

# Analysis of Relationship between AQI Measurement and Car Sales Growth Rate

Gang Wang<sup>1,a, †</sup>, Shuo Wang<sup>2, b, †</sup>

<sup>1</sup>University of Birmingham Birmingham, The United Kingdom

<sup>2</sup>University of Toronto Toronto, Canada

<sup>a</sup>GXW991@student.bham.ac.uk, <sup>b</sup>twinkle.wang@mail.utoronto.ca

†These authors contributed equally

**Abstract.** The development of industry has had a great impact on the natural environment, especially on air, which is important for human survival. The introduction of carbon neutral policies in recent years has shown that mankind is taking air quality seriously again. So far, because transportation has benefited greatly from the upgrade of the industrial revolution, cars are often considered the primary source of emissions. Therefore, this paper focuses on the relationship between the outcome growth rate of car sales and three factors in Air Quality Index (AQI) measurement, ground-level ozone, average annual PM2.5 level and AQI scores of nitrogen dioxide in seven major regions/countries. First, data are visualized to observe the potential patterns between these three factors affecting air quality and the growth rate of car sales. Then, using statistical regression, the separate relationship between them was modeled and the estimates of parameters in the model were calculated through the computer software. Finally, the model is evaluated by analyzing the statistical significance of these parameters and utilizing the model selection criteria for the proposed models. As a result, a simple linear relationship is found between ground-level ozone concentration and car sales growth rate. An approximately cubic function could be simulated the correlation between the PM2.5 level and the negative car sales growth rate, while the regions/countries with positive car sales growth rate remain undetermined with the given data. Besides, scores of nitrogen dioxide may have no obvious pattern with respect to car sales growth rate in those regions/countries.

**Keywords:** Air Quality Index; Sales Growth Rate; Particle Pollution; Ground-level Ozone; Nitrogen Oxide.

## 1. Introduction

The World Health Organization (WHO) estimates that air pollution causes close to seven million deaths worldwide each year. Currently, nine out of ten people breathe air that contains more contaminants than the WHO's recommended levels, with those in low- and middle-income nations suffering the most [1]. It has long been known that respiratory illnesses like asthma and chronic pulmonary obstructive disease are made worse by air pollution. The once-surprising finding that air pollution is linked to cardiovascular disease is now widely accepted in the twenty-first century [2].

As people's standard of living and economic ability has increased in recent years, people tend to have their private transportation tools to facilitate their daily commuting. Studies have shown that recent vehicle traffic exhaust emissions potentially contribute to negatively impacting human health and contribute to the formation of ground-level ozone, particle pollution, carbon monoxide, sulfur dioxide [3], etc. Other research from the United Kingdom [4] revealed that residents living next to certain streets that are heavily exposed to traffic-related air pollutants experienced a significant reduction in rhinitis and rhinoconjunctivitis after reducing their exposure to traffic-related air pollutants. Although the scope of this study is not universal and may have regional limitations, it is reasonable to assume that variables related to vehicles may be related to the quality of air based on the reduction of transportation exhaust and the improvement of a particular health indicator of the population shown in this research. The reason is that these variables may affect the air quality index by changing the level of vehicle exhaust in the near-ground air.

In today's world where various industries of transportation are growing, if a more specific bridge can be identified between the transportation-related variable and factors affecting air quality, it will be likely to provide appropriate suggestions for environmental management in the near future.

In this paper, we use linear regression and polynomial regression to explore the relationship between ground-level ozone concentration, average annual PM2.5 level, average AQI scores of nitrogen oxides and car sales growth rate in seven major regions/countries. We selected one of the transportation-related variables which is the growth rate in vehicle sales from 2019 to 2022 and extended the regional research scope to seven main regions/countries around the world, primarily aiming to find out the more specific link between the traffic-related variables and ground-level ozone concentration, average annual PM2.5 level and AQI scores of nitrogen oxides, which are three major factors in AQI measurement. The primary mathematical and statistical methods utilized in this research are to model the pattern of the sample points involved. We applied statistical regression models to measure the performance of the models by hypothesis test and Pearson-Wilke test. The simple linear regression model could be used to model the ground-level ozone concentration and the PM2.5 has a cubic relation with car sales growth rate. However, AQI scores of nitrogen oxides show no patterns with the growth rate of car sales.

## 2. Method

In this section, the process of data collection is introduced first (Sec. A), and we conducted the data visualization based on three factors respectively (Sec. B). After attaining a basic understanding of the given variables, i.e., ground-level ozone concentration, average annual PM2.5 level, and the average AQI scores of nitrogen oxides, we proposed different kinds of models to simulate the sample data for these three factors separately and identified several statistical methods to exam the goodness of fit of the models (Sec. C).

### 2.1. Data Collection

To gather the data on three factors being responsible for quantifying the AQI index, we found suitable datasets ranging from 2019 to 2022 on three different websites. For factors recording the average ozone concentration, it is known that the ground-level ozone is the specific kind of ozone that contributes to air pollution, so we found the average seasonal 8-hour maximum ground-level ozone [5] for this analysis. And the data indicating the average annual PM2.5 level could be found on a website designing for recording the PM2.5 level in the most polluted countries around the world [6].

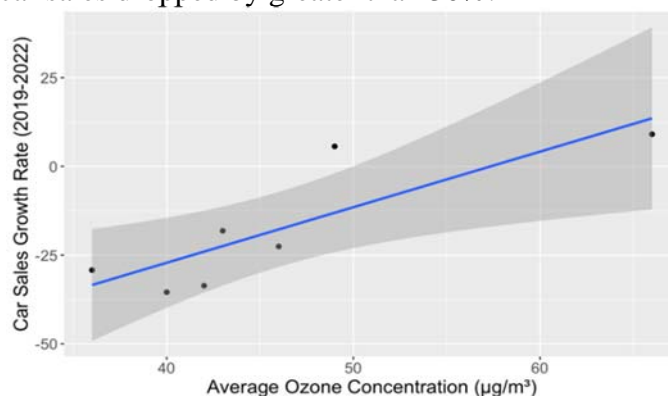
Data on nitrogen oxide are from CWOP/APRS (The Citizen Weather Observer Program). This data source uses the AQI index standard to measure pollution. The maximum score is 100 and the minimum score is 1. Higher scores mean more pollution. In this experiment, we collected the average data of seven countries/regions, among which Europe, China and the United States collected the data of each capital for average calculation. Brazil, India, Japan and Russia respectively calculated the average AQI scores of nitrogen oxides in 15 to 25 cities with high populations and high economic volume. In addition, we found that a few cities scored more than 30 AQI scores.

### 2.2. Data Visualization

#### 2.2.1. Ozone Concentration

In terms of the overall situation, the average ozone concentrations of seven countries/regions are all within the 70ppb threshold stipulated by the National Ambient Air Quality Standards (NAAQS) for ground-level ozone [7], a standard based on substantial scientific proof of the negative impacts of ozone on the human health condition. It implies that these countries/regions are not exposed to the ozone environment that is harmful to human being. For car sales growth rate, annual car sales in China and India increase from 2019 to 2022, while five other countries/regions have negative growth rates.

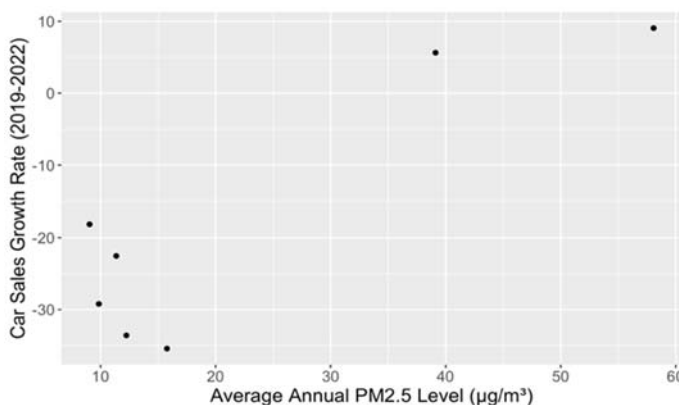
Specifically, the car sales market in Europe and Brazil seems to encounter a recession during the given period since their car sales dropped by greater than 30%.



**Figure 1.** Average Ozone Concentration VS. Car Sales Growth

Fig.1 presents the rough observation of the relationship between average seasonal ozone concentration and the outcome variable recording car sales growth rate from 2019 to 2022. It shows a potential trend that countries/regions with a higher level of ozone concentration are likely to attain higher car sales growth rates. Based on this interpretation, an assumption could be made that six countries other than Europe tend to have a positive correlation between car sales growth rate and average ozone concentration.

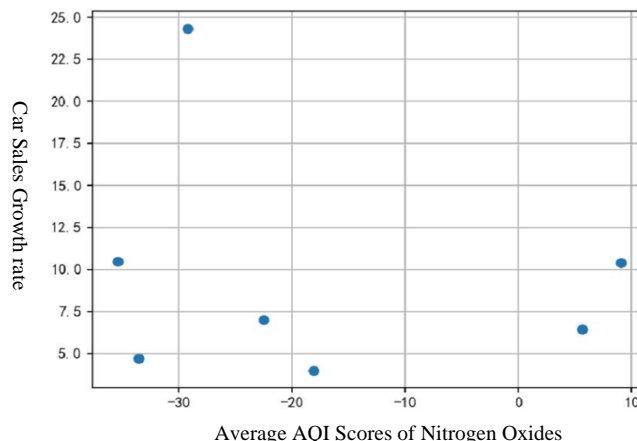
### 2.2.2. PM2.5 Level



**Figure 2.** Verage Annual PM2.5 VS. Car Sales Growth Rate

For the relationship between the average annual PM2.5 and the car sales growth rate, situations might be separated in terms of the positive and negative values of the growth rate. As is shown in Fig. 2, only two countries, China and India, have relatively high PM2.5 concentrations, whereas five countries/regions with decreasing car sales from 2019 to 2022 all get ozone concentrations less than 20 micrograms per cubic meter.

For five countries/regions with non-positive growth rates, three countries meet the primary health standard for particle exposure defined by the Environmental Protection Agency (EPA), which is the annual concentration of PM2.5 should be less than 12 micrograms per cubic meter ( $\mu\text{g}/\text{m}^3$ ) [8]. Based on this fine particle standard, a conclusion could be made that four out of seven countries/regions tend to suffer from the deleterious influence of long-term PM2.5 concentration. Also, a negative trend between the growth rate and average annual PM2.5 concentration is exhibited among the countries/regions with negative growth rates. In other words, countries/regions with the more obvious drop in car sales amount may link to a larger concentration of PM2.5 but within an appropriate interval that fails to attain dramatically harmful effects on human beings. For China and India whose car sales growth rates are positive from 2019 to 2022, they both far exceed the threshold of the EPA standards, implying a serious problem of PM2.5 particle pollution currently.



**Figure 3.** Average AQI scores of nitrogen oxides VS. Car Sales Growth Rate

It can be observed from the Fig. 3 that only one point is distributed outside a banded area and the upper limit of banded area is 12.5 AQI scores for nitrogen oxides.

Even in the ribbon region, we can observe that not only countries with positive growth rate have relatively high AQI scores for nitrogen oxides, but also some countries with negative growth rate still have relatively high AQI scores for nitrogen oxides. Therefore, through the display of Fig. 3, it can be determined that for countries with high AQI scores, reducing the sales growth rate of cars will not help them reduce nitrogen oxide pollution in a short time.

### 2.3. Model Construction

**Ozone Concentration:** To verify the potential positive relationship between ozone concentration and our outcome variable, a simple linear regression model is likely to have high-level goodness of fit since the sample points in the above output tend to follow an upward linear trend. In the model analysis process, the selection techniques and criteria, P-values and hypothesis testing, and Coefficient of determination, are included to evaluate the performance of the model.

To evaluate the performance of the model, hypothesis testing is conducted on both intercepts. The "Null Hypothesis" is that the parameter estimate is different from zero. On the contrary, the alternative hypothesis is the estimate is exactly zero. In this way, both of the parameter estimates on the presented model are highly statistically significant.

Besides, the "coefficient of determination"[9], known as the  $R^2$  values, is included in our model assessment procedure. Note that the  $R^2$  value in our hypothetical model is approximately 0.70, which indicates that a 70% variation of the outcome variable, car sales growth rate in this context, could be explained by the above-selected model.

**PM2.5 Level:** It is obvious that the simple linear regression model does not fit the data well, so it is no longer proper in this case to simulate the sample data with simple linear regression. Therefore, the simple linear regression related to variable x is raised to a degree from two to five respectively to select the best model that fits the sample point. Then, the variances captured by each model of different degrees are examined to determine the best model. To be specific, if the model with the greatest degrees four is chosen, the standard form of the relationship between PM2.5 and car sales growth rate is:

$$y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 \tag{1}$$

Where y stands for the outcome variable, car sales growth rate and x is our variable of interest, which is PM2.5 concentration. And  $\beta_0, \beta_1, \dots, \beta_4$  serve as the parameters that should be measured in the model.

**Nitrogen Oxides:** Fig. 3 was drawn by the obtained AQI score data to determine whether there was a correlation. It can be found that the trend in AQI scores for nitrogen oxides is not visible as sales increase.

To find the correlation and since our data belong to small sample size, we first use Pearson-Wilke [10] with an Alpha of 0.05 to verify whether the data conform to the normal distribution. It can be concluded that the growth rate of automobile sales conforms to normal distribution. Then we calculate the explicitness of the Pearson coefficient and Pearson correlation coefficient to determine whether there is a linear correlation and the test of explicitness. In addition, the OLS regression model is tried. The parameters of OLS show that this may be a primary function with a maximum constant, the overall slope of the function is -0.0731. Therefore, we deleted the isolated point of Russia to detect the processed scatter diagram and found that the slope decreased to 0.0283 at this time, meanwhile the equation is below:

$$y = 0.0283x + 7.6181, \quad (2)$$

Where  $y$  represents the growth of automobile sales, and  $X$  is the score of AQI index of nitrogen oxides.

### 3. Results and Discussion

This section shows our detailed results and analysis, including results of Ozone concentration (Sec. A), PM2.5 level (Sec. B), and Nitrogen Oxides (Sec. C).

#### 3.1. Ozone Concentration

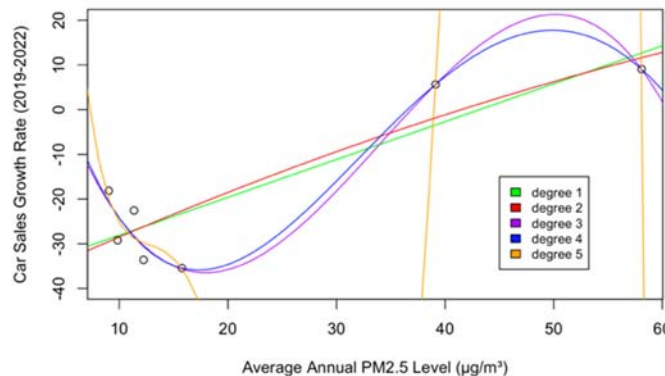
The results show that the P-value of the intercept parameter is significantly less than the 0.05 threshold. While the parameter, presenting the amount of change in sales growth rate for each microgram per cubic meter increase in ozone level, has a P-value slightly greater than 0.01, but this could also be considered highly statistically significant to some extent since a P-value should be less than 0.01 to achieve a high level of statistical significance. For other criteria, note that the  $R^2$  value in our hypothetical model is approximately 0.70, which indicates that a 70% variation of the outcome variable, car sales growth rate in this context, could be explained by the selected model.

$$Y_{Growth\ Rate} = -89.70 + 1.56x_{Ozone} \quad (3)$$

Therefore, each parameter estimate cannot be eliminated from (3) so that it can be a strong candidate to simulate the relationship between the two targets based on the observation sample point whose parameters are highly statistically significant.

It is known that the predictive model has two functions: simulating the current observative dataset and providing future predictions. The presented model implies that for the seven given countries/regions, there is a statistically significant model to capture the correlation between their car sales growth rate and their average seasonal ozone concentration. To be specific in (3), every one microgram per cubic meter increase in ground-level ozone concentration will be expected to associate with a corresponding 1.56% increase in its car sales growth rate. Thus, (3) shows a potential tendency that a country/region with a higher standard of ozone concentration probably attained a more dramatic rise or a more moderate drop in its car sales market in general.

### 3.2. PM2.5 Level



**Figure 4.** Polynomial Regressions of Different Degree

Fig. 4 shows the linear models to different degrees. It is clear that the blue and purple trajectories are more suitable for the sample point. After examining the  $R^2$  values in Fig. 4, it claims that the purple plot representing the polynomial with degree3 is the best model needed.

The five countries/regions with negative car sales growth rates have a decent fit in the proposed model, which is evenly distributed on both sides of the tail of the curve as showcased in Fig.4. Therefore, the relationship between the average annual PM2.5 level in these regions/countries could be considered a cubic relationship in this case and the mathematical expression is:

$$Y_{Growth\ Rate} = 37.19 - 9.33x_{PM2.5} + 0.35x_{PM2.5}^2 - 0.003\beta_3x_{PM2.5}^3 \quad (4)$$

On the other hand, the only two countries, India and China, whose car sales quantities experience upward trends and average PM2.5 concentrations are much higher than the other five regions/countries. It is important to note that the cubic formula exactly passing through these two points cannot guarantee the true relationship between two variables for countries with positive sales growth rates due to a lack of sample points. In particular, only two sample points have positive car sales growth rates, and they vary greatly on average PM2.5 level, the trajectory between them has big uncertainty, which may not follow this particular pattern. Thus, even though these two countries fit well in cubic polynomials, the real relationship between the average PM2.5 level and the growth rate in these countries still cannot be defined as (4).

### 3.3. Nitrogen Oxides

For nitrogen oxides, we consider that is the small sample size, we choose to make the Pearson-Wilke test to determine whether the data conforms to the results sampled from the normal distribution. The results show that the vehicle growth rate is in line with this test. Therefore, we decided to calculate the Pearson coefficient to judge whether there is a linear correlation.

The surface correlation of the Pearson coefficient is weak, and the significance is not high. then the regression model of the least square estimation method is used to determine the coefficient of the function. Thus, it is further verified that there is no strong correlation between the growth rate of automobile sales and the AQI score of nitrogen oxides. The result of the formula shows and verifies that there is no strong correlation between the two variables. It should be noted that the three tests all have P-test and correlation tests.

## 4. Conclusion

In general, according to our study, there is a strong positive linear relationship between ground-level ozone concentration and car sales growth rate for the given seven regions/countries. We could conclude that based on our sample data, the increasing amount of ground-level ozone could be partly contributed to the additional cars entering our everyday life. Besides, for the PM2.5 level, the situation

should be divided into two parts. In countries/regions with negative car sales growth rate, the PM<sub>2.5</sub> level has cubic relation, whereas, for countries/regions whose car sales are increasing, we do not have enough sample points to determine its pattern. Finally, improving the sales growth rate of cars does not effectively reduce the rate of nitrogen dioxide pollution, which means that controlling the sales growth rate of cars does not improve the quality of nitrogen dioxide in the air in a short period of time, based on the AQI index. However, this does not mean that reducing the number of cars does not improve the AQI of nitrogen dioxide in the air.

In the future, we will consider more factors in AQI measurement in our model to determine today's AQI values around the world. Further, three factors, as well as AQI, are all dynamic variables, it will be helpful if the real-time data capture could be employed to get instant information, which could reduce the bias of our models from the time difference between data.

## References

- [1] NRDC.(2021) Air Pollution: Everything You Need to Know. <https://www.nrdc.org/stories/air-pollution-everything-you-need-know>.
- [2] Mark R Miller, David E Newby, Air pollution and cardiovascular disease: car sick, *Cardiovascular Research*, Volume 116, Issue 2, 1 February 2020, Pages 279–294, <https://doi.org/10.1093/cvr/cvz228>
- [3] AirNow.(2021) Air Quality Index (AQI) Basics. <https://www.airnow.gov/aqi/aqi-basics/>
- [4] Burr, Michael Leslie, et al. (2004) "Effects on respiratory health of a reduction in air pollution from vehicle exhaust emissions." *Occupational and Environmental Medicine*.
- [5] State of Global Air.(2020) Ozone Exposure. <https://www.stateofglobalair.org/air/ozone>
- [6] World Population Review.(2022) Most Polluted Countries. <https://worldpopulationreview.com/country-rankings/most-polluted-countries>
- [7] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [8] Carbon Neutrality Coalition.(2022) <https://carbon-neutrality.global/members/>
- [9] Particulate Matter (PM) Pollution, Final Decision to Retain the National Ambient Air Quality Standards for Particulate Matter (PM). <https://www.epa.gov/pm-pollution/final-decision-retain-national-ambient-air-quality-standards-particulate-matter-pm>
- [10] Lewis-Beck, C., & Lewis-Beck, M. (2015). *Applied regression: An introduction* (Vol. 22). Sage publications.