# How Do Pollution, Rainfall and Location Impact Extreme Weather Conditions for Cities in China?

Table 1: The Determinants of Temperature Index in China

| Explanatory Variables: | Dependent Variable: Log Temperature Index | | |
|---|---|---|---|
| | Regression #1 | Regression #2 | Regression #3 |
| Constant | -1.163*** (0.24) | 4.295*** (0.191) | -1.16*** (0.238) |
| Log Air Pollution | -0.065*** (0.022) | 0.197*** (0.026) | -0.065*** (0.021) |
| Log Rainfall | -0.02 (0.016) | -0.355*** (0.015) | -0.020 (0.016) |
| Log Latitude | 1.28*** (0.046) | - | 1.28*** (0.045) |
| Time Trend | - | - | -0.001 (0.002) |
| Shenzen | - | 0.71*** | - |
| Guangzhou | - | -0.097 | - |
| Tianjin | - | -0.062 | - |
| Shanghai | - | 0.12* | - |
| Year Dummies | Yes | Yes | No |
| Observations | 846 | 846 | 846 |
| $R^2$ | 0.725 | 0.527 | 0.725 |

*Note:* Each column reports results from a separate OLS regression. Standard errors are in parentheses for non-dummy variables. Reference category for regression#2 is Beijing. Four cities are missing PM10 data in 2003. * Estimated coefficient is statistically different from 0 at the 10% level, ** significant at 5%, and *** significant at 1% or better.

Global Warming has had a significant impact on weather conditions across the world. Our table displays three regression models using data from China's *Statistical Yearbooks* and *City Statistical Yearbooks.* From our first regression, we find overwhelming evidence to prove that a 10% increase in latitudes, i.e., cities located further north from the equator, on average, observe a 12.8% increase in colder temperatures and hotter summers after controlling for rainfall, air pollution, and differences across time. Surprisingly, in China, after controlling for rainfall, locations, and differences across time, we find strong evidence that a 10% increase in air pollution is on average associated with a 0.7% decrease in extreme climate conditions. We know air pollution to be a generally large contributor to global warming, however, we might be seeing this decrease due to other factors, such as cloud coverage, i.e., lurking variables, having a stronger impact on the temperature index for cities in China instead. Further, we find no evidence to prove a relationship between rainfall and the temperature index in Chinese cities after controlling for air pollution, location and differences across time.

We focus on five cities in our second regression and find that when we control for the different cities, there's overwhelming evidence for a 2% increase on average in the temperature index associated with a 10% increase in air pollution, while controlling for rainfall and differences across time and the five cities. This confirms our suspicion about lurking variables. Further, we see strong evidence for an association between a 10% increase in rainfall and a 4% decrease on average, in the temperature index, holding air pollution, difference across times and the five cities constant. Further, there is overwhelming evidence that Shenzen experiences a temperature index a vast 71% higher than Beijing, on average, holding rainfall, air pollution and differences across time constant. This makes sense as Shenzen is widely known for experiencing heatwaves, along with heavy rainfall during the summer months. Similarly, we see weak evidence of Shanghai experiencing on average, a 12% higher temperature index than Beijing after controlling for air pollution, rainfall and differences across time. In our third regression, we include a time trend, but do not find any significant evidence to prove time having any effect on the temperature index in cities.

Replication Steps:

1. Open **pol_chn.xlsx**.
2. Select all the data and Filter out any missing observation (blanks) from the following variables: *temp_index, pm10, rainfall, latitude, year, shanghai, beijing, tianjin, guangzhou and shenzen*.
3. Copy and paste these filtered variables into a new worksheet called **pol_data.**
4. Create a new variable called *ln_pm10* and Calculate the ln (e.g., **=ln(B2)**) of the *pm10* variable for the first observation.
5. Double click the bottom right corner of the first cell under *ln_pm10* to fill the entire column.
6. Repeat steps **(5.)** and **(6.)** for the variables *temp_index, rainfall* and *latitude* and name them *ln_temp_index, ln_rainfall* and *ln_latitude*, respectively.
7. Create a new variable (empty) called *time_trend* in the worksheet **pol_data**.
8. Select the entire data and Filter the **pol_data** worksheet for the year 2003 (the lowest year) only.
9. Put the value '1' in the first cell under the *time_trend* variable.
10. Double click the bottom right corner of the first cell under the *time_trend* variable to fill in the value '1' for all the rows with 2003 data.
11. Then Remove the Filter and Re-filter for the next year (e.g., in this case the next year is 2004).
12. Repeat steps (**9).** to (**12.)** for all the years (2003 - 2012), increasing the input value from '1' to '10', as the year we filter our data for increases (i.e., from '1' for 2003 to '10' for 2012).
13. Using the same concept as steps (9.) to (13.), create 9 new variables for each year (2004-2012, we omit 2003 and use it as our reference category for our time fixed effect) and using the Filter option, filter through the years, adding the value '1' for the filtered year and '0' for all the other years in the new dummy variables we created.
14. From the **pol_data** worksheet, Copy the variables *ln_temp_index, ln_pm10, ln_rainfall, ln_latitude* and our *time fixed effects* and Paste the Values (**using Paste Special > Values**) into a new work sheet and name it **Reg 1**.
15. Create a Multiple Regression with the variable *ln_temp_index* as the y-variable (dependent variable) and the variables *ln_pm10, ln_rainfall, ln_latitude* and *time fixed effects* as the x-variables (predictor variables) using the Regression option in the Data Analysis Tool-pack (ensure all x-variables are placed next to each other for every regression).
16. Then, from the **pol_data** worksheet, copy the variables *ln_temp_index, ln_pm10, ln_rainfall*, and the dummy variables *shanghai, tianjin, guangzhou, shenzen*(we use *beijing* as our reference category), and *time fixed effects* and Paste the Values (**using Paste Special > Values**) into a new work sheet and name it **Reg 2**.
17. Create a Multiple Regression with the variable *ln_temp_index* as the dependent variable and the variables *ln_pm10, ln_rainfall, fixed time effects, shanghai, tianjin, guangzhou and shenzen* as the predictor variables (we don't include *ln_latitude* due to Independence Condition issues) using the Regression option in the Data Analysis Tool-pack.
18. Then, from the **pol_data** worksheet, copy the variables *ln_temp_index, ln_pm10, ln_rainfall, ln_latitude and time_trend* and Paste the Values (**using Paste Special > Values**) into a new work sheet and name it **Reg 3**.
19. Create a Multiple Regression with the variable *ln_temp_index* as the y-variable (dependent variable) and the variables *ln_pm10, ln_rainfall, ln_latitude and time_trend* as the x-variables (predictor variables) using the Regression option in the Data Analysis Tool-pack.

Tables for Inspiration:

1. April 2022 Exam, Questions (4): "Women's Downhill Skiing Final: 2022 Beijing Olympics"
2. April 2022 Exam, Questions (6): "Table 4. OLS regression results, with dummy variables for blue and red states" from Bachmann et al. (2021)
3. April 2017 Exam, Question (3): "Table 1: Estimates of Learning by Doing" from Levitt et al. (2013)