The Impact of COVID-19 Policy on Ambient Tropospheric Nitrogen Dioxide Concentrations in the United States and India

Student ID: 219021422

Supervisor: Professor John Remedios

1. Abstract

This project utilised nitrogen dioxide concentration data as recorded by the Ozone Monitoring Instrument aboard the NASA Aura Satellite, alongside detailed policy indices found in the Oxford Covid-19 Government Response Tracker (OxCGRT) database, to assess the relationship between COVID-19 containment policies and tropospheric ambient nitrogen dioxide concentrations in the United States and India.

In particular, this project aimed to assess whether quantitative measures of COVID-19 containment policy could explain and predict differences between COVID period nitrogen dioxide records and the tropospheric nitrogen dioxide profile which would be expected considering the previous decade of data, with the further goal of assessing the relative impact of each containment policy area on the nitrogen dioxide records.

In addition, this project more generally aimed to assess the viability of using quantitative measures of policy and satellite-collected nitrogen dioxide concentrations to gain insights in relation to policy to address pollution.

Multivariate regressions, with time series policy indices as the independent variables and time series de-seasonalised nationally area-averaged tropospheric nitrogen dioxide concentrations as the dependent variable, were employed as the primary means of analysis.

It was found that in the US, workplace closure was by far the best predictor of nitrogen dioxide variation (coefficient = -0.9091, P-value < 0.001), whereas in India, confinement orders were the best predictor (coefficient = -0.7684, P-value < 0.001). In both cases, those respective indices were the sole statistically significant indices.

Accordingly, single independent variable linear regressions were performed for the US (coefficient = -0.1504, P-value < 0.001) and India (coefficient = -0.1092, P-value = 0.001) using only the workplace closure and confinement orders indices respectively:

Because of how the indices were normalised, these suggest an associated 15% and 11% drops in nitrogen dioxide concentrations respectively associated with the maximum implementation of these policy areas in each country.

It should be noted however that high random variation in the nitrogen dioxide profiles and high correlation between policy indices in both cases limited the reliability of the results in drawing precise causal conclusions. Evidence was also found to suggest the existence of a confounding variable/variables increasing nitrogen dioxide concentrations.

For these reasons, suggestions for improvements in future work have been made.

2. Introduction

Air pollution is a serious mortality risk, being associated with over 500,000 excess deaths in 2015 in Europe alone[1], and considering how effective a proxy nitrogen dioxide is for air pollution more generally[2], a deeper understanding of the relationship between policy choices and nitrogen dioxide levels will develop and strengthen the body of knowledge available to make optimal choices to improve air quality and save lives.

The unprecedented natural experiment that was the coronavirus pandemic is an invaluable opportunity for investigation in this regard, with countries across the world rapidly implementing policies of unprecedented scale, with dramatic improvements in air quality seen during the strictest months of the lockdowns[3][4].

The fact that lockdown policies improved air quality (largely through their impact on travel, as revealed by Google and Apple movement data[5][6]) is well documented[4][5][6], and M.J. Cooper et al (2022)[4] even showed that cities experiencing stricter lockdowns saw greater decreases in nitrogen dioxide concentrations. However, detail in specific policy impact beyond that is lacking in the existing literature.

M.J. Cooper et al (2022)[4] used a threshold of a time average of the Stringency Index found in the Oxford Covid-19 Government Response Tracker database to sort cities into those considered strict or not strict. This Stringency Index quantifies a general sense of policy strictness on a scale of 0 to 100, and is available as a daily time series for 175 countries (and includes sub-national data for 7 countries). It is calculated as a weighted average of 9 more specific sub-indices:

- Closing of schools and universities
- Closing of workplaces
- Cancelling of public events
- Gathering size limits
- Public transport closure
- Confinement orders
- Restrictions on internal movement
- Restrictions on international travel for foreigners
- Presence of public information campaigns

The availability of these more detailed indices in an extensive and comprehensive database presented an invaluable opportunity for a more in-depth analysis that so far had not been explored in previous literature.

In particular, this project aimed to assess whether these sub-indices could successfully predict the nitrogen dioxide profiles, and further aimed to assess the relative impact of each policy area. Furthermore, this project aimed to assess the viability of using quantitative measures of policy and satellite-collected nitrogen dioxide concentrations to gain insights in relation to policy to tackle pollution.

Multivariate linear regressions were the chosen avenue of analysis, with the 8 time series containment sub-indices as the independent variables, and the de-seasonalised time series nitrogen dioxide concentration profiles as the dependent variable. The 8 containment sub-indices are those listed above, with the exception of 'Presence of public information campaigns', which the OxCGRT database did not classify as a containment measure but included in the Stringency Index regardless.

From now on in this report, the sub-indices will be referred to simply as indices for ease of parsing, as the combined indices like the Stringency Index will no longer be relevant.

The countries chosen for analysis were the United States and India. As large and populous countries with oceans forming a substantial portion of their borders, the US and India are less prone to atmospheric transport issues than smaller or landlocked countries, as nitrogen dioxide originating from outside their borders will form only a very small portion of that recorded in the troposphere.

The final indices used for the analyses were not those listed as the country-level indices in the OxCGRT database, and instead the state-level indices were summed, weighted by population. This formed a more accurate measure of the intensity of the policy areas' implementation, as the country-level indices describe the most strict level of policy implemented anywhere in the country.

The multivariate regressions were run using weekly-averaged data from the first week of 2020 to the last week of 2022. Single-year regressions were also run for all three years for each country, but the usefulness of these analyses was limited.

3. Theory

3.1. Determining tropospheric nitrogen dioxide concentrations

Tropospheric nitrogen dioxide concentration data for this project was collected by the Ozone Monitoring Instrument aboard the NASA Aura satellite, which is in a sun-synchronous orbit, collecting data at 13:30 local time[7].

It measures the presence of pollutants like nitrogen dioxide via spectroscopy. It first records the presence of nitrogen dioxide along it's line of sight to the surface (Slant Column Densi**ty)** before converting to a Vertical Column Density. In our case, the tropospheric density must then be calculated from the Vertical Column Density, which is a complex calculation that uses factors like surface and cloud reflectance[7]**.**

The pre-calculated tropospheric nitrogen dioxide concentration data is publicly available online via NASA EarthData's GIOVANNI service[8], which was used as the data source for this project.

3.2. The Oxford Covid-19 Government Response Tracker

The Oxford Covid-19 Government Response Tracker is an extensive information repository published by the Blavatnik School of Government in Oxford, tracking many aspects of policy and other indicators relating to the COVID-19 pandemic[9].

The data of interest for this project were the 8 containment policy indices, tracked daily from the start of 2020 to the end of 2022 for 175 countries, and for the state-level governments of 7 countries, including the United States and India[9].

The indices, and the requirements for reaching each index value, are given below, as taken from the OxCGRT working paper[9]:

Table 1: Containment sub-indices used in the Oxford COVID Government Response Tracker, and the policy thresholds for assigning index values

Sub-index	Value requirements
Closing of schools and universities	0 - no measures 1 - recommend closing or all schools open with alterations resulting in significant differences compared to non-Covid-19 operations 2 - require closing (only some levels or categories, eg just high school, or just public schools)

	3 - require closing all levels
Closing of workplaces	0 - no measures 1 - recommend closing (or recommend work from home) or all businesses open with alterations resulting in significant differences compared to non-Covid-19 operation 2 - require closing (or work from home) for some sectors or categories of workers 3 - require closing (or work from home) for all-but-essential workplaces (eg grocery stores, doctors)
Cancelling of public events	0- no measures 1 - recommend cancelling 2 - require cancelling
Gathering size limits	0 - no restrictions 1 - restrictions on very large gatherings (the limit is above 1000 people) 2 - restrictions on gatherings between 101-1000 people 3 - restrictions on gatherings between 11-100 people 4 - restrictions on gatherings of 10 people or less
Public transport closure	0 - no measures 1 - recommend closing (or significantly reduce volume/route/means of transport available) 2 - require closing (or prohibit most citizens from using it)
Confinement orders	0 - no measures 1 - recommend not leaving house 2 - require not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips

	3 - require not leaving house with minimal exceptions (eg allowed to leave once a week, or only one person can leave at a time, etc)
Restrictions on internal movement	0 - no measures 1 - recommend not to travel between regions/cities 2 – internal movement restrictions in place
Restrictions on international travel for foreigners	0 - no restrictions 1 - screening arrivals 2 - quarantine arrivals from some or all regions 3 - ban arrivals from some regions 4 - ban on all regions or total border closure

It should be noted that the score given to a country or region is based on the most extreme policy in place for that country or region. This is why population-weighted sums of state-level scores were used to create new national indices for the purpose of this project, reducing the impact of small sections of the population being subject to more extreme measures than the average experience.

4. Experimental

The following methods are simplified from what precisely was done during the project, for the sake of ease of replication. In reality, the processes involved more sporadic experimentation, while in the process of learning the data handling techniques eventually employed. The full process can be found in the associated lab book.

4.1. Policy indices data handling

The raw data was first downloaded from the OxCGRT GitHub[10]. The specific file was OxCGRT_simplified_v1.csv. Along the rows are the various indices and indicators recorded by the Blavatnik School, and along the columns are each day recorded for each country or state alphabetically (i.e. the first data row lists the indices for 01/01/2020 for Aruba, followed by the 02/01/2020 values for Aruba, continuing until 31/12/2022, at which point the next row is the 01/01/2020 values for Afghanistan).

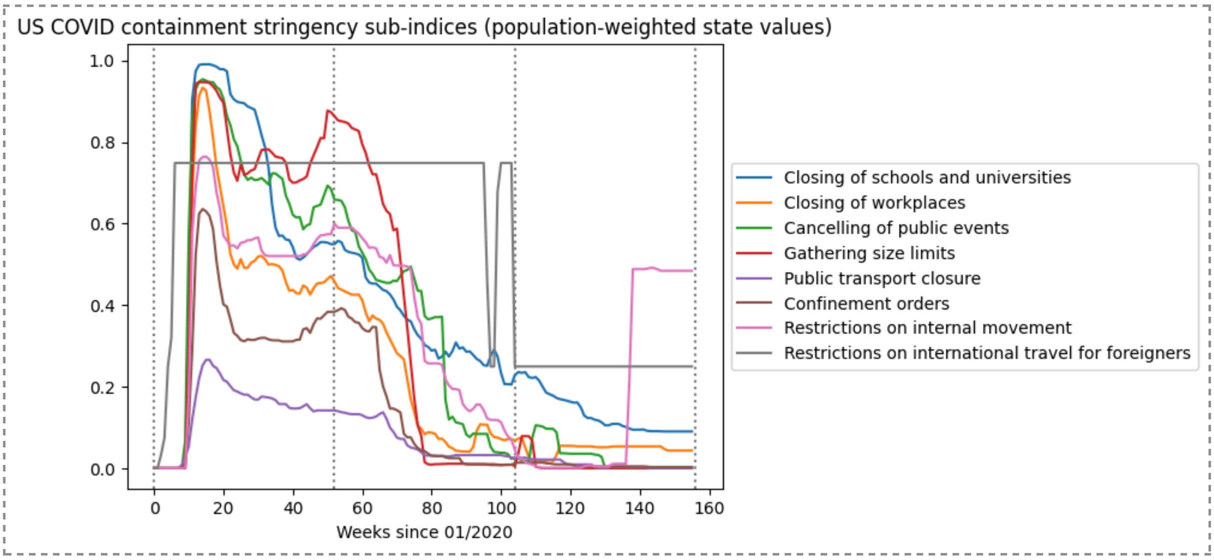
For both countries to be analysed, new CSV files were made in Excel, in which the cells were trimmed to include only the rows of the state–level data of the relevant countries, and only the columns with the relevant containment indices.

As this data would later be week-averaged, the years would have to be normalised to exact 52 week years (364 days). With 2020 being a leap year, this involved removing all rows describing the dates 30/12/2020, 31/12/2020, 31/12/2021 and 31/12/2022. This would very slightly shift the data, but in a manner negligible when averaged by week.

9 new blank columns were then inserted. The first stored the 2020 state populations[11][12] of the state described in each row, and the other 8 used Excel functions to multiply each Index by the state's proportion of population within the country.

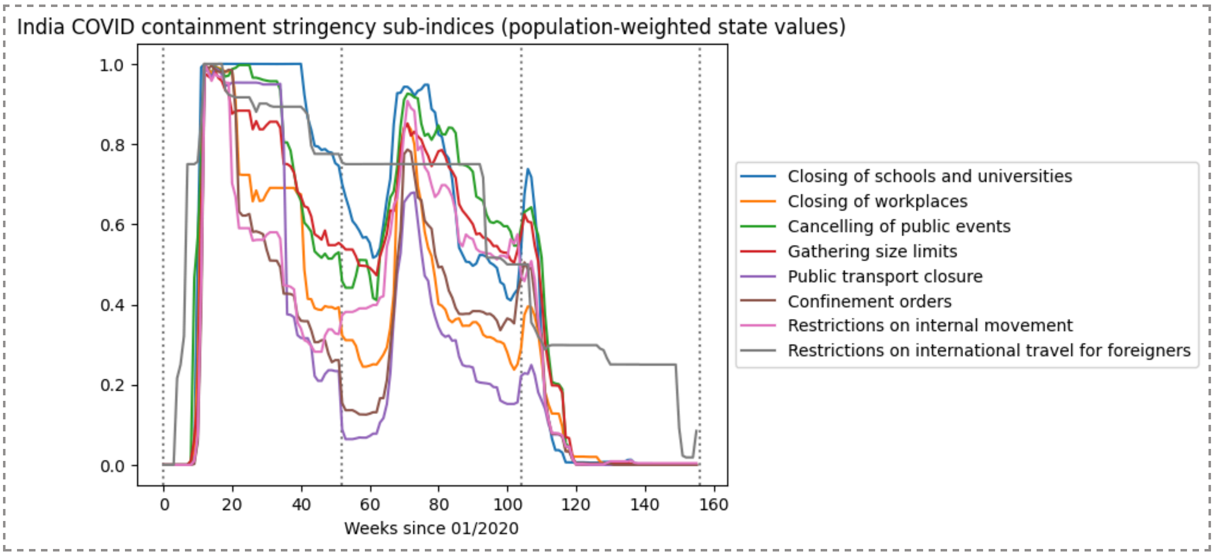
The new CSVs were then loaded into Python, the data then weekly-averaged, and each day in the COVID period summed over every state's population weighted entry for that day, and divided by each index's maximum value. This created 8 time series arrays for each country, varying between 0 and 1, where 0 describes no policy measures in place for the entire population, and 1 describes the maximum measures in place for the whole population.

Figure 1:



Grey dashed lines indicate the start and end of each year

Figure 2:



Grey dashed lines indicate the start and end of each year

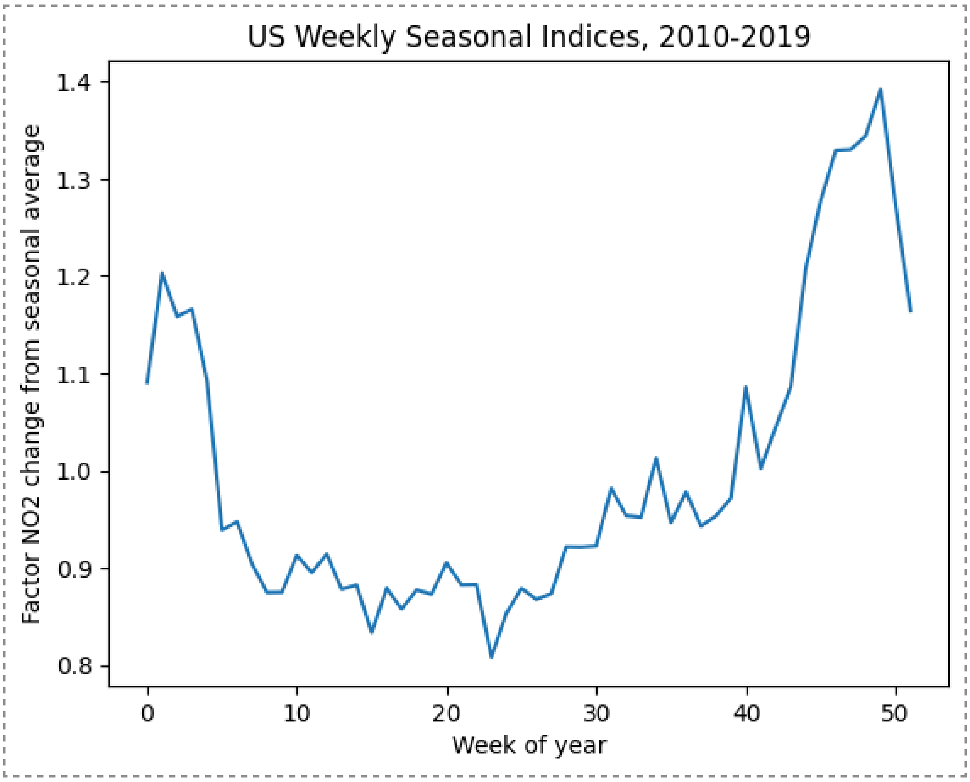
4.2. Nitrogen dioxide concentration cleaning and de-seasonalisation

The OMI tropospheric nitrogen dioxide concentration data was sourced from NASA EarthData's GIOVANNI[8] service. Each country's time-series and area-averaged datasets from 01/01/2010 to 31/13/2022 were generated and downloaded using GIOVANNI's built-in tools and shapefiles.

In both cases, there were missing and anomalous (in this case negative) entries, which were replaced with linearly interpolated values. As before, the data was then normalised to exact 52 week years.

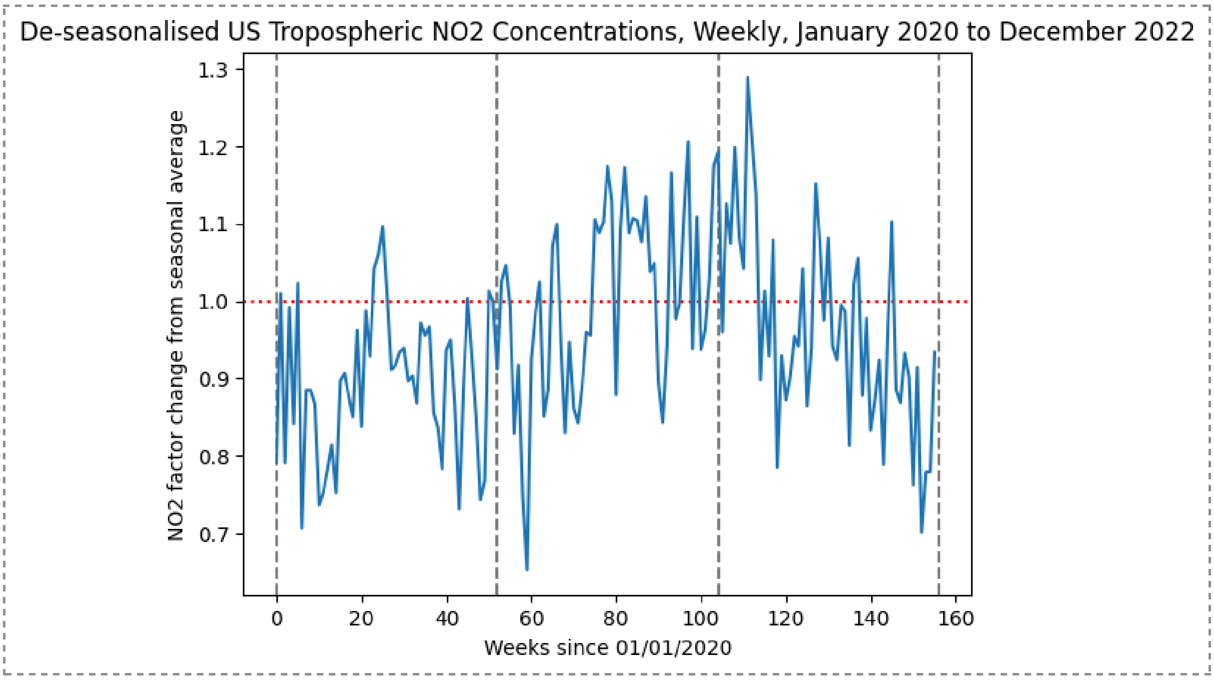
The cleaned data was then loaded into Python and weekly-averaged. To de-seasonalise the data, arrays were created describing the average year profile considering the data from the first week of 2010 to the last week of 2019. In these arrays, the first entry contained the average concentration for the first week of the year, considering every year in the 2010s, and so on, such that the full array described the average yearly cycle.

Figure 3: Average seasonal profile of tropospheric nitrogen dioxide concentrations in the US from 2010 to 2019



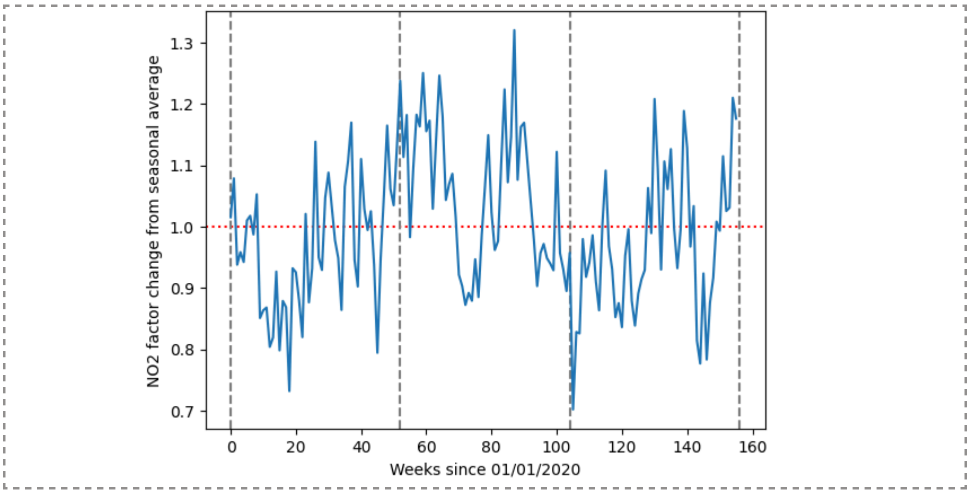
The 2020 to 2022 data was then de-seasonalised from these profiles, each entry divided by the relevant week's average, such that a 1 in the de-seasonalised data represented a week with tropospheric nitrogen concentrations exactly that of the 2010s average, and a 1.5 represented a week where concentrations were 50% higher than the 2010s average.

Figure 4:



Grey dashed lines indicate the start and end of each year
The red dashed line indicates the seasonal average

Figure 5: De-seasonalised India Tropospheric Nitrogen Dioxide Concentrations, Weekly, January 2020 to December 2022



Grey dashed lines indicate the start and end of each year
The red dashed line indicates the seasonal average

4.3. Multivariate analysis

The multivariate regressions were performed in Python using the statsmodels library. For each country, the indices were selected as the independent variables, and the de-seasonalised tropospheric nitrogen dioxide concentrations were chosen as the dependent variables. The regressions were run, and the statistical summaries printed, which included the calculated coefficients, their significance levels, the R^2 value (a measure of how much of the dependent variable the independent variables account for), and other useful statistics.

This was then repeated for each country using only the data for one year at a time, yielding six more sets of results giving insights into the trends in individual years.

Bar chart plots of the calculated coefficients were then produced to visualise the results.

4.4. Model diagnostics

Measures were taken to further assess the quality of the models.

Firstly, correlation matrices were produced, heatmap grids displaying the correlation between indices in every analysis. These were produced for both countries, for the entire COVID period and for each individual year for both.

Also, the model predictions were plotted visually, overlayed on the real nitrogen dioxide data. The visual predictions were produced by multiplying the coefficients by each Index array, and adding the calculated intercept.

Then, regressions were repeated using only the coefficients that were found to be statistically significant in the initial regressions, to mitigate overfitting. The predictions of these limited variable multivariate regressions were also plotted visually as described above.

5. Results

Only results of P-value < 0.01 will be considered statistically significant. This choice is due to the large number (64) of multivariate coefficient results, in that it is expected that some coefficients will have a P-value < 0.1 by coincidence, not having any true causal relation.

Results of P-value less than 0.1 but greater than 0.01 will at times be referred to as ‘weakly statistically significant’.

It should also be noted that due to the precision limit of the statsmodels library used to produce these results, the most extremely statistically significant results will be listed as P-value < 0.001 rather than an exact value, as the exact value could not be determined.

5.1. 3-year multivariate regressions

5.1.1. US 3-year multivariate regression

Table 2: Multivariate regression results for US data spanning 01/2020 to 12/2022 using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	0.3694	0.012
Closing of workplaces	-0.9091	<0.001
Cancelling of public events	0.1243	0.311
Gathering size limits	-0.2136	0.047
Public transport closure	0.7200	0.344
Confinement orders	0.4122	0.171
Restrictions on internal movement	-0.0456	0.453
Restrictions on international travel for foreigners	0.0300	0.574

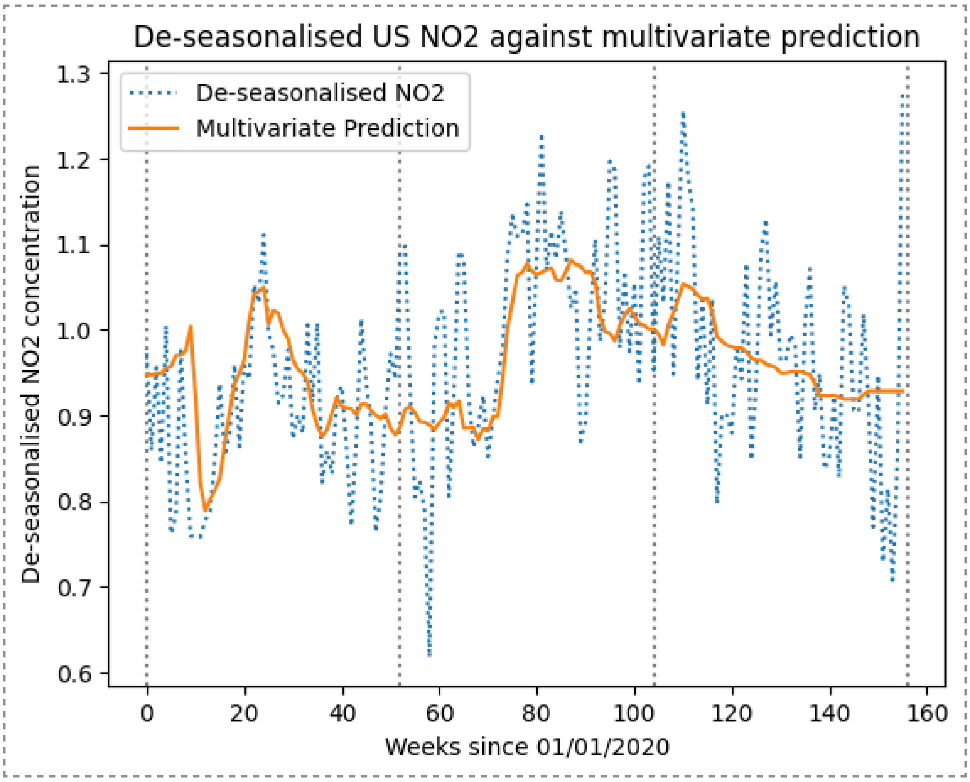
R^2: 0.283

Intercept: 0.9476

Here, workplace closure is the only statistically significant result, while also having the greatest magnitude coefficient, and in the negative direction.

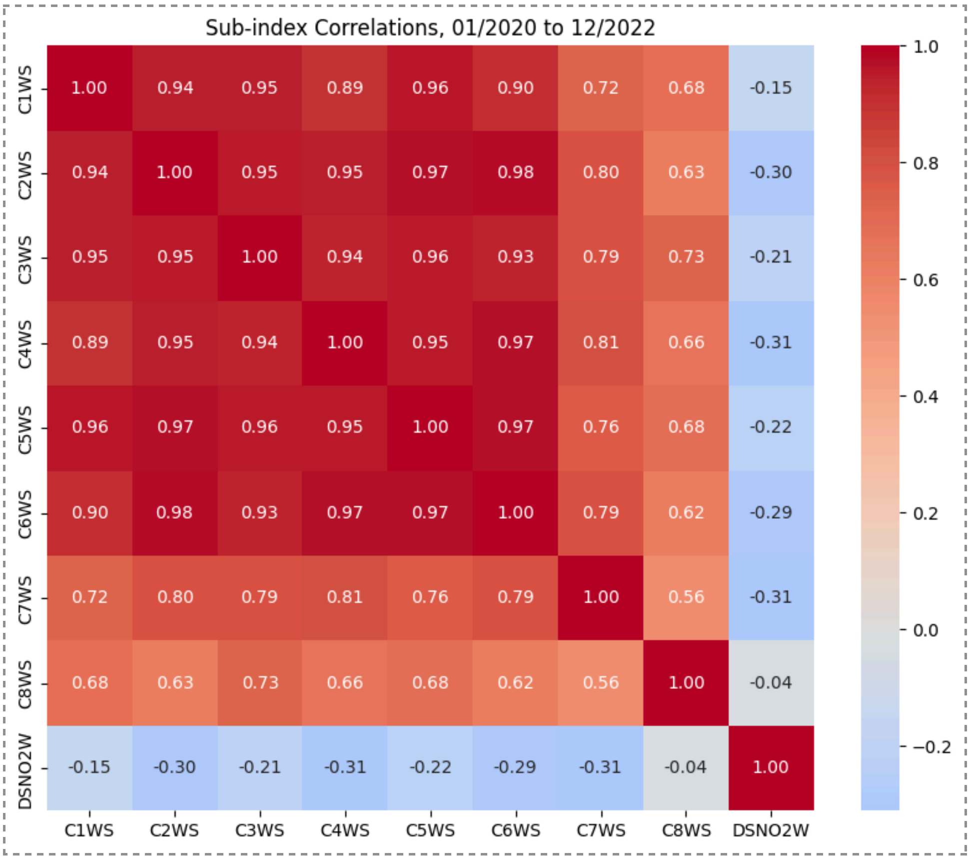
The fit of the coefficients and intercept are plotted against the real data here:

Figure 6:



The correlation between variables is plotted here:

Figure 7: Correlation between variables in the US 3-year multivariate analysis



Key:

- C1WS: Closing of schools and universities
- C2WS: Closing of workplaces
- C3WS: Cancelling of public events
- C4WS: Gathering size limits
- C5WS: Public transport closure
- C6WS: Confinement orders
- C7WS: Restrictions on internal movement
- C8WS: Restrictions on international travel for foreigners
- DSNO2W: De-seasonalised nitrogen dioxide concentrations

5.1.2. India 3-year multivariate regression

Table 3: Multivariate regression results for India data spanning 01/2020 to 12/2022 using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	-0.0734	0.536
Closing of workplaces	0.0562	0.717
Cancelling of public events	-0.1006	0.450
Gathering size limits	0.4062	0.070
Public transport closure	0.0373	0.610
Confinement orders	-0.7684	<0.001
Restrictions on internal movement	0.3085	0.017
Restrictions on international travel for foreigners	0.0291	0.677

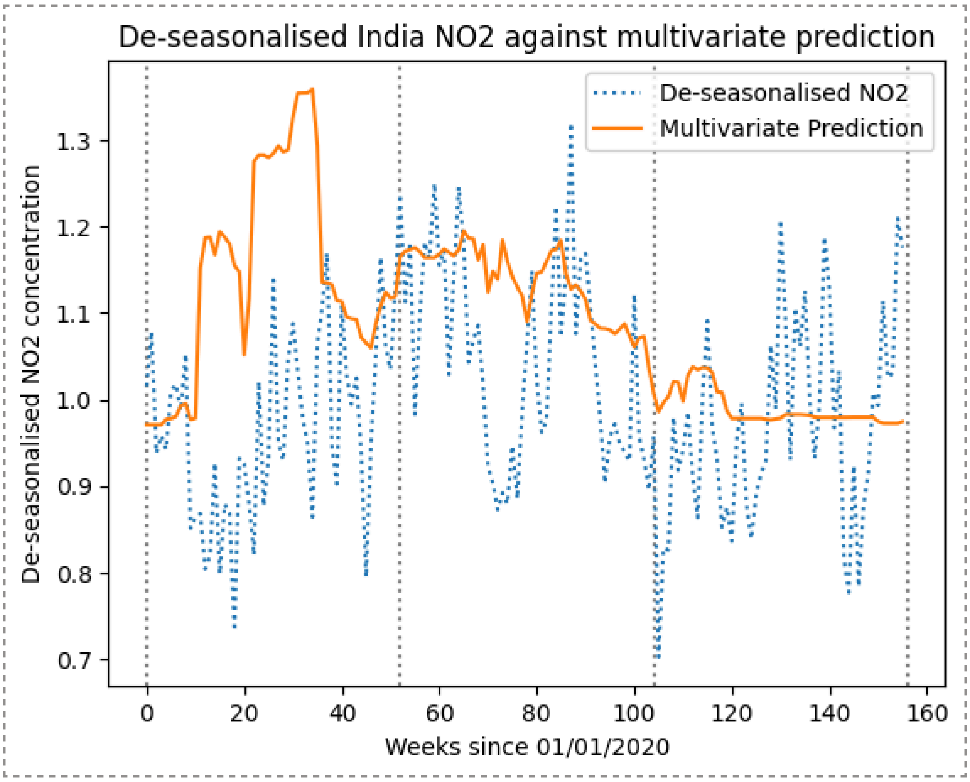
R^2: 0.283

Intercept: 0.9709

Here, confinement orders are the only statistically significant result, while also having the greatest magnitude coefficient, and in the negative direction.

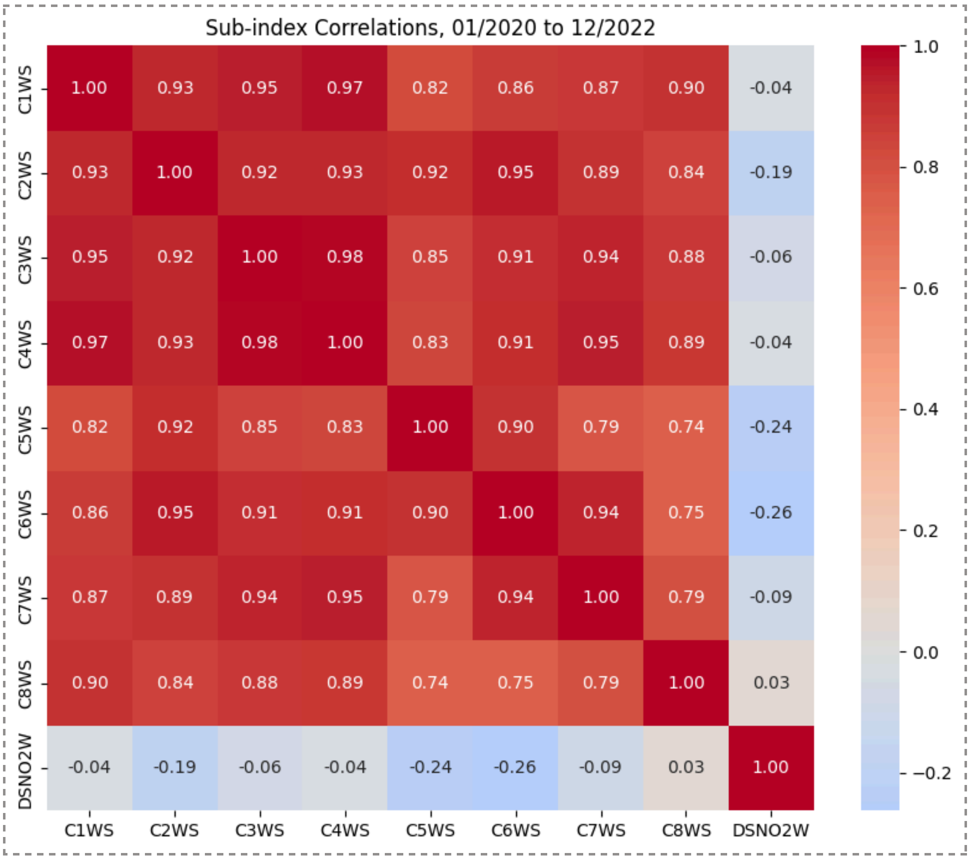
The fit of the coefficients and intercept are plotted against the real data here:

Figure 8:



The correlation between variables is plotted here:

Figure 9: Correlation between variables in the India 3-year multivariate analysis



Key:

- C1WS: Closing of schools and universities
- C2WS: Closing of workplaces
- C3WS: Cancelling of public events
- C4WS: Gathering size limits
- C5WS: Public transport closure
- C6WS: Confinement orders
- C7WS: Restrictions on internal movement
- C8WS: Restrictions on international travel for foreigners
- DSNO2W: De-seasonalised nitrogen dioxide concentrations

5.2.1 US annual multivariate regressions

2020:

Table 4: Multivariate regression results for US data spanning 2020 only, using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	0.2479	0.121
Closing of workplaces	-0.8230	0.016
Cancelling of public events	0.0321	0.907
Gathering size limits	0.0334	0.912
Public transport closure	1.2900	0.073
Confinement orders	0.3303	0.485
Restrictions on internal movement	-0.0878	0.841
Restrictions on international travel for foreigners	-0.1208	0.091

R^2: 0.419

Intercept: 0.9157

This regression yielded no statistically significant results.

2021:

Table 5: Multivariate regression results for US data spanning 2021 only using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	-1.0558	0.163
Closing of workplaces	1.3403	0.190
Cancelling of public events	0.0351	0.920
Gathering size limits	-1.2739	0.024
Public transport closure	1.9997	0.345
Confinement orders	0..7730	0.135
Restrictions on internal movement	0.4763	0.435
Restrictions on international travel for foreigners	-0.0188	0.910

R^2**:** 0.432

Intercept: 1.1204

This analysis yielded no statistically significant results.

2022:

Table 6: Multivariate regression results for US data spanning 01/2020 to 12/2022 using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	1.4299	0.401
Closing of workplaces	1.5014	0.515
Cancelling of public events	2.6281	0.099
Gathering size limits	2.1086	0.159
Public transport closure	-12.7544	0.079
Confinement orders	1.3409	0.934
Restrictions on internal movement	-0.0676	0.503
Restrictions on international travel for foreigners	2.9433	<0.001

R^2**:** 0.302

Here, restrictions on international travel for foreigners is the only statistically significant result.

5.2.2 India annual multivariate regressions

2020:

Table 7: Multivariate regression results for India data spanning 2020 only using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	0.0240	0.943
Closing of workplaces	0.0712	0.807
Cancelling of public events	-0.3516	0.081
Gathering size limits	0.4775	0.323
Public transport closure	0.0946	0.485
Confinement orders	-0.3314	0.211
Restrictions on internal movement	-0.1275	0.654
Restrictions on international travel for foreigners	0.0132	0.896

R^2**:** 0.367

Intercept: 0.9925

This analysis yielded no statistically significant results.

2021:

Table 8: Multivariate regression results for India data spanning 2021 only using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	0.2371	0.249
Closing of workplaces	0.1972	0.531
Cancelling of public events	1.2827	0.001
Gathering size limits	-1.0310	0.074
Public transport closure	-0.3743	0.520
Confinement orders	-0.3521	0.275
Restrictions on internal movement	-0.5137	0.255
Restrictions on international travel for foreigners	0.4217	0.038

R^2**:** 0.701

Intercept: 0.8359

Here, cancelling of public events is the only statistically significant result, while also having the greatest magnitude coefficient, though in the positive direction.

2022:

Table 9: Multivariate regression results for India data spanning 2022 only using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	-0.2304	0.758
Closing of workplaces	-1.5936	0.311
Cancelling of public events	0.8858	0.452
Gathering size limits	1.2345	0.430
Public transport closure	-2.8351	0.276
Confinement orders	-2.7443	0.086
Restrictions on internal movement	2.8695	0.098
Restrictions on international travel for foreigners	-0.4896	0.016

R^2: 0.370

Intercept: 1.0975

This analysis yielded no statistically significant results.

5.2.3. Single index analyses

In the case of the 3-year results, regressions were repeated using only indices previously found to be statistically significant as independent variables. In both cases, only one variable was statistically significant, so these are both single variable linear regressions.

5.2.3.1. 3-year US workplace closure model

Workplace closure was the only statistically significant coefficient in the 3-year US multivariate regression, so a single independent variable linear regression was run.

With the workplace closure index as the only independent variable, and the de-seasonalised nitrogen dioxide profile as the dependent variable, the following values were found:

Intercept: 0.9938

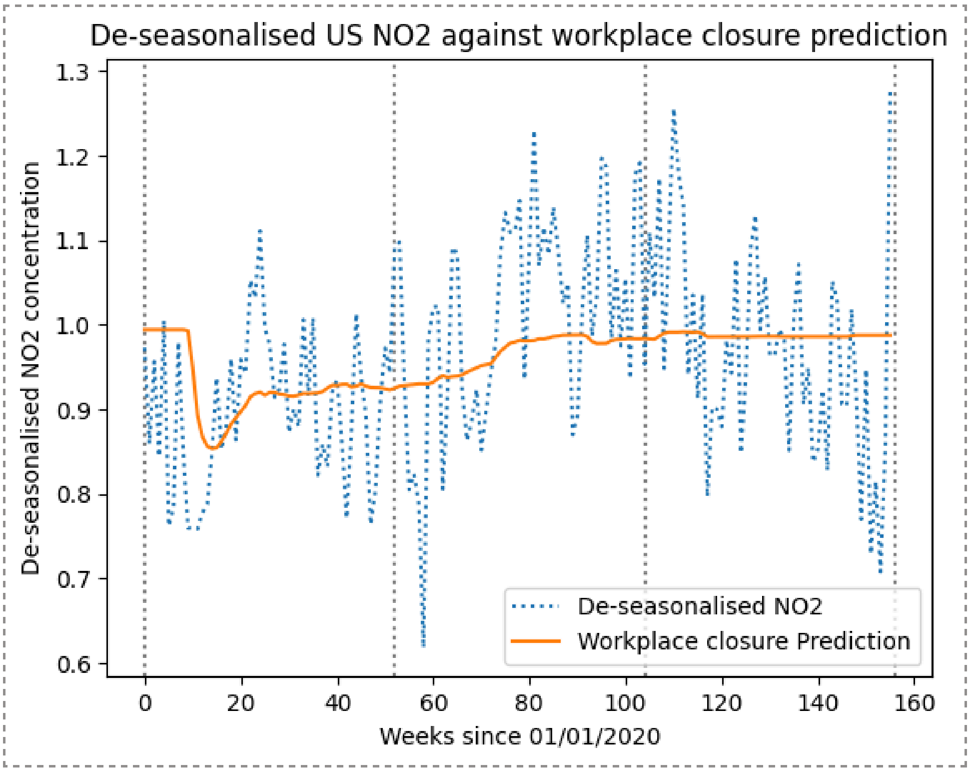
Coefficient: -0.1504

P-value: < 0.001

R^2: 0.088

The fit of the coefficient and intercept are plotted against the real data here:

Figure 10:



5.2.3.2. 3-year India confinement orders model

Confinement orders was the only statistically significant coefficient in the 3-year India multivariate regression, so a single independent variable regression was run.

With the confinement orders index as the only independent variable, and the de-seasonalised nitrogen dioxide profile as the dependent variable, the following values were found:

Intercept: 1.0287

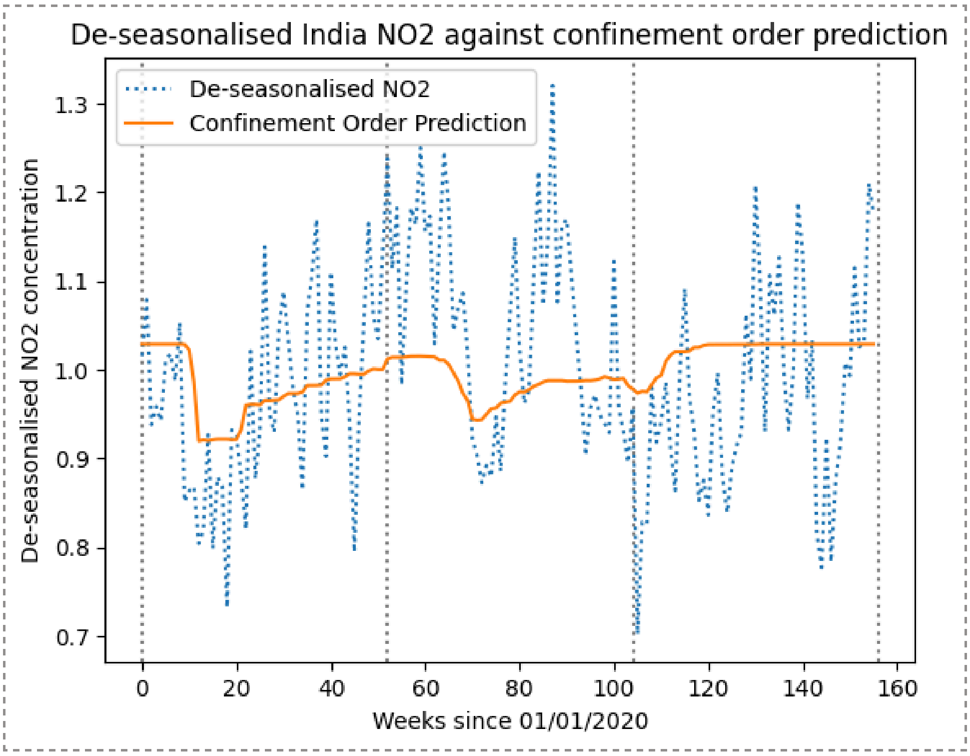
Coefficient: -0.1092

P-value: 0.001

R^2: 0.069

The fit of the coefficient and intercept are plotted against the real data here:

Figure 11:



6. Discussion

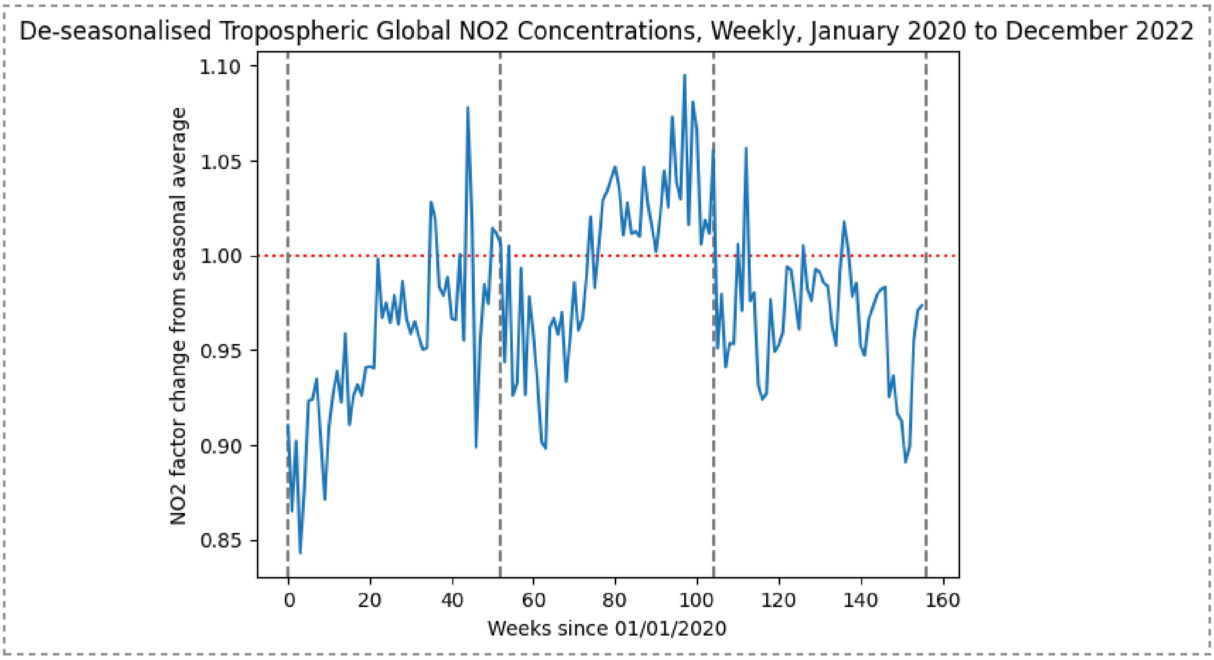
6.1. Nitrogen dioxide profiles

The plots of the de-seasonalised nitrogen dioxide profiles alone yielded some insights of note, so they will be discussed here.

6.1.1. General trends

In the initial exploratory phase of the project, a global analysis was considered, and so a de-seasonalised plot of the global nitrogen dioxide profile was produced using the same method as was later used for the specific countries.

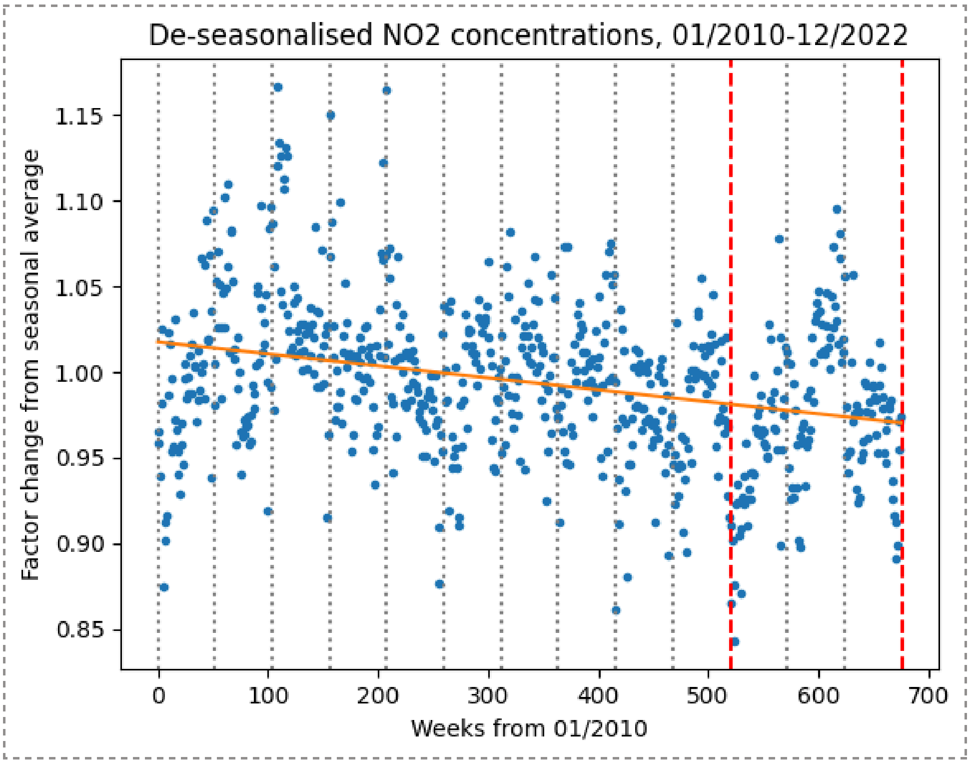
Figure 12:



Grey dashed lines indicate the start and end of each year
The red dashed line indicates the seasonal average

Here is the de-seasonalised plot from 2010 onward, the COVID period being studied demarcated in red.

Figure 13:



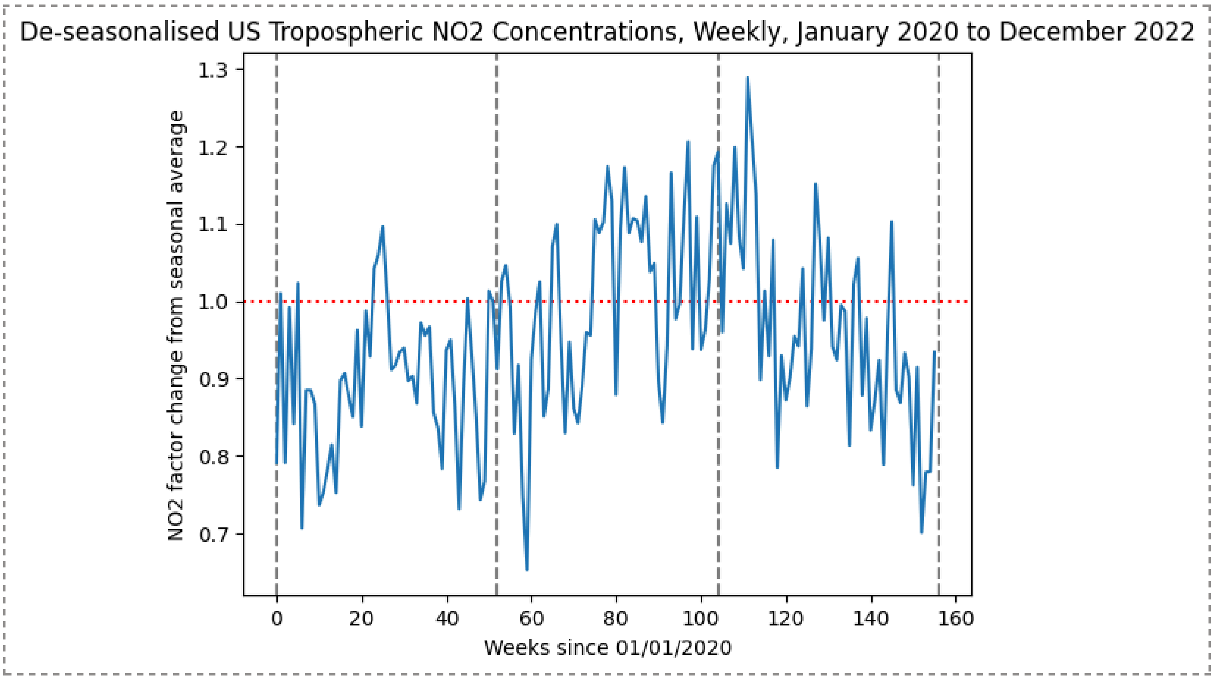
From this plot, it is apparent that the clearest deviations between the decade trend and that of the COVID period are the lower-than-trend cluster in the first half of 2020, and the higher-than-trend cluster in the second half of 2021.

The lower 2020 cluster was expected, considering that lockdowns generally took place at this time[9], but the higher cluster in the second half of 2021 was not expected. There was worry that this implied a confounding variable unaccounted for by containment policy, and for this reason, attempts were made to include oil, coal and gas production data[13] in the analysis (being large nitrogen dioxide contributors[14]), so as to isolate the policy contributions and perhaps also find an explanation for this 2021 anomaly. Unfortunately however there were problems in accessing the necessary data, and this avenue of research was unsuccessful.

There was also some hope that the anomaly could be explained by policy. Though one would expect COVID containment policy to only result in nitrogen dioxide levels being lower than the decade average, it would be possible for public transport closure to have the opposite effect, as demand is driven for less combustion-efficient modes of transport, but unfortunately no such results relating to public transport closure were found.

Of further note is that the US and to a lesser extent the India de-seasonalised nitrogen dioxide profiles follow broadly the same pattern as that of the global data, with levels being at first broadly lower than the decade average, particularly in the first half of 2020, until around the midpoint of 2021, at which point the concentrations are strangely above the decade average, before returning to below the decade average in 2022.

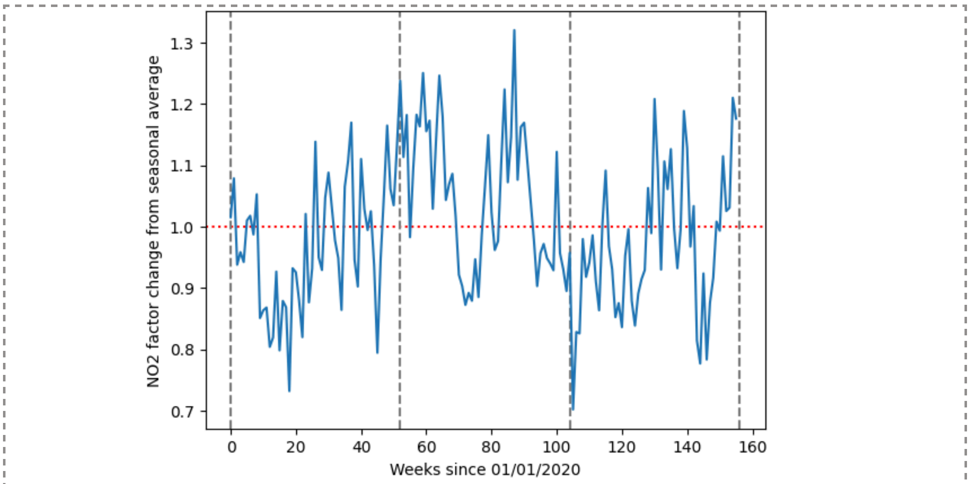
Figure 4:



Grey dashed lines indicate the start and end of each year
The red dashed line indicates the seasonal average

Though the broad trend aligns, notable differences in the US profile include a less even increase in 2020 with a slightly delayed minimum, and the above-decade-average section extending into 2022.

Figure 5: De-seasonalised India Tropospheric Nitrogen Dioxide Concentrations, Weekly, January 2020 to December 2022



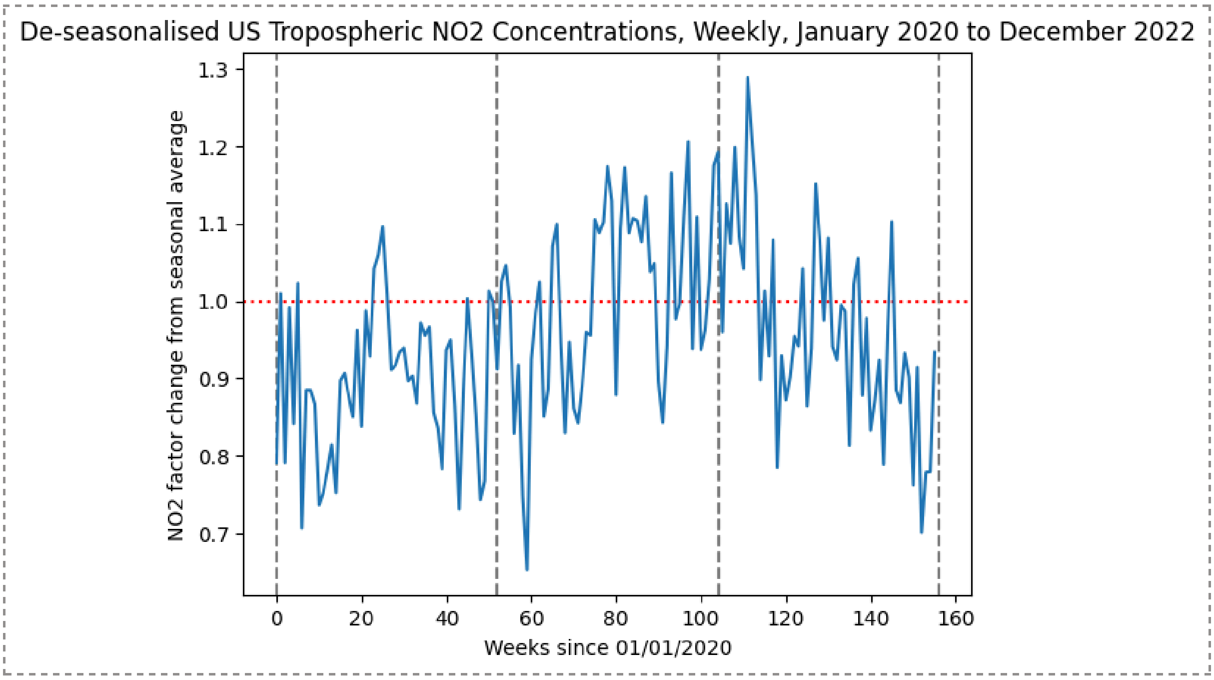
Grey dashed lines indicate the start and end of each year
The red dashed line indicates the seasonal average

Once again the very broad trend aligns, but notable differences once more include a less even increase in 2020 with a slightly delayed minimum, and in this case the above-decade-average sections are more sporadic.

6.1.2. Random variation

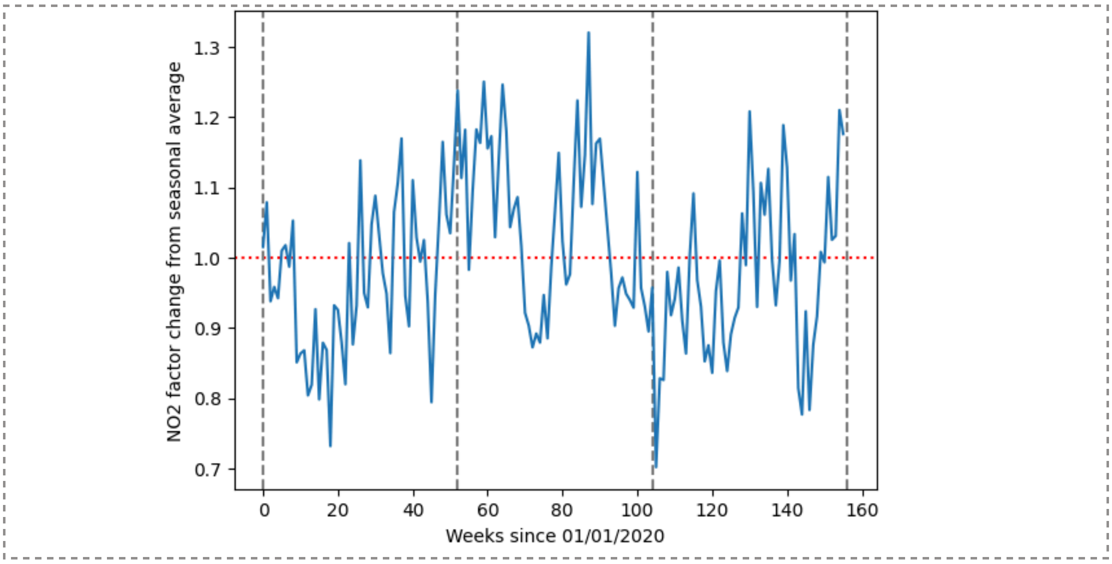
One serious limitation in this project was the very high random variation in the nitrogen dioxide concentration data. Even after averaging by week, there was significant ‘fuzz’ to the data, appearing to vary randomly about a trend.

Figure 4:



Grey dashed lines indicate the start and end of each year
The red dashed line indicates the seasonal average

Figure 5: De-seasonalised India Tropospheric Nitrogen Dioxide Concentrations, Weekly, January 2020 to December 2022



Grey dashed lines indicate the start and end of each year
The red dashed line indicates the seasonal average

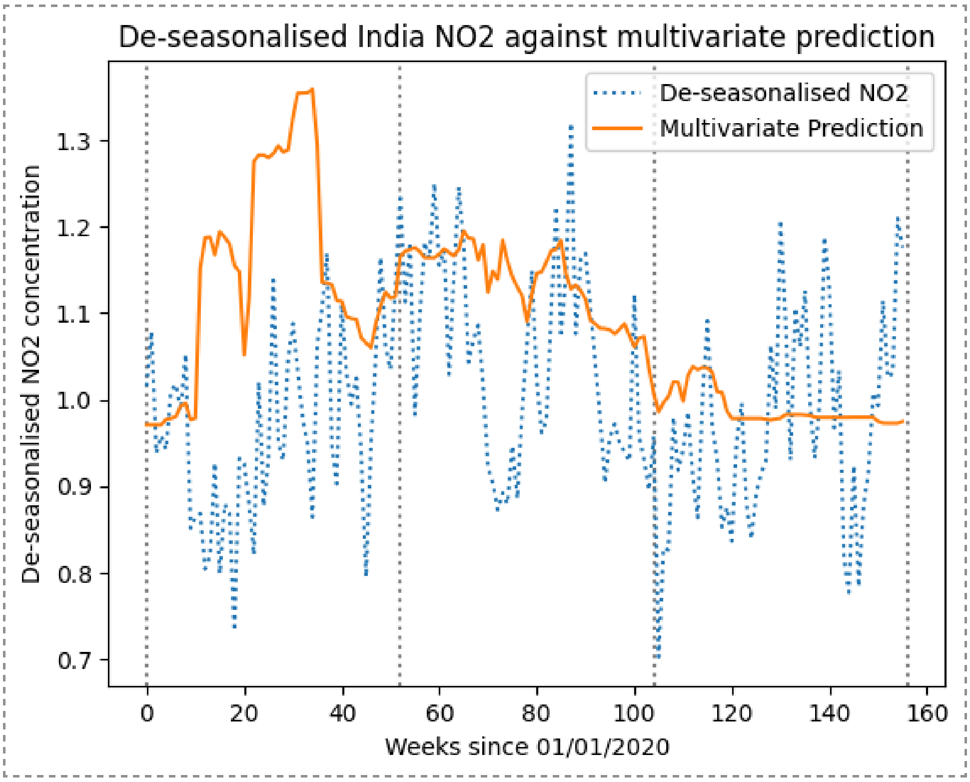
This could have been partially mitigated by averaging the data over a greater timeframe, such as averaging monthly, but with the data already being over a fairly small period from which to differentiate the influence of 8 variables, this was not done for fear that it would leave a dataset too small to yield meaningful results, particularly in the individual year analyses, for which only 12 data points would remain.

This limited the efficacy of the multivariate regressions, as the models attempted to set coefficients that would have the independent variables cause the random variation, which runs counter to the expectation that the influence of the policies should only relate to the more general profile.

Related to this is the issue that the R^2 statistic became unreliable as a measure of a model's fit. One would expect that a higher R^2 value would imply a fit closer to the general trend of the data, but in some cases it was found that in attempting to account for random variation, this was no longer always the case.

For example, here is the plot of the full 3-year 8 index regression for India:

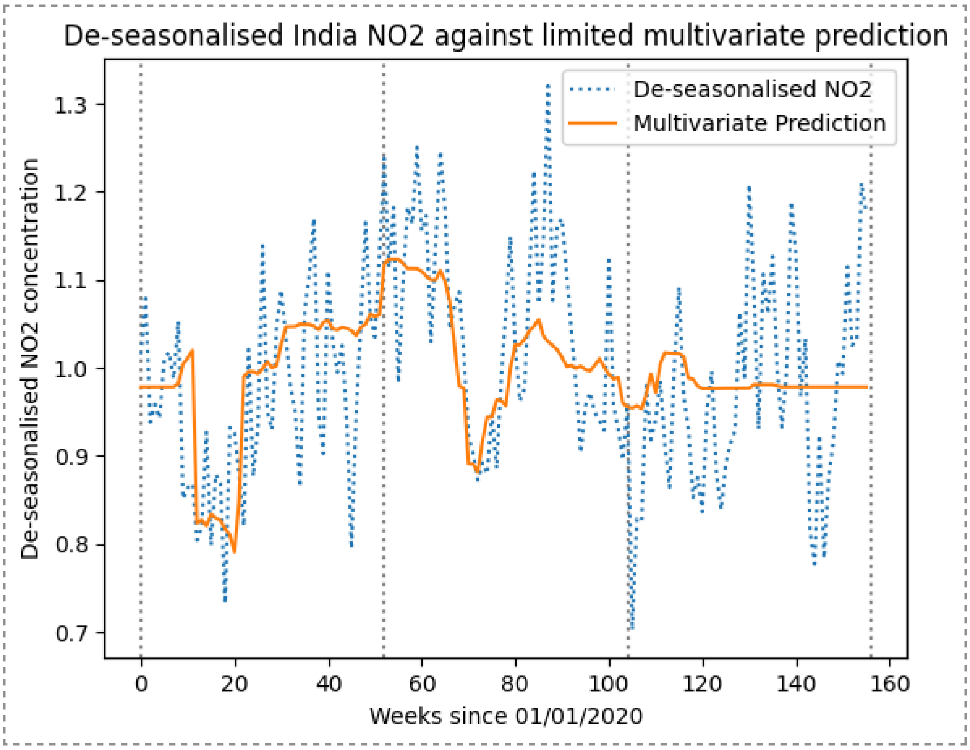
Figure 8:



This regression had an R^2 value of 0.343, with the visual fit appearing quite poor, particularly in 2020.

To contrast, while experimenting with different limited variable models, this fit was produced, using only confinement orders and gathering size limits as the independent variables:

Figure 14: De-seasonalised India nitrogen dioxide concentrations against the fitted multivariate regression using the confinement orders and gathering size limit indices as independent variables



As is always the case when using fewer of the same set of independent variables, this regression yielded a lower R^2 value, in this case 0.300. Despite that, it is strikingly visually apparent that it fits the general trend of the data far better than that of the 8 index regression with a higher R^2 value.

This confirms that in this project, the R^2 statistic unfortunately can't be fully relied on as a measure of a model's quality in fitting the general nitrogen dioxide profile.

Furthermore, as the models aim to maximise the R^2 value, this implies that the coefficients were being set in a manner sub-optimal for predicting the general nitrogen dioxide trends.

6.2. Variable correlation

Another serious limitation to the multivariate regressions was the very high correlation between the indices, for both the US and India.

Figure 7: Correlation between variables in the US 3-year multivariate analysis

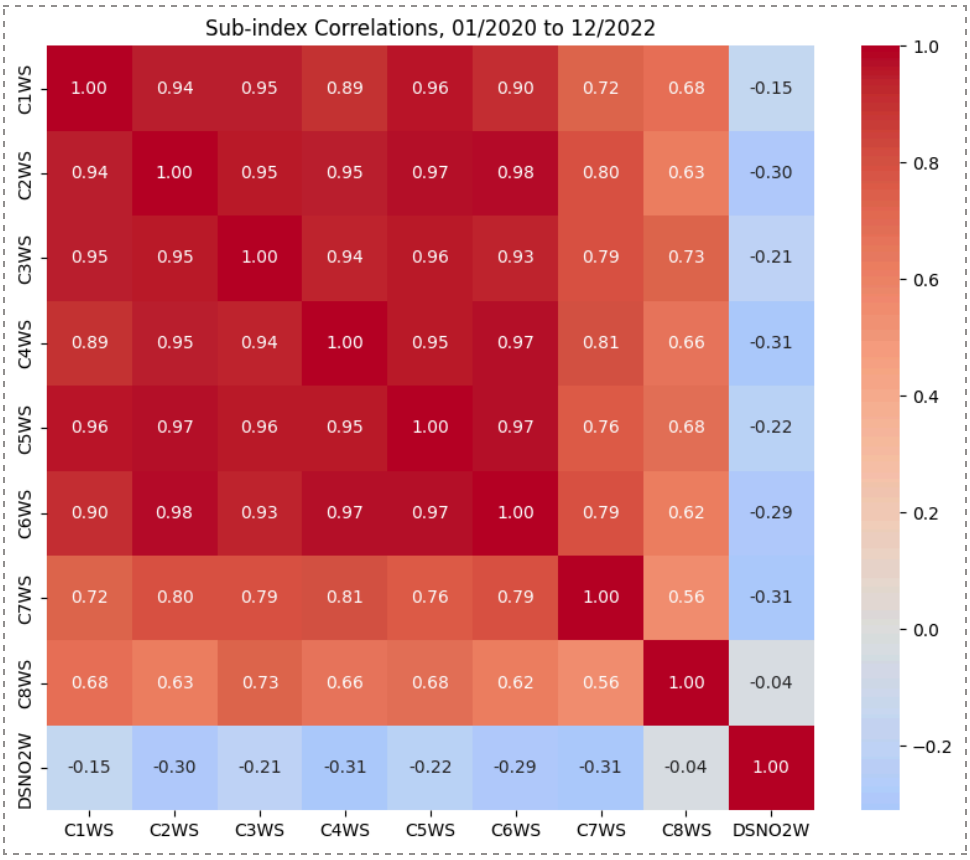
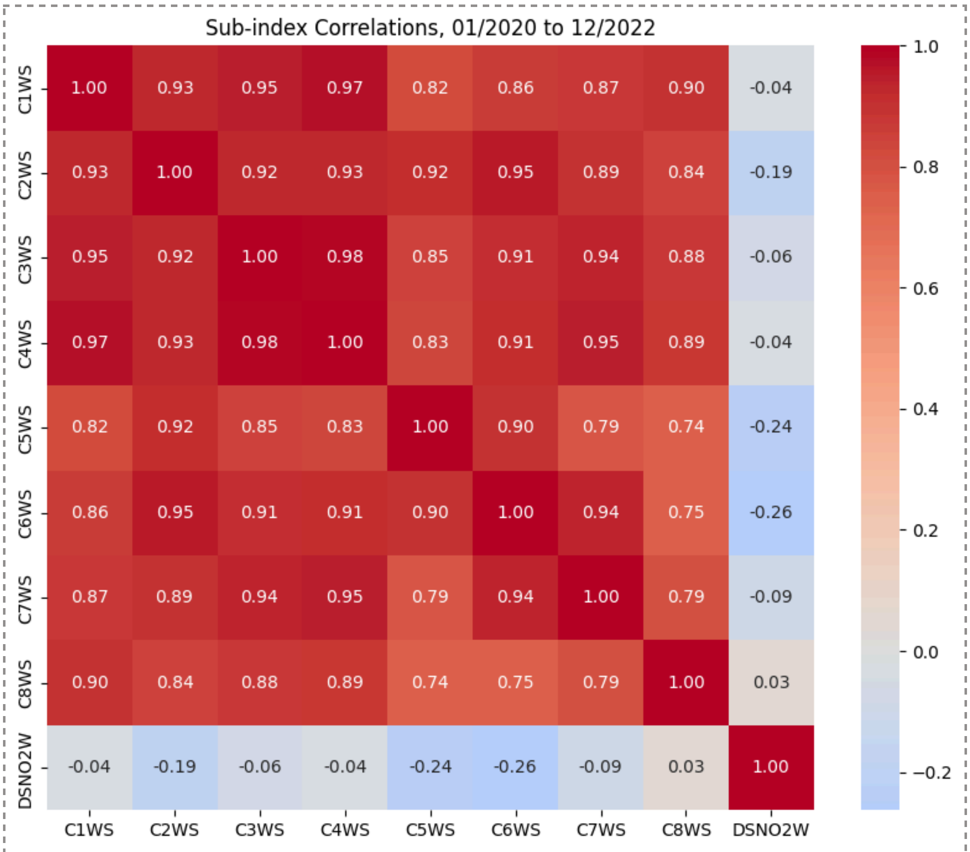


Figure 9: Correlation between variables in the India 3-year multivariate analysis



Key:

- C1WS: Closing of schools and universities
- C2WS: Closing of workplaces
- C3WS: Cancelling of public events
- C4WS: Gathering size limits
- C5WS: Public transport closure
- C6WS: Confinement orders
- C7WS: Restrictions on internal movement
- C8WS: Restrictions on international travel for foreigners
- DSNO2W: De-seasonalised nitrogen dioxide concentrations

High independent variable correlation poses a serious problem for multivariate regressions, in that it becomes extremely difficult for the model to differentiate between the influence of each independent variable's impact on the dependent variable. As such, these high correlations are a likely cause of the sparsity of statistically significant results in the regressions performed for this project.

6.3. 3-year results

6.3.1. 3-year US results

Table 2: Multivariate regression results for US data spanning 01/2020 to 12/2022 using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	0.3694	0.012
Closing of workplaces	-0.9091	<0.001
Cancelling of public events	0.1243	0.311
Gathering size limits	-0.2136	0.047
Public transport closure	0.7200	0.344
Confinement orders	0.4122	0.171
Restrictions on internal movement	-0.0456	0.453
Restrictions on international travel for foreigners	0.0300	0.574

R^2**:** 0.283

Intercept: 0.9476

With the workplace closure coefficient being so strongly negative and so strongly statistically significant, it seems that the closure of workplaces had the most impact on nitrogen dioxide levels in the US out of COVID containment policy factors. However, unfortunately this result isn't entirely reliable due to the model quality issues discussed earlier (high random variation in nitrogen dioxide concentrations and high index correlation).

This is further apparent in that there is only one statistically significant result, when one would expect that at the very least, containment orders should have an impact on the nitrogen dioxide profile considering that such policies are the most directly restrictive on movement. If we can assume this to be the case, the fact that the model couldn't detect this to a significant degree is worrying.

If one considers more weakly significant results as significant (P-value < 0.1), then this regression also suggests that the closure of schools and universities (coefficient = 0.3694, P-value = 0.012) and gathering size limits (coefficient = -0.2136, P-value = 0.047) have measurable impact on the nitrogen dioxide profile. However, the fact that the closure of schools and universities coefficient is positive brings the causal meaningfulness of these results into doubt, as it would be highly unusual for the closing of educational institutions to somehow result in greater nitrogen dioxide concentrations, with fewer people travelling to and from educational institutions.

The repeat of the model using only workplace closure as an independent variable is likely a better source for meaningful inferences, as it is not distorted by the influence of statistically insignificant indices:

Intercept: 0.9938

Coefficient: -0.1504

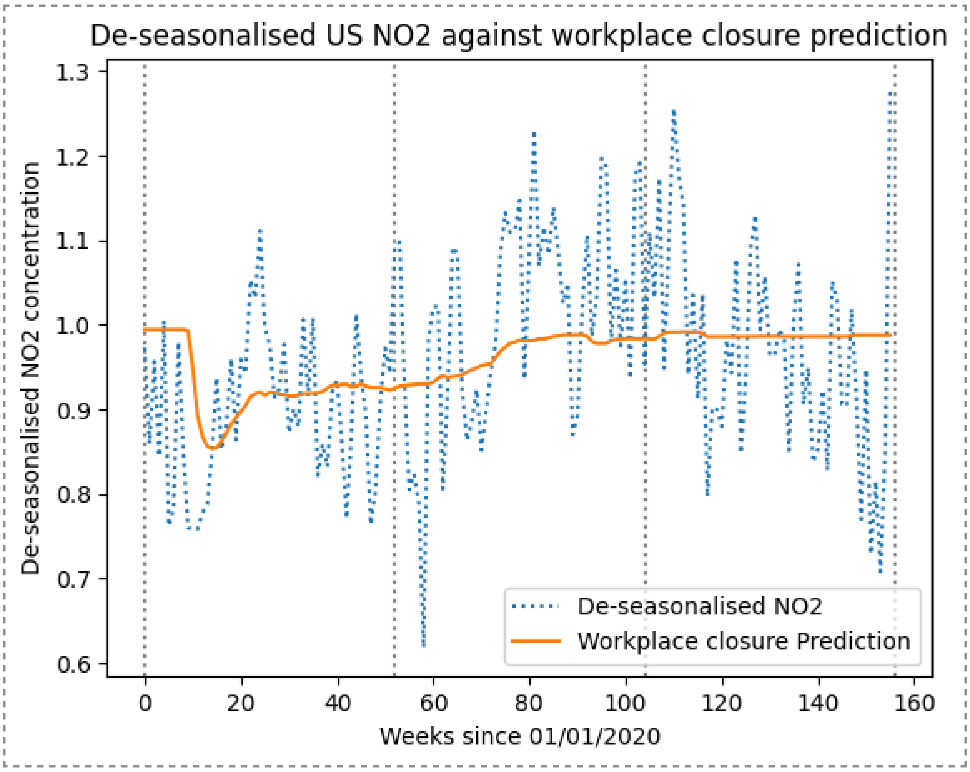
P-value: < 0.001

R^2: 0.088

Considering how the index was normalised between 0 and 1, the coefficient suggests a 15% drop in nitrogen dioxide concentrations from the decade average associated with maximum implementation, though this should not be taken as a causal relation.

The exceptionally low R^2 value suggests that this index didn't successfully predict much of the nitrogen dioxide profile, but as discussed earlier, the R^2 statistic can be an underestimate of an index's predictivity due to the high random variation in the nitrogen dioxide profile.

Figure 10:



From the visual plot it is apparent that the workplace closure at least vaguely tracks with the more general trend of the nitrogen dioxide profile, though this is hardly sufficient to draw any confident conclusions unfortunately.

6.3.2. 3-year India results

Table 3: Multivariate regression results for India data spanning 01/2020 to 12/2022 using all containment indices as independent variables and the de-seasonalised tropospheric nitrogen dioxide concentration profile as the dependent variable

Index	Coefficient	P-value
Closing of schools and universities	-0.0734	0.536
Closing of workplaces	0.0562	0.717
Cancelling of public events	-0.1006	0.450
Gathering size limits	0.4062	0.070
Public transport closure	0.0373	0.610
Confinement orders	-0.7684	<0.001
Restrictions on internal movement	0.3085	0.017
Restrictions on international travel for foreigners	0.0291	0.677

R^2**:** 0.283

Intercept: 0.9709

Similarly to the US results, though with a different index, the confinement orders coefficient is strongly negative and strongly statistically significant. So, it seems that confinement orders had the most impact on nitrogen dioxide levels in India out of COVID-19 containment policy factors. However, as before, unfortunately this result isn't reliable due to the model quality issues of high random variation in nitrogen dioxide concentrations and high index correlation.

As in the US 3-year analysis, this is further apparent in that there is only one statistically significant result, when one would expect that multiple policy areas would have an impact, particularly considering that workplace closure seems to have had such an impact in the US. The fact that the model couldn't detect the influence of other factors to a significant degree brings more doubt as to the quality of the model.

If one considers more weakly significant results as significant (P-value < 0.1), then this regression also suggests that gathering size limits (coefficient = 0.4062, P-value = 0.070) and restrictions on internal movement (coefficient = 0.3085, P-value = 0.017) have measurable impact on the nitrogen dioxide profile. However, both of these coefficients are positive, bringing their causal meaning into serious doubt, as it would be highly unusual for these measures that both presumably would reduce automotive travel to then somehow cause greater nitrogen dioxide emissions.

The repeat of the model using only confinement orders as an independent variable is likely a better source for meaningful inferences, as it is not distorted by the influence of statistically insignificant indices:

Intercept: 1.0287

Coefficient: -0.1092

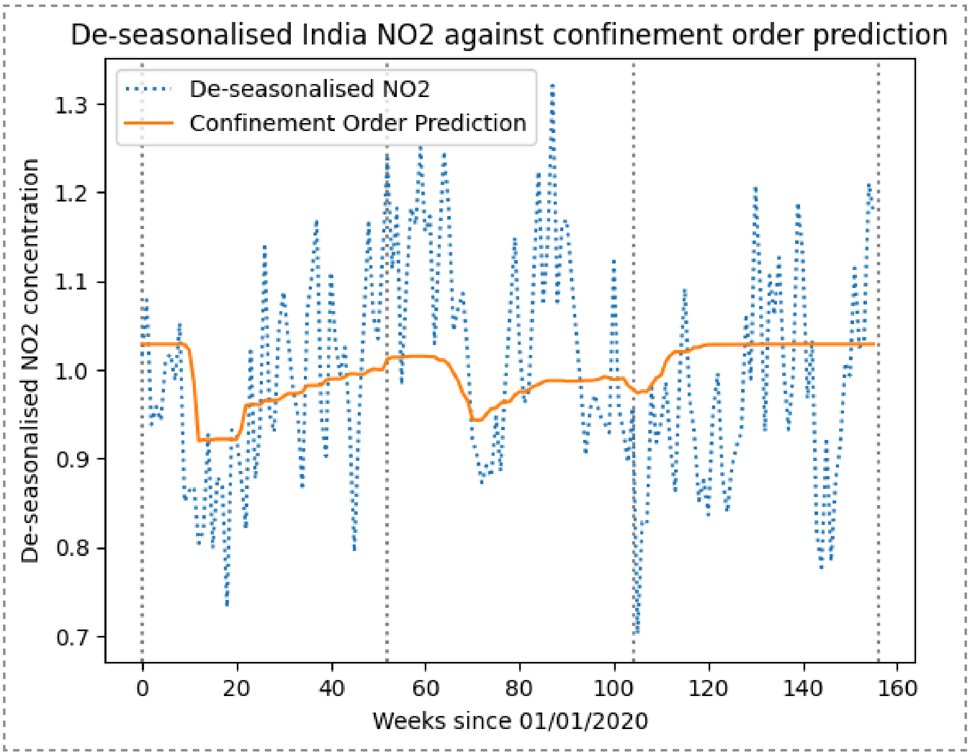
P-value: 0.001

R^2: 0.069

Once more considering the index normalisation, the coefficient suggests an 11% drop in nitrogen dioxide concentrations from the decade average associated with maximum implementation of confinement orders, though as before this should not be taken as a causal relation.

Similarly to the US workplace closure model, the exceptionally low R^2 value would ordinarily suggest that this index was not predictive of much of the nitrogen dioxide profile, but we know that in this case the R^2 statistic can be an underestimate of predictivity due to the random variation issue discussed earlier.

Figure 11:



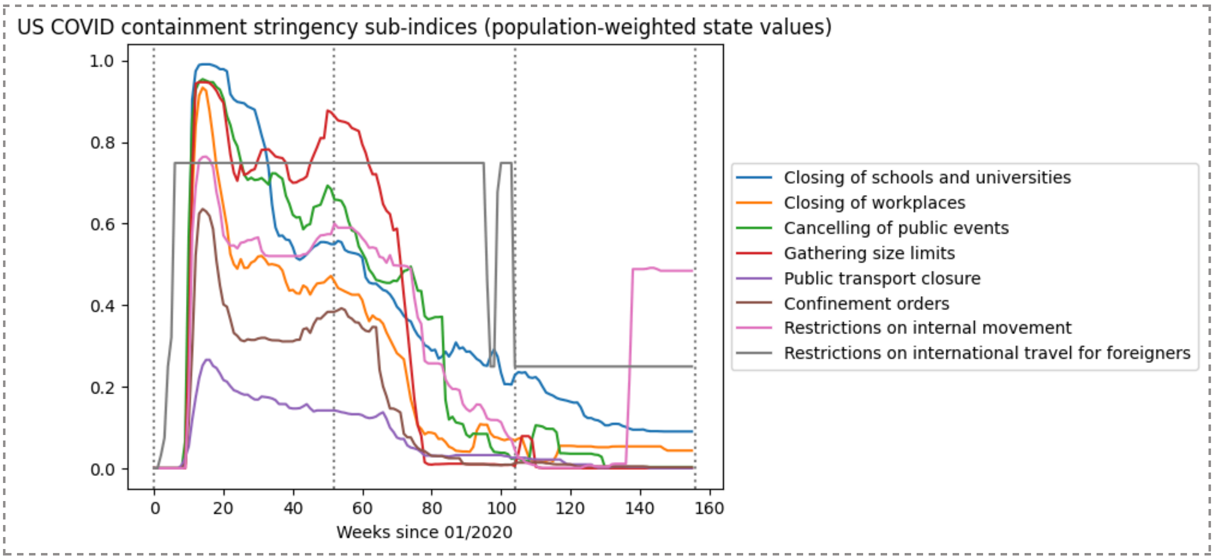
From this visual plot, it is apparent that the confinement orders index at least vaguely tracks with the more general trend of the nitrogen dioxide profile, though with the notable exception of the spike in the second half of 2021. Unfortunately, once again, this is not sufficient to draw confident causal conclusions, though it does support earlier speculation of a confounding variable being responsible for the higher than decade-average nitrogen dioxide levels in the second half of 2021.

6.4. Single year results

Unfortunately, the single year results are not of sufficient quality to draw conclusions. This conclusion is drawn from the fact that there were only two statistically significant results, and those result didn't make any sense as a potential causal relationship.

The first statistically significant result was for the 2022 US analysis, for the restrictions on international travel for foreigners index (coefficient = 2.9433, P-value < 0.001). This does not make sense as a causal relation, as fewer people travelling by plane, a nitrogen dioxide emitting[15] form of transport, should not result in more nitrogen dioxide in the atmosphere. A likely cause of this anomalous statistical significance becomes apparent when viewing the US index plot:

Figure 1:



Grey dashed lines indicate the start and end of each year

The plot shows that the index describing restrictions on international travel for foreigners is constant throughout 2022. Extremely low variance in independent variables in regressions can result in decreased P-values not representative of a meaningful relation, as the incredibly simple relation can be integrated easily into any fit regardless of predictivity.

The other statistically significant result was for the 2021 India regression, wherein the cancelling of public events was statistically significant (coefficient = 1.2827, P-value = 0.001), though with a positive coefficient, which makes little sense as being indicative of a causal relationship.

So, it seems unfortunately that the existing quality issues, when combined with the fact that only a third of every dataset was used in each case, have critically limited the usefulness of the single year analyses.

6.5. Improvements and future work

6.5.1. Improvements for a similar study

6.5.1.1. Mitigating random variation in tropospheric nitrogen dioxide concentrations

The main limitations of this project have been the high random variation in the tropospheric nitrogen dioxide concentration data, and the high correlation between the policy indices.

In the first case, averaging all data by month would result in an improvement to the random variation issue. This was not done during this project in large part because it was thought that running regressions for individual years would be crucial for the project, and averaging by month would result in regressions over only 12 data-points, which would be unlikely to yield meaningful results. However, in large part due to the high random variation, the single year regressions yielded no useful results anyway, and so the monthly averaging would clearly be worth doing.

A more challenging but potentially effective approach to mitigating the random variation problem would be to run regressions on the global nitrogen dioxide data rather than individual countries. Because the global data is area-averaged over far greater satellite coverage, the random variation is lower. The challenge here would be forming a set of global policy indices that would yield meaningful results. One approach would be to sum the indices for each country at each time slot weighted by that country's average proportional contribution to global nitrogen dioxide presence.

This could be done using NASA EarthData's GIOVANNI service[8], finding every country's average nitrogen dioxide concentration in, say, 2015 through to 2019, and weighting the sum of policy indices accordingly. This would result in a usable set of indices, but would be very time consuming.This could partially be mitigated by omitting small countries that would not be expected to substantially contribute to the global nitrogen dioxide profile.

6.5.1.2. Mitigating high index correlation

The problem of high index correlation may be a more difficult issue to address, as policy reactions during the COVID-19 pandemic were generally scaled with a sense of risk severity of the virus to that country, and so it is expected that policy indices generally rise and fall in some degree of alignment.

However, this could have partially been mitigated by first searching through a handful of candidate countries, finding their index correlations, and choosing those with the lowest index correlations as those to analyse.

6.5.1.3. Confounding variables

Another improvement would be to include energy production data as variables in the multivariate regressions. This is because they are likely to be confounding variables, considering that COVID-19 containment policies target the movement and contact of people, which can be expected to strongly influence transport related emissions. This leaves the other major contributors to nitrogen dioxide emissions unaccounted for, that is energy production from oil, coal and gas[14]. Accounting for changes in energy production from these sources will likely help isolate the influence of containment policies, as well as potentially explain anomalies like the higher-than-trend nitrogen dioxide concentrations seen in the second half of 2021.

6.5.2. Alternative methods

One related piece of future work that could hold significant social and academic value is a similar multivariate study of policy indices and their relation to tropospheric nitrogen dioxide levels, but using a newly created set of policy indices that quantify more deliberate policies in addressing air pollution.

Since anthropogenic impact on nitrogen dioxide levels is more pronounced in cities[16], and that many pollution mitigating policies are implemented on a city level[17][18], it may be best to create the index dataset for cities rather than countries. This would also allow for ground-level nitrogen dioxide monitoring, for more consistent coverage.

Policy indices could include:

- Implementation of congestion charges
- Public transport prioritisation
- Pedestrianisation
- Bike lane coverage

And variables to include to eliminate as confounding variables would include:

- Population
- Population density
- GDP per capita

Using a different set of policy indices would mean that the time period to analyse could be far greater, meaning that the data could be averaged but over months or even years depending on the scope of the index database, greatly mitigating the random variation problem.

Additionally, there is great diversity in policy throughout the world, and even within cities of the same country, so unlike in the case of COVID-19 containment policies which all generally track with a country's COVID-19 risk, it is likely that these policy indices would correlate far less, enhancing the performance of models.

Such a study would be of far greater scope than this project, with the development of the OxCGRT database taking the dedicated work of **x academics**, and so this study would similarly be labour-intensive, but could yield invaluable insights as to the efficacy of policies to address air pollution.

7. Conclusions

This project found through multivariate regressions, with COVID-19 containment policy indices as the independent variables, that in the US, workplace closure was by far the best predictor of de-seasonalised tropospheric nitrogen dioxide concentrations (coefficient = -0.9091, P-value < 0.001) during the COVID-19 pandemic (01/2020 to 12/2022), whereas in India, confinement orders were the best predictor (coefficient = -0.7684, P-value < 0.001).

Single independent variable linear regressions for these indices yielded the following results:

US: coefficient = -0.1504, P-value < 0.001

India: coefficient = -0.1092, P-value = 0.001

Because of how the indices were normalised between 0 and 1, these suggest an associated 15% and 11% drops in nitrogen dioxide concentrations respectively associated with the maximum implementation of these policy areas in each country.

However, high random variation in the nitrogen dioxide profiles and high correlation between policy indices in the analyses of both countries seriously limited the reliability of the results.

Also, there appears to be a confounding variable issue, with higher-than-decade-average nitrogen dioxide concentrations in the second half of 2021 in both the global trends and those of both countries studies, that could not be explained by the analysed indices.

Accordingly, recommendations have been made to mitigate these issues in future work:

- Data should be averaged over months rather than weeks
- An analysis of the global data, which has lower random variation, should be performed, using a set of indices summed over countries/states weighted by average proportional nitrogen dioxide emissions or concentrations
- When performing individual country analyses, candidate countries should first have their index correlations calculated, and those with the lowest correlations between indices should be prioritised
- Where the data is available, oil, coal and gas energy production data should be included in multivariate regressions, to eliminate these as confounding variables

This project was partially successful in the aim of assessing whether quantitative measures of COVID-19 containment policy could explain and predict changes in the nitrogen dioxide records during the COVID-19 pandemic, in that the fits of the two indices discussed generally tracked with the de-seasonalised profiles.

The further goal of assessing the relative impact of each containment policy area on the nitrogen dioxide records was not so successful, with the models being unsuccessful in finding statistically significant relations between the nitrogen dioxide profiles and the multiple indices expected to have an impact.

However, this project provides valuable information in assessing the viability and challenges of using quantitative measures of policy and satellite-collected nitrogen dioxide concentrations to gain insights in relation to policy to tackle pollution, being a valuable example from which future studies can guide their development.

8. Statement of personal contributions

Due to the work of myself (219021422) and my partner diverging from an early stage, all work described in this report was undertaken by myself.

9. References

1. H. Carvalho, *Air pollution-related deaths in Europe – time for action*, Journal of Global Health, 2019

2. H. Moshammer et al, *Nitrogen-Dioxide Remains a Valid Air Quality Indicator*, Int. J. Environ. Res. Public Health, 2020

3. M. Pathak, V.K. Patel, J. Kuttippurath, *Spatial heterogeneity in global atmospheric CO during the COVID–19 lockdown: Implications for global and regional air quality policies*, Environmental Pollution Volume 335, 2023

4. M.J. Cooper et al, *Global fine-scale changes in ambient NO2 during COVID-19 lockdowns*, Nature, Volume 601, pages 380-387, 2022

5. V. Fioletov et al, *Quantifying urban, industrial, and background changes in NO2 during the COVID-19 lockdown period based on TROPOMI satellite observations*, Atmospheric Chemistry and Physics, Volume 22 issue 6, pages 4201-4236, 2022

6. G.D. Jacobsen, K.H. Jacobsen, *Statewide COVID‐19 Stay‐at‐Home Orders and Population Mobility in the United States*, World Medical Health Policy, Volume 12, pages 347-356, Dec 2020

7. 'OMI Team', *Ozone Monitoring Instrument (OMI) Data User's Guide*, 2012

8. *NASA EarthData GIOVANNI*, https://giovanni.gsfc.nasa.gov/giovanni/, accessed January to March 2024

9. T. Hale et al, *Variation in government responses to COVID-19*, Blavatnik School of Government WP-2020/032, Version 15, 2023

10. T. Phillips et al, *covid-policy-dataset*, https://github.com/OxCGRT/covid-policy-dataset, accessed January to April 2024

11. United States Census Bureau, *State Population Totals and Components of Change, 2020-2023*, https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html, accessed February 2024

12. Unique Identification Authority of India, *State=wise Aadhaar Saturation*, https://uidai.gov.in/images/state-wise-aadhaar-saturation.pdf, accessed March 2024

13. U.S. Energy Information Administration, *US Energy Database*, https://www.eia.gov/opendata/browser/electricity/rto/daily-fuel-type-data?frequency=daily&data=value;&facets=fueltype;&fueltype=COL;NG;OIL;&start=2019-01-01&end=2022-12-31&sortColumn=period;&sortDirection=desc;, accessed March 2024

14.D.L. Goldberg et al, *TROPOMI NO2 in the United States: A Detailed Look at the Annual Averages, Weekly Cycles, Effects of Temperature, and Correlation With Surface NO2 Concentrations*, Earth's Future, Volume 9, 2021

15. K.L. Brubaker et al, *Impact of Aircraft Emissions on Air Quality in the Vicinity of Airports. Volume 4. Nitrogen Dioxide and Hydrocarbons*, Defense Technical Information Center, 1984

16. D.A. Potts et al, *Diagnosing air quality changes in the UK during the COVID-19 lockdown using TROPOMI and GEOS-Chem*, Environmental Research Letters, Volume 16, Number 5, 2021

17. Greater London Authority, *Clearing the air: The Mayor's Air Quality Strategy*, 2010

18. Mairie de Paris, *Paris Resiliency Strategy*, 2018