How Has Air Pollution Been Impacted by COVID Restrictions in Canada

Vincent Tang

Abstract

Vehicle emissions are a major component of air pollution. The appearance of COVID-19 in Canada decreased vehicle use by discouraging human interaction. We investigate the impact of the COVID-19 lockdown on air pollution in Canada and other variables that impact vehicle use such as transit use and gasoline prices. Using difference-in-difference and linear regression models, we demonstrate the lockdown negatively impacted air pollution, the number of transit trips, and gasoline prices. However, transit use and gasoline price demonstrated a negative correlation with air pollution. Therefore, unknown factors present during the lockdown mitigated the effect and decreased air pollution.

1 Introduction

Air pollution has been an issue humanity has struggled to resolve for centuries. Defined as the contamination of the air due to the presence of harmful substances, air pollution can negatively impact animal health and the planet as a whole. In addition to causing numerous health concerns in humans such as respiratory infections, air pollution can also damage the environment by reducing the growth of crops and increasing susceptibility to diseases. Despite these negative repercussions, human activity has continued to release air pollutants into the atmosphere throughout the years. One of the largest contributors to air pollution is vehicle usage, as harmful emissions are produced and emitted into the atmosphere by operating a vehicle. However, the reliance on vehicles to travel within a city suggests that there will be great difficulty in convincing society to stop the usage of vehicles in favour of a more environmentally friendly alternative. Therefore, attempts to reduce air pollution by limiting vehicle usage should aim to alter factors that can impact one's vehicle usage instead.

In January 2020, Canada received the first reported case of COVID-19. Due to COVID-19 being spread through contact with others, human interaction became limited and people were discouraged from exiting their homes unless necessary. Although this was a difficult time for everyone in the world, these restrictions placed due to the sudden appearance of COVID-19 may have impacted air pollution by discouraging individuals from travelling and therefore reducing vehicle usage. Determining the effect COVID-19 had on air pollution is important as future policies regarding pollution and transportation could utilize this information to further reduce pollutants in the atmosphere. To establish this, Chang, Meyerhoefer, and Yang (2021) studied the effect of COVID-19 on air pollution in the two largest cities in Taiwan. Through their research, they found that there was an overall 3-7% increase in air pollution during the COVID-19 lockdown, with an increase in pollution during working days and a decrease during non-working days. This was because although the threat of COVID-19 was present, individuals still needed to commute to work. To limit interaction with others, many chose to use personal vehicles rather than use public transportation such as buses or metro trains.

However, this study was performed using data from Taiwan. As such, the results may not apply to Canada. Thus, this paper aims to develop models to estimate the effect of COVID-19 on air pollution in Canada. We will first review the literature and the models used to estimate the effect of COVID-19 on air pollution in Taiwan. Difference-in-difference models based on the previous literature will then be created for the six main pollutants from vehicle emissions to estimate the effect of the COVID-19 lockdown on air pollution in Canada, as well as the effect on transit usage and gas prices. Linear regressions will then be performed to estimate the individual correlation of each variable on the amount of air pollution. Furthermore, air pollution in rural areas will be compared to that in urban areas. Finally, all models will be compared and the results of each will be used to estimate the overall effect the COVID-19 restrictions had on air pollution in Canada.

2 Literature Review

In the paper *COVID-19 prevention, air pollution and transportation patterns in the absence of a lockdown*, Chang, Meyerhoefer, and Yang investigate the impact that the COVID-19 lockdown had on air quality in the two largest cities in Taiwan: Taipei and New Taipei City. Data was compiled on air quality, confirmed COVID-19 cases, and transportation to conduct a difference-in-difference estimation on the amount of air pollution from local sources. As stated previously, they found an overall 3-7% increase in air pollution, specifically CO, O_3, SO_2, PM_{10}, and $PM_{2.5}$. The results from their paper demonstrate the importance of public transit and encouraging bicycle usage as the increase in pollution came largely from individuals preferring to use their vehicles over sharing transportation with others. In addition to estimating the amount of air pollution, they also determine the effect the COVID-19 lockdown had on metro usage and shared-bicycle usage using difference-in-difference estimators. Their paper encourages policymakers to take action and limit the substitution of personal vehicles for public transit to mitigate air pollution and assist the population in becoming accustomed to once again interacting with each other.

Difference-in-Difference

The difference-in-difference estimator is used to estimate the average treatment effect (ATE) of a given treatment by splitting the data into a control group and a treatment group. A regression is performed, and the predicted value of the control group is subtracted from the treatment group to estimate the ATE. Chang, Meyerhoefer, and Yang split their groups based on date, with the pre-treatment period being January 1 - January 21 and the post-treatment period being January 22 - March 31, with observations ranging from 2017-2019. An observation is considered treated if there was a confirmed case of COVID-19 on that day. Two models were implemented for their difference-in-difference estimator. Due to how their data showed that outcome variables were right-skewed, they implemented the following model:

$$\log(y_{1ijt}) = \alpha_1 + \gamma_1 \cdot COVID_{jt} + \beta_1' X_{1ijt} + \upsilon_j + t_m + t_y + \varepsilon_{1ijt}$$

where *y* refers to the outcome variable (air pollution, metro usage, shared-bicycle usage) for data station *i* in city *j* at time *t*, and COVID is either a discrete or continuous variable of confirmed

COVID-19 cases in city *j* at time *t*. *X* is a vector of other explanatory variables, *v* is fixed effects for the city, and t_m and t_y are month and year respectively. The coefficient γ measures the effect COVID-19 had on the outcome variable in percentage terms.

However, the traffic data they received showed some zero values. As such, they also implemented the following difference-in-difference model:

$$y_{2it}= \alpha_2+\gamma_2 \cdot COVID_t+\beta_2' X_{2it}+t_m+t_y+\varepsilon_{2it}$$

where *y* refers to the amount of traffic flow detected at station *i* at time *t*. All other variables are the same as in the previous model, however, γ now measures the effect of COVID-19 on how many 100s of vehicles travel between the two cities per hour. Both models estimate γ by comparing differences in the outcome variable before and after the first reported case of COVID-19.

3 Methodology

a) Difference-in-difference Estimator

A difference-in-difference estimator uses a control group and a treatment group to estimate the average treatment effect (ATE). For the model in this paper, an observation will be considered treated if a positive number of COVID-19 cases were reported that day. The pre-treatment period will be January 1 - January 25, and the post-treatment period will be January 26 - December 31 because the first recorded case of COVID-19 in Canada occurred on January 26, 2020. Furthermore, observations will be used with data from 2018-2020. This model will then be used to estimate air pollution, the number of transit trips, and gas prices in a day. 4 assumptions are required to use the difference-in-difference estimator:

1. Stable Unit Treatment Value Assumption (SUTVA)

   Assumption 1 assumes that the outcome of one observation is independent of the outcome of others. This is satisfied as the pollution, the number of transit trips, and gas prices of one day are independent of other days.

2. Conditional Independence

   Assumption 2 assumes that the assignment of treatment was not assigned based on the outcome. This is satisfied as whether a case of COVID-19 was reported or not in a day was not determined by the air pollution, the number of transit trips, or gas prices for that day.

3. Common Support

   Assumption 3 assumes that the probability of an observation being treated is positive given any value of the confounder variables. This is satisfied as none of the confounder variables change the probability of a day having a reported COVID-19 case.

4. Parallel Trends

   Assumption 4 assumes that if the treatment group had not received treatment, then it would have followed a similar trend to that of the control group. This is impossible to prove, but we can provide a counterfactual by looking at the trends before the treatment period. By observing this, we can see that 2020 begins with a similar trend to the previous years before deviating. For example, the amount of CO emissions in 2018 and 2019 can be observed to dip, forming a “U” shape over each year. As such, the beginning of each year sees a majority of points forming a peak but beginning to trend downwards. A similar trend can be seen at the beginning of 2020. Plots to support this claim will be available in the Appendix.

In the difference-in-difference models for this paper, the following two models were used:

$$Y_1 = \beta_{1,0} + \tau D_i + \lambda D_t + \delta (D_i \times D_t) + \beta_{1,1}X + u$$

$$\log(Y_2) = \beta_{2,0} + \tau D_i + \lambda D_t + \delta (D_i \times D_t) + \beta_{2,1}X + u$$

where D_i and D_t are dummy variables representing treated (1) or not treated (0) and post-treatment (1) or pre-treatment(0), X includes other variables such as the number of reported cases, gasoline price, and the number of transit trips depending on which outcome variable is being estimated, and u is the error term. δ in model 1 represents ATE in the change in the number of emitted pollutants, gasoline price, or the number of transit trips. δ in model *2* represents the ATE in percent change. Every variation of the model is put into a function similar to the one below, with the regression formula changing for each.

```r
DiD_Regression = function(data){
  subset1 <- data %>% filter(Treated == 1 & After_Treatment == 1)
  subset2 <- data %>% filter(Treated == 1 & After_Treatment == 0)
  subset3 <- data %>% filter(Treated == 0 & After_Treatment == 1)
  subset4 <- data %>% filter(Treated == 0 & After_Treatment == 0)

  r <- lm(formula = Pollutants ~ Treated + After_Treatment + (Treated * After_Treatment) + (Price + Trip
s), data = data)

  p1 <- mean(predict(r,subset1))
  p2 <- mean(predict(r,subset2))
  p3 <- mean(predict(r,subset3))
  p4 <- mean(predict(r,subset4))

  estimate <- ((p1 - p2) - (p3 - p4))
}
```

Difference-in-difference is a useful technique to use when needing to compare changes in outcome over a period of time when considering a treatment and is capable of considering impacts from other potential differences other than the treatment by subtracting out this difference.

b) Linear Regression

A linear regression model is used to determine the marginal effect of independent variables on a dependent variable. 5 Assumptions must be met to use a linear regression model.

1. Linearity

   Assumption 1 assumes that the model must be linear in the parameters. Meaning it follows the model:

$Y_i = X'_i\beta + u_i$

where Y_i is the dependent variable, X_i is the vector of independent/explanatory variables, β is the vector of parameters of interest, and u_i is the error term. This will be satisfied through the models displayed later.

2. Independent and Identically Distributed Sample (i.i.d. sample)

   Assumption 2 assumes that the sample is distributed independently from one another and identically. This is satisfied as the sample is distributed based on date and each observation is independent of the other.

3. Exogeneity

   Assumption 3 assumes that the error term is independent of X. This cannot be tested but must be assumed to ensure consistency and unbiasedness.

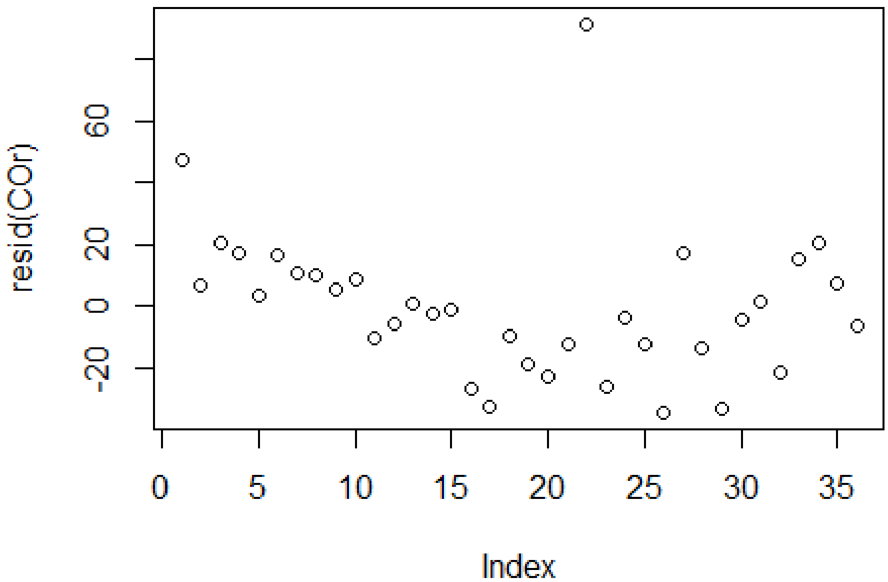
4. No Collinearity

   Assumption 4 assumes that explanatory variables cannot be expressed using a linear combination of each other, as this would mean using those two variables is equivalent to using the same information.

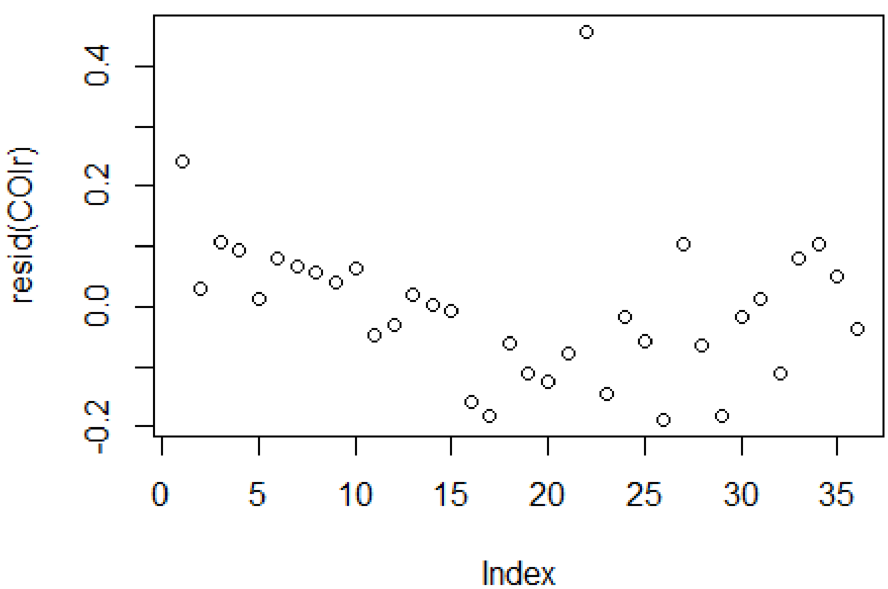
5. Homoskedasticity

   Assumption 5 assumes that the variance of the error term is independent of x. This is important as satisfying this assumption allows for the regression to have asymptotic normality and be the best linear unbiased estimator (BLUE). This can be proved by plotting the residuals of the regression and observing the points. The points spreading into a fan shape indicate heteroskedasticity. Below are two plots from the regressions performed on the CO pollutant. We do not observe a fan-shaped distribution in the residuals, so we can assume homoskedasticity.

```
plot(resid(COr))
```




```
plot(resid(COlr))
```



Regressions were used to estimate the amount of carbon monoxide (CO), ozone pollutants (O3), particular matter with concentration under 10 or 2.5 micrograms per cubic meter of air (PM10, PM2.5), sulphur dioxide (SO2), and nitrogen dioxide (NO2) emitted, as these are the main pollutants emitted from vehicle use. Two regression models were used to estimate each pollutant:

$$Y_i = X'_i\beta + u_i$$

$$\log(Y_i) = X'_i\beta + u_i$$

Both models estimate the coefficients in the vector β, with the first model estimating the marginal effect and the second estimating the effect in percentage terms. However, due to constraints from the available data, regressions were performed using months for observations instead of dates. The functions used can be found below:

```
Regression <- function(data){
  return_data <- data %>% lm(formula = Pollutants ~ Cases + Price + Trips)
  return(return_data)
}

LRegression <- function(data){
  return_data <- data %>% lm(formula = log(Pollutants) ~ Cases + Price + Trips)
  return(return_data)
}
```

4 Data Description

Numerous data sources were used to collect data for this paper. COVID-19 case data was retrieved from the World Health Organization's (WHO) website, which compiles the number of reported cases and deaths from COVID-19 daily. From this data, we took the number of reported COVID-19 cases per day as a variable. Emissions data for all pollutants included in this paper were retrieved from Canada's Environmental Statistics website, where they utilize the National Air Pollution Surveillance (NAPS) system to compile data on the number of emissions for each pollutant by the hour. The number of emissions for each pollutant, measured in parts per billion (ppm), was taken as the dependent variable. Gasoline prices and transit data were retrieved from Statistics Canada through a census and a survey respectively. Unfortunately, only monthly data was available for these data sets. Therefore, gasoline prices and the number of transit trips per month were taken as variables, and some values were repeated to fit into the combined data set. While this does not affect the difference-in-difference estimator as much because the estimation of the ATE subtracts out the coefficients of other variables, this does impact the outcome of the linear regressions. To resolve this issue, the regressions were performed using monthly averages rather than daily data. Furthermore, to compare pollution in rural areas with pollution in urban areas, a list of cities in Canada was retrieved from the World Population Review website. There are many definitions for what constitutes a rural area in Canada, but we will be assuming that any city that is not on the provided list, that is any city with a population less than 10,000, will be considered rural.

To better estimate the lockdown's effect on vehicle usage, gathering data on traffic flow in Canada would have been excellent. However, this was not possible as Canada does not have stations to monitor traffic data like those present in Taiwan. As such, that data is missing from the models in this paper.

From all data sets, data was gathered from January 2018 - December 2020 and manipulated such that all data sets could be combined. The first 6 observations of the data sets used for the difference-in-difference estimator and the linear regression are provided below:

```
## # A tibble: 1,096 x 13
##    date       Pollutants Cases Treated Year  Month   Day After_Treatment Rural
##    <date>          <dbl> <dbl>   <dbl> <chr> <dbl> <int>           <dbl> <lgl>
##  1 2018-01-01       221.     0       0 18        1     1               0 FALSE
##  2 2018-01-02       256.     0       0 18        1     2               0 FALSE
##  3 2018-01-03       287.     0       0 18        1     3               0 FALSE
##  4 2018-01-04       283.     0       0 18        1     4               0 FALSE
##  5 2018-01-05       232.     0       0 18        1     5               0 FALSE
##  6 2018-01-06       203.     0       0 18        1     6               0 FALSE
##  7 2018-01-07       215.     0       0 18        1     7               0 FALSE
##  8 2018-01-08       299.     0       0 18        1     8               0 FALSE
##  9 2018-01-09       223.     0       0 18        1     9               0 FALSE
## 10 2018-01-10       231.     0       0 18        1    10               0 FALSE
## # ... with 1,086 more rows, and 4 more variables: Date.x <chr>, Trips <dbl>,
## #   Date.y <chr>, Price <dbl>

## # A tibble: 36 x 6
## # Groups:   Month [12]
##    Month Year  Pollutants    Cases Trips Price
##    <dbl> <chr>      <dbl>    <dbl> <dbl> <dbl>
##  1     1 18          232.    0      4.99 121.
##  2     1 19          213.    0      5.19 103.
##  3     1 20          216.    0.387  5.29 115.
##  4     2 18          211.    0      5.54 120.
##  5     2 19          213.    0      5.49 105.
##  6     2 20          219.    7.90   5.56 112.
##  7     3 18          198.    0      5.37 123.
##  8     3 19          202.    0      5.31 118.
##  9     3 20          186. 1116.     2.85  91.3
## 10     4 18          183.    0      5.14 132.
## # ... with 26 more rows
```

Below are the functions used to manipulate each data set:

```
CovidData <- read.csv("WHO-COVID-19-global-data.csv") %>% filter(Country == "Canada") %>% mut
ate(date = as_date(ï..Date_reported)) %>% select(date,Country,Cumulative_cases)

Can_Pop <- read.csv("Canadian_Cities.csv")
Cities <- Can_Pop$city

data_prep <- function(data){
  data[is.na(data)] <- 0
  data[data == -999] <- 0
```

```
  return_data <- data %>% mutate(date = mdy(Date..Date)) %>% mutate(Daily_Sum = H01..H01 + H02..
H02 + H03..H03 + H04..H04 + H05..H05 + H06..H06 + H07..H07 + H08..H08 + H09..H09 + H10..H10
+ H11..H11 + H12..H12 + H13..H13 + H14..H14 + H15..H15 + H16..H16 + H17..H17 + H18..H18 + H1
9..H19 + H20..H20 + H21..H21 + H22..H22 + H23..H23 + H24..H24) %>% mutate(Cumulative_cases =
0) %>% group_by(date) %>% summarise(Pollutants = sum(Daily_Sum),Cases = mean(Cumulative_cases
),) %>% mutate(Treated = 0) %>% mutate(Year = substring(year(date),3,4)) %>% mutate(Month = mont
h(date)) %>% mutate(Day = day(date)) %>% mutate(After_Treatment = if_else(Month==1 & Day<=25,0
,1))

  return_data[is.na(return_data)] <- 0
  return(return_data)
}

data_prep2 <- function(data){
  data[is.na(data)] <- 0
  data[data == -999] <- 0

  return_data <- data %>% mutate(date = mdy(Date..Date)) %>% mutate(Daily_Sum = H01..H01 + H02..
H02 + H03..H03 + H04..H04 + H05..H05 + H06..H06 + H07..H07 + H08..H08 + H09..H09 + H10..H10
+ H11..H11 + H12..H12 + H13..H13 + H14..H14 + H15..H15 + H16..H16 + H17..H17 + H18..H18 + H1
9..H19 + H20..H20 + H21..H21 + H22..H22 + H23..H23 + H24..H24) %>% left_join(CovidData, by = "d
ate") %>% group_by(date) %>% summarise(Pollutants = sum(Daily_Sum),Cases = mean(Cumulative_ca
ses)) %>% mutate(Treated = 1) %>% mutate(Year = substring(year(date),3,4)) %>% mutate(Month = mo
nth(date)) %>% mutate(Day = day(date)) %>% mutate(After_Treatment = if_else(Month==1 & Day<=25
,0,1))

  return_data[is.na(return_data)] <- 0
  return(return_data)
}

PollCO_2018 <- data_prep(urban(read.csv("CO_2018(Edited).csv"))) %>% mutate(Rural = FALSE)

PollCO_2020 <- data_prep2(urban(read.csv("CO_2020(Edited).csv"))) %>% mutate(Rural = FALSE)

PollCO_2018r <- data_prep(rural(read.csv("CO_2018(Edited).csv"))) %>% mutate(Rural = TRUE)

PollCO_2020r <- data_prep2(rural(read.csv("CO_2020(Edited).csv"))) %>% mutate(Rural = TRUE)

data_prep3 <- function(data){
  month_list <- c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec")

  return_data <- data %>% mutate(Date = ï..Date) %>% mutate(Year = substring(Date,1,2)) %>% mutate(
Month = match(substring(Date,4,6),month_list))
}

month_length <- c(31,28,31,30,31,30,31,31,30,31,30,31)
TransitStats <- data_prep3(read.csv("Urban-Transit-Stats(Edited).csv")) %>% mutate(Trips = Total.passe
nger.trips/if_else(Month == 2 & Year == 20,29,month_length[Month])) %>% select(Date,Year,Month,Tr
ips)
GasPrices <- data_prep3(read.csv("Gasoline-Price-Stats(Edited).csv")) %>% mutate(Price = Canada) %>
```

```
% select(Date,Year,Month,Price)

data_prep4 <- function(data){

  return_data <- data %>% left_join(TransitStats,by = c('Year' = 'Year','Month' = 'Month')) %>% left_join
(GasPrices,by = c('Year' = 'Year','Month' = 'Month'))
}

PollCO <- data_prep4(rbind(PollCO_2018,PollCO_2019,PollCO_2020))

Regression_prep <- function(data){
  return_data <- data %>% group_by(Month,Year) %>% summarise(po = mean(Pollutants),c = mean(Cas
es),t = mean(Trips), pr = mean(Price))

  colnames(return_data) <- c("Month","Year","Pollutants","Cases","Trips","Price")

  return(return_data)
}

COr <- Regression(Regression_prep(PollCO))
```

Only the CO data set manipulation is displayed above, as the data set for the other 5 pollutants follow the same procedure.

5 Estimation Results

Using the models described above, the results can be estimated and compared. As mentioned previously, the difference-in-difference model is a useful model that allows for an estimation of the average treatment effect (ATE). However, although the ATE can be estimated, the significance of the estimation needs to be considered as well. This is represented by the number of stars in the significance column in the following results:

```
##       Variable   Difference Significance
## 1      CO(ppb)    13.910681          ***
## 2      O3(ppb)   347.060320
## 3  PM10(μg/m3)   -35.633178
## 4 PM2.5(μg/m3)  -105.308341
## 5     SO2(ppb)   -95.030937
## 6     NO2(ppb) -1101.866195           **
## 7        Trips    -3.247652          ***
## 8    Gas Price   -25.577273          ***

##       Variable Percent_Difference Significance
## 1      CO(ppb)        0.072721312          ***
## 2      O3(ppb)        0.013050415
## 3  PM10(μg/m3)        0.121187130          ***
## 4 PM2.5(μg/m3)        0.007766509
## 5     SO2(ppb)       -0.101794146
## 6     NO2(ppb)       -0.073504437          ***
## 7        Trips       -1.076796497          ***
## 8    Gas Price       -0.235498522          ***
```

From these results, we can observe that while the estimation for the ATE on the number of transit trips and gas prices was significant, only a few pollutants saw significant estimations. In particular, the linear regression model produced significant estimations for CO and NO2, while the logarithmic regression model produced significant estimations for CO, NO2, and PM10. As such, it is not appropriate to interpret the results for other pollutants. However, we can look at the statistically significant estimates to determine the overall effect of the lockdown. Looking at the estimates from the logarithmic regression model, we can observe an overall increase in the number of pollutants, with the percentage change in CO and NO2 being very close, and the percentage change in PM10 being positive, for a total significant percentage change of around 0.12% after the treatment. However, the actual change in NO2 is significantly larger compared to that of CO, with a change almost 100 times greater. This suggests that the treatment, or the introduction of confirmed COVID-19 cases in Canada, did not impact a majority of the pollutants included in the model but did have an overall negative impact.

We can also observe the significant impact the treatment had on the number of transit trips and gas prices. Looking at the impact on gas prices, we can observe a negative percentage decrease of around 0.24% and a decrease of around 25 cents per Litre. We can also observe the number of trips decreasing by 3 trips per month during the post-treatment period. Furthermore, the percentage change in the number of trips is estimated to decrease by 1.06%. Both of these results suggest that the COVID-19 lockdown caused gas prices to decrease while also discouraging the use of public transit. However, unlike the results found by Chang, Meyerhoefer, and Yang in Taiwan, the average number of pollutants in Canada decreased during the post-treatment period. This could be because although gas prices were lowered and the usage of public transit decreased, the overall number of individuals travelling outside their homes was low. One possible reason for this is that many companies and organizations began to allow employees to work from home, removing the need for them to use their vehicles to go to work.

We can then compare these results to the results of the linear regressions:

Linear:

```
##                       CO CO_Signif            O3 O3_Signif
## (Intercept)  2.327058e+02       ***  5.933731e+04        **
## Cases        1.193477e-04         * -7.593118e-02         *
## Price       -1.086628e+00         *  2.903311e+02
## Trips        1.646282e+01        ** -5.346131e+03         *

##                     PM10 PM10_Signif        PM2.5 PM2.5_Signif
## (Intercept) -4.273423e+03           . 1.080536e+04
## Cases        3.455280e-04             6.011565e-03
## Price        7.724375e+01          ** 6.493647e+01
## Trips       -4.961912e+02           . 3.697869e+02

##                      SO2 SO2_Signif           NO2 NO2_Signif
## (Intercept)  9.216531e+02         **  3.844626e+04        ***
## Cases       -9.379187e-04          .  3.448492e-02        ***
## Price        4.309562e+00            -3.856909e+02        ***
## Trips       -3.580547e+01             5.333564e+03        ***
```

```
##             Total Change in Pollutants
## (Intercept)              9.466451e+04
## Cases                   -4.226483e-02
## Price                   -3.095337e+02
## Trips                   -4.922947e+02
```

Logarithmic:

```
##                       CO CO_Signif            O3 O3_Signif
## (Intercept)  5.497045e+00       ***  1.098218e+01       ***
## Cases        6.931648e-07         * -1.128510e-06         *
## Price       -6.340186e-03         *  4.351534e-03
## Trips        9.405200e-02       *** -8.139531e-02         *

##                     PM10 PM10_Signif        PM2.5 PM2.5_Signif
## (Intercept)  5.391935e+00         *** 9.634405e+00          ***
## Cases        5.500551e-07             4.884508e-07
## Price        2.597447e-02          ** 3.247589e-04
## Trips       -1.641937e-01           . 4.393252e-02

##                      SO2 SO2_Signif           NO2 NO2_Signif
## (Intercept)  6.882045e+00        ***  1.080204e+01        ***
## Cases       -8.132489e-07          *  2.027680e-06        ***
## Price        3.276686e-03            -2.100658e-02        ***
## Trips       -2.987146e-02             3.017963e-01        ***
```

From these regressions, we can observe the significance each variable has on each pollutant. Overall, the significance levels are similar to what was seen in the difference-in-difference regression. Significance levels for CO and NO2 are high are the linear regression, and significance levels for CO, PM10, and NO2 are high in the logarithmic regression. However, the number of COVID-19 cases and the number of transit trips also demonstrate a slight significance level for the emissions of O3 and SO2. Considering the significance and magnitude levels for each estimate from both linear regression models, we can determine that number of COVID-19 cases, gasoline prices, and transit use have a negative correlation with air pollution, as proven by the negative value when calculating the sum of the significant change in pollutants.

Additionally, we can compare the average pollution found in urban areas to that of rural areas. The difference-in-difference model was applied to rural areas in addition to urban areas, producing the following results:

```
##        Variable  Difference Significance
## 1       CO(ppb)    25.53976           **
## 2       O3(ppb) -2604.56245
## 3  PM10(µg/m3)  -652.16701
## 4 PM2.5(µg/m3)  -529.90898
## 5      SO2(ppb)    25.08310
## 6      NO2(ppb)  -804.35447
```

```
##       Variable  Difference Significance
## 1      CO(ppb)  0.17468928           **
## 2      O3(ppb) -0.03303261
## 3  PM10(µg/m3)  0.10064052
## 4 PM2.5(µg/m3) -0.01404020
## 5     SO2(ppb) -0.02310578
## 6     NO2(ppb) -0.07886756
```

The results from the rural data demonstrate a large lack of significance in all pollutants except for CO. CO emissions are not exclusive to vehicle usage, and can also be emitted by burning wood. Therefore, the threat of COVID-19 in rural areas did not significantly impact the use of vehicles but may have encouraged more burning of wood to keep homes warm. We can further investigate the effect COVID-19 had on rural areas by using the linear regression model to produce the following results:

Linear:

```
##                        CO CO_Signif           O3 O3_Signif
## (Intercept) 1.472855e+02       ***  6.825956e+04       ***
## Cases       4.668724e-05       *** -3.394123e-02       ***

##                      PM10 PM10_Signif         PM2.5 PM2.5_Signif
## (Intercept)  3.726065e+03         ***  1.745198e+04          ***
## Cases       -2.798363e-03         *** -3.194475e-03

##                       SO2 SO2_Signif           NO2 NO2_Signif
## (Intercept)  1.369774e+03        ***  1.420282e+04        ***
## Cases       -5.663127e-04        *** -1.199478e-03

##             Total Change in Pollutants
## (Intercept)               1.051575e+05
## Cases                    -3.725922e-02
```

Logarithmic:

```
##                      CO CO_Signif           O3 O3_Signif
## (Intercept) 4.937790332       ***  1.110459e+01       ***
## Cases       0.000000359       *** -4.896560e-07       ***

##                      PM10 PM10_Signif         PM2.5 PM2.5_Signif
## (Intercept)  7.967134e+00         ***  9.664125e+00          ***
## Cases       -1.956782e-07             -1.376419e-07

##                       SO2 SO2_Signif           NO2 NO2_Signif
## (Intercept)  7.162179e+00        ***  9.446715e+00        ***
## Cases       -4.700599e-07        *** -7.105664e-08
```

Due to a lack of data on transit use and gasoline prices, only the number of confirmed COVID-19 cases could be used as a variable for the linear regression models. From these results, we can observe that the number of reported COVID-19 cases had a significant impact on the amount of emitted CO, O3, and SO2, with an overall negative effect. When adding the significant results together, we can see that rural areas experienced a significant decrease of around 0.037 ppb. When compared to urban areas, we can observe that this is slightly lower in magnitude as the decrease in urban areas was around 0.042 ppb. Furthermore, we can get a better understanding of the difference in pollution between rural and urban areas by comparing the two directly.

By separating the dataset into rural and urban and calculating the average emissions of each pollutant, we get the following results:

```
##   Pollutant      Rural      Urban Rural_Difference Rural_Percentage_Diff
## 1        CO   116.6383   181.9003        -65.26207           -35.8779257
## 2        O3 66613.7974 67020.9124       -407.11496            -0.6074447
## 3      PM10  4687.0474  2527.2719       2159.77555            85.4587727
## 4     PM2.5 14577.8312 20054.6460      -5476.81478           -27.3094563
## 5       SO2  1465.5582  1225.8630        239.69526            19.5531853
## 6       NO2 10360.4188 17943.2810      -7582.86223           -42.2601765
```

Summarizing the percentage difference for each pollutant demonstrates that rural areas see a total decrease in emitted air pollutants of around 1.04% on average compared to urban areas. Furthermore, summarizing the actual difference for each pollutant demonstrates a total decrease of 11132.53828 parts per billion (ppb) compared to urban areas. As a result, we can conclude from this data that rural areas experience less air pollution than urban areas on average despite experiencing less impact from the COVID-19 lockdown. A possible reason for this is that rural areas often have a lower reliance on motor vehicles for transportation as well as a lower population than urban areas. This further supports the idea that air pollution can be reduced in the future by mitigating vehicle use. Additional plots to visualize the difference in emission levels for each pollutant are available in the Appendix.

Both the difference-in-difference model and the linear regression model have their benefits, and by utilizing both we can observe the overall impact the COVID-19 lockdown had and the impact other variables had on air pollution. The results above have demonstrated that an increase in the number of reported COVID-19 cases, the number of transit trips, or gas prices can decrease the amount of air pollution. This is because an increase in the number of COVID-19 cases or gas prices discourages individuals from frequently operating their vehicles or exiting their homes. Furthermore, an increase in the number of transit trips suggests that more individuals are willing to share a vehicle with others, reducing air pollution by reducing the total number of vehicles on the road. In addition, the results also demonstrate a lack of impact from the treatment. In other words, the lockdown and how society reacted to the knowledge that COVID-19 was present in Canada did not significantly impact air pollution. However, total air pollution still decreased during the lockdown, perhaps due to other variables such as the increase in online employment and the resulting reduction in traffic offsetting the effect of decreasing gas prices and transit use. Overall, both models provided valuable information regarding how the arrival of COVID-19 in Canada impacted air pollution, but the linear regression model provided more statistically significant results. As a result, that model is capable of providing more information and is therefore a better fit for the data in this paper.

6 Conclusion

In this paper, a difference-in-difference model and a linear regression model were used to estimate the effect that the COVID-19 lockdown had on air pollution. After observing the results, it is clear that the COVID-19 lockdown had a small impact on reducing the number of air pollutants, with no magnitude of change greater than 0.12%. This is significantly different from the results found by Chang, Meyerhoefer, and Yang in their paper *Covid-19 prevention, air pollution and transportation patterns in the absence of a lockdown*, in which they found a 3-7% overall increase in air pollution during the lockdown in Taiwan. Furthermore, the results showed a significant negative impact on the number of transit trips and gas prices after the appearance of COVID-19 in Canada. From the results of the linear regression, we can see that both of these variables have a negative correlation with the number of pollutants emitted from vehicle use. This is likely because an increase in gas prices discourages individuals from using their vehicles, and more use of public transit results in more shared travel. Therefore, future policies looking to improve air pollution should consider methods to discourage the unnecessary use of personal vehicles and encourage the use of shared travel methods such as public transit. Overall, these results suggest unknown factors present during the lockdown mitigated the effect of decreased transit use and gasoline prices, allowing for the number of air pollutants to decrease.

In future works on this topic, other methods and data should be considered as well. The difference-in-difference model showed a lack of significant results for a majority of the included pollutants. As such, a different treatment variable other than the appearance of COVID-19 may need to be considered in future works. A regression discontinuity design was considered for this paper, however, there existed no confounder variable X that determined whether a date would receive treatment or not in the dataset. As such, it became difficult to create an accurate estimation model. Furthermore, additional data would have been useful to gather, as daily data was missing for both transit and gas price data, the accuracy of those estimations likely suffered as a result. Chang, Meyerhoefer, and Yang utilized traffic flow data to determine how often vehicles were used during the lockdown in Taiwan. However, Canada does not have complete traffic monitoring station data from 2018-2020, resulting in the necessary removal of that variable. If that variable were to have been included, the results may have been more accurate and significant. In addition, no data was found for transit use or gasoline price in rural areas, resulting in models with few variables and potentially causing omitted variable bias. If air pollution should be compared between rural and urban areas in future works, additional data should be gathered to better understand the reason why rural areas experience less air pollution than urban areas on average. As a result, future works on this topic should consider additional variables and alternative models to better capture what factors had the largest impact on air pollution in Canada during the COVID-19 lockdown.

7 References

Chang, H.H., Meyerhoefer, C.D., & Yang, F.A. (2021). Covid-19 prevention, air pollution and transportation patterns in the absence of a lockdown. Journal of Environmental Management. 298(2021).https://doi.org/10.1016/j.jenvman.2021.113522

CO_2018.(2018). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2018/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

CO_2019.(2019). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2019/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

CO_2020.(2020). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2020/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

Daily cases and deaths by data reported to WHO. (n.d.). World Health Organization. Retrieved from: https://covid19.who.int/data

Monthly average retail prices for gasoline and fuel oil, by geography. (2022, November 16). Statistics Canada. Retrieved from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000101&pickMembers%5B0%5D=2.2&cubeTimeFrame.startMonth=01&cubeTimeFrame.startYear=2006&cubeTimeFrame.endMonth=09&cubeTimeFrame.endYear=2022&referencePeriods=20060101%2C20220901

NO2_2018.(2018). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2018/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

NO2_2019.(2019). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2019/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

NO2_2020.(2020). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2020/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

O3_2018.(2018). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2018/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

O3_2019.(2019). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2019/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

O3_2020.(2020). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2020/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

Passenger bus and urban transit statistics, by the North American Industry Classification System (NAICS). (2022, November 21). Statistics Canada. Retrieved from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2310025101&pickMembers%5B0%5D=2.1&cubeTimeFrame.startMonth=01&cubeTimeFrame.startYear=2017&cubeTimeFrame.endMonth=12&cubeTimeFrame.endYear=2022&referencePeriods=20170101%2C20221201

PM10_2018.(2018). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2018/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

PM10_2019.(2019). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2019/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

PM10_2020.(2020). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2020/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

PM25_2018.(2018). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2018/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

PM25_2019.(2019). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2019/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

PM25_2020.(2020). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2020/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

Population of cities in Canada. (2022). Retrieved from: https://worldpopulationreview.com/countries/cities/canada

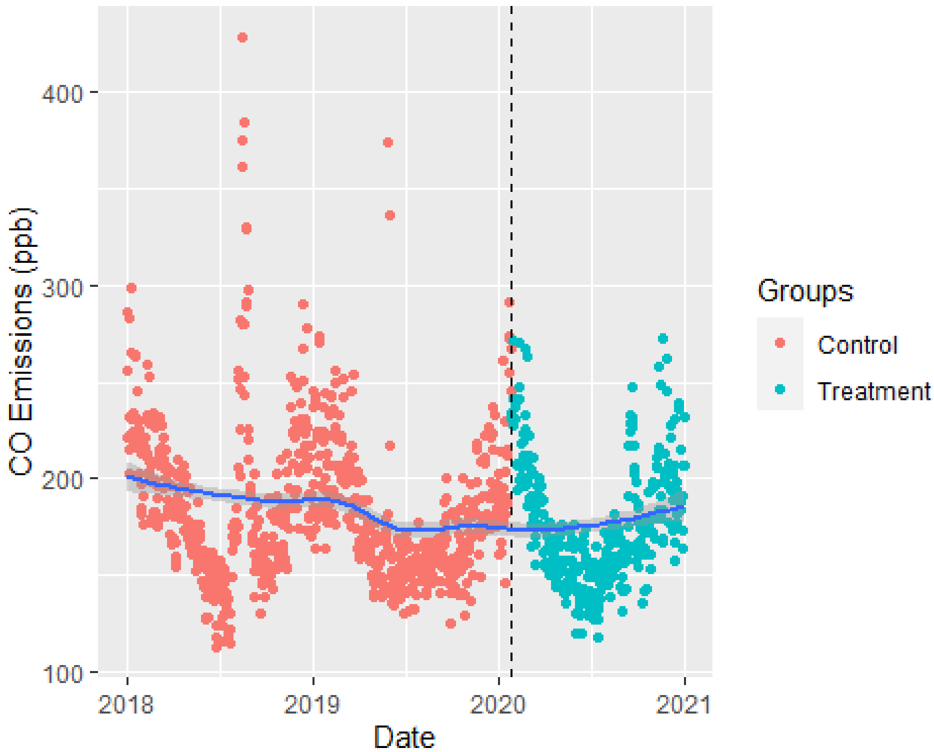
SO2_2018.(2018). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2018/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

SO2_2019.(2019). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2019/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

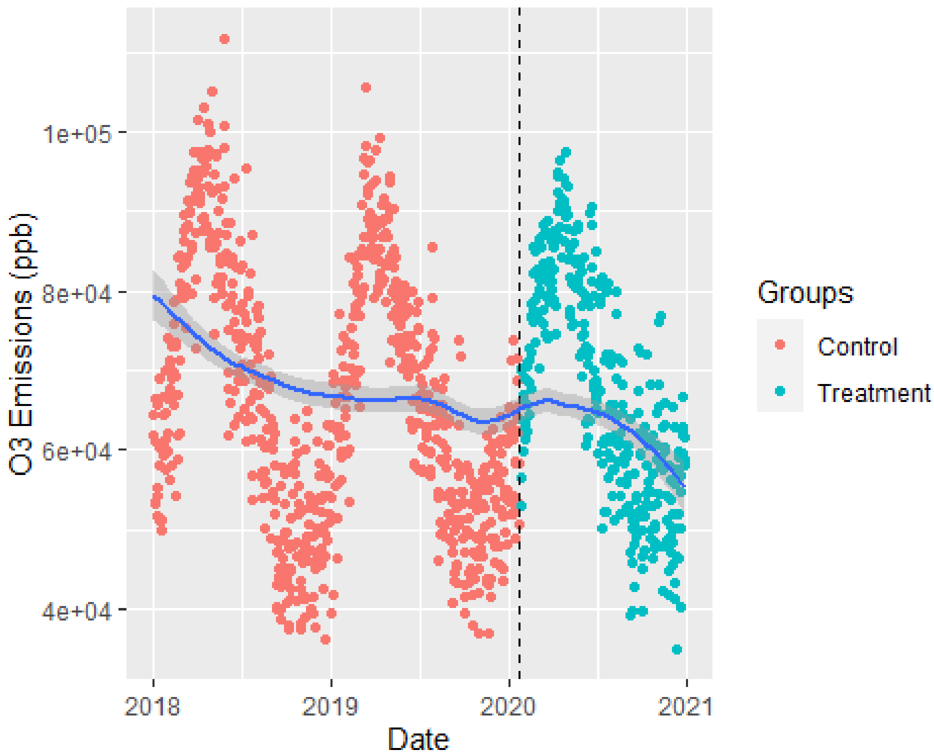
SO2_2020.(2020). National Air Pollution Surveillance (NAPS) Program. Retrieved from: https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2020/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en

8 Appendix

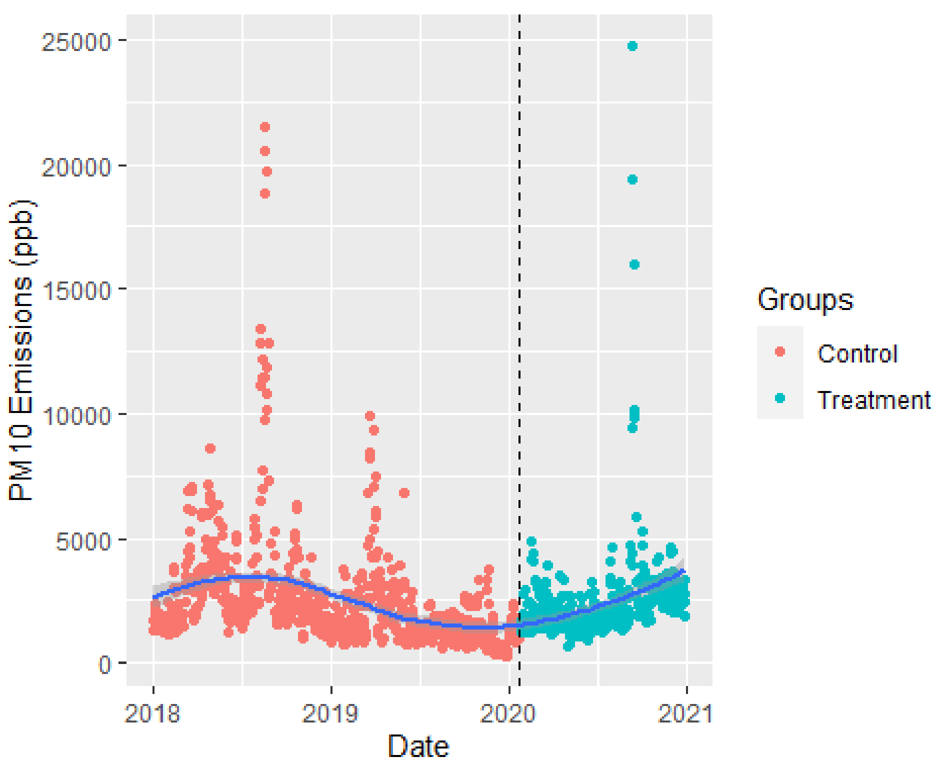
```
PT_Plot(PollCO,"CO")
```



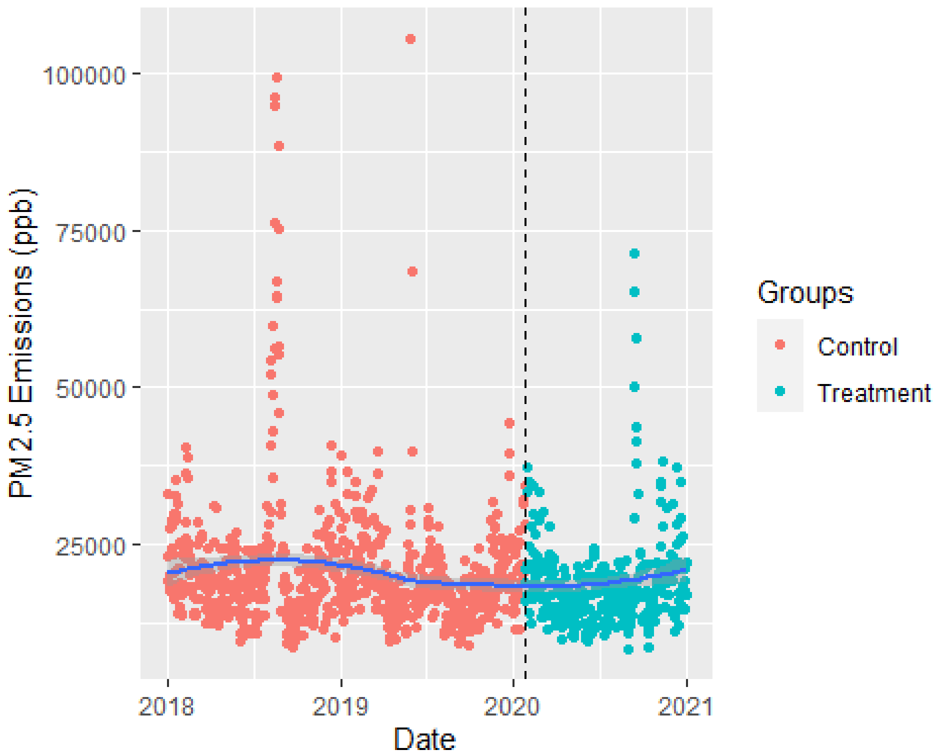
```
PT_Plot(PollO3,"O3")
```



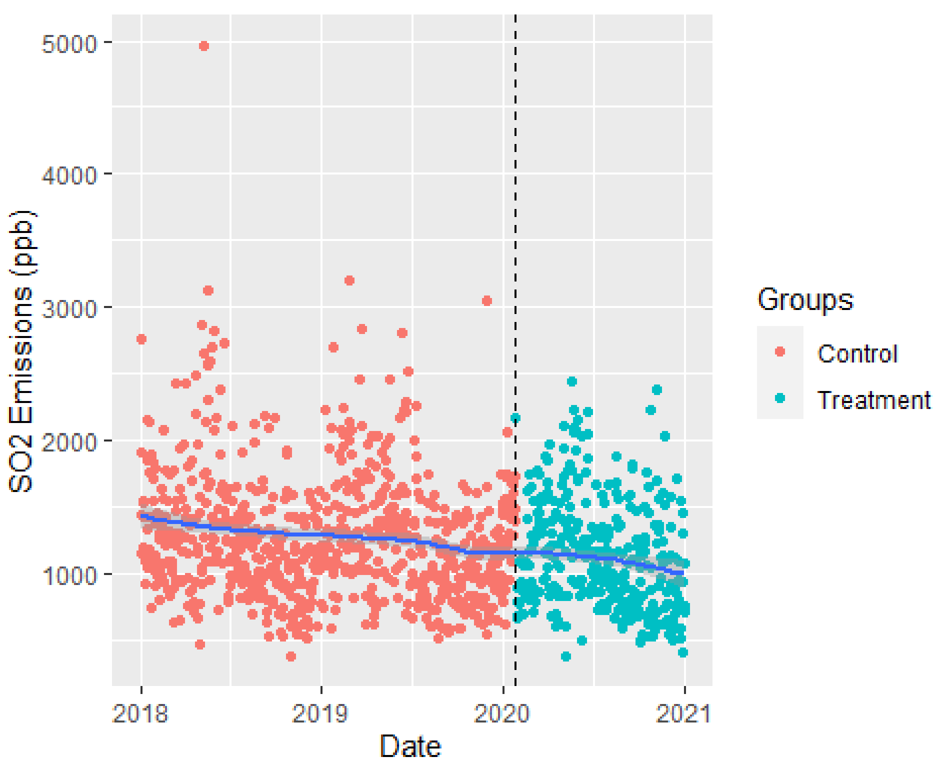

```
PT_Plot(PollPM10,"PM10")
```



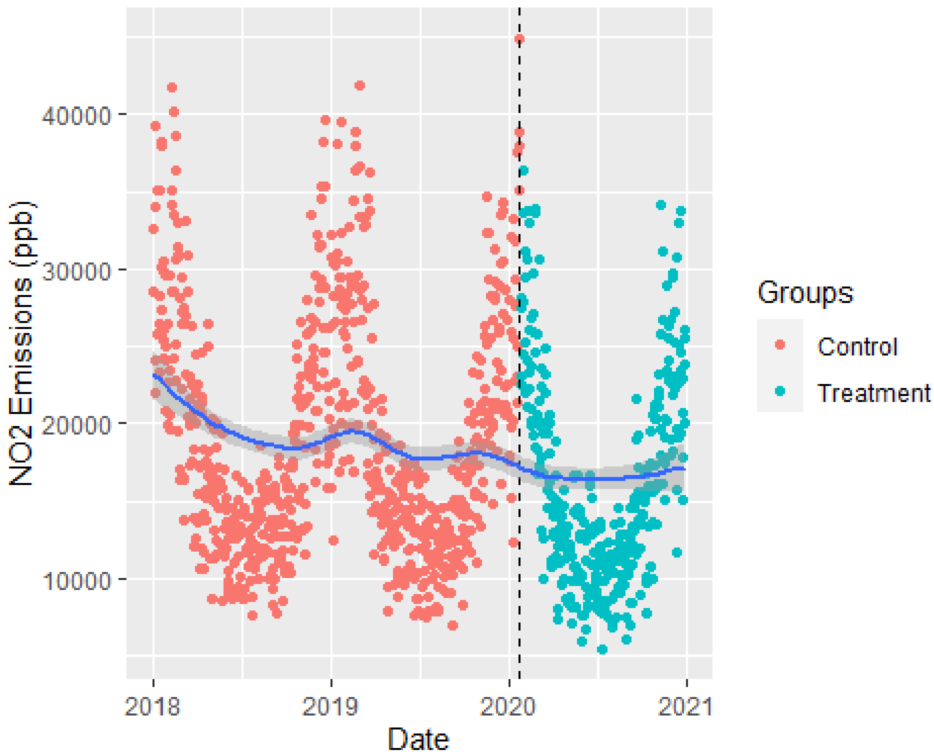
```
PT_Plot(PollPM25,"PM2.5")
```

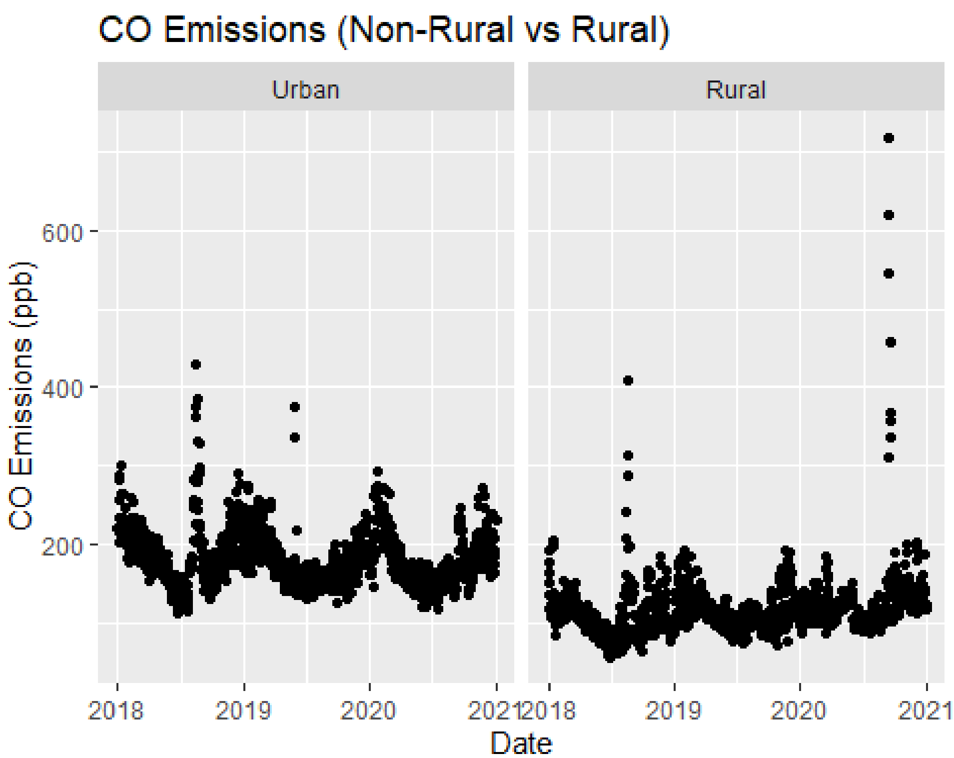


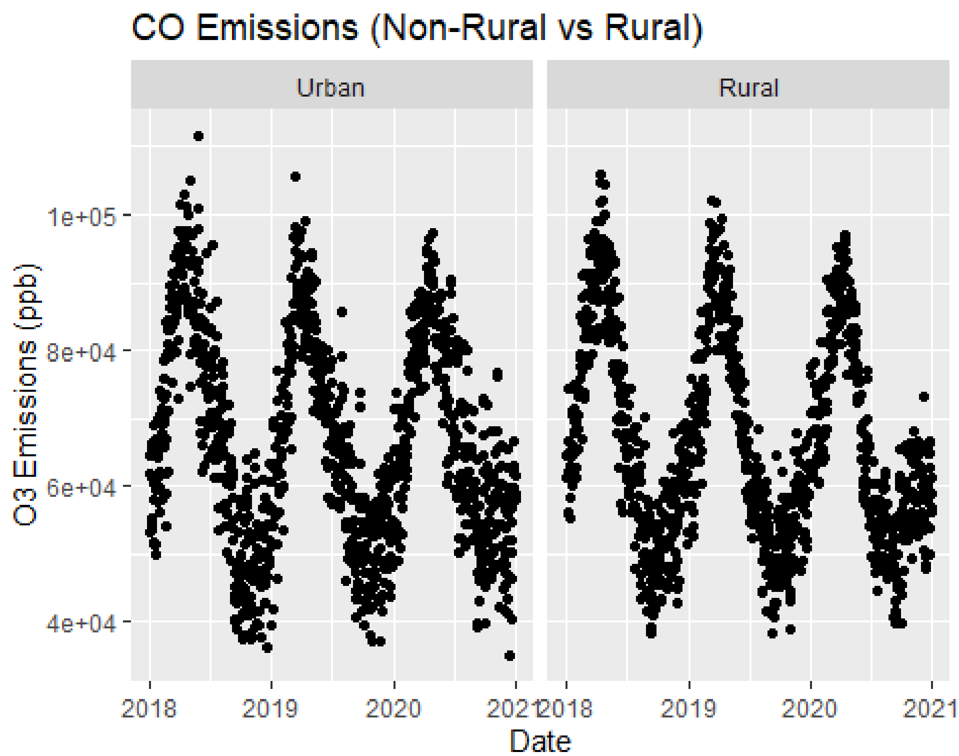

```
PT_Plot(PollSO2,"SO2")
```

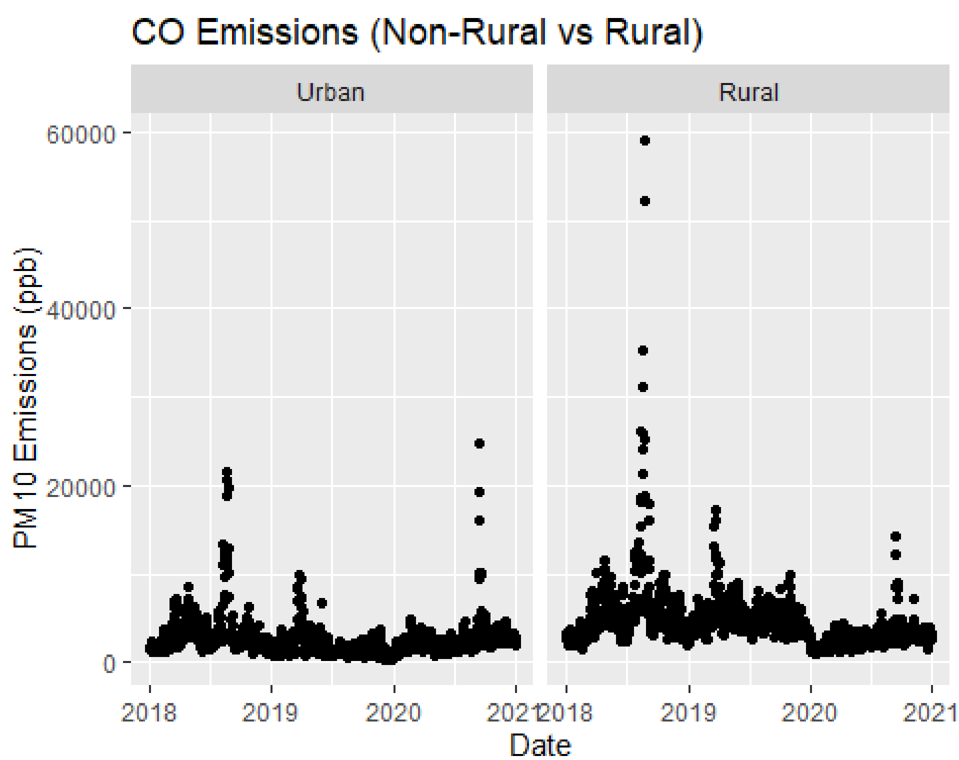


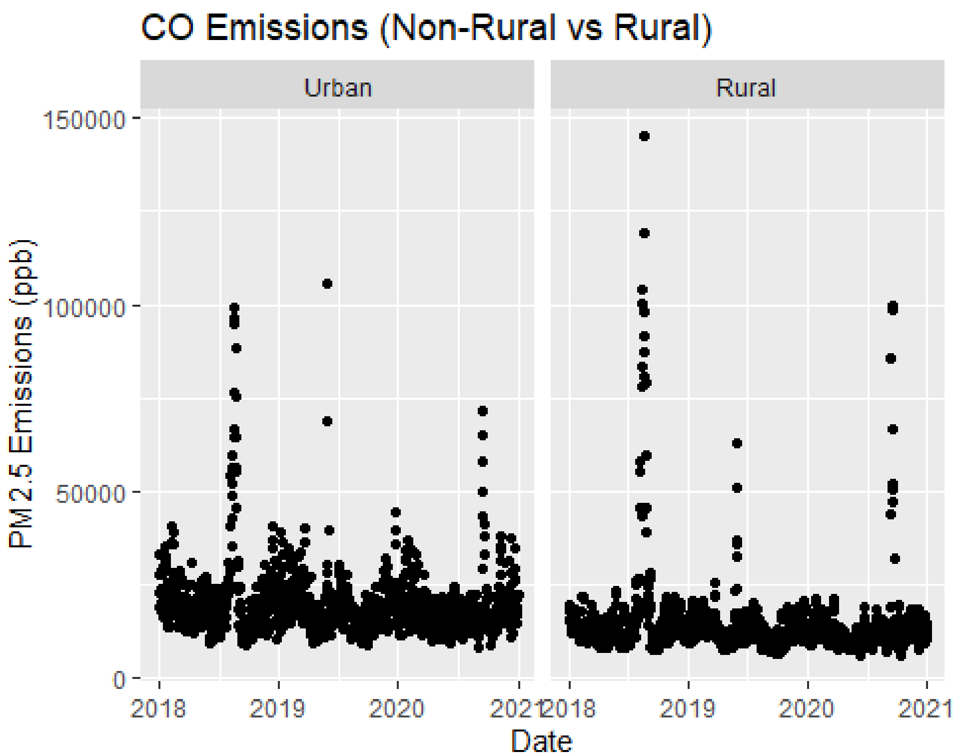
```
PT_Plot(PollNO2,"NO2")
```

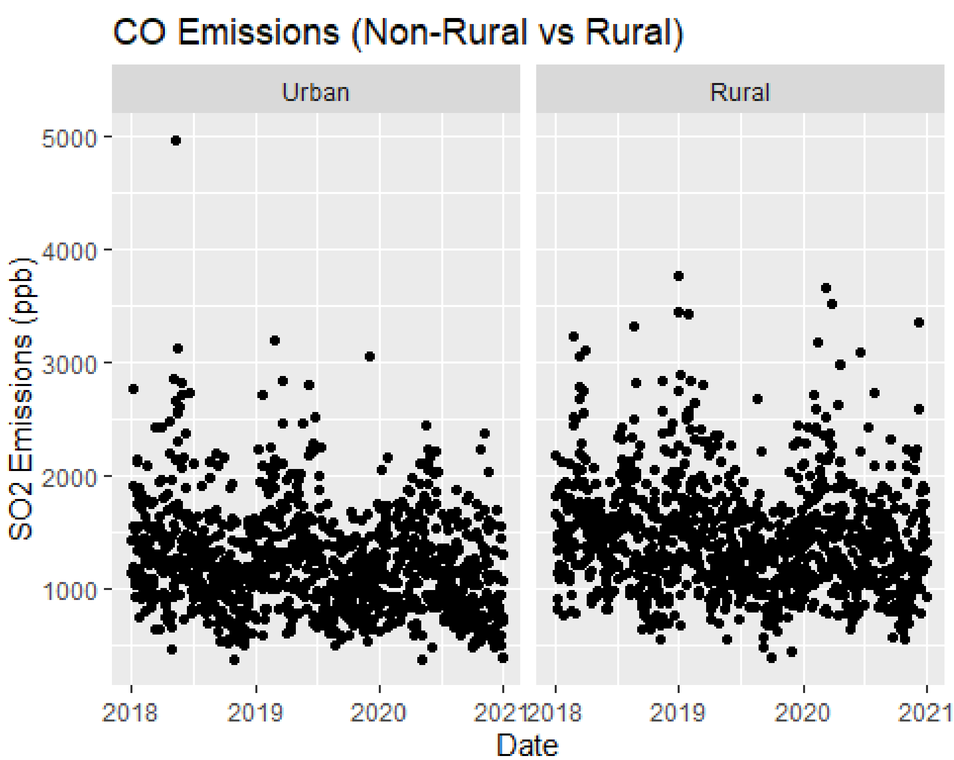


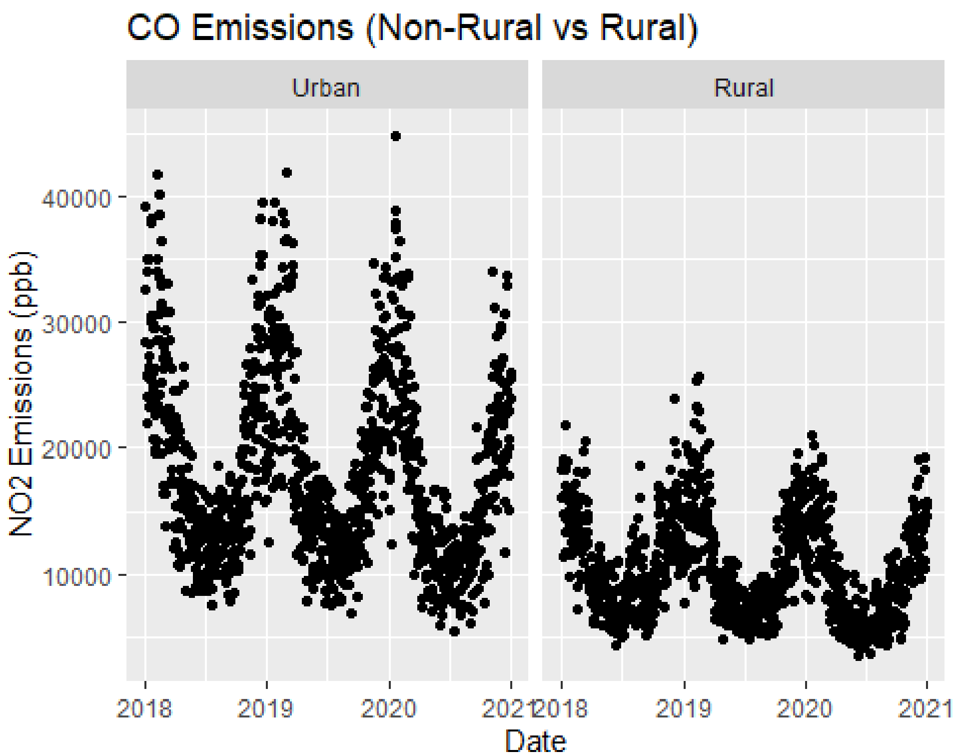



CO Emissions (Non-Rural vs Rural)
Urban
Rural
O3 Emissions (ppb)
1e+05
8e+04
6e+04
4e+04
2018
2019
2020
2021
Date


CO Emissions (Non-Rural vs Rural)
Urban
Rural
PM 10 Emissions (ppb)
60000
40000
20000
0
2018
2019
2020
2021
Date


CO Emissions (Non-Rural vs Rural)
Urban
Rural
PM2.5 Emissions (ppb)
150000
100000
50000
0
2018
2019
2020
2021
Date


CO Emissions (Non-Rural vs Rural)
Urban
Rural
SO2 Emissions (ppb)
5000
4000
3000
2000
1000
2018
2019
2020
2021
Date



CO Emissions (Non-Rural vs Rural)
Urban
Rural
NO2 Emissions (ppb)
40000
30000
20000
10000
2018
2019
2020
2021
Date