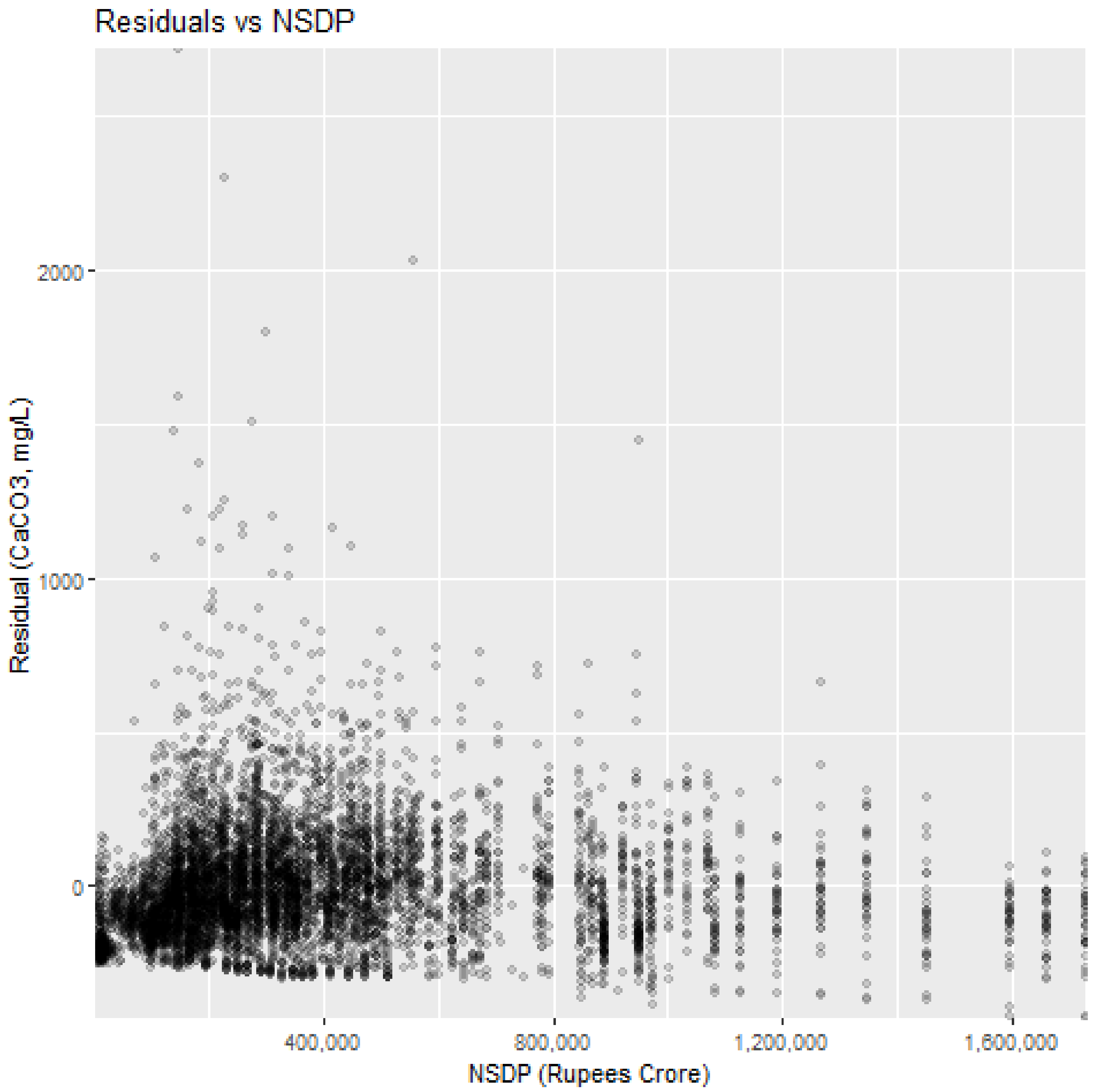
Sample size	7012	F-statistic	348.4
R^2	**0.04735**	p-value	< 2.2e-16



First model: GWQ ~ NSDP



Sum of residuals: -4.911849e-11 ≈ 0



First model: GWQ ~ NSDP

Some general notes:

1. Assumption: When a district splits, the new district(s) inequality is the same as in the old district. *Also, inequality did not change during the time period concerned.
2. Assumption: NSDP captures the effects of economic growth correctly. Districts in a state may have uniform economic growth, or people in a state may demand better conditions for all districts in the state, etc.
3. Assumption: Exactly zero hardness is a measurement error.
4. Assumption: Hardness measurements are MCAR. Seems to be mostly MAR, depending on the state.
5. The p-value is extremely small (i.e. confident) for all models tested. In passing, we note that the large sample size inherently leads to low p-values and limits the inference from hypothesis testing [3].

Results:

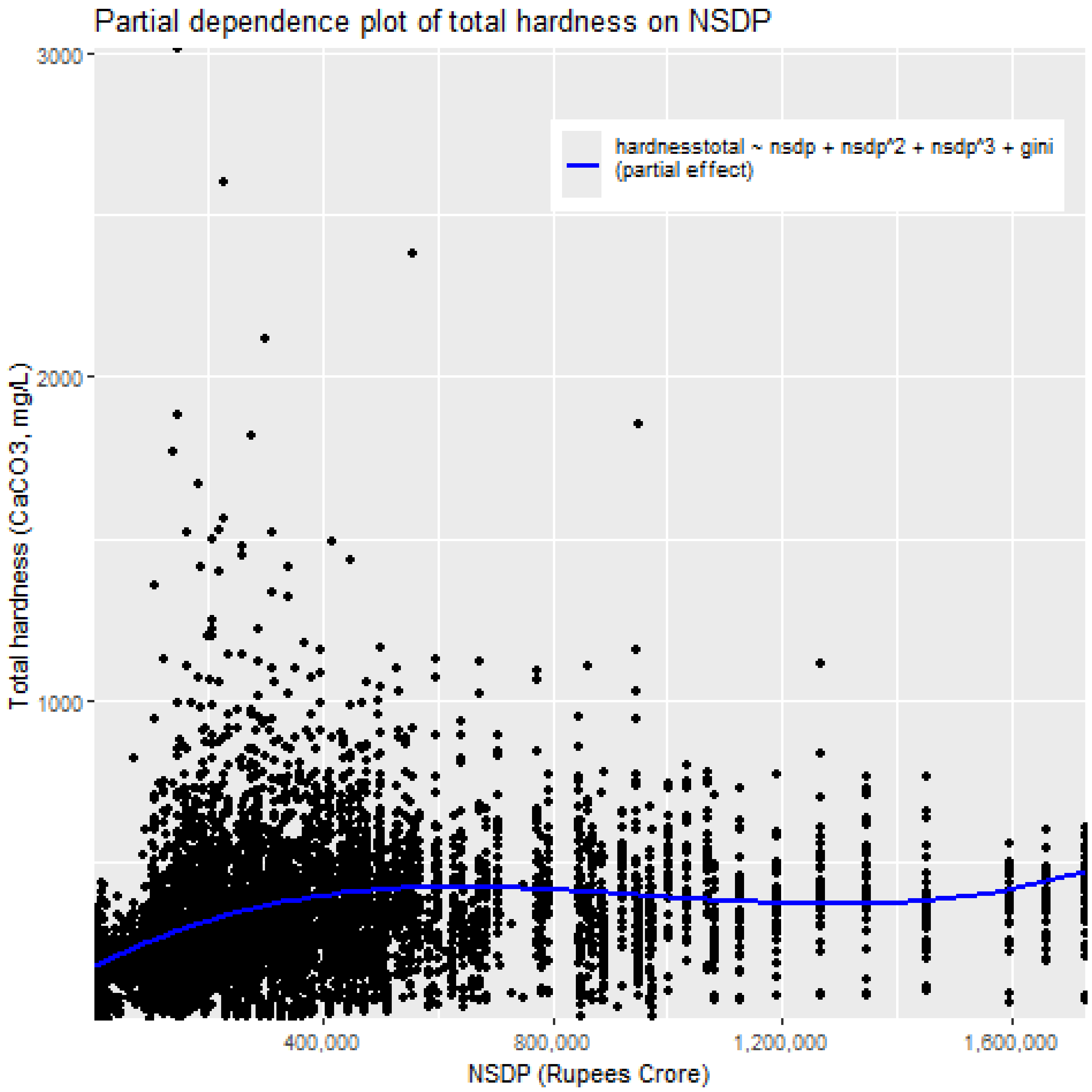
1. Economic growth and environmental pollution are positively correlated.
2. No indication of peaks due to linearity.

*Invoked in the following models to infer Gini indices from the 2010-11 estimates.

[3] Lin, Mingfeng, et al. "Research Commentary: Too Big to Fail: Large Samples and the p-Value Problem." Information Systems Research, vol. 24, no. 4, 2013, pp. 906–17. JSTOR, http://www.jstor.org/stable/24700283. Accessed 28 Apr. 2024.

Extended model: GWQ ~ NSDP + $NSDP^2$ + $NSDP^3$ + Gini

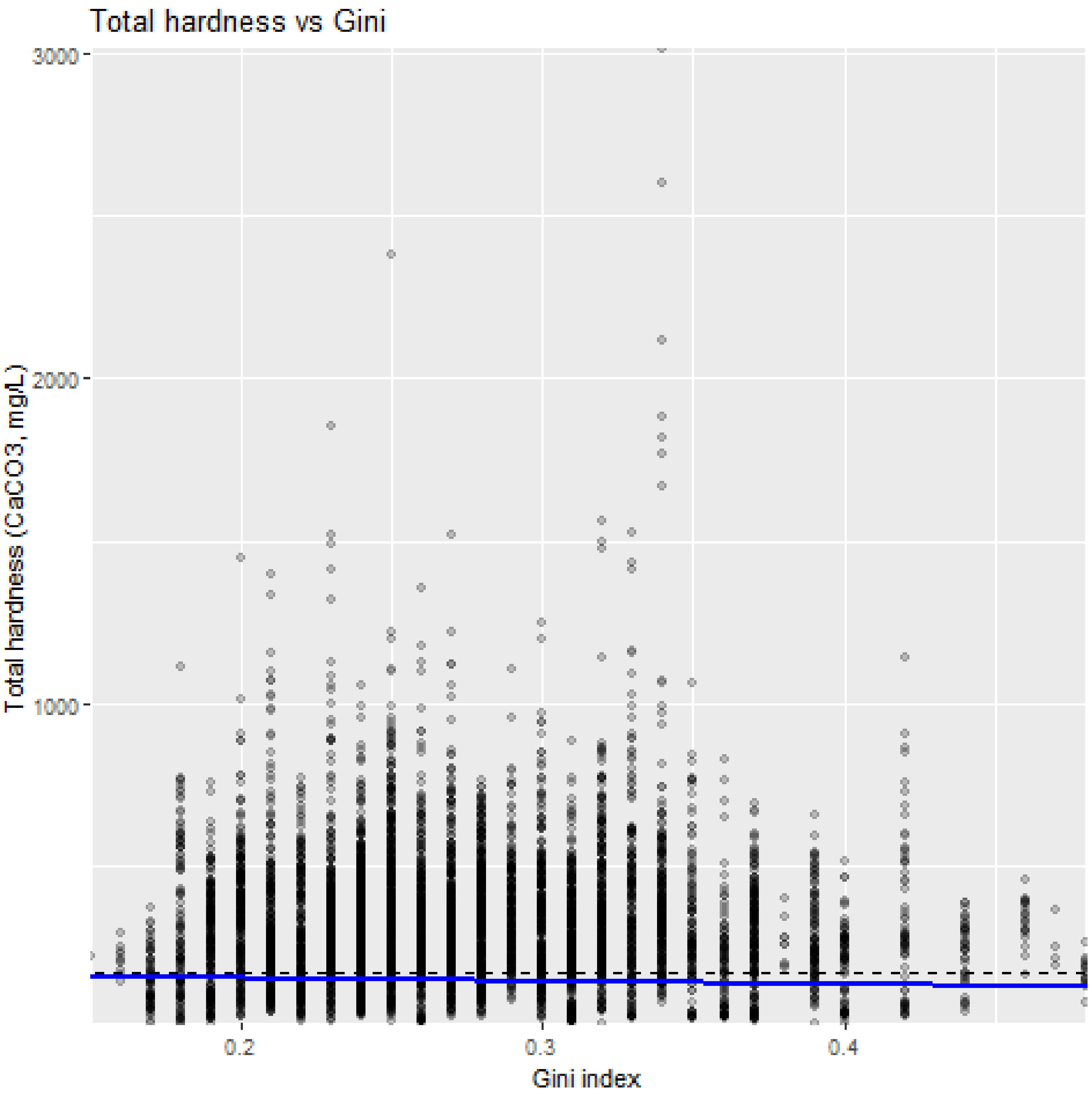
Variable	Coefficient (Standard Error)		
Intercept	177.7 (12.56)		
NSDP	0.00093 (0.00004489) (mg/L)/Crore 0.93 (0.04489) (µg/L)/Crore		
$NSDP^2$	-1.090e-09 (7.184e-11) (mg/L)/Crore		
$NSDP^3$	3.770e-16 (3.078e-17) (mg/L)/Crore		
Gini	-88.91 (39.78) (mg/L)		
Sample size	7012	F-statistic	190.5
R^2	**0.0981**	p-value	< 2.2e-16



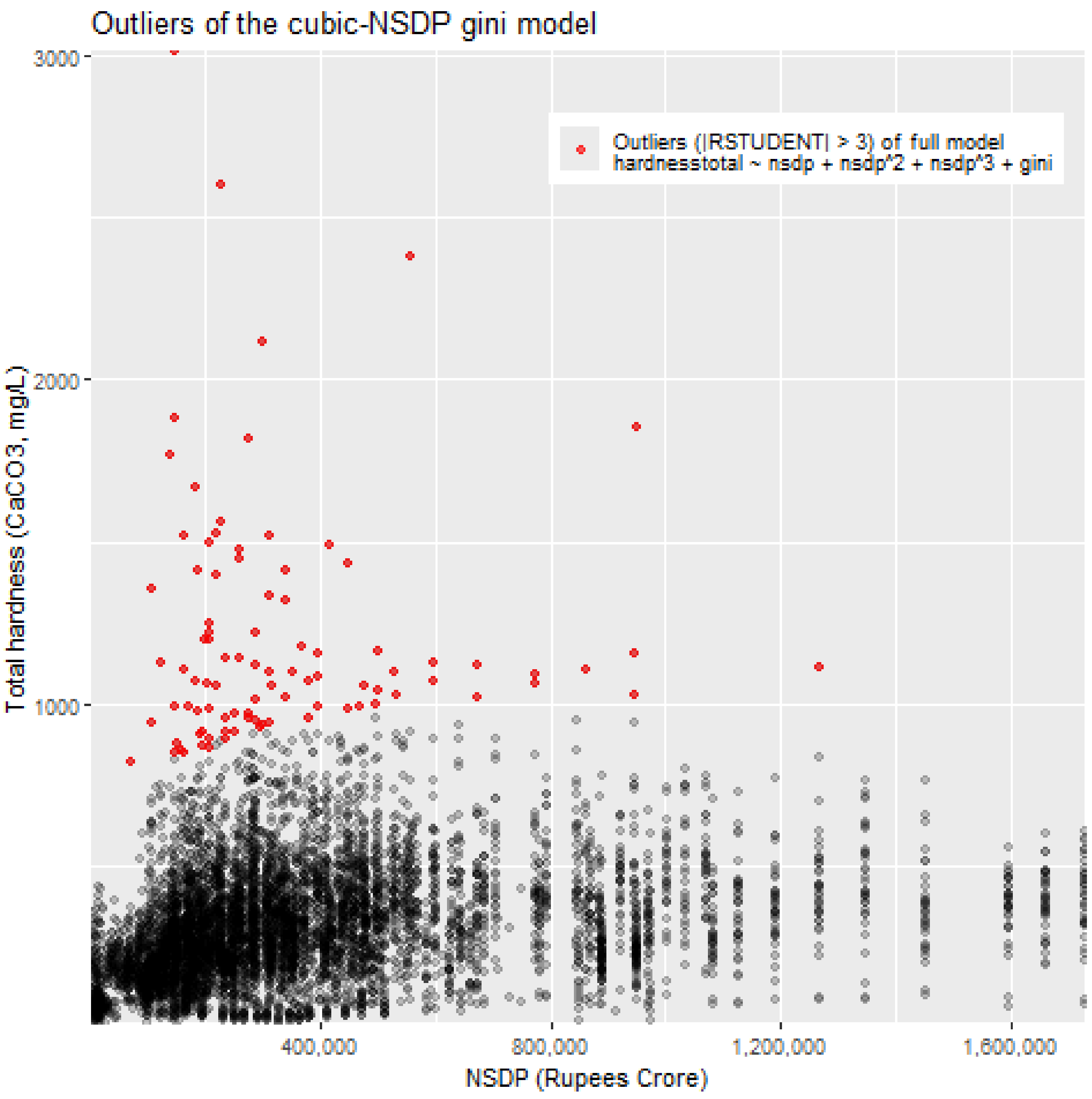
IV

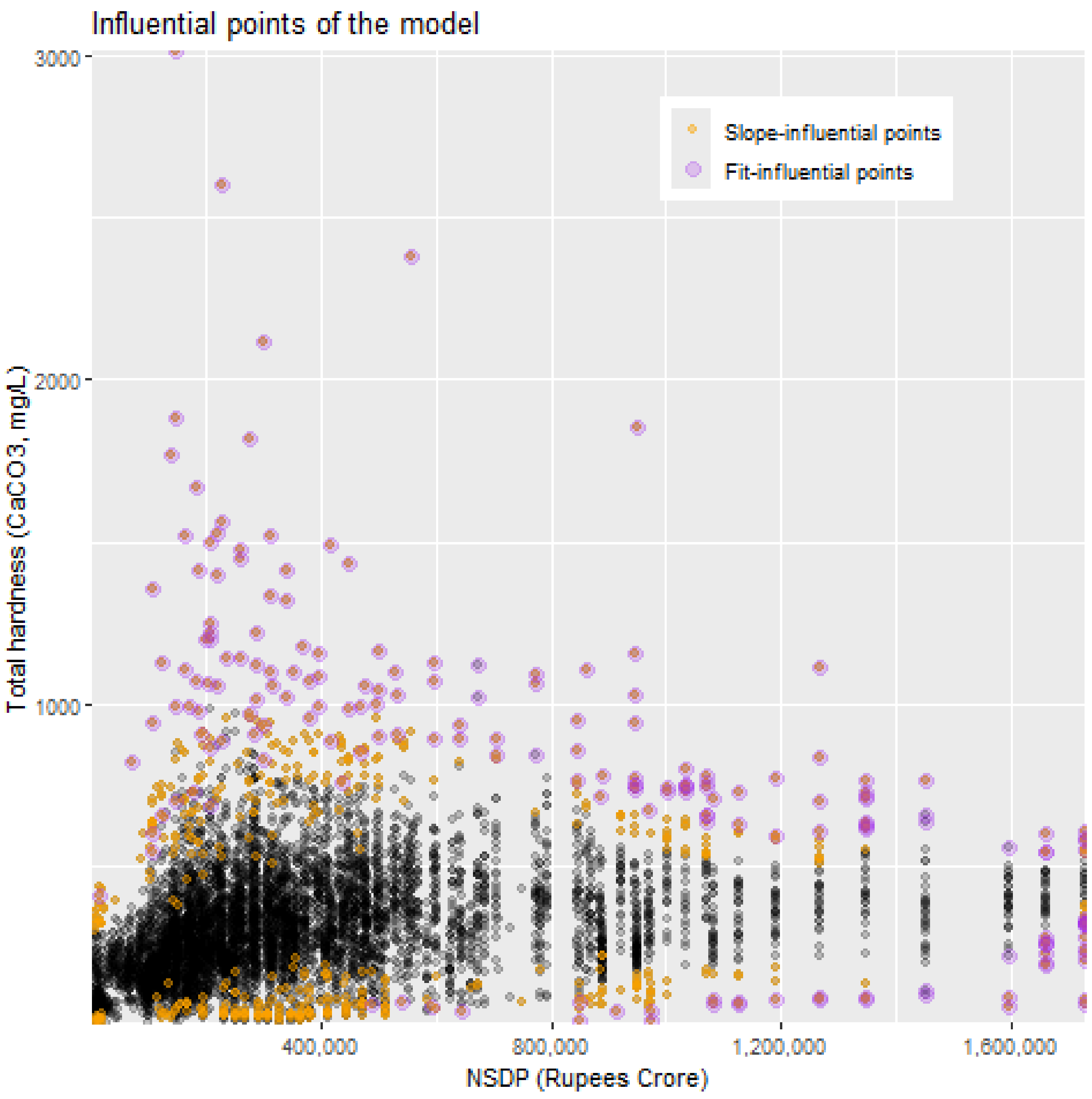
Extended model: GWQ ~ NSDP + $NSDP^2$ + $NSDP^3$ + Gini

Variable	Coefficient (Standard Error)		
Intercept	177.7 (12.56)		
NSDP	0.00093 (0.00004489) (mg/L)/Crore 0.93 (0.04489) (µg/L)/Crore		
$NSDP^2$	-1.090e-09 (7.184e-11) (mg/L)/Crore		
$NSDP^3$	3.770e-16 (3.078e-17) (mg/L)/Crore		
<u>Gini</u>	-88.91 (39.78) (mg/L)		
Sample size	7012	F-statistic	190.5
R^2	**0.0981**	p-value	< 2.2e-16



Extended model: GWQ ~ NSDP + NSDP² + NSDP³ + Gini





Extended model: GWQ ~ NSDP + $NSDP^2$ + $NSDP^3$ + Gini

Notes:

- Several districts have impermissible hardness, but some show abnormal measurement trends (possibly measurement errors).
- It should also be noted that as the NSDP increases, the number of outlier point observations decreases, signifying that as NSDP increases, the areas with very high hardness values decrease.

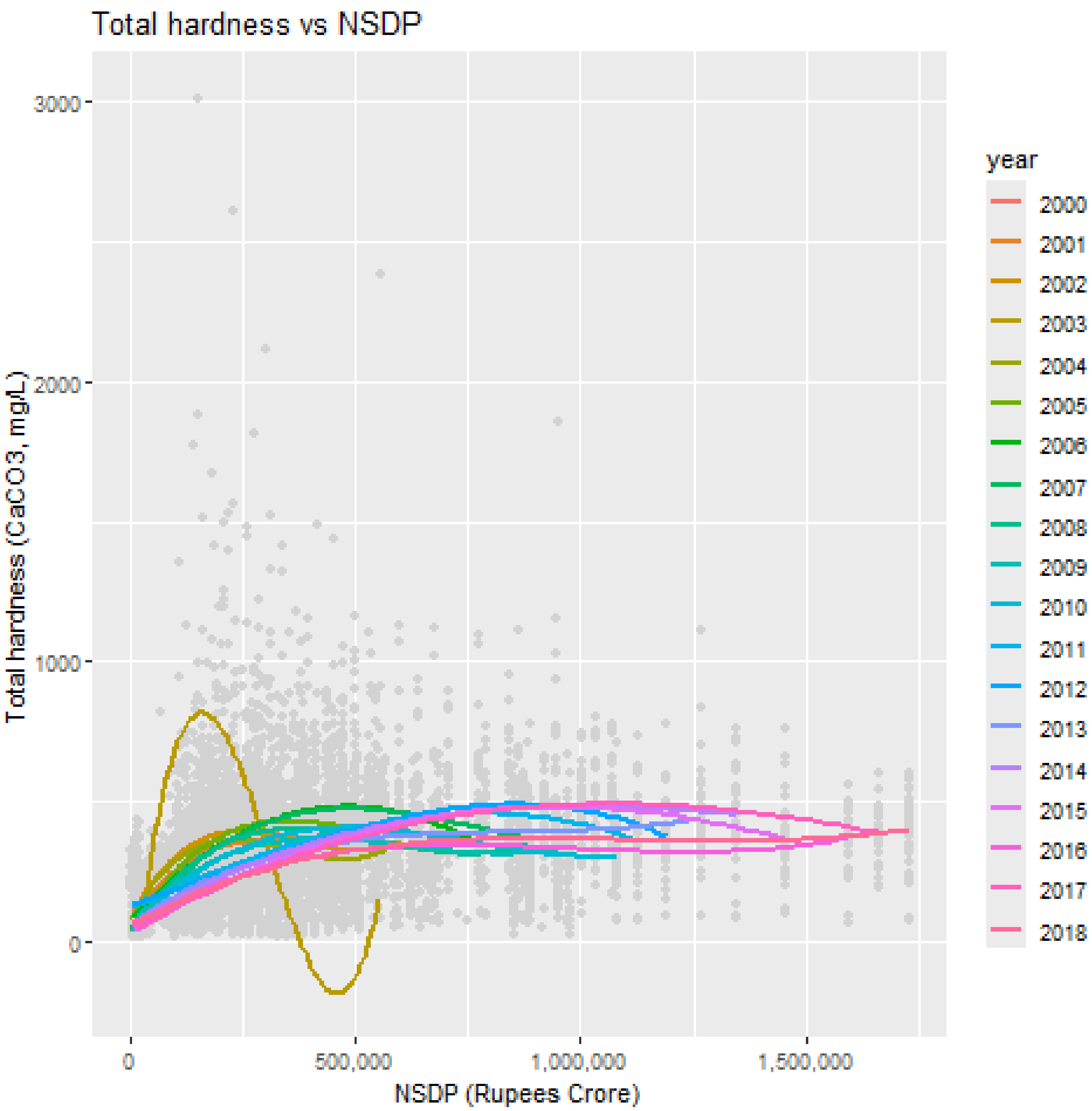
Extended model: GWQ ~ NSDP + $NSDP^2$ + $NSDP^3$ + Gini

Results:

1. Pollution increases till it reaches a peak at a certain amount of economic progress, after which it (barely) decreases for some amount of progress before seemingly ramping up again.
2. The quadratic NSDP model shows the expected EKC of a single turning point, a similar F-statistic but a lower R-squared.
3. The Gini index negatively correlates with environmental pollution, but the coefficient is insignificant. If introduced quadratically, the Gini coefficients become significant, and an initial increase is again followed by decreasing pollution with increasing inequality. A possible explanation is that the correlation between a high Gini index and a high NSDP [4] allows such districts, constituting those with economic power, to affect change in environmental pollution better.

[4] Mohanty, S. K., Govil, D., Chauhan, R. K., Kim, R., & Subramanian, S. V. (2016). Estimates of Poverty and Inequality in the Districts of India, 2011–2012. Journal of Development Policy and Practice, 1(2), 142-202. https://doi.org/10.1177/2455133316642338

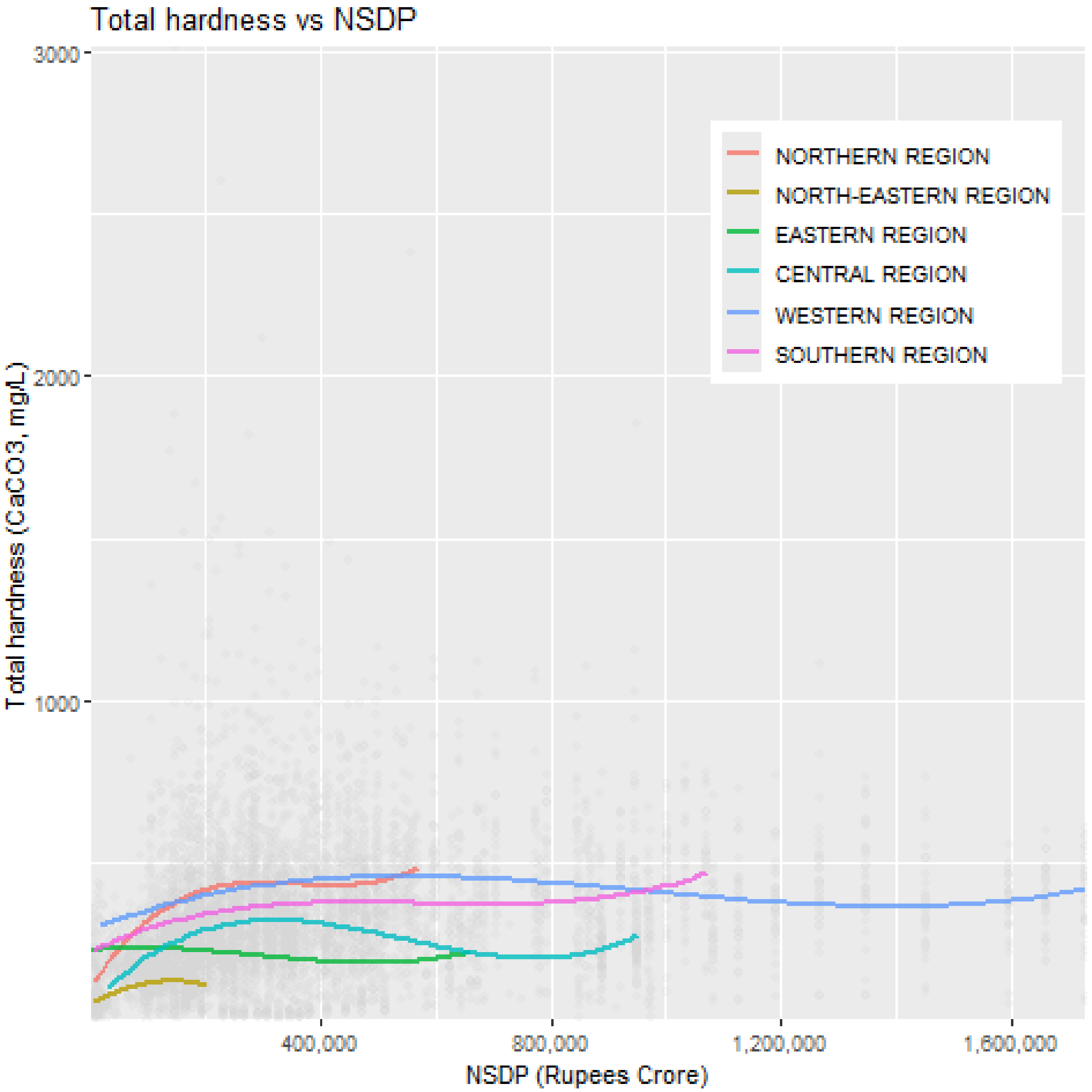
Extended model: GWQ ~ year·NSDP + year·$NSDP^2$ + year·$NSDP^3$ + year·Gini



Notes

1. 2003 has very few data points, thus the odd shape.
2. Several coefficients are insignificant, but the model R^2 of 0.1925 has a p-value $< 2.2e-16$ with a sample size of 7012.
3. As the year increases, the peaks have shifted to the right, but the general trend remains the same.

Model: GWQ ~ region·NSDP + region·$NSDP^2$ + region·$NSDP^3$ + region·Gini

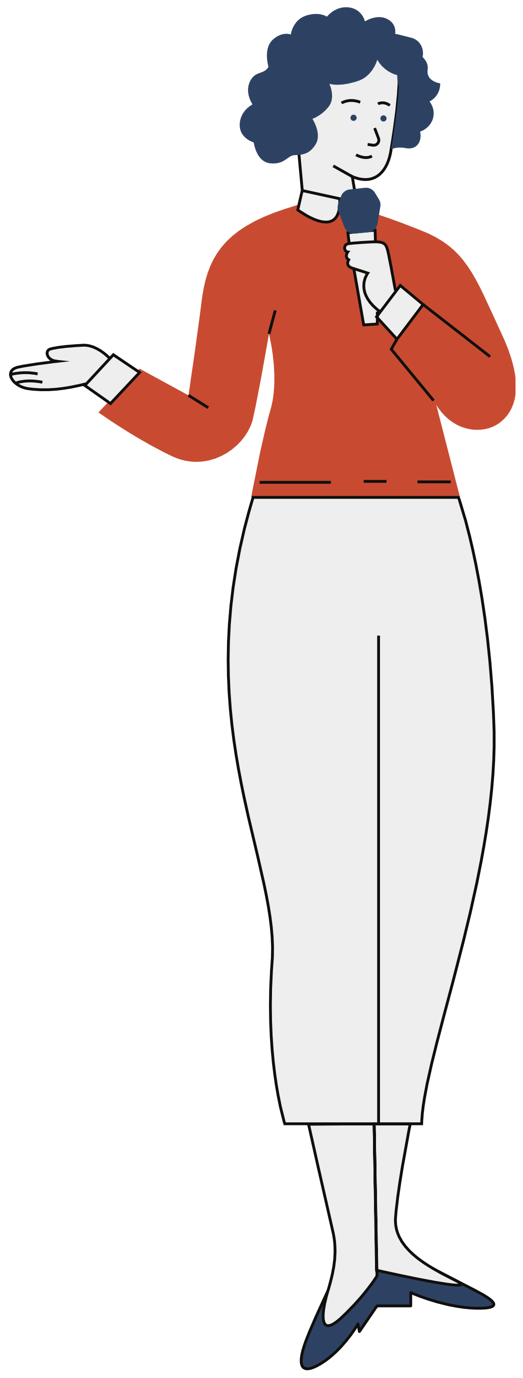


Notes

1. Most coefficients are significant, and the R-squared of 0.2685 is also significant with 7012 observations.
2. The differences in average NSDP and average pollution are evident; the NE, E, and central regions seem to have less pollution.
3. Regions with higher economic growth have more pollution.
4. Several of the Gini coefficients are insignificant and have confidence intervals including zero, but at first glance, NE, E and central regions have positive Gini coefficients.
5. The southern region has the most negative Gini coefficient, possibly indicating that people affect change most in the South.
6. Economic progress most dramatically affects pollution in the central and northern regions. Industrial?
7. Telangana has been assigned the appropriate region ourselves.

Limitations

1. Data was missing for certain years for certain districts so there may be unequal representation.
2. The aggregation level of the data (e.g., district-level for ground-water quality and state-level for economic output) might not be optimal for capturing local variations. This could potentially obscure important relationships or introduce ecological fallacy.
3. The analysis might not account for all relevant factors that could influence hardness of water quality and economic output simultaneously.
4.The relationship between total hardness of water and NSDP could be bidirectional, with changes in economic activities influencing groundwater quality and vice versa. Failure to address endogeneity or reverse causality in the analysis could lead to biased estimates of the relationship.



The End

Thank you for listening