King's Business School, King's College London

Cover Sheet for [7QQMM906] 25/26

Candidate ID:	**AG37642, AG 33419**
Module Code:	**7QQMM906**
Module Name:	**Environmental Economics**
Word Count:	

DECLARATION BY STUDENT

I understand what is meant by plagiarism/collusion and have signed at enrolment the declaration concerning the avoidance of plagiarism/collusion.
I understand that plagiarism and collusion is a serious examinations offence that may result in disciplinary action being taken.
I understand that I must submit work BEFORE the deadline, and that failure to do so may result in late submission penalties.
I understand that where a word limit is indicated, I should adhere to the word limit and failure to do so may result in penalties.
I understand that King's requires students to acknowledge any use of generative AI tools in coursework.

Please delete one of the statements below and attach this document to your assignment before uploading:

I declare that parts of this submission has contributions from AI software and that it aligns with acceptable use as specified as part of the assignment brief/ guidance and is consistent with good academic practice. The content can still be considered as my own words. I understand that as long as my use falls within the scope of appropriate use as defined in the assessment brief/guidance then this declaration will not have any direct impact on the grades awarded.

I acknowledge use of software to:
(i) Generate ideas or structure suggestions, for assistance with understanding core concepts, or other substantial foundational and preparatory activity.

[ChatGPT for research assistance and explainations]

(iii) Generate some other aspect of the submitted assessment.

By completing this assessment, you acknowledge that you have read and understand the above. Please do not sign your name.

Data Report Road Map - Group 6

1. Part I: Motivation & Data Context

We will be investigating the change in NO2 concentration in the London area that was added to the Ultra Low Emission Zone (ULEZ) in 2021. We will compare the NO2 concentration from before 2021 to 2023, to see the effects of the ULEZ-Policy. We will be using the statistical Software „Stata“ and Google Earth Engine to work with these inputs. The motivation behind our Data choice is to see whether environmental policies have an impact on GHG-emissions and are therefore effective instruments for mitigation.

a. Dataset Selection and Research Questions

Chosen Databases:

- **Sentinel-5P OFFL NO2** (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_OFFL_L3_NO2?hl=de)

This Dataset published by Google Earth Engine (Copernicus European space agency) was chosen as the base information, since it specifically collects data concerning NO2 levels in the air. We chose the following band from the possible options “*tropospheric_NO2_column_number_density*” out of the dataset because of its connection to human activity.

- **ERA5 Land Hourly - ECMWF Climate Reanalysis** (https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_HOURLY?hl=de)

We chose this Dataset additionally, to get information on different bands like wind speed, rainfall (read: accumulated rainfall), temperature,etc.

We chose the following bands:

'temperature_2m',

'u_component_of_wind_10m',

'v_component_of_wind_10m',

'surface_pressure',

'dewpoint_temperature_2m',

'total_precipitation_sum',

Research question 1:

Did the NO2 concentration in the Air measurement in all of London change over the course of 2021 to 2023?

Research question 2:

Did the NO2 concentration in the Air measurement in the area of the expansion of the ULEZ change over the course of 2021 to 2023?

Research

Relationship between research questions and environmental economics theory:

Our Data report connects to the theoretical comparison between the impact of environmental tax and environmental policies as it investigates the real-life impact of said policies.

b. Data Significance and Relevance

Why is the dataset of interest to the world and environmental economics specifically?

NO2 emissions directly affect the public's wellbeing, air quality and therefore influence political instrument choices, especially for the traffic sector. It also affects the automobile industry. For environmental economics, NO2 measurements can provide a reliable source for mitigation/emission tracking.

Connect the data to current environmental policy debates or economic challenges?

- Effectiveness of environmental policy, specifically low-emissions zones.
- GHG mitigation of the transport sector
- Economic fairness in terms of externalities

Explain potential policy implications or academic contributions of your analysis?

- Effectiveness of environmental policy, specifically low-emissions zones.
- Impact of electric vehicles
- Behavioural change through policies

Identify the target audience for your findings (policymakers, researchers, NGOs, etc.)?

- Government/policy makers
- Researchers & Academic field
- The public

2. Part II: Technical Implementation

a. Coding Setup and Documentation

- **Master file code structure**: We will be using stata and Github for this.
- **Folder organization**: We will make sure to have the same path in our OneDrive and to keep the folder clean.
- **Version control**: We have set up a Github repository
- **Reproducibility**: For this we will make sure that we work meticulously on Stata (Do-files) and fetch changes into Github

b. Data Preparation and Management

- **Cleaning processes**

Variables		Definition	Code

Dependent variable	NO2	Daily mean NO_2 concentration
		
		(mol/m²) in a zone or grid cell
			`//ln_NO2`
			`gen ln_NO2= ln(no2) if no2>0`
			`drop ln_NO2`
			`gen ln_NO2= ln(no2)`

Group Dummy	treatment	1 if observation is inside LEZ, 0 if in control zone	`//Treat`
			`gen Inner_ULEZ=1 if region=="Inner"`
			`drop Inner_ULEZ`
			`gen treatment=1 if region=="Inner"`
			`replace treatment=0 if treatment==.`

Time Dummy	post	1 if date is after LEZ/ULEZ implementation, 0 if before	`//Post`
			`gen post=0`
			`replace post=1 if date >= date("2023-08-29", "YMD")`

Independent variable (interaction term)	lez_policy (did)	Main DID estimator: 1 if in LEZ group AND date is after policy (0 otherwise)	`//Interation term` `gen treatment_post= treatment*post`
Controls	temp	Daily average air temperature (°C)	/

	wind	Daily average wind speed (m/s)	/

	cloud	Daily average or fraction cloud cover	`//ln_cloud`
			`gen ln_cloudfraction=ln(cloudfraction)`
			`drop if cloudfraction==.`
			`drop if ln_cloudfraction ==.`

	holiday	Public holiday dummy (1 if holiday, 0 otherwise), day-of-week dummies	`//Is holiday?` `gen is_holiday = 0` `replace is_holiday = 1 if date == date("2023-01-02", "YMD") |date == date("2023-04-10", "YMD") | date == date("2023-05-01", "YMD") |date == date("2023-05-08", "YMD") | date == date("2023-05-29", "YMD") |date == date("2023-08-28", "YMD") |date == date("2023-12-25", "YMD") |date == date("2023-12-26", "YMD")`  `replace is_holiday = 1 if date == date("2024-01-01", "YMD") |date == date("2024-03-29", "YMD") |date == date("2024-04-01", "YMD") |date == date("2024-05-06", "YMD") |date == date("2024-05-27", "YMD") |date == date("2024-08-26", "YMD") |date == date("2024-12-25", "YMD") |date == date("2024-12-26", "YMD")`  `replace is_holiday = 1 if date == date("2022-08-29", "YMD") |date == date("2022-09-19", "YMD") |date == date("2022-12-26", "YMD") |date == date("2022-12-27", "YMD")`  `label variable is_holiday "UK Bank Holiday (0=No, 1=Yes)"`

	time dummy	Year, month,weekday	`//time dummy`
			`gen year = year(date)`
			`gen month = month(date)`
			`gen dow = dow(date)`

Fixed effect	area_fe ()	Area fixed effects	`//Sitenum` `encode sitecode, generate(sitenum)`

	date_fe ()	Time fixed effects	//numeric date gen date_numeric = date(date, "YMD") if !missing(date) format date_numeric %td drop date rename date_numeric date

Handling missing values, outliers, and inconsistencies:

```
//Winsor
ssc install winsor2
winsor2 ln_NO2 temp windspeed
winddirection pressure rh precipitation
 ln_cloudfraction, cuts(1 99) replace

//Drop
drop if missing(ln_NO2)|missing(
temperature)|missing(windspeed)|missing
(winddirection)|missing(pressure)|
missing(rh)|missing(precipitation)|
missing(ln_cloudfraction)
```

Problems encountered and solutions implemented (this far):

Figuring out which areas exactly were part of the 2021 ULEZ expansion and which datasets corresponded to these was a challenge. We have therefore decided not to use the LAQN at this point as it would complicate our process. We have instead opted for the 5SENTINEL database.

Data aggregation:

- Aggregate data to appropriate temporal/spatial units for analysis
- Create summary statistics by relevant groupings (country, year, sector, etc.)

For this we will focus on the site codes and respective dates (e.g. holidays)

Data merging IN GEE

- Connect multiple datasets using common identifiers √
- Handle non-perfect matches and document decisions √

```
    // merge all bands
    var combinedImage = no2Image
      .addBands(era5Day)
      .addBands(windSpeed)
      .addBands(windDirection)
      .addBands(rh);

    var extractPoints = function(points) {
      return combinedImage.reduceRegions({
        collection: points,
        reducer: ee.Reducer.first(),
        scale: 1000,
        tileScale: 4
      }).map(function(feature) {
        return feature.set({
          'Date': currentDate.format('yyyy-MM-dd'),
          'Region': points === pointsFC ? 'Inner' : 'Outer'
        });
      });
    };
    var innerPointsData = extractPoints(pointsFC);
    var outerPointsData = extractPoints(ptsWithCode);
    return innerPointsData.merge(outerPointsData);
  };

  var allData = ee.FeatureCollection(
    days.map(extractBandValues)
  ).flatten();
```

    - Validate merge success and identify any data loss √
- **Variable transformation**:
    - Generate derived variables relevant to research questions √
    - Create categorical variables and binary indicators √
    - Apply appropriate transformations (log, winsorizing, standardization) √
    - **Document theoretical justification for transformations**

3. Part III: Descriptive Analysis and Export of Results for Presentation

As learnt in the Stata Introduction class, we will be using the "import" as well as the "generate" codes that were provided to us.

a. Summary Statistics Table (will vary depending on the data set you choose)

- **Table 1**: Create comprehensive descriptive statistics table
  - Include relevant measures (mean, median, standard deviation, min/max, observations)

```
tabstat ln_NO2 treatment post treatment_post
cloudfraction temperature windspeed winddirection
pressure rh precipitation ln_cloudfraction is_holiday,


logout,save(mytable) word replace: tabstat ln_NO2
treatment post treatment_post cloudfraction temperature
 windspeed winddirection pressure rh precipitation
ln_cloudfraction is_holiday, s(N mean p50 sd min max
range) f(%12.3f) c(s)
```

Variable	N	Mean	p50	SD	Min	Max	Range
ln NO2	63564	-9.440	-9.427	0.608	-11.24	-7.980	3.261
treatment	63564	0.900	1	0.301	0	1	1
post	63564	0.468	0	0.499	0	1	1
treatment*post	63564	0.420	0	0.494	0	1	1
cloudfraction	63564	0.121	0.0930	0.109	0	0.685	0.685
temperature	63564	284.7	284.9	6.004	271.6	295.2	23.65
windspeed	63564	3.175	3.140	1.402	0.562	7.271	6.709
winddirect	63564	178.2	198.9	89.48	7.650	352.6	344.9
pressure	63564	101026	101133	1087	97666	103317	5652
rh	63564	77.77	77.59	8.955	57.99	94.41	36.42
Precipitation	63564	0.00100	0	0.00200	0	0.0160	0.0160
ln cloudfraction	63564	-2.708	-2.371	1.332	-6.437	-0.378	6.059
is holiday	63564	0.0250	0	0.157	0	1	1

We've completed the main regression section, and the results show that the treatment_post(interaction term) coefficient (-0.0984729) indicates that NO2 levels decreased by approximately 9.85% in treated sites after the policy intervention, with this effect being highly significant (t=-23.24, p<0.001).

b. Data Visualization and Exploration (the below will vary depending on what data set you choose)

- **Distribution analysis**: (Density plots, histograms, or boxplots for key outcome and explanatory variables)

We will use density plots for the change in NO2 level in the air for the different areas of the ULEZ expansion.

For the visualisation of the change in NO2 levels over the years we will most likely use a density plot.

We will include a SCREENSHOT GEE + CODE HOW, then:

    - Identify skewness, multimodality, or other distributional features
- **Temporal and spatial patterns**:
    - Time series plots showing evolution of variables over time
    - Cross-sectional comparisons across countries/regions
    - Two-way dot plots or bar charts for categorical breakdowns
    - Facet grid/wrap plots for multiple group comparisons
- **Comparative analysis**:
    - Bar plots comparing different groups, policies, or time periods
    - Scatter plots exploring relationships between key variables
- **Geospatial visualization** (if applicable):
    - Maps showing geographic distribution of key variables
    - Choropleth maps for country/region-level data

c. Regression Analysis

- Simple bivariate or multivariate regression exploring relationships suggested by descriptive analysis √

C-1 Parallel Trend Test

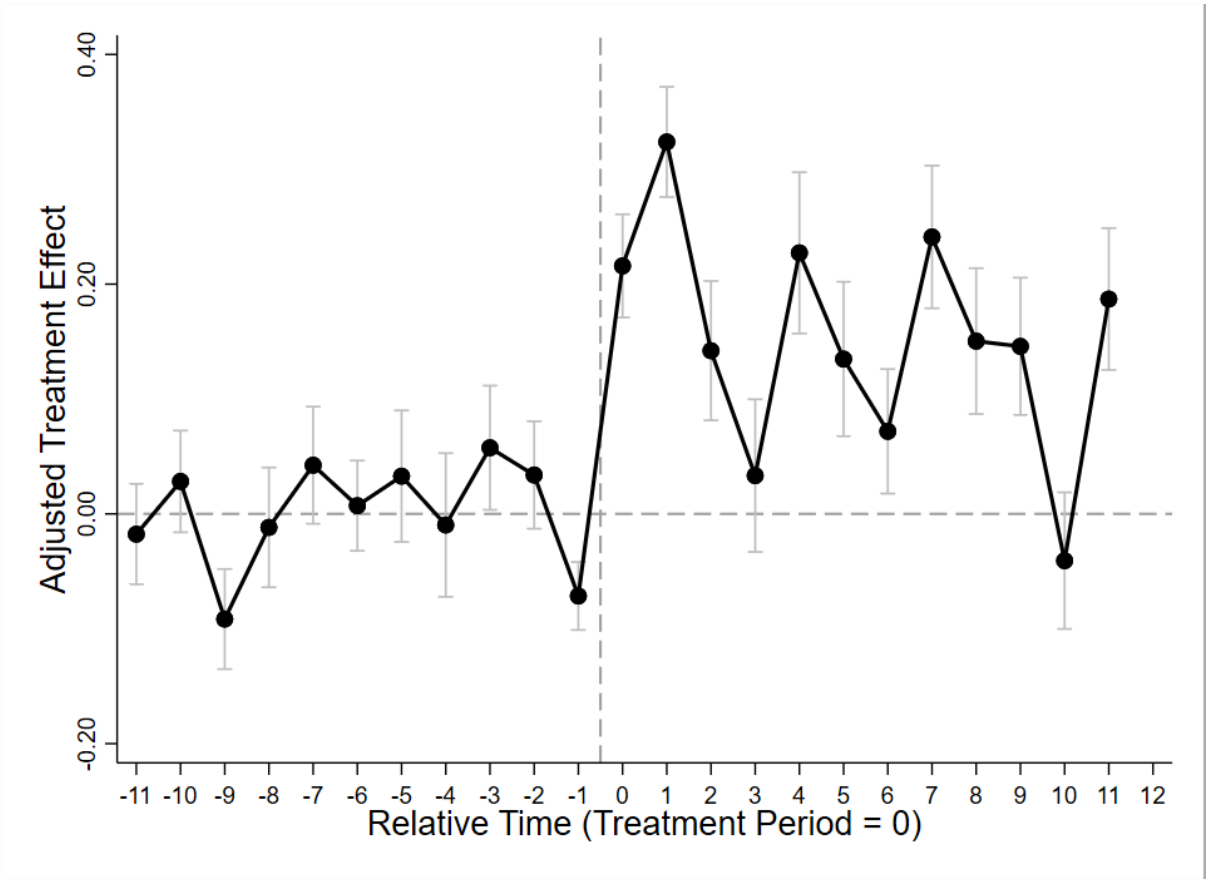
Code

```
/*============demeaned parallel trends test===============*/
//Set the conditional variable
gen judgment= ym(year(date), month(date)) - ym(2023, 9)
gen pre_1 = ( judgment < 0 & treatment  ==1)
drop pre_1
ssc install coefplot

//Variables generated 11 periods prior to policy
implementation
forvalues i = 11(-1)1 {
    gen pre_`i'=(judgment ==-`i' & treatment ==1)
}

//current variable
gen current = ( judgment ==0 & treatment ==1)

//Variables in the 11 periods following the policy
forvalues i = 11(-1)1 {
    gen las_`i' = (judgment ==`i' & treatment ==1)
}

//regression analysis
reghdfe ln_NO2 treatment_post temp windspeed winddirection
pressure rh precipitation ln_cloudfraction is_holiday pre_*
current las_* ,absorb(sitenum dow month) vce(cluster sitenum)
```

```
//regression analysis
reghdfe ln_NO2 treatment_post temp windspeed winddirection
pressure rh precipitation ln_cloudfraction is_holiday pre_*
current las_* ,absorb(sitenum dow month) vce(cluster sitenum)

//demeaned
//Install command
ssc install parmest, replace
ssc install fillmissing, replace

//Extract regression coefficients
parmest, format(estimate min95 max95 %8.3f p %8.3f) saving(
"temp1.dta", replace)

//Read temporary files and filter coefficients before and
after the policy.
use "temp1.dta", clear
keep if ustrregexm(parm, "pre_*|las_*|current")
drop if ustrregexm(parm, "pressure")
drop if ustrregexm(parm, "precipitation")
//Only retain the coefficients before and after the policy
and during the policy period.

//Mark the current position of the policy, calculate the
relative time, and id is the adjusted relative time.
gen num = _n

gen minus = num if ustrregexm(parm, "current") //Find the
policy number for the current period
fillmissing minus //Fill in the current period number to all
rows
gen id = num - minus

//Core: Calculate the mean of the coefficients before the
policy, and subtract the mean from all coefficients.
egen average =mean(estimate)if id<0 //Only the mean
coefficient before the policy was implemented (id<0) was
calculated.
fillmissing average // Fill all rows with the mean (adjust
using the same mean after the policy is implemented).
replace estimate=estimate-average //Subtract the pre-policy
average from each coefficient to complete the adjustment.

//Step4-Recalculate the confidence intervals and plot the
trend of "meeting the criteria".

*1. Recalculate the confidence interval (ul = upper limit,
11 = lower limit)
gen ul=estimate+stderr*1.65 //Upper limit of 90% confidence
interval
gen ll=estimate-stderr*1.65 //90% confidence interval lower
limit

*2. Base period (1 period before the policy, id=-1) is set
to 0 (base period effect is 0 when plotting).
for var estimate ul ll:replace X=0 if mi(estimate)& id == -1
```

```
sort id
twoway (rcap ul ll id, lcolor(gs12) lwidth(medthin)) ///
       (connected estimate id, msymbol(O) msize(medium) ///
        mcolor(black) lcolor(black) lwidth(medthick)), ///
       yline(0, lp(dash) lc(gs10)) ///
       xline(-0.5, lp(dash) lc(gs10)) ///
       xlabel(-11(1)12, nogrid labsize(small)) ///
       ylabel(, nogrid format(%4.2f) labsize(small)) ///
       xtitle("Relative Time (Treatment Period = 0)", size(
medium)) ///
       ytitle("Adjusted Treatment Effect", size(medium)) ///
       legend(off) graphregion(color(white))
```

Result


Adjusted Treatment Effect
0.40
0.20
0.00
-0.20
-11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12
Relative Time (Treatment Period = 0)

C-2 Regression Analysis

$$NO_{2,it} = \alpha + \beta 1(treatment_i \times post_t) + \beta_2 treatment_i + \beta_3 post_t + \gamma' X_{it} + \mu_i + \tau_t + \epsilon_{it}$$

```
use "F:\Onedrive映射\1kcl\ESG\7QQMM906 Environmental Economics\Group
Assessment\data after cleaning\data_after_cleaning.dta", clear


* 1.Main DID
reghdfe ln_NO2 treatment_post temperature windspeed winddirection pressure rh
precipitation ln_cloudfraction is_holiday, absorb(sitenum dow month) vce(
cluster sitenum)
estimates store main_did

* 2. Heterogeneity by Site Type
foreach type in "Roadside" "Urban Background" "Kerbside" "Suburban"
"Industrial" {
    quietly reghdfe ln_NO2 treatment_post temp windspeed winddirection pressure
 rh precipitation ln_cloudfraction is_holiday,  absorb(sitenum dow month) vce(
cluster sitenum), if sitetype == "`type'"

    estimates store `=subinstr("`type'", " ", "_", .)'
}

* 3. Export
esttab main_did Roadside Urban_Background Kerbside Suburban Industrial ///
    using "full_regression_table1.rtf", ///
    replace rtf ///
    keep(treatment_post temperature windspeed winddirection pressure rh
precipitation ln_cloudfraction ls_holiday) ///
    order(treatment_post temperature windspeed winddirection pressure rh
precipitation ln_cloudfraction ls_holiday) ///
    b(4) se(4) star(* 0.10 ** 0.05 *** 0.01) ///
    mtitles("All Sites" "Roadside" "Urban BG" "Kerbside" "Suburban"
"Industrial") ///
    mgroups("Main did" "Heterogeneity by Site Type", ///
            pattern(1 1 0 0 0 0) ///
            prefix(\multicolumn{@span}{c}{) suffix(}) ///
            span erepeat(\cmidrule(lr){@span})) ///
    title("Table: DID Estimates and Heterogeneity by Site Type") ///
    addnote("Standard errors clustered at site level in parentheses." ///
            "All models include site, month, and day-of-week fixed effects." //
            "* p<0.10, ** p<0.05, *** p<0.01") ///
    stats(N r2, fmt(%9.0fc %9.3f) ///
          labels("Observations" "R-squared")) ///
    label
```

Table: DID Estimates and Heterogeneity by Site Type

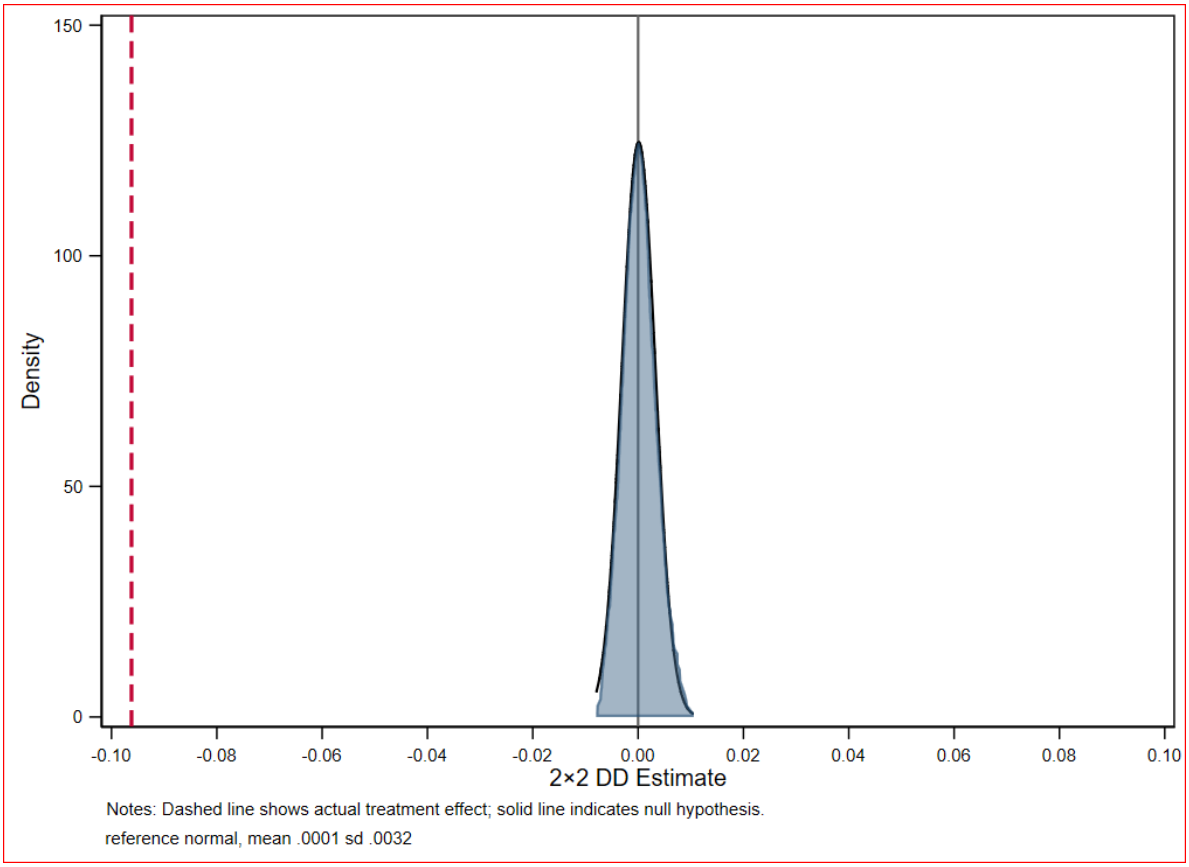
	Main DID	Heterogeneity by Site Type				
	(1)	(2)	(3)	(4)	(5)	(6)
	All Sites	Roadside	Urban BG	Kerbside	Suburban	Industrial
treatment_post	-0.0962***	-0.1008***	-0.0878***	-0.0987***	-0.1116***	-0.0943***
	(0.0037)	(0.0045)	(0.0069)	(0.0122)	(0.0124)	(0.0198)
Temperature	-0.0048***	-0.0049***	-0.0034***	-0.0065***	-0.0055**	-0.0035**
	(0.0005)	(0.0006)	(0.0009)	(0.0014)	(0.0025)	(0.0016)
WindSpeed	-0.2281***	-0.2342***	-0.2249***	-0.2337***	-0.2209***	-0.2421***
	(0.0021)	(0.0021)	(0.0062)	(0.0045)	(0.0055)	(0.0058)
WindDirection	-0.0004***	-0.0005***	-0.0002**	-0.0004**	-0.0005**	-0.0006***
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Pressure	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
RH	0.0033***	0.0029***	0.0042***	0.0032***	0.0015	0.0014
	(0.0004)	(0.0006)	(0.0009)	(0.0010)	(0.0015)	(0.0015)
Precipitation	-8.9875***	-9.7127***	-8.7241***	-13.2592***	-10.2959**	-5.9291*
	(0.8444)	(1.0367)	(1.9301)	(2.4193)	(4.4020)	(2.7857)
ln_cloudfraction	-0.0177***	-0.0150***	-0.0173***	-0.0191***	-0.0239***	-0.0224***
	(0.0012)	(0.0017)	(0.0019)	(0.0043)	(0.0057)	(0.0060)
ls_holiday	Control	Control	Control	Control	Control	Control
Observations	63,564	28,204	14,153	5,862	5,329	3,413
R-squared	0.539	0.549	0.553	0.551	0.500	0.567

Standard errors in parentheses
Standard errors clustered at site level in parentheses.
All models include site, month, and day-of-week fixed effects.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

C-3 Placebo Test: Permutation-based Methods

https://library.virginia.edu/data/articles/testing-significance-permutation-based-methods

Result



Code

```
/*==============================Placebo Test====================================*/
use "data_after_cleaning.dta", clear
reghdfe ln_NO2 treatment_post temp windspeed winddirection pressure rh
precipitation ln_cloudfraction is_holiday, absorb(sitenum dow month) vce(
cluster sitenum)
estimates store main_did

//ssc install permute
//help permute

**# Bookmark #1
permute treatment_post beta=_b[treatment_post] se=_se[treatment_post] df=e(df_r
), reps(588) seed(1008) saving("stimulations.dta"): reghdfe ln_NO2
treatment_post temp windspeed winddirection pressure rh precipitation
ln_cloudfraction is_holiday, absorb(sitenum dow month ) vce(cluster sitenum)

use "stimulations.dta", clear
```

```
#delimit ;

dpplot beta,

    xline(-0.0961978, lcolor(cranberry) lpattern(dash) lwidth(medthick))
    xline(0, lcolor(gs6) lpattern(solid) lwidth(medium))


    color(navy%50) recast(area)
    lcolor(navy) lwidth(medium)


    xtitle("2×2 DD Estimate", size(small))
    xlabel(-0.10(0.02)0.10, format(%4.2f) labsize(vsmall))


    ytitle("Density", size(small))
    ylabel(0(50)150, format(%3.0f) labsize(vsmall) angle(horizontal) nogrid)


    legend(order(1 "Placebo estimates"
                 2 "True effect (β̂=-0.096)"
                 3 "Null (β=0)")
           position(2) ring(0) cols(1) size(vsmall)
           region(lcolor(gs12) fcolor(white%90) lwidth(vthin))
           symxsize(*.6) rowgap(*.5))


    note("Notes: Dashed line shows actual treatment effect; solid line
indicates null hypothesis.",size(vsmall))

    graphregion(color(white) margin(small))
    plotregion(lcolor(black) lwidth(thin) margin(small))

    scheme(s1mono) ;

#delimit cr

graph export "placebo_paper.png", replace width(2400) height(1800)
graph export "placebo_paper.eps", replace
```

After the Regression Analysis, we will interpret the results. Here, we will discuss limitations of our findings and connect them with the learnt theory from class

4. Part IV: Discussion and Conclusions

a. Key Findings Summary

This section of the Data Report will focus on answering the proposed research questions mentioned in the beginning of this road map. We will combine the findings and discuss the limitations.

We aim to visualise our main findings through graphs and charts to make them easily understandable. Additionally, we will try to explain patterns we identify through the regression analysis.

b. Policy and Research Implications

In regards to our target audience and our possible addresses, we will try to suggest policy recommendations or improvements for the existing ULEZ-Policy.

c. Technical Reflection

We will try to evaluate this through comparing our work with other existing research.