

Project:

You are provided the data for multiple indicators of groundwater quality (GWQ) measured in milligrams per liter for all Indian districts from 2000 – 2019 [here](#). As a first step, please check the group assignment from the ‘Group Allocation’ sheet. Each group is assigned one indicator, please check the indicator assigned to your group in the ‘Dependent Variable Assignment’ sheet. Please merge the district-year level groundwater quality data with the corresponding state-year wise economic output data, i.e., the net state domestic product (SDP) at constant prices (shared [here](#)) provided by the Reserve Bank of India accessed on the [Database for the Indian Economy \(DBIE\)](#) portal. Finally, merge your dataset with the district-level Gini index from the following paper by Mohanty et al. (2016). PDF linked [here](#).

- 1) Now, using this data, estimate the following regression, summarize the results in a table and interpret. Note that i indexes districts, t indexes years, and $u_{i,t}$ is random error.

$$GWQ_{i,t} = \beta_0 + \beta_1 SDP_{i,t} + u_{i,t}$$

- 2) Visualize the model residuals (i.e., $\hat{u}_{i,t}$) on a plot having the groundwater quality indicator on Y-axis and SDP on the X-axis. Now, construct a second plot having $\hat{u}_{i,t}$ on Y-axis and SDP on x-axis. Are the plots, what you would expect? Explain.
- 3) Plot a histogram of $\hat{u}_{i,t}$ and verify that $\sum_{i,t} \hat{u}_{i,t} = 0$.
- 4) Existing empirical evidence with regards to the environmental Kuznets curve postulates a non-linear relationship between environmental quality and economic growth. Enhance the model discussed in question 1 to reflect this understanding. Estimate the regression and summarize the results in a table. Additionally, prepare a detailed summary statistics table for all the variables. Are there any outliers and/or influential observations? If so, how will you tackle them?
- 5) Articulate the relationship between economic growth (as measured by SDP) and groundwater quality as predicted from the regression above. Are the results what you would expect?
- 6) Do the results for the previous question differ by year?
- 7) Enhance the model in question 4 to examine if there are differences in the estimates of the Kuznets curve across regions within India. The definition of the states belonging to each region are given by RBI – linked [here](#).