Effect of Air Quality Regulations on PM

Sai Omkar Kandukuri

18/05/2022

What is the impact of provincial air quality regulations on local particulate matter (PM 2.5)? We start by presenting the ideal experiment. in a completely unconstrained world, and describe the ideal dataset to have to perform the analysis.

Ideal Experiment:

For the same province i, we need to observe the following

1) Local particulate matter (PM 2.5) in air if the Air Quality regulation was introduced
2) Local particulate matter (PM 2.5) in air if the Air Quality regulation was not introduced

Difference between 1 and 2 gives us the effect of "Air Quality regulations" on the local particulate matter (PM 2.5) in air in province i.

Potential outcomes framework: Let i be the individual province where $i \in \{1, 2, ...N\}$. Treatment indicator D_i where $D_i \in \{0, 1\}$ Treated: $D_i = 1$: Air Quality pollution regulations introduced
Untreated: $D_i = 0$: Air Quality pollution regulations not introduced

Outcome treated: $Y_i(D_i = 1)$: Total local particulate matter in air for province i, when Air Quality regulation was introduced - Treatment Outcome untreated: $Y_i(D_i = 0)$: Total local particulate matter in air for province i, when Air Quality regulation was not introduced - Control

We get the impact of treatment(i.e disconnecting household's electricity) τ_i from the difference between the above outcomes.

- $\tau_i = Y_i(D_i = 1) - Y_i(D_i = 0)$

The impact of treatment τ_i is the difference between the two outcomes, the difference between total local particulate matter in air for province i when Air Quality pollution regulations was introduced vs total local particulate matter in air for province i when Air Quality pollution regulations was not introduced where all other factors are kept constant.

From above: - $\tau_i = Y_i(D_i = 1) - Y_i(D_i = 0)$

While we need both the outcomes at a given time to compute the impact of treatment, the problem is that at a given time, we cannot observe both the outcomes \$ or \$ we can only observe either \$Y_i(D_i = 1) \$ or \$ Y_i(D_i = 0)\$ at a given time.

In case a province is treated (i.e Air quality regulation introduced), then the observed outcome would be \$Y_i(D_i = 1) \$(Total local particulate matter in air for province i, when Air Quality regulation was introduced), and \$ Y_i(D_i = 0)\$ (Total local particulate matter in air for province i, when Air Quality regulation was not introduced) would become an unobserved outcome. Due to the un-observable outcome *or* not being able to observe both the outcomes at a given time, measuring τ_i is impossible.

Average Treatment Effect \$ tau^{ATE}\$

- $ tau^ATE = E[Y__i(D__i = 1)] - E[Y__i(D__i = 0)]$

ATE measures the average effect of treatment across a population of provinces. ATE measures the effect of Air Quality regulations on total local particulate matter. The problem is same in case of ATE. At the same time, for a province i, we cannot observe both outcomes. Hence it is impossible to measure ATE.

How would a realistic experiment look like?

An RCT where the treatment is assigned to provinces randomly. Then we can calculate the effect of treatment i.e Air Quality regulation on total local particulate matter. When the treatment assigned randomly and the distribution of the observables and the unobservables are same across the treated and untreated, we can take that there is no selection problem by design.

Hence we get, $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$ and $E[Y_i(0)|D_i = 0] = E[Y_i(0)]$

As a result, $tau^{ATE} = E[Y_i(D_i = 1)] - E[Y_i(D_i = 0)]$

Then the ATE will be equal to Naive estimator $\tau_N = \bar{Y}(D = 1) - \bar{Y}(D = 0)$

For this to workout, we assume that the outcome is solely affected by the treatment and there is 100% compliance and there are no spillover effects among the treated or control groups.

We have some data to work with. A single temporal snapshot of air quality across many municipalities. We'd like to look at average differences in air quality between municipalities with and without air quality regulations to get a sense of what these regulations do to air quality.

ATE measures the average effect of treatment across a population of provinces. ATE measures the effect of Air Quality regulations on total local particulate matter. The problem is same in case of ATE. At the same time, for a muncipality i, we cannot observe both outcomes. Hence it is impossible to measure ATE.

$\tau^{ATE} = E[Y_i(D_i = 1)] - E[Y_i(D_i = 0)]$

The problem is same in case of ATE. At the same time, for a muncipality i, we cannot observe both outcomes. Hence it is impossible to measure ATE.

If we compare the average differences in total particulate matter of muncipalities with and without air quality regulations, we are determinng a Naive Estimator

$\tau_N = \bar{Y}(D = 1) - \bar{Y}(D = 0)$

Here we observe $Y_i(D_i = 1)$ and $Y_j(D_j = 0)$, where i is not equal to j. It would've benefitted us if we observed $Y_i(D_i = 1)$ and $Y_i(D_i = 0)$ for the same i. The difference between ATE and the Naive estimator is that the prior is based on the potential outcomes and the latter is based on observed outcomes. And moreover, while calculating Naive estimator, it can be so that the treated muncipalities and the untreated muncipalities can differ substantially.

This brings us to the assumptions that the expectation of Y is same as (conditional expectation of Y that D_i is 1)and same as the (conditional expectation of Y given D_i is 0). We are assuming that the average of Y given $D_i = 1$ is a good counterfactual for when $D_i = 0$. Below in mathematical expression form:

$E[Y_i(1)] = E[Y_i(1)|D_i = 1] = E[Y_i(1)|D_i = 0]$
and
$E[Y_i(0)] = E[Y_i(0)|D_i = 0] = E[Y_i(0)|D_i = 1]$

Lets discuss three examples why this is not the ideal way to go about it.

There can be a problem when i and j significantly differ from each other. Which is, the units that receive treatment differ a lot from the units that donot receive any treatment on observables and the unobservables.Thus this brings us to the assumption stated above. This can lead to various problems as stated below:

1) Selection of Observables case Lets consider Air quality regulation was introduced i.e the treatment group in the muncipalities where stubble burning is high and Air quality regulation was not introduced i.e the control group in the muncipalities where stubble burning is low In a year of high stubble burning, the total particulate matter in air in the treatment muncipalities will be comparitively very high than that of the total particulate matter in air in the control group muncipalities. This is due to the effect of stubble burning in the treatment group muncipalities and no effect of the same in the control group muncipalities. This will result in a selection problem while determining the Naive estimator where the observable characteristic will result in an unestimation of the average effect of treatment

τ_N given by $\tau_N = \bar{Y}(D = 1) - \bar{Y}(D = 0)$

2) Selection of Unobservables case Lets consider that all the muncipalities have same levels of total particulate matter to start with. But the citizens of the muncipalities where the Air Quality regulations are introduced i.e the treatment group are passionate to rede their pollution levels than the citizens of the muncipalities where the Air QUality regulations are not introduced i.e the control group. In this case. the naive estimate as above will underestimate the average effect of treatment.

τ_N given by $\tau_N = \bar{Y}(D = 1) - \bar{Y}(D = 0)$

An overestimation case can also be constructed similarly to show that the Naive estimator would overestimate the average effect of the treatment.

3) Case of Non compliance Lets consider that some of the muncipalities in the treatment have administrations that are running on low budget. In these cases, there may be a situation of non-compliance where, even thugh the muncipality comes under treatment group, the administration doesnt take initiative to implement the air quality regulations. In this case, the Naive estimator provides an inaccurate calculation of the average effect of the treatment.

τ_N given by $\tau_N = \bar{Y}(D = 1) - \bar{Y}(D = 0)$

The same case can be constructed with the muncipalities in control group with proactive administrations running on high budget, implementing the Air quality regulations even though they are in the control group. This again leads to an inaccurate estimate of the average effect of the treatment.

The above are a few reasons why the Naive Estimator(τ_N) may not be a good estimate of the Average Treatment Effect (ATE) τ^{ATE}.

4) Spill overs case Spill overs can happen when a change in one of the muncipalities can affect other muncipalities. Lets say that there are two muncipalities that are not very distant to each other. And one is in treatment group and one is in control group. As a result of air quality regulations in the treatment muncipality made the air purer or with reduced local particulate matter (PM 2.5) can improve the air quality in the muncipality from control group. This is because both are not very distant from each other. This can be re worded as that the effect of treatment in the treatment group muncipality has spilled over to the control group muncipality. In a case like this, the Naive estimator τ_N estimates innacurately the ATE of air quality regulations.

$\tau_N = \bar{Y}(D = 1) - \bar{Y}(D = 0)$

Thus, the "Naive Estimator" τ_N, may not be a good estimate of the Average Treatment Effect (ATE) τ^{ATE}

What is the benefit of being able to observe municipalities at multiple points in time. The available database goes from 2001 to 2019.

In RCTs, the treated and control group have the same distribution of observables and the unobservables. Hence we can assume that there is no selection problem by design. But undertaking an RCT may not be practical in nature and may not be feasible. Thus to solve this problem, we can take a muncipality i and compare it with itself across different times such as t, t-1.

t-1 is before treatment, no air quality regulation i.e before 2004 t is after treatment, air quality regulation i.e after 2004

Consider a muncipality i. The outcome when treated is Y_t(D_t = 1) and when untreated is Y_t(D_t = 0).

Average Treatment Effect of Air Quality regulations: $\tau^{ATE} = E[Y_t(D_t = 1) - Y_t(D_t = 0)]$

Lets say t = 0 is period before 2004, i.e $D_{t=0} = 0$ and the period after 2004 t =1 when treatment is introduced $D_{t=1} = 1$, the ATE can be estimated using the difference estimator $\hat{\tau}^{TS}$, which is

$\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0}$

This is how through the difference estimator we compare a muncipality i with itself over time.

THe regression model for the same would be:

$Y_{it} = \tau.D_{it} + \beta.X_i$

Y_{it}: Outcome i.e the local particulate matter (PM 2.5), for municipality i in period t
τ: Estimate of $\tau_A TE$
$D_i t$: Treatment status, for municipality i in period t
$\beta.X_i$: Covariate X_i which has time-invariant observable characteristics, for muncipality i and its coefficient β

Now we can estimate ATE using the difference estimator as follows:

$\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0}$ gives

$$\hat{\tau}^{TS} = \tau(D_{i,t=1} - D_{i,t=0}) + \beta(X_i - X_i)$$

which transforms to $\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0} = \tau(1 - 0) + 0 = \tau$

Through this way,by taking differences over time for effect, we remove any time-invariant unobservable characteristics of a muncipality i and thus reach to an unbiased estimator of τ^{ATE} i.e the difference estimator $\hat{\tau}^{TS}$

Assumptions:

We assume that there is no change in muncipality i over the time period in question that affects the local particulate matter in air other than the air quality regulation introduction i.e the treatment. In other words, the counterfactual that is any observable and unobservable time variant characteristics are equal to 0. Also we assume that there are no time variant unobservables for muncipality i which affect the local particulate matter in air.

$Y_{it} = \tau.D_{it} + \beta.X_i + \gamma.U_i + \delta.V_{it}$

Y_{it}: Outcome - Local particulate matter (PM 2.5), for municipality i in period t
τ: Estimate of ATE $\tau_A TE$
$D_i t$: Treatment status, for municipality i in period t
$\beta.X_i$: Covariate X_i which has time-invariant observable characteristics, for muncipality i and its coefficient β
U_i: Time invariant unobservable characteristics, for muncipality i - something that doesnt change over time
V_{it}: Time variant observable and unobservable characteristics, for muncipality i - something that changes over time

Now we take the difference over time

$\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0} = \tau + \delta(V_{i,t=1} - V_{i,t=0})$

We need to have $\delta = 0$ or $(V_{i,t=1} - V_{i,t=0}) = 0$ for the $\hat{\tau}^{TS}$ to be equal to τ

Thus having time variant characteristics i.e something that changes over time will introduce bias in our estimation resulting for the $\hat{\tau}^{TS}$ to be a biased estimator of τ^{ATE}

Example of such occurances can be as follows:

1) Consider a case like whats happening in India now, a terrible heat wave. Not many people would go out and thus the emission from vehicles and industries reduce due to the reduced use and reduced work casued by the heatwave. In such a case where a muncipality i in 2004 had faced a situation like heatwave throughout, this would produce a characteristic that is observable and at the same time not seen or is different prior to 2004. Thus there will be a bias in the estimation of $\hat{\tau}^{TS}$. Also the counter factual $(V_{i,t=1} - V_{i,t=0}) \neq 0$. Due to the bias introduced by such a situation, $\hat{\tau}^{TS}$ will be a biased estimator for the estimation of ATE of treatment i.e the average effect of air quality regulation on the local particulate matter (PM 2.5).

2) Consider a muncipality i with a profound industrial sector. And most of the pollution is due to the emissions from these industries. Now lets say that the administration of a particular group of industries decided to move to a green energy based industry in 2004. THis will result in reduced emissions that is observable for this muncipality i. In this case the entry of this green energy industry would affect the outcome and will create a bias is in the $\hat{\tau}^{TS}$. Due to this bias, $\hat{\tau}^{TS}$ can be a biased estimator of τ^{ATE} ATE of outcome, i.e average effect of air quality regulations on the local particulate matter.

3) Lets assume a muncipality i where the temperatures are very cold in general. And the cold has increased in 2004 due to the change in weather. We know that due to the density of air in cold temperatures , the local particulate matter increases in absolute values. Thus for the muncipality i, the observable characteristic is different from prior to 2004 i.e before the colder temperatures. Then there will be a bias introduced in the $\hat{\tau}^{TS}$ and hence will be biased in estimating the ATE of outcome, i.e average effect of air quality regulations on the local particulate matter. Also the trend in the counter factual will be $(V_{i,t=1} - V_{i,t=0}) \neq 0$

Would it be even better to have data on multiple municipalities, divided into two groups: municipalities that never imposed air quality regulations, and municipalities that imposed air quality regulations in 2004? Would there be more concerns left?

We have the equation from above:

$\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0} = \tau + \delta(V_{i,t=1} - V_{i,t=0})$

We have seen that we need to have $\delta = 0$ or $(V_{i,t=1} - V_{i,t=0}) = 0$ for the $\hat{\tau}^{TS}$ to be equal to τ. This says that we dont have idea of the outcome for a treated muncipality i in t = 1, in case of no treatment. TO solve this , we can take approximate using the data from the group of muncipalities that never the said introduced air quality regulations.

This means we need data on the muncipalities that never introduced air quality regulations. Once we have the same, we can employ a difference in difference (DiD) estimator $\hat{\tau}^{DD}$, that helps to find differences across muncipalities, within time and within muncipality and across time comparisions from the data on muncipalities that never introduced air quality regulations.

$\hat{\tau}^{DD}$ compares treated to untreated muncipalities over time.

The equation for the same is:

$\hat{\tau}^{DD} = \hat{\tau}^{TS}_{D_i=1} - \hat{\tau}^{TS}_{D_i=0}$

and

$Y_{it} = \tau.D_{it} + \beta.X_i + \delta.S_t$ - treatment for muncipality i $Y_{jt} = \beta.X_i + \delta.S_t$

Y_{it}: Outcome - Local particulate matter , for municipality i in period t Y_{it}: Outcome - Local particulate matter , for municipality j in period t τ: Estimate of ATE $\tau_A TE$
$D_i t$: Treatment status, for municipality i in period t
$\beta.X_i$: Covariate X_i which has time-invariant observable characteristics, for muncipality i and its coefficient β S_t: Time variant observable and unobservable characteristics and delta

Thus the DiD estimator compares a (muncipality i with itself over time (t, t-1)) to a (muncipality j with itself over time (t, t-1).

The equation for muncipality i gives, $Y_{i,t=1} - Y_{i,t=0} = \tau.(D_{i,t=1} - D_{i,t=0}) + \beta.(X_i - X_i) + \delta.(S_{t=1} - S_{t=0}) = \tau.(D_{i,t=1} - D_{i,t=0}) + \delta.(S_{t=1} - S_{t=0})$

Similarly for muncipality j,, we have $Y_{j,t=1} - Y_{j,t=0} = \beta.(X_i - X_i) + \delta.(S_{t=1} - S_{t=0}) = \delta.(S_{t=1} - S_{t=0})$

We take difference of the above two equations to determine the DiD estimator:

$\hat{\tau}^{DD} = Y_{i,t=1} - Y_{i,t=0} - Y_{j,t=1} - Y_{j,t=0} = \tau.(D_{i,t=1} - D_{i,t=0}) = \tau(1 - 0) = \tau$

From the above, we can say that the DiD estimator is a good estimator of the ATE.

Assumptions:

1) Parallel counterfactual trends: We assume that treatment is as good as randomly assigned ,as we assume that in absence of treatment time trend shoudl be same across muncipalities i and muncipalities j.

This assumption will make the DiD a good estimator for ATE, otherwise as we saw above, there would be bias in the estimation of ATE.

Example: Lets say that muncipality i which introduced air quality regulations has also introduced anothe program where citizens start community approaches to reduce emmisions through switch to low emmission alternatives. Due to these community approach programs the outcome i.e the total particulate matter is reduced significantly for muncipality i. Thus the parallel trends assumption is not satisfield as the trend changed in 2004 due to introduction of another program.

Similar example for muncipality j which did not introduce air quality regulation. But a new low emmision alternative started to be available in the market in 2004. The citizens who are aware of the effects of high pollution started to switch to this new low emmission alternative available in the market such as affordable electric vehicles. Due to this change in trend, the outcome i.e total particulate matter in air for the muncipality j is significantly reduced. Thus the parallel trends assumption is not satisfied as the trend changed in 2004 due to new trend.

In both of these above cases, we observed that there isn't same trend across the muncipalities i and muncipalities j in the absence of treatment. Thus we need the parallel counterfactual trends assumption to have DiD estimator τ^{DD} become an good unbiased estimator of ATE τ^{ATE}.

Now we are given data on the universe of consumers from 2003 to 2007. This includes municipalities that imposed air quality regulations across several different years. How can we use to this to find our estimates?

The simple panel data regression equation is

$Y_{it} = X_{it}\beta + \tau D_{it} + \epsilon_{it}$

and $\epsilon_{it} = \alpha_i + \delta_t + v_{it}$

α_i comprises the municipality variant time invariant characteristics effects and δ_t is the municipality invariant time variant effects and $v_i t$ is the muncipality and time variant characteristics effects

Given that the data includes muncipalities that imposed air quality regulations across several different years. Hence the simple Fixed Effects Model which is a general case of DiD model can be used to estimate treatment effect.

$Y_{it} = X_{it}\beta + \tau D_{it} + \sum_{i=1}^{N} 1[unit = i]_i + \sum_{t=2003}^{2007} 1[time = t]_t + v_{it}$

$Y_i t$: Outcome - the total particulate matter level, for municipality i in time t
$D_i t$: Treatment status, for municipality i, time period t in year
$X_i t$: Muncipality varying and time varying observables effects
$v_i t$: Unobservable municipality specific time varyiant effects which, each muncipality i for each year t $\sum_{i=1}^{N} 1[unit = i]_i$: Dummy variables to determine the municipality specific time invariant fixed effects
$\sum_{t=1}^{T} 1[time = t]_t$: Dummy variables to determine the time variant fixed effects same across all the municipalities

Thus by introducing these effects from the Fixed effects model to the DiD model , we get the new regression equation as

$Y_i = \alpha + \tau Treat * Post_{it} + \beta Treat_i + \delta Post_t + \beta X_{it} + \epsilon_{it}$

Reduced to

$Y_i = \tau D_{it} + \alpha_i + \delta_t + \beta X_{it} + \epsilon_{it}$

We can also add pre-treatment effects to the above model and perform an Event study design. The equation for the same will be as follows:

$Y_{it} = \sum_{s=2003}^{2007} \tau_s D_i.1[periods\ to\ treatment = s]_{it} + \alpha_i + \delta_t + \beta.X_{it} + \epsilon_{it}$

Here

Y_{it}: Outcome - the total particulate matter level, for municipality i in time t α_i: Individual fixed effects
δ_t: Time fixed effects
ϵ_t: Error term which comprises time variant and time invariant observables and unobservables. $\tau_s D_i.1[periods\ to\ treatment = s]_{it}$: ATE estimate for municipality i with treatment status $D_i t$, that imposed the regulation in time periods s X_i: Covariate that comprises the time invariant observable characteristics, for the municipality i

Dataset: ps4_data.csv. Dataset to implement a simple comparison of average particulate matter between municipalities with and without air quality regulations.

Lets use regression analysis to perform a time-series analysis of the effect of air quality regulations on particulate matter, using only municipalities who introduced regulations in 2004.

Lets answer the following questions:

1) How does this differ from what you estimated using the initial estimator?

Plot particulate matter against time for municipalities that imposed air quality restrictions in 2004. What do we see?

Does this plot affect how we interpret the estimates?

```
data <- read_csv('ps_4_data.csv')
```

```
## Rows: 22000 Columns: 4
## -- Column specification --------------------------------------------------------
```

```
## Delimiter: ","
## dbl (4): year, municipality_id, air_quality_regulation_year, particulate_matter
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#We have year and particulate matter numeric values, lets mutate type to numeric
data <- data %>%
          mutate(year = as.numeric(year)) %>%
          mutate(particulate_matter = as.numeric(particulate_matter))
```

```
summary(data)
```

```
##       year      municipality_id  air_quality_regulation_year particulate_matter
##  Min.   :2003   Min.   :   1.0   Min.   :2002                Min.   :-67.42
##  1st Qu.:2005   1st Qu.: 550.8   1st Qu.:2002                1st Qu.: 20.90
##  Median :2008   Median :1100.5   Median :2002                Median : 53.06
##  Mean   :2008   Mean   :1100.5   Mean   :2003                Mean   : 54.05
##  3rd Qu.:2010   3rd Qu.:1650.2   3rd Qu.:2005                3rd Qu.: 85.37
##  Max.   :2012   Max.   :2200.0   Max.   :2007                Max.   :321.40
##                                  NA's   :10990
```

```
#We have na's in the data, replace them with 0
data[is.na(data)] = 0
summary(data)
```

```
##       year      municipality_id  air_quality_regulation_year particulate_matter
##  Min.   :2003   Min.   :   1.0   Min.   :   0                Min.   :-67.42
##  1st Qu.:2005   1st Qu.: 550.8   1st Qu.:   0                1st Qu.: 20.90
##  Median :2008   Median :1100.5   Median :2002                Median : 53.06
##  Mean   :2008   Mean   :1100.5   Mean   :1003                Mean   : 54.05
##  3rd Qu.:2010   3rd Qu.:1650.2   3rd Qu.:2002                3rd Qu.: 85.37
##  Max.   :2012   Max.   :2200.0   Max.   :2007                Max.   :321.40
```

```
data <-
  data %>%
  mutate(ever_regulated = ifelse(air_quality_regulation_year != 0, 1, 0))

data %>%
  group_by(ever_regulated) %>%
  summarise_at(vars(particulate_matter), list(name = mean))
```

```
## # A tibble: 2 x 2
##   ever_regulated  name
##            <dbl> <dbl>
## 1              0  66.3
## 2              1  41.8
```

```
reg1 <- lm(particulate_matter ~ ever_regulated, data = data)
summary(reg1)
```

```
##
## Call:
## lm(formula = particulate_matter ~ ever_regulated, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -109.870  -25.773    3.053   24.072  279.570
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     66.2814     0.4943  134.10   <2e-16 ***
## ever_regulated -24.4486     0.6987  -34.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.81 on 21998 degrees of freedom
## Multiple R-squared:  0.05273,    Adjusted R-squared:  0.05269
## F-statistic:  1224 on 1 and 21998 DF,  p-value: < 2.2e-16
```

From the comparision of average particulate matter between muncipalities with and without air quality regulation
Average particulate matter in municipalities with air quality regulations is 41.83 and the Average particulate matter in municipalities without air quality regulations is 66.28. The difference in average being -24.44.

```
data_2004 <-
  data %>%
  filter(air_quality_regulation_year==2004)

data_2004 <-
  data_2004 %>%
  mutate(is_regulated2004=ifelse(year<2004,0,1))

summary(data_2004)
```

Regression to perform time series analysis of effect of treatment using only the muncipalities that introduced regulations in 2004.

```
##       year      municipality_id air_quality_regulation_year particulate_matter
##  Min.   :2003   Min.   :1800    Min.   :2004                Min.   :-58.27
##  1st Qu.:2005   1st Qu.:1825    1st Qu.:2004                1st Qu.: 10.36
##  Median :2008   Median :1850    Median :2004                Median : 39.14
##  Mean   :2008   Mean   :1850    Mean   :2004                Mean   : 26.81
##  3rd Qu.:2010   3rd Qu.:1874    3rd Qu.:2004                3rd Qu.: 53.54
##  Max.   :2012   Max.   :1899    Max.   :2004                Max.   : 81.05
##  ever_regulated is_regulated2004
##  Min.   :1      Min.   :0.0
##  1st Qu.:1      1st Qu.:1.0
##  Median :1      Median :1.0
##  Mean   :1      Mean   :0.9
##  3rd Qu.:1      3rd Qu.:1.0
##  Max.   :1      Max.   :1.0
```

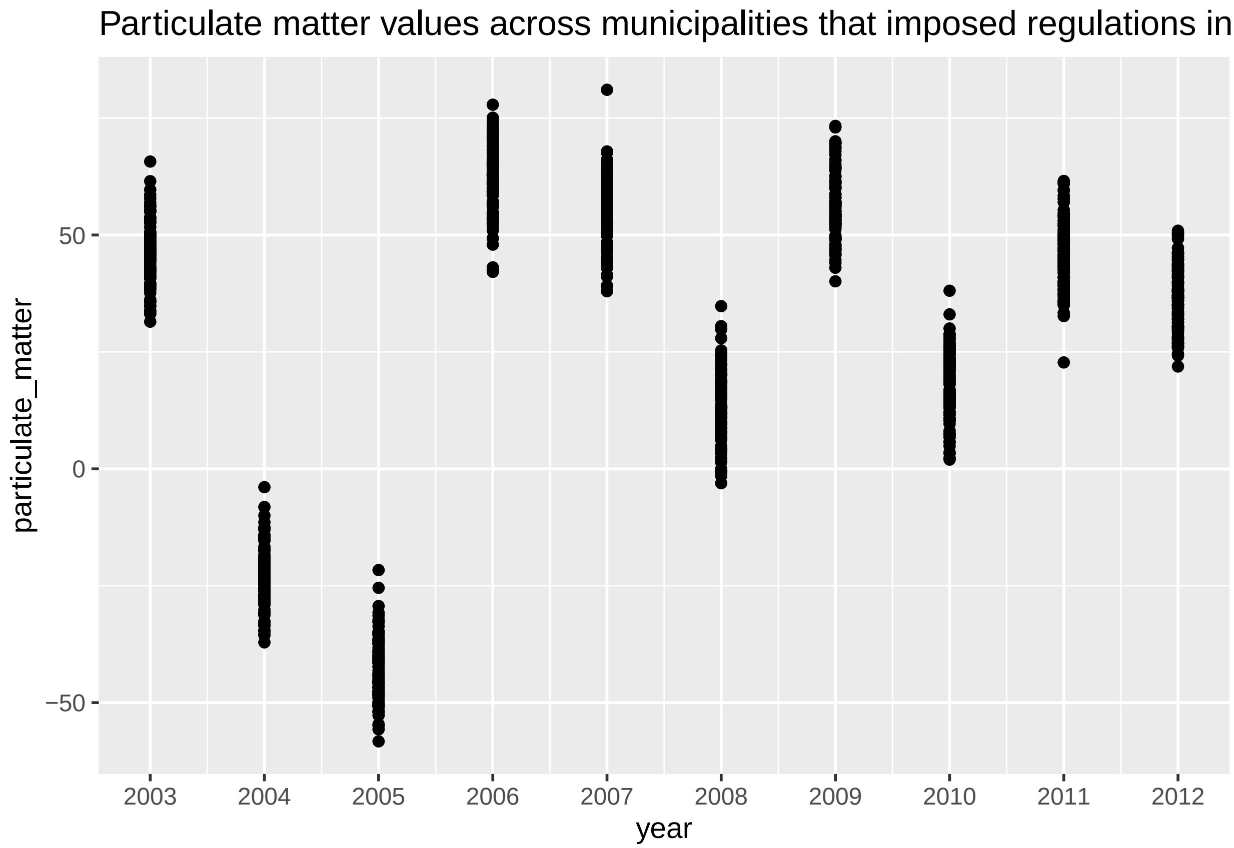
```r
reg2 <- lm(particulate_matter ~ is_regulated2004, data_2004)
summary(reg2)
```

```
##
## Call:
## lm(formula = particulate_matter ~ is_regulated2004, data = data_2004)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.937 -14.320   5.748  28.010  56.389
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        46.132      3.403  13.556  < 2e-16 ***
## is_regulated2004  -21.470      3.587  -5.985 3.01e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.03 on 998 degrees of freedom
## Multiple R-squared:  0.03465,    Adjusted R-squared:  0.03369
## F-statistic: 35.83 on 1 and 998 DF,  p-value: 3.006e-09
```

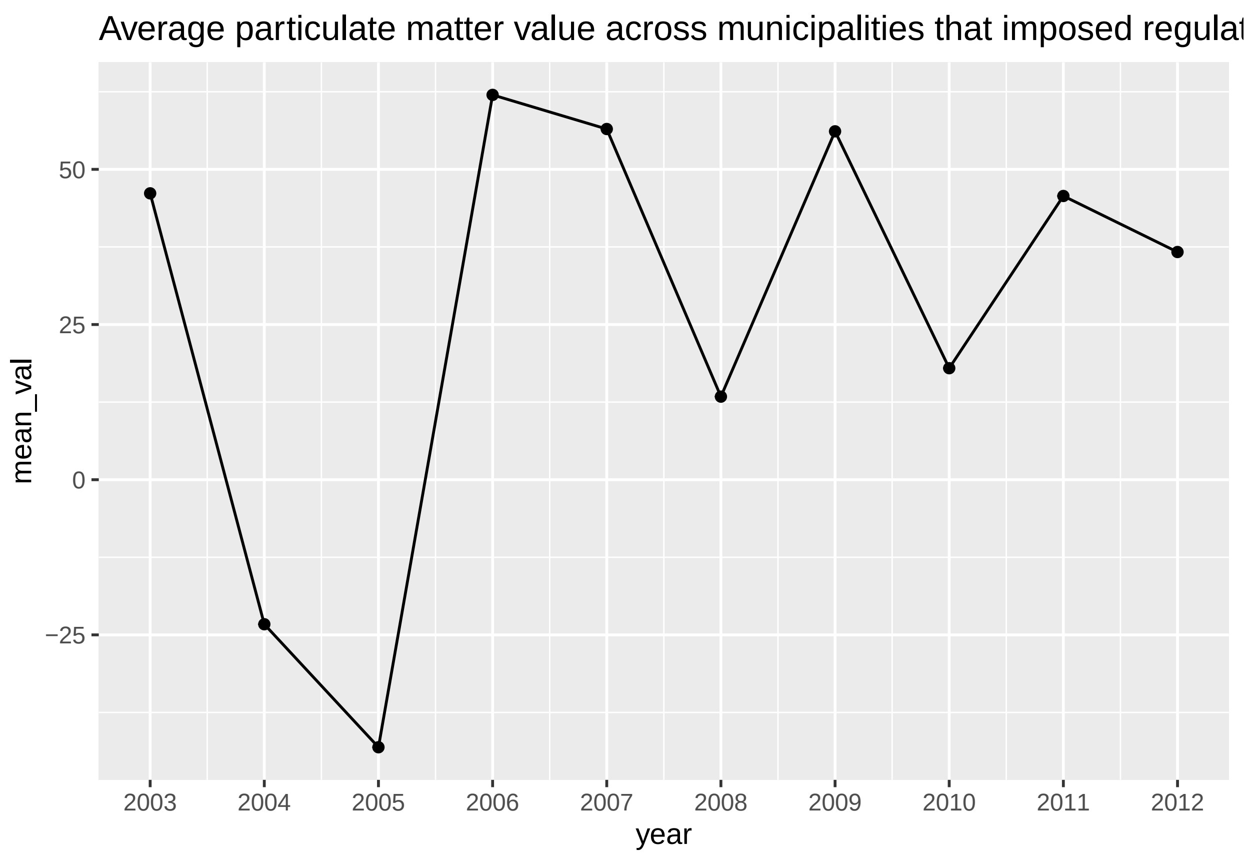
Average effect of air quality regulations on particulate matter using only municipalities who introduced regulations in 2004 = -21.470 Difference in averages of particulate matter in municipalities before introduction of air quality regulations in 2004 and after introduction of air quality regulations in 2004 = -21.470

From the analysis above, we observed that the difference in averages of particulate matter between municipalities with and without air quality regulations is -24.4486 and from here we observe that the difference in averages of particulate matter before 2004 and after 2004 for municipalities who introduced the regulations in 2004 is -21.470

```r
pm_mean_values <-
  data_2004 %>%
  group_by(year) %>%
  summarise(mean_val = mean(particulate_matter))


plot1 <- ggplot(data_2004, aes(x=year, y=particulate_matter)) + geom_point() +
  scale_x_continuous(breaks=c(2003,2004,2005,2006,2007,2008,2009,2010,2011,2012)) +
  ggtitle("Particulate matter values across municipalities that imposed regulations in 2004")
plot1
```




```r
mean_plot2 <- ggplot(pm_mean_values, aes(x=year, y=mean_val)) + geom_point() + geom_line()+
  scale_x_continuous(breaks=c(2003,2004,2005,2006,2007,2008,2009,2010,2011,2012))+
  ggtitle("Average particulate matter value across municipalities that imposed regulations in 2004")
mean_plot2
```



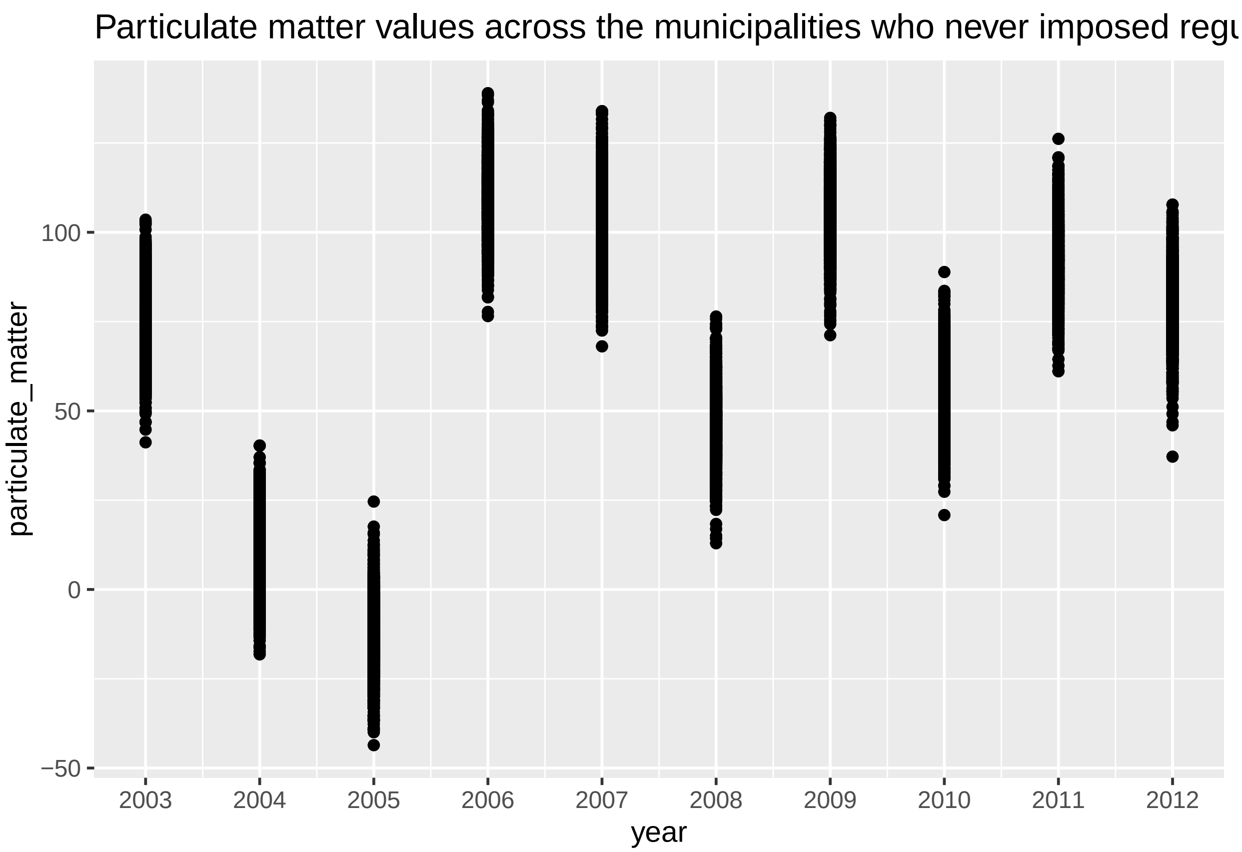
We observe a downward trend in the total particulate matter from 2003 till 2005 across the muncipalities that had implemented the air quality regulations in 2004. The same is the case for the mean particulate matter values. However, we also see that there is a steep increase to the pre-2004 levels in 2006 for both the total particulate matter and mean particulate values. Post to that the mean particulate matter varies with no continuous trend. This couldve have resulted in the lower difference in averages of particulate matter for the muncipalities that introduced the regulations in 2004, before and after 2004.From the plots, we cannot say anything in respect to the time variant and invariant characteristics regarding the trend we see in the plots.

Let's plot (average) particulate matter against time for municipalities who never imposed air quality regulations and assess the viability of using these municipalities as a control group for the 2004 regulators. Also, plot (average) particulate matter against time for municipalities who passed air quality regulation in 2006 and assess the viability of using the non-regulating municipalities as a control group for the 2006 regulators.

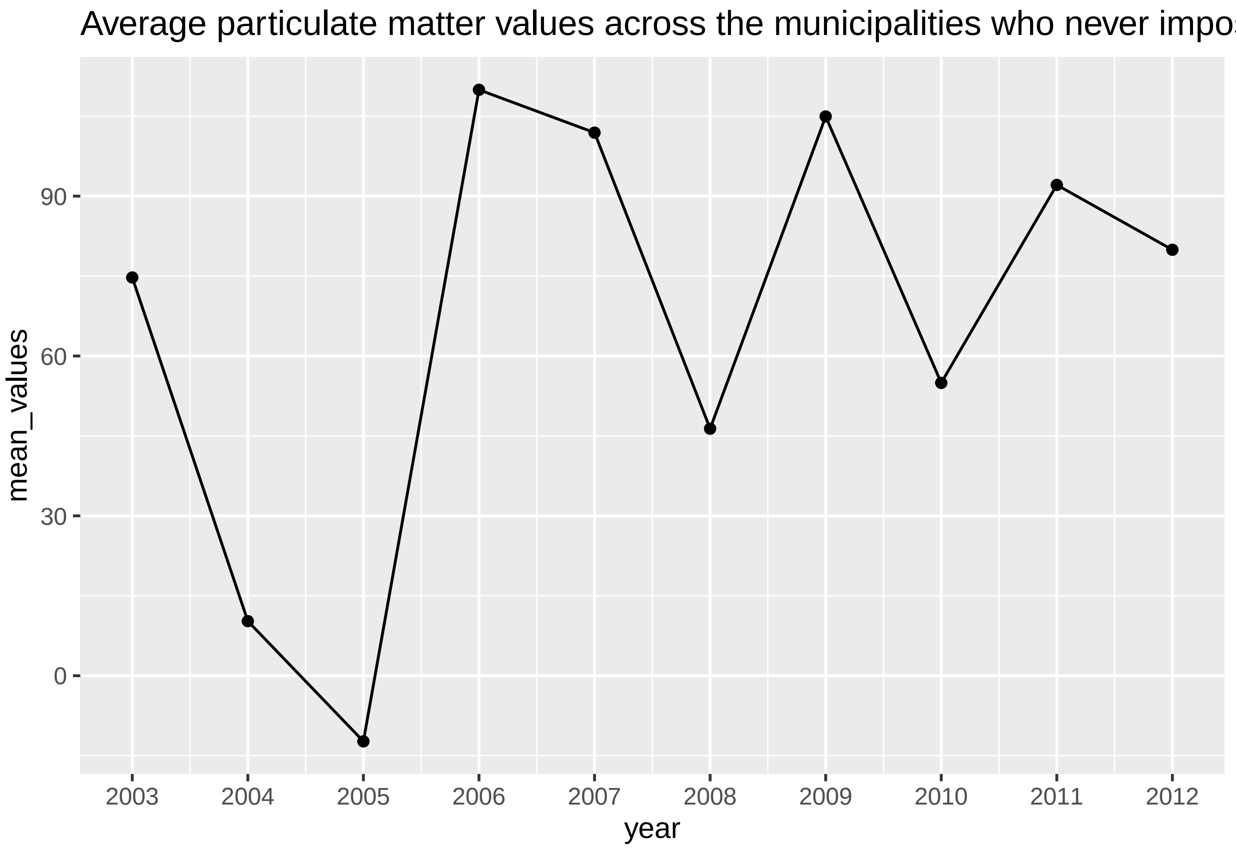
Plot (average) particulate matter against time for municipalities who never imposed air quality regulations

```
data_no_regulation <-
  data %>%
  filter(air_quality_regulation_year==0)

pm_mean_values2 <-
  data_no_regulation %>%
  group_by(year) %>%
  summarise(mean_values = mean(particulate_matter))
```

```r
plot3 <- ggplot(data_no_regulation, aes(x=year, y=particulate_matter)) + geom_point() +
  scale_x_continuous(breaks=c(2003,2004,2005,2006,2007,2008,2009,2010,2011,2012)) +
  ggtitle("Particulate matter values across the municipalities who never imposed regulations")
plot3
```



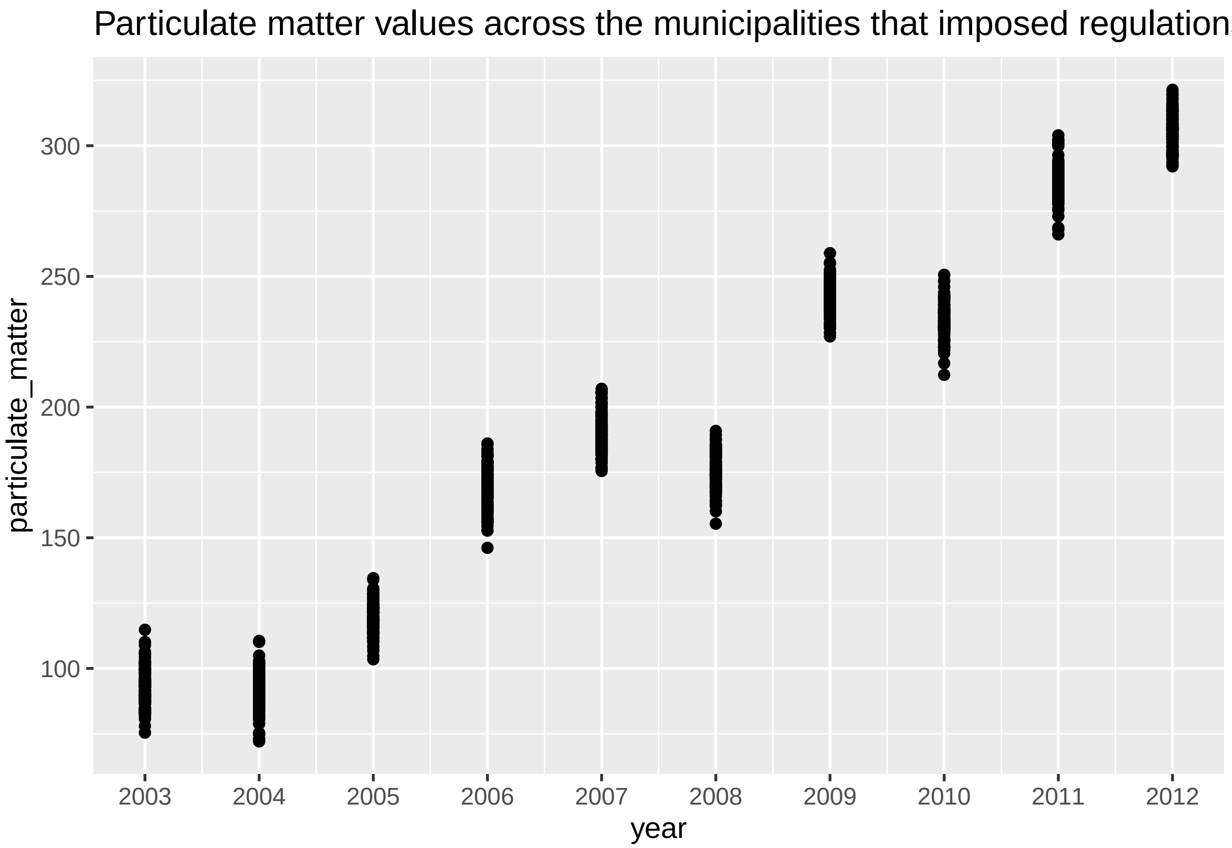

```r
mean_plot4 <- ggplot(pm_mean_values2, aes(x=year, y=mean_values)) +
  geom_point() +
  geom_line()+
  scale_x_continuous(breaks=c(2003,2004,2005,2006,2007,2008,2009,2010,2011,2012))+
  ggtitle("Average particulate matter values across the municipalities who never imposed regulations")
mean_plot4
```



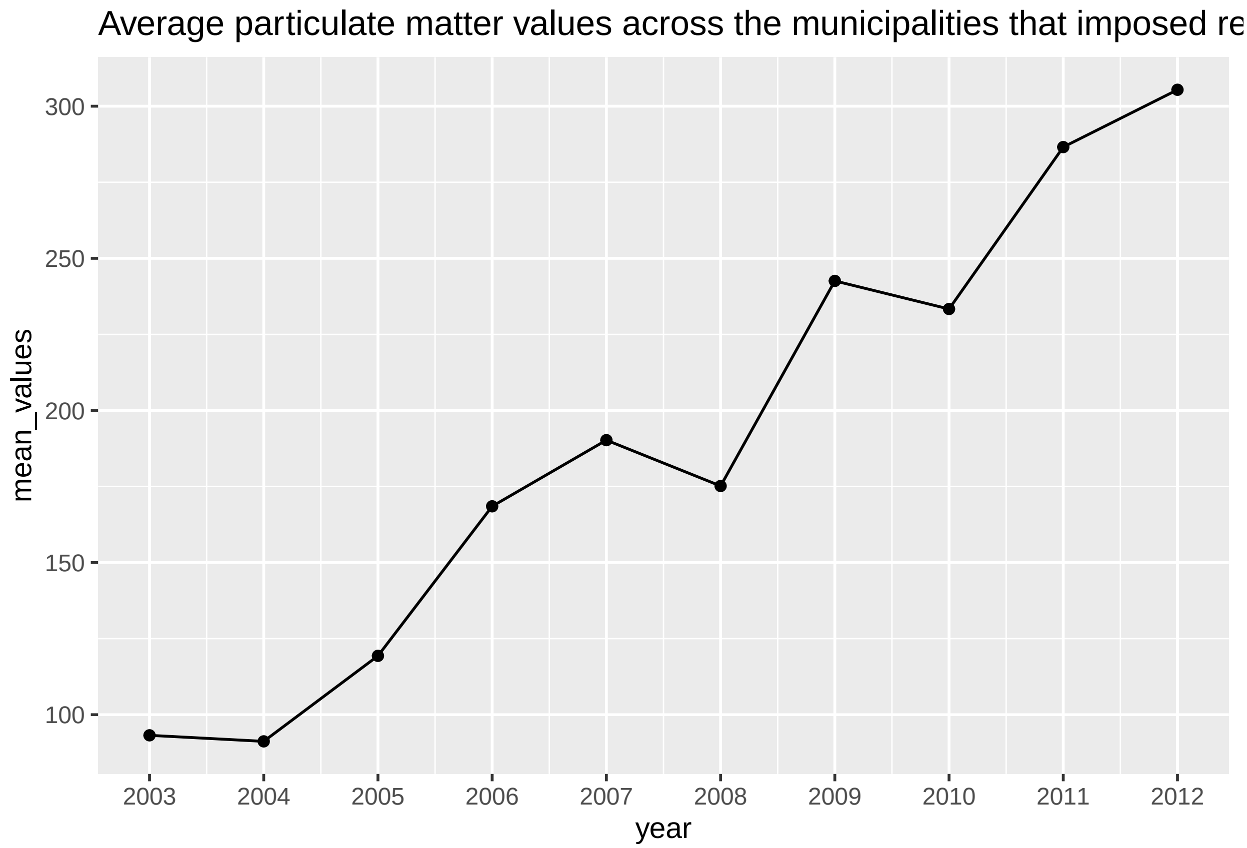
The total and mean particulate matter for the muncipalities that had never imposed regulations over the years 2003 to 2012 is similar to the muncipalities that imposed the regulations. We observe a parallel trend between them. Thus, assuming parallel trend is satisfied, we can analyse by keeping these muncipalities group as control group to determine the Did estimator $\hat{\tau}^{DD}$ for the 2004 regulators.

2006 Plot:

```
data_2006 <-
  data %>%
  filter(air_quality_regulation_year==2006)

pm_mean_values3 <-
  data_2006 %>%
  group_by(year) %>%
  summarise(mean_values = mean(particulate_matter))


plot5 <- ggplot(data_2006, aes(x=year, y=particulate_matter)) + geom_point() +
  scale_x_continuous(breaks=c(2003,2004,2005,2006,2007,2008,2009,2010,2011,2012)) +
  ggtitle("Particulate matter values across the municipalities that imposed regulations in 2006")

plot5
```




```r
mean_plot6 <- ggplot(pm_mean_values3, aes(x=year, y=mean_values)) + geom_point() + geom_line()+
  scale_x_continuous(breaks=c(2003,2004,2005,2006,2007,2008,2009,2010,2011,2012))+
  ggtitle("Average particulate matter values across the municipalities that imposed regulations in 2006

mean_plot6
```



We see a different trend from above. The total particulate matter values and the mean particulate values for the muncipalities that had never imposed regulations throughout the years 2003 to 2012 has a increasing trend and is different to what we observed with the muncipalities that imposed regulations in 2006. Hence there is no parallel trend observed. As we see that the parallel trend assumption is not satisfied, we cannot use these muncipalities group to determine the DiD estimator for the ATE for 2006 regulators.

Using just the non-regulators and the 2006 regulators, let's estimate the causal impact of imposing air quality regulation on particulate matter. To do this, we begin with a simple difference in means (rather than regression). Next, we use a simple regression (no fixed effects). Finally, we use fixed effects to control for common time shocks and time-invariant municipality characteristics (we can do this either via dummy variables or de-meaning). Always, be sure to adjust the standard errors appropriately in the regression-based estimates. Describe how this compares to what you estimated above.

```
data_no_regulation_2006 <-
  data %>%
  filter(air_quality_regulation_year==0 | air_quality_regulation_year==2006)


data_no_regulation_2006 <-
  data_no_regulation_2006 %>%
  mutate(is_treated = ifelse((year>= 2006), 1, 0))

pm_mean_values4 <-
  data_no_regulation_2006 %>%
```

```r
  group_by(air_quality_regulation_year,is_treated) %>%
  summarize(mean = mean(particulate_matter))
```

```
## `summarise()` has grouped output by 'air_quality_regulation_year'. You can
## override using the `.groups` argument.
```

```r
pm_mean_values4
```

```
## # A tibble: 4 x 3
## # Groups:   air_quality_regulation_year [2]
##   air_quality_regulation_year is_treated  mean
##                         <dbl>      <dbl> <dbl>
## 1                           0          0  24.2
## 2                           0          1  84.3
## 3                        2006          0 101.
## 4                        2006          1 229.
```

Difference in Means:

$Differenceinmeans = \bar{Y}_{i=treat,t=post} - Y_{i=treat,t=pre} - Y_{j=control,t=post} - Y_{j=control,t=pre}$

Calculation Difference in means $Differenceinmeans = (228.827 - 101.265) - (84.309 - 24.216) = 67.46895$

Regression without fixed effects:

```r
data_no_regulation_2006_2 <-
  data %>%
  filter(air_quality_regulation_year==0 | air_quality_regulation_year==2006)

# Defining treatment and control groups using dummy variable 'is_after_2006'
data_no_regulation_2006_2 <-
  data_no_regulation_2006_2 %>%
  mutate(is_treated = ifelse((year>= 2006), 1, 0))

# Defining Individual fixed affects
data_no_regulation_2006_2 <-
  data_no_regulation_2006_2 %>%
  mutate(is_regulated_2006 = ifelse((air_quality_regulation_year == 2006), 1, 0))

reg_4 <- lm(particulate_matter ~ is_regulated_2006 + is_treated + is_regulated_2006 * is_treated,
                  data = data_no_regulation_2006_2)
summary(reg_4)
```

```
##
## Call:
## lm(formula = particulate_matter ~ is_regulated_2006 + is_treated +
##     is_regulated_2006 * is_treated, data = data_no_regulation_2006_2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.681 -26.395   1.227  21.816  92.576
##
## Coefficients:
```

```
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    24.2163     0.5407   44.79   <2e-16 ***
## is_regulated_2006              77.0488     1.8722   41.15   <2e-16 ***
## is_treated                     60.0929     0.6462   92.99   <2e-16 ***
## is_regulated_2006:is_treated   67.4689     2.2377   30.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31.05 on 11986 degrees of freedom
## Multiple R-squared:  0.6916, Adjusted R-squared:  0.6915
## F-statistic:  8960 on 3 and 11986 DF,  p-value: < 2.2e-16
```

The difference of means from using DiD an estimator to determine effect for regression is 67.4689

Using Fixed Effects:

```
#Using felm
data_no_regulation_2006_2$municipality_id <- as.factor(data_no_regulation_2006_2$municipality_id)

reg6 <- felm(particulate_matter ~ is_regulated_2006 + is_treated
             + is_regulated_2006 * is_treated | municipality_id + year|0|0, cluster ="municipality_id",
             data = data_no_regulation_2006_2)
```

```
## Warning: Argument(s) clustervar are deprecated and will be removed, use
## multipart formula instead
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or not positive definite
```

```
summary(reg6)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or not positive definite
```

```
##
## Call:
##    felm(formula = particulate_matter ~ is_regulated_2006 + is_treated +      is_regulated_2006 * is_t
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -100.596   -7.562   -0.004    7.523   90.483
##
## Coefficients:
##                               Estimate Cluster s.e. t value Pr(>|t|)
## is_regulated_2006                  NaN       0.0000     NaN      NaN
## is_treated                         NaN       0.0000     NaN      NaN
## is_regulated_2006:is_treated   67.4689       0.5446   123.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.09 on 10781 degrees of freedom
## Multiple R-squared(full model): 0.9058   Adjusted R-squared: 0.8953
```

```
## Multiple R-squared(proj model): 0.199   Adjusted R-squared: 0.1092
## F-statistic(full model, *iid*):85.86 on 1208 and 10781 DF, p-value: < 2.2e-16
## F-statistic(proj model):  5115 on 3 and 1198 DF, p-value: < 2.2e-16
```

```r
#plm

reg7 <- plm(particulate_matter ~ is_regulated_2006 + is_treated +
                    is_regulated_2006 * is_treated, data = data_no_regulation_2006_2,
                    index = c("municipality_id","year"),
                    model = "within")

summary(reg7)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = particulate_matter ~ is_regulated_2006 + is_treated +
##     is_regulated_2006 * is_treated, data = data_no_regulation_2006_2,
##     model = "within", index = c("municipality_id", "year"))
##
## Balanced Panel: n = 1199, T = 10, N = 11990
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -82.1166 -26.6034   1.3567  21.6080  92.8581
##
## Coefficients:
##                              Estimate Std. Error t-value  Pr(>|t|)
## is_treated                   60.09291    0.67771  88.670 < 2.2e-16 ***
## is_regulated_2006:is_treated 67.46895    2.34669  28.751 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    23188000
## Residual Sum of Squares: 11436000
## R-Squared:      0.50679
## Adj. R-Squared: 0.45193
## F-statistic: 5543.04 on 2 and 10789 DF, p-value: < 2.22e-16
```

```r
data_demean_fixed <- with(data_no_regulation_2006_2,
            data.frame(particulate_matter = particulate_matter -
                         ave(particulate_matter, municipality_id), is_treated, is_regulated_2006))

reg8 <- lm(particulate_matter ~ is_treated + is_regulated_2006 +
             is_treated * is_regulated_2006, data = data_demean_fixed)

summary(reg8)
```

```
##
## Call:
## lm(formula = particulate_matter ~ is_treated + is_regulated_2006 +
##     is_treated * is_regulated_2006, data = data_demean_fixed)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.117 -26.603   1.357  21.608  92.858
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -42.065      0.538  -78.19   <2e-16 ***
## is_treated                     60.093      0.643   93.46   <2e-16 ***
## is_regulated_2006             -47.228      1.863  -25.35   <2e-16 ***
## is_treated:is_regulated_2006   67.469      2.226   30.30   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.89 on 11986 degrees of freedom
## Multiple R-squared:  0.5068, Adjusted R-squared:  0.5067
## F-statistic:  4105 on 3 and 11986 DF,  p-value: < 2.2e-16
```

```r
#Coefficient Test
coeftest(reg8)
```

```
##
## t test of coefficients:
##
##                               Estimate Std. Error t value  Pr(>|t|)
## (Intercept)                  -42.06504    0.53796 -78.194 < 2.2e-16 ***
## is_treated                    60.09291    0.64298  93.460 < 2.2e-16 ***
## is_regulated_2006            -47.22826    1.86276 -25.354 < 2.2e-16 ***
## is_treated:is_regulated_2006  67.46895    2.22643  30.304 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observations from above Average effect of air quality regulations on particulate matter using only municipalities who introduced regulations in 2004 = -21.470 Difference in averages of particulate matter in municipalities before introduction of air quality regulations in 2004 and after introduction of air quality regulations in 2004 = -21.470

We observed a downward trend in the total particulate matter from 2003 till 2005 across the muncipalities that had implemented the air quality regulations in 2004. The same is the case for the mean particulate matter values. However, we also see that there is a steep increase to the pre-2004 levels in 2006 for both the total particulate matter and mean particulate values. Post to that the mean particulate matter varies with no continuous trend. This couldve have resulted in the lower difference in averages of particulate matter for the muncipalities that introduced the regulations in 2004, before and after 2004.From the plots, we cannot say anything in respect to the time variant and invariant characteristics regarding the trend we see in the plots.

Observations from above The total and mean particulate matter for the muncipalities that had never imposed regulations over the years 2003 to 2012 is similar to the muncipalities that imposed the regulations. We observe a parallel trend between them. Thus, assuming parallel trend is satisfied, we can analyse by keeping these muncipalities group as control group to determine the Did estimator $\hat{\tau}^{DD}$ for the 2004 regulators.

The total particulate matter values and the mean particulate values for the muncipalities that had never imposed regulations throughout the years 2003 to 2012 has a increasing trend and is different to what we observed with the muncipalities that imposed regulations in 2006. Hence there is no parallel trend observed. As we see that the parallel trend assumption is not satisfied, we cannot use these muncipalities group to determine the DiD estimator for the ATE for 2006 regulators.

Observations from above

We saw the estimated τ^{DiD} is 67.46 in the simple regression without fixed effects. This means that the effect of introducing air quality regulation has effected the local particulate matter to increase by 67.468. We also saw the same results when we applied fixed effects to control time invariant muncipality characteristics and the time changes. We saw an estimated τ^{DiD} values of 67.468. This means introducing air quality regulation resulted in an average increase of local particulate matter by 67.468.

As we saw above observations, the parallel trends assumption is not satisfied for the muncipalities gorup that never introduced regulations. Also the muncipalities group that has never introduced regulations cannot be used as a control group for the 2006 regulators when using the DiD estimator $\hat{\tau}^{DD}$ for estimating the ATE τ^{ATE}. This also holds true from what we saw here.

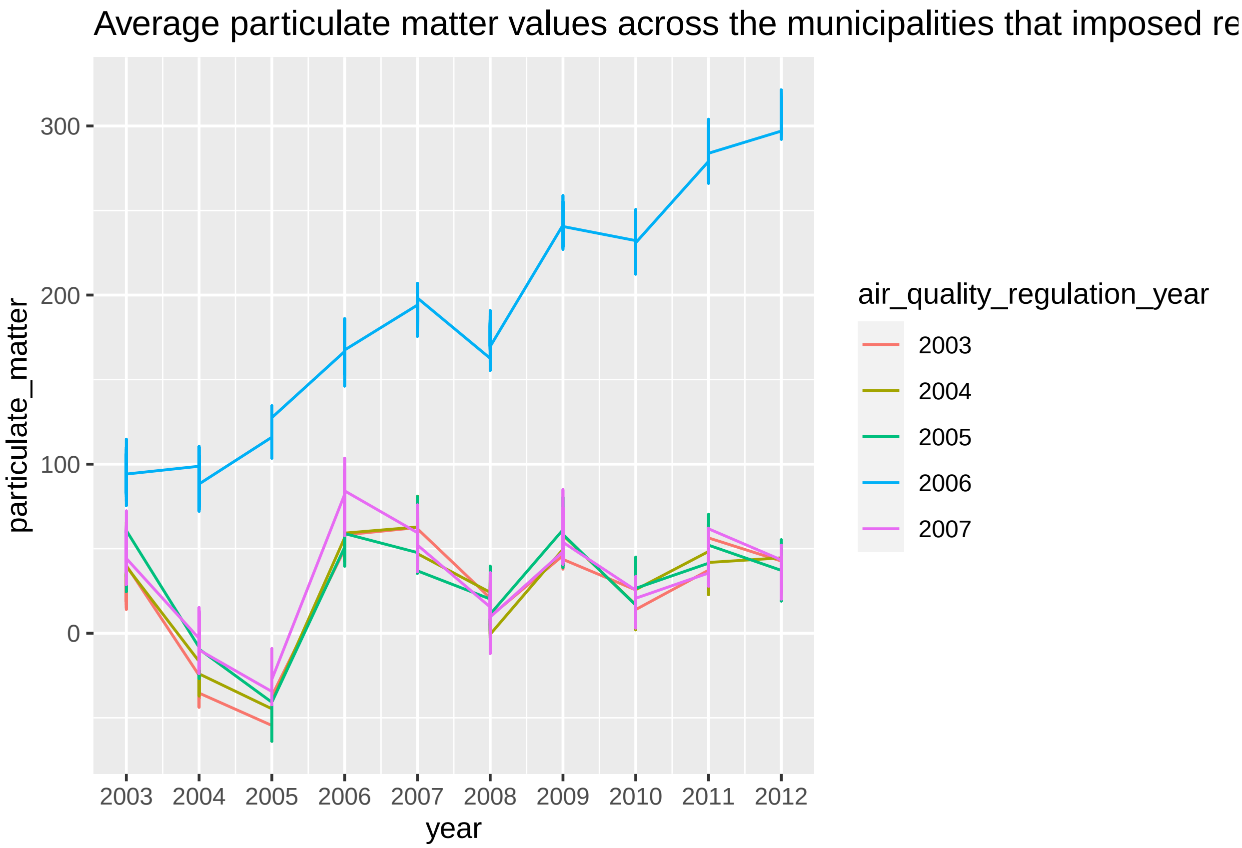
We also saw in the FELM model, regression with fixed effects, there was the lowest clustered standard error for τ^{DD} with a value of 0.54. THe same in the plm model is 2.34 and in the demeaning method is 2.22

How does the average particulate matter over time, for municipalities that imposed air quality regulations in each of the years from 2003 to 2007.

Let's drop the municipalities that regulated air quality in the year that looks different from the rest of the years, and describe why you can't estimate a credible causal effect for these municipalities.

Let's use the remaining municipalities to estimate a panel fixed effects regression to identify the causal effect of air quality regulation on particulate matter. Note that, here we will have to omit one of the event study treatment dummies (otherwise everything will be collinear). Standard practice is to leave out the T-1 dummy.

```
data$'air_quality_regulation_year' <- as.character(data$'air_quality_regulation_year')
plot6 <- data %>%
  filter(air_quality_regulation_year >= 2003 & air_quality_regulation_year <= 2007) %>%
  ggplot(aes(x= year, y = particulate_matter, color = air_quality_regulation_year)) + geom_line() +
  scale_x_continuous(breaks=c(2003,2004,2005,2006,2007,2008,2009,2010,2011,2012))+
  ggtitle("Average particulate matter values across the municipalities that imposed regulations between

plot6
```



Observations from above ?

The total particulate matter values and the mean particulate values for the muncipalities that had never imposed regulations throughout the years 2003 to 2012 has a increasing trend and is different to what we observed with the muncipalities that imposed regulations in 2006. Hence there is no parallel trend observed. As we see that the parallel trend assumption is not satisfied, we cannot use these muncipalities group to determine the DiD estimator for the ATE for 2006 regulators.

We saw the above observation from above to be holding true here, the parallel trends assumption doesnt seem to hold and we cannot use the muncipalities group that never imposed regulations as a control for the 2006 regulators. Also the regressions with Fixed effects and without the fixed effects resulted in the estimated τ^{DiD} vaues of 67.4689. Thus without the parallel trend observed, we cannot estimate the credible causal effect for these muncipalities that were regulated in 2006 using the DiD estimator.

Lets go ahead with analysis, filter for the wanted muncipalities

```
data_drop_2006 <-
  data %>% filter(air_quality_regulation_year %in% c(0,2003, 2004, 2005, 2007))
```

```
#data_drop_2006 <- pdata.frame(data_drop_2006, index=c("municipality_id"))

#data_drop_2006$is_treated <-
#  ifelse(data_drop_2006$year < data_drop_2006$air_quality_regulation_year,0,1)

#plm on remaining

#reg8 <- plm(particulate_matter ~ air_quality_regulation_year + is_treated + air_quality_regulation_yea
#                        data = data_drop_2006,
```

```r
#                         index = c("municipality_id","year"),
#                         model = "within")

#summary(reg8)



#The DiD estimator to determine the ATE of the air quality regulations is 9.2592. This means that there

#However we say, that the parallel trends assumption was not satisfying and thus we cannot use the grou

#As we see that the parallel trend is observed for the muncipalities that never introduced the air qual
```

```r
data_drop_2006$air_quality_regulation_year <- as.numeric(data_drop_2006$air_quality_regulation_year)
data_drop_2006$year <- as.numeric(data_drop_2006$year)

data_drop_2006$is_treated <- ifelse((data_drop_2006$year < data_drop_2006$air_quality_regulation_year)|
                                    (data_drop_2006$air_quality_regulation_year==0),0,1)
data_drop_2006$is_treated <- as.factor(data_drop_2006$is_treated)

data_drop_2006$municipality_id <- as.factor(data_drop_2006$municipality_id)

reg_10 <- felm(particulate_matter ~ is_treated |
municipality_id + year|
0|
0,
cluster ="municipality_id",
data = data_drop_2006)
```

```
## Warning: Argument(s) clustervar are deprecated and will be removed, use
## multipart formula instead
```

```r
#fe_ols1 = feols(particulate_matter ~ i(time_since_treatment, is_treated, ref = -1)
#          |  municipality_id + year, cluster = ~municipality_id, data = data_drop_2006)

#summary(fe_ols1)
summary(reg_10)
```

```
##
## Call:
##    felm(formula = particulate_matter ~ is_treated | municipality_id +      year | 0 | 0, data = data_
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.159  -6.270  -0.024   6.210  35.603
##
## Coefficients:
##             Estimate Cluster s.e. t value Pr(>|t|)
## is_treated1 -16.9102       0.3758  -44.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 9.775 on 13490 degrees of freedom
## Multiple R-squared(full model): 0.9517   Adjusted R-squared: 0.9463
## Multiple R-squared(proj model): 0.08606   Adjusted R-squared: -0.01617
## F-statistic(full model, *iid*):176.1 on 1509 and 13490 DF, p-value: < 2.2e-16
## F-statistic(proj model):  2024 on 1 and 1499 DF, p-value: < 2.2e-16
```

Event study regresiion

```
fe_ols2 = feols(particulate_matter ~ sunab(air_quality_regulation_year, year)
           |   municipality_id + year, cluster = ~municipality_id,data = data_drop_2006)
```

```
## NOTE: 1 observation removed because of NA values (RHS: 1).
```

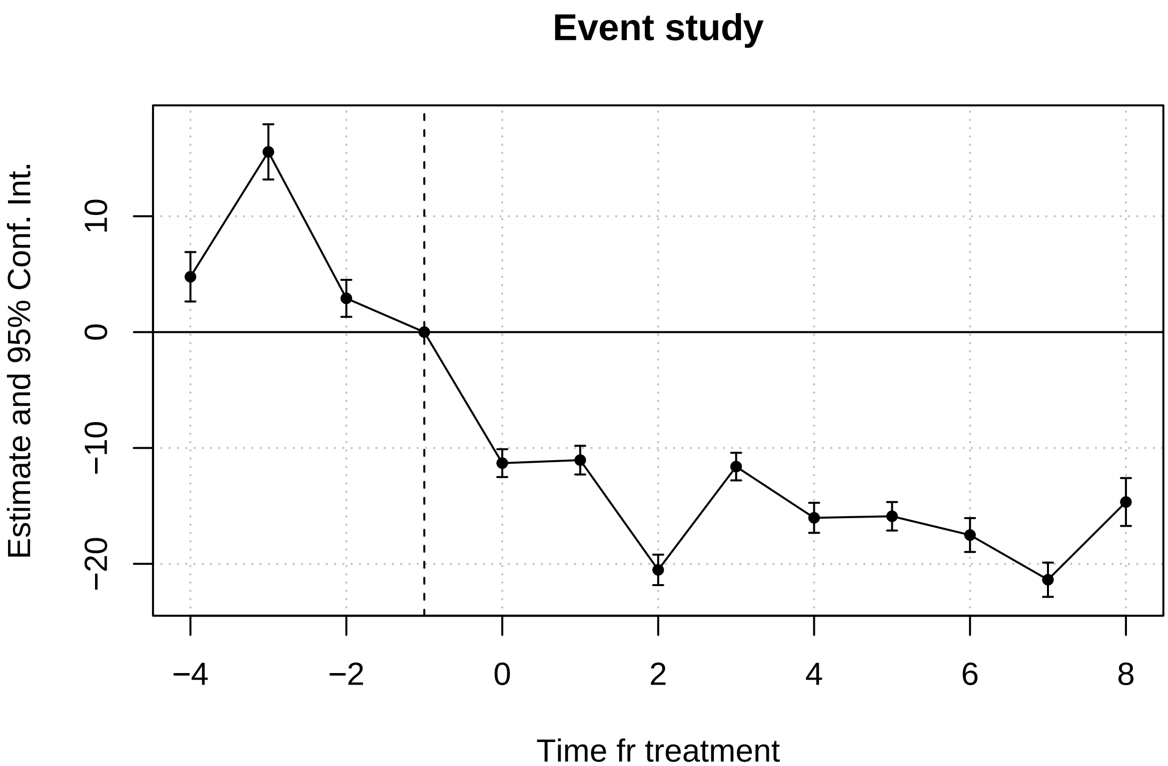
```
etable(fe_ols2)
```

```
##                              fe_ols2
## Dependent Var.: particulate_matter
##
## year = -4         4.773*** (1.087)
## year = -3         15.55*** (1.215)
## year = -2        2.914*** (0.8139)
## year = 0        -11.31*** (0.6122)
## year = 1        -11.05*** (0.6317)
## year = 2        -20.52*** (0.6702)
## year = 3        -11.61*** (0.6067)
## year = 4        -16.03*** (0.6606)
## year = 5        -15.90*** (0.6265)
## year = 6        -17.51*** (0.7451)
## year = 7        -21.37*** (0.7536)
## year = 8         -14.66*** (1.053)
## Fixed-Effects:  ------------------
## municipality_id                Yes
## year                           Yes
## _______________ __________________
## S.E.: Clustered by: municipality..
## Observations                14,999
## R2                         0.95497
## Within R2                  0.14796
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#summary(fe_ols2)
```

Plot event study point estimates and 95% interval

```
iplot(fe_ols2, sep = 0.5, ref.line = -1, pt.join = TRUE,
      xlab = 'Time fr treatment',
      main = 'Event study')
```



Describe the treatement effect varies over time

From what we observed above in the plot and the results of fe_ols, the air quality regulations have resulted in reduction of the local particulate matter by -16.91 where the maximum decrease stands at -21.37. The decrease can be deemed significant as this is the general trend after introduction of the air quality regulation years after the introduction. The maximum reduction also happened at 7 years after the introduction of the regulation. This means that the air quality regulations are effective in reducing the local particulate matter.

Let's explain all the results. Any shortcomings? Also a recommendation based on results ?

Summary

1) In the muncipalities that introduced the air quality regulations in the years from 2003 - 2007 except for the year 2006, we observed to have a parallel trend in the counterfactual in the muncipalities that never introduced the air quality regulations.
2) The naive estimator that calculated the difference in means of local particulate matter in air between muncipalities with air quality regulations and without air quality regulations is -24.44. The same difference in means of particulate matter for the muncipalities who introduced the air quality regulations prior to 2004 and after 2004 stands at -21.47.
3) We also saw that the muncipalities group that never introduced the air quality regulations can be used as the control group for the other set of muncipalities for determining the DiD estimator to estimate the ATE.
4) The set of muncipalities that regulated in the year 2006, we found that they dont observe a parallel trend in the counterfcatual, hence the set of the muncipalities that never introduced the regulations cannot be used as a control group for the above set to determine the DiD estimator $\hat{\tau}^{DD}$ for the ATE. 4)In the regression analysis of panel data, with and without fixed effects, we saw that introduction of

air quality regulations have reduced the local particulate matter in air by -16.91 across the muncipalites that introduced air quality regulations in 2003-2007 except for 2006, with a maximum reduction of local particulate matter in air by -21.37, that after 7 years of treatment i.e introduction of air quality regulations.Thus we concluded that the air quality regulations are beneficial to reduce the local particulate matter in air in the muncipalities.

5) The simple regression without fixed effects resulted in the τ^{DiD} of 67.4689. This is same in the case of when we applied fixed effects into the regression.

Describe at least one remaining potential shortcoming with these results.

From the results in Q9, we can say that there is no selection in treatment problem by the observation of parallel trend in counterfactuals for the event design. But what if there are coincident treatments, where a completely different policy has been introduced in the treated group of muncipalites with the air quality regulation. In such a case, we cannot accurately estimate the ATE.

Also we need to analyze the "Anticipatory effects and the Ashenfelter dip" in the plot of the event study and 95 percent confidence intervals. If we see 3 yuears and 2 years before the treatment, we see a reduction of 12.64 and 14.22 respectively. This mightve created a bias in our estimation of DiD estimator and further the ATE. This decrease however we dont know the cause, can be said due to the Ashenfelter dip i.e a sudden dip in trend just before the treatment i.e introduction of air quality reduction.

The above shortcomings with the results are remaining and we need to analyze the same before we determine the causal effect of the air quality regulation on the local particulate matter (PM 2.5) in air using the DiD estimator for the ATE.

Interpret the magnitude of your estimated effects: do the results suggest that we should be strongly promoting air quality regulations?

We observed that the introduction of air quality regulations has reduced the local particulate matter (PM 2.5) in air for the muncipaliteis that introduced the air quality regulations in 2003-2007 except for 2006 by -16.91. Also we observe a reduction of local particulate matter not only immediately but also after nearly 7 years of the introduction of these regulations. Thus, in general after observing a reducing trend in the local particulate matter (PM 2.5) in the muncipalities that introduced the regulations, our results suggest that we should be strongly promoting air quality regulations, assuming that theese shortcomings provided above are not significant in effecting the results.