SCHOOL OF ECONOMICS AND MANAGEMENT

A blessing in disguise? – The effect of COVID-19 restrictions on air pollution in India

Author: Jonatan Pupp

Department: Economics

Supervisor: Petra Thiemann

Essay seminar period: August 24th – August 26th, 2022

Table of contents

Acknowledgements

I would like to extend my gratitude to my supervisor Petra Thiemann for her efforts to support me during the writing process of this master's thesis.

Abstract

The outbreak of COVID-19 plunged the world economy into a pandemic with severe consequences. However, the pandemic may very well have been a blessing in disguise for the mitigation of another killer: ambient air pollution. This study investigates how the COVID-19 lockdown restrictions in India affected the ambient air pollution concentrations in a sample of Indian cities. Based on air pollution data from the World Air Quality Index, and within the framework of a sharp non-parametric regression discontinuity design, a treatment effect of the lockdown restrictions on the concentrations of $PM_{2.5}$, PM_{10}, and NO_2 is estimated through local linear regressions. The results imply that the lockdown restrictions did not have a direct significant impact on air pollution levels in the full city sample. However, when accounting for heterogeneity in subsamples based on the presence of polluting industries and power stations, significant, negative reductions in $PM_{2.5}$ concentrations of -29.40 and -11.97 µg/m^3 are observed. The results from the local linear regressions on the full city sample are robust to changes in covariates at the cut-off where lockdown is initiated. They are not robust to a change of the interval around the cut-off and the use of a linear regression model, possibly due to the presence of a low number of mass points in the forcing variable.

Key words: air pollution, COVID-19, lockdown, local linear regression, non-parametric regression discontinuity design

Introduction

From the discovery of COVID-19 in December 2019 in Wuhan, China, up until August 15th, 2022, more than 587 million confirmed cases have been reported from the pandemic worldwide and more than 6.4 million deaths (Centres for Disease Control and Prevention, 2021; WHO, 2022). The disease has prompted urgent responses from governments to contain the disease, including measures such as stay-at-home restrictions, school – and workplace closures, cancelled public events, and public information campaigns (Ritchie et al., 2020). The impact on the global economy from the virus has been severe in its negative effect on sector gross domestic product (GDP) in various industries, its tole on household incomes, and the potential long-run downturns in asset returns based on historical evidence (World Development Report, 2022; Dua, Mahajan, Oyer & Ramaswamy, 2020; Jordà, Singh & Taylor, 2020) In contrast to these negative issues stemming from the pandemic restrictions, there are studies which suggest that they have proven to be beneficial for another health concern: ambient air pollution (Dang & Trinh, 2020; Zhao, Cheng & Jian, 2021; Liu, Wang & Zheng, 2021).

Ambient, or outdoor, air pollution is a phenomenon which claimed approximately 4.2 million people's lives across the globe in 2016, based on estimates from the World Health Organization (WHO). According to the WHO, in 2016 the most common diseases related to air pollution deaths were cardiovascular or heart-related disease and strike (58 per cent), followed by lung-related diseases such as chronic pulmonary disease and respiratory infections (18 per cent), and lung cancer (6 per cent) (WHO, 2021). In addition, there is a massive, disproportionate exposure to ambient air pollution in low- and middle-income countries. This is partly due to weak legislation, lower vehicle emission standards, and a stronger presence of coal power stations (UNEP, 2019). Of the 4.2 million premature deaths from ambient air pollution, 91 per cent of these take place in low – and middle–income countries, mainly those located in South-East Asia and Western Pacific regions (WHO, 2021).

One country which suffers from high ambient air pollution is India. In 2016, WHO reported an annual mean level of particulate matter of a diameter equal to or smaller than 2.5 µm ($PM_{2.5}$) in Indian cities, weighted by population, at 68.76 micrograms per cubic meter ($\mu g/m^3$). This was the sixth poorest global ambient air quality level measured by particulate matter, the air pollutant which affects most people, and a common one used to indicate air pollution (WHO, 2021).

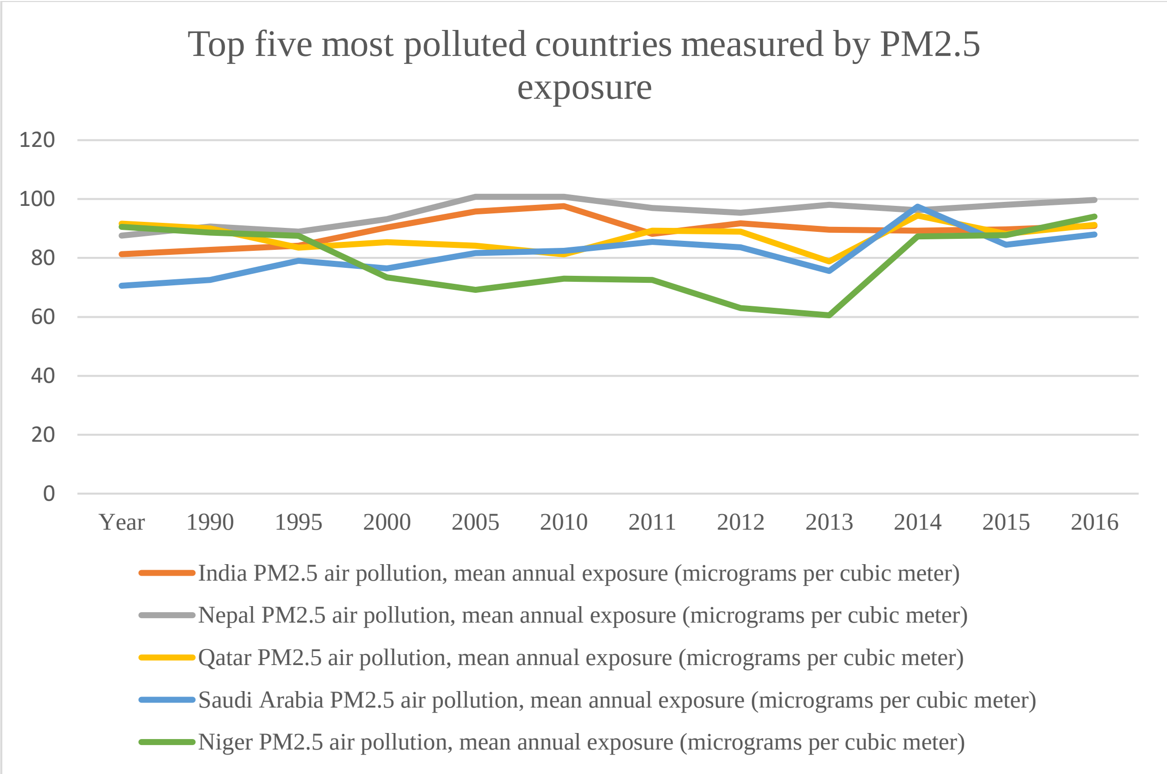

Figure 1.1. In 2017, India was ranked as the country with the world's fourth strongest total mean annual exposure of $PM_{2.5}$ µg/m^3. Source: Our World in Data (2017).

Furthermore, by inspecting *Table 1.1* one can observe that India's ambient air pollution levels strongly exceed the WHO guidelines (WHO 2021; WHO 2021). In contrast to these guidelines, India is placed far above their thresholds, with regard to $PM_{2.5}$, a common indicator of air pollution. Its annual mean pollution levels in urban areas exceed the recommended guidelines with more than six times as much in 2016, as presented in *Table 1.2*.

Table 1.1. 2005 WHO Air Quality Guidelines. Source: WHO (2021).

Air pollutant specie	**Time period (means)**	**2005 WHO Air Quality Guidelines**
$PM_{2.5}$	Annual concentration	10
	Daily 24-hour concentration	25
PM_{10}	Annual concentration	20
	Daily 24-hour concentration	50
NO_2	Annual concentration	40
	Daily 24-hour concentration	-

Table 1.2. India's annual mean level, $PM_{2.5}$, measured on urban level and in total ($\mu g/m^3$). Source: WHO (2021).

Year	Residence area type	Annual mean level ($PM_{2.5}$)
2016	Total	68.76
2016	Urban	78.16
2015	Total	67.22
2015	Urban	75.92
2014	Total	71.7
2014	Urban	81.07
2013	Total	65.93
2013	Urban	74.19
2012	Total	62.68
2012	Urban	70.18
2011	Total	58.26
2011	Urban	65.11
2010	Total	60.51
2010	Urban	68.06

In late January 2020, the first positive case of COVID-19 was reported in Kerala, India. The subsequent spread of the virus generated a response from prime minister Narendra Modi who announced a national lockdown for 21 days, between March 25th and April 15th, 2020, which consisted of strict social distancing for Indian citizens. This lockdown, with the objective to curb the national spread of COVID-19, was executed with restricted mobility and economic activity. The guidelines for the restrictions made exceptions regarding the closing of various entities, such as hospitals and medical establishments, manufacturing units of essential commodities, production units in need of continuous process, power generation units, and transportation of essential goods (Press Information Bureau, 2020; Gautam & Hens, 2020; Soni, 2021; Ministry of Home Affairs, 2020).

On April 15th, the lockdown period was extended to May 3rd, with conditional relaxations allowed from April 20th and onwards for certain activities such as all goods traffic, mining activity, coal production, construction activities, and people's movement for procuring essential goods. This re-opening of the economy came with a condition that the concerned geographical area was designed as a non-hotspot in terms of the spread of COVID-19. If there was a significant spread of the virus within a geographical area, this area would be considered a hot spot, where the relaxation of the restrictions would not apply from April 20th (Ministry of Home Affairs, 2020; IMF, 2022).

Overall, India's poor record for ambient air pollution, and its national lockdown affecting mobility and economic activity initiated on March 25th, 2020, are factors which make it an interesting country to study regarding air pollution responses to lockdown restrictions. This relationship will therefore constitute the general topic of this thesis.

Another issue in the relationship between lockdown restrictions and air pollution concerns the heterogeneity in air pollution responses to the lockdown: do the responses vary depending on the presence of air pollution sources in cities, including coal power stations, and the presence of heavy industries such as iron ore, cement, and steel (Gurjar, Ravindra & Nagpure, 2016; Gurjar, 2021)? Studying the heterogeneity of the air pollution responses is relevant to this study given that India occupied second place in the world in 2020 with regards to its share of domestic electricity production which stems from coal power (Ministry of Coal, 2022). It was also the country with the third largest capacity for coal power, measured in megawatts (MW), in 2019 (Carbon Brief, 2020). Moreover, India was the world's second-largest cement producer in 2022, the fourth largest iron ore producer in 2022, and the second-largest crude steel producer in 2021 (IBEF, 2022; World Population Review, 2022; World Steel Association, 2022). India's status as one of the world's most polluted countries, its international ranking among polluting industries, and its national lockdown imposed on March 25th, 2020, jointly support a study with regards to heterogeneity in air pollution responses from lockdown restrictions in Indian cities.

A third reason to study the relationship between lockdown restrictions and air pollution in India is the relevance of this relationship to policymakers. The findings on this topic may provide valuable information to policymakers in their efforts to limit air pollution through the lens of environmental governance. The findings are made even more relevant given claims for reform of India's air quality legislation with regards to the Air Act (1981). The current criticism of the current policies intended to control air pollution and the demand for policy reform and the imposition of the lockdown restrictions to inhibit the spread of COVID-19 on March 25th consists of an opportunity for this study to inform Indian policymakers on how to meet these demands (Mathew & Uppal 2021; Abraham & Rosencrantz, 1986).

Based on the reasons above, the national lockdown restrictions imposed by the Indian government constitute a unique opportunity to study how, and in what variety, restrictions on mobility and economic activity affected air pollution concentration. Despite the relatively short time frame, this opportunity has been explored in the related literature to a considerable degree. India, followed by China, has been the country included in the largest share of studies on the effect of lockdown restrictions on air quality (Addas & Maghrabi, 2021).

However, despite the relatively strong coverage of India regarding the lockdown restrictions' impact on air pollution, this study is using an estimation strategy which, to this study's knowledge, has not yet been deployed in a study of the impact of lockdown restrictions on air pollution in India in the existing literature. Therefore, the purpose of this study is to use a non-parametric regression discontinuity design to study the relationship between COVID-19 restrictions in India and air pollution in a sample of Indian cities (Angrist & Pischke, 2009). This estimation strategy provides the study with necessary scientific relevance for contributing to new findings in the expanding literature on the impact of lockdown restrictions on air pollution. Consequently, based on the purpose of this study, the main research question for this study can be formulated as:

Research question: What impact did COVID-19 restrictions have on air pollution concentration in Indian cities during the national lockdown period between March 25th – April 20th, 2020?

Moreover, the research question is broken into two research hypotheses which this study will test separately using a non-parametric regression discontinuity design. The two research hypotheses are formulated as follows:

H1: The national lockdown period in India between March 25th – April 20th, 2020, had a significant negative impact on air pollution concentrations in Indian cities.

H2: The national lockdown period in India between March 25th – April 20th, 2020, had a significant heterogeneous impact on air pollution concentration in Indian cities due to city characteristics such as polluting industries and the presence of energy plants.

Note that the reason for limiting the study of the impact of lockdown on air pollution to March 25th – April 20th, 2020, in the city sample is founded on the structure of the lockdown phases. The initiation of lockdown began on March 25th, and on April 20th, the restrictions received conditional relaxations. Including a longer time period may create omitted variable bias in the results if new potential sources of air pollution from the relaxed restrictions are not controlled for (Angrist & Pischke, 2009). This is the reason for restricting the treatment period to March 25th – April 20th, to obtain as similar observations near the threshold as possible.

A common problem with testing these hypotheses is endogeneity in the relationship between air pollution and lockdown restrictions. This endogeneity may be present in underlying traits affecting the air pollution outcomes and the lockdown restrictions. For example, the lockdown restrictions and air pollution may have been affected simultaneously by some factors. Such traits may be unobserved and could contribute to endogeneity regarding the relationship between lockdown restrictions and air pollution (Dang & Trinh, 2020).

The threat of endogeneity is a key reason why this study deploys a non-parametric estimation strategy. The non-parametric regression discontinuity design in this study consists of comparing the air pollution in a sample of 22 Indian cities within a limited neighbourhood of 26 days before and 26 days after a certain cut-off, March 25th, 2020. The reason for choosing 26 days before and after the cut-off is founded in that the lockdown period considered in this study lasted 26 days, from March 25th – April 20th, 2020, which means a symmetrical period before treatment should be consisting of 26 days. The cut-off is set on the first day of the lockdown restrictions, which function as a treatment for the cities. As the treatment is applied, a discontinuous "jump" in the air pollution outcomes explained by the time variable would be an indication of a treatment effect of the lockdown restrictions on air pollution (Angrist & Pischke, 2009). The simplest way to estimate the treatment effect is then to calculate the difference in the values of the regression function on each side of the cut-off (Lee & Lemieux, 2010; Angrist & Pischke, 2009; Calonico, Cattaneo & Titiunik, 2014). The design deals with the endogeneity problem through the assumption that around the cut-off, i.e. the lockdown date, March 25th, 2020, the treatment assignment is randomized between the observations, similar to a randomized experiment. This requires that the observations are unable to manipulate and affect the variable assigning treatment (which in our case is the time variable) at the cut-off (March 25th). If the treatment assignment is randomized between the observations, then the unobserved traits which affect lockdown restrictions and air pollution simultaneously in each city should be kept constant around the cut-off (Lee & Lemieux, 2020).

Moreover, the city-level data in the study for air pollution species and meteorological control variables were imported from the World Air Quality Index Project (2022). Additional data for variables related to heterogeneity in the air pollution outcomes to lockdown restrictions such as industry structure and the presence of energy plants were imported from the Bureau of

Mines, the Government of India, and the Energy Map of India, respectively (World Air Quality Index Project, 2022; Bureau of Mines, 2021; 2022; Energy Map of India, 2022).

For the main results obtained from the non-parametric regression discontinuity design, the testing of H1 suggests that there is not a significant decrease in air pollution due to the lockdown restrictions, based on the interpretation of the local linear regressions. There is a negative, insignificant treatment effect in the air pollutants around the cut-off point on March 25th, for $PM_{2.5}$ and PM_{10} and a positive, insignificant treatment effect for NO_2. Graphically, it is possible to observe a negative discontinuity in the regression function for $PM_{2.5}$ and a positive one for NO_2.

Furthermore, the testing of H2 yielded results suggesting significant negative air pollution changes in $PM_{2.5}$ and PM_{10} from lockdown restrictions in cities located in states with iron ore industry. In addition, cities located in states with above-median cement production across all represented states in the sample showed a significant negative treatment effect in $PM_{2.5}$. The graphical analysis for H2 suggests negative discontinuities for cities without iron ore industry in their state as well as for cities with below-median steel production. A positive discontinuity emerges for $PM_{2.5}$ in cities located in states with above-median steel production, accompanied by a small positive discontinuity for cities in states with above-median cement production. The findings are robust to a test which regresses covariates on the forcing variable. They are not robust, however, to a test for H1 which included a decreased interval around the cut-off point and the use of a linear regression model to estimate the treatment effect of lockdown restrictions on air pollution outcomes.

2. Literature review

The literature on the topic of COVID-19 lockdown restrictions and air pollution can be separated into two theoretical and empirical segments. This literature review starts by describing the former, and how it constitutes the theoretical background of this study. This is followed by an encompassing review of the empirical literature on the topic of COVID-19 lockdown restrictions and air pollution.

2.1. Theoretical literature

The theoretical roots of the relationship between air pollution and lockdown restrictions can be traced to the literature covering the microeconomic theory about market failures. Externalities are the most relevant market failure in this setting and are defined as the indirect effects which stem from when the market produces more of a certain good than what is socially optimal due to private utility maximisation. Negative externalities are one form of externalities, and they are often represented by pollution which makes them the most relevant externalities category to discuss in the context of air pollution and lockdown restrictions. The pollution may stem from the activity of a producer, who optimises his/her behaviour with regard to direct costs and profits, ignoring the indirect costs which constitute the pollution. The firm may maximise its profits at the expense of generating pollution as indirect costs for society, which means that the social costs of production will exceed the private costs of production. The polluting firm does not consider the indirect costs of its production on society, since it does not bear these indirect costs and hence does not account for these when making their production decision based on the profits and costs of production. The result is pollution from the firm that inflicts a difference between the social costs and the firm's private costs, where the former is larger, for example, due to larger health economic costs on a social level (Helbling, 2020).

The difference in social and private costs from the negative externality in the form of pollution leads to ineffective market outcomes on a social level. This in turn motivates governmental intervention to rectify the negative externality to maximise utility on a social level. The intervention needs to ensure that all costs are internalised by all agents in the economy. One such governmental intervention is Pigouvian taxation which stems from the work of Arthur Pigou (1920), who argued for the government to set a tax equal to the social cost imposed by the externality, i.e., pollution. This tax would make the responsible firms internalise the externality of pollution in their costs and hence lead to a market outcome where social costs equal private costs (Edenhofer, Franks & Kalkuhl, 2020; Helbling, 2020).

However, Pigouvian taxes are subject to criticism with regard to their implementation, partly because it is very difficult to calculate the social costs for externalities to which they are to be calibrated. For instance, regarding the social cost of carbon (SCC), the social cost of an additional unit of carbon dioxide in the long-term, the IPCC does not include this cost measure in their report partly due to the large uncertainty from the range in its estimated values (EPA, 2016; Edenhofer et al., 2020). In fact, estimates of the SSC lie within the interval of $7 to $100 based on estimates from integrated assessment models (Waldhoff et al., 2011), dynamic integrated assessment models (DICE) (Nordhaus, 2017), and additional panel models and regressions (Kalkuhl & Wenz, 2020; Ricke et al., 2018). This contributes to the difficulty of successfully implementing Pigouvian taxes. Additional problems with Pigouvian taxes include distributional issues with the tax burden among heterogeneous households in the economy, political motives which may hamper commitments, and disharmonious pricing of emissions due to fragmented responsibilities among authorities (Edenhofer et al., 2020).

One method in the theoretical economic literature used to address the issue of Pigouvian taxes to quantify the social cost of the externality is the bargaining among agents to internalise the externalities. This was a suggested way of handling the externalities problem suggested by Ronald Coase (1960) and is laid out through the Coase theorem which states that internalisation of the negative externalities through bargaining in a market setting, is possible given zero transaction costs, complete information among agents, and well-defined property rights (Zamagni, 2020 pp. 227-269; Helbling, 2020).

Unfortunately, there are challenges with the Coase theorem as the establishment of property rights gets increasingly challenging when the externality takes the form of a public good. Public goods are goods which are non-excludable and non-rival, and a good example of such a good is clean air. Property rights for clean air are inherently not well-defined and hence this means the bargaining between agents to achieve a socially optimal amount of clean air is not feasible under the Coase theorem. This leads to an outcome with more air pollution than what would have been socially optimal since clean air has the properties of a public good (Helbling, 2020).

2.2. Theoretical background

Based on the above concepts in the theoretical literature, this may establish a theoretical background for its purpose to study how lockdown restrictions for slowing COVID-19 infection rates impacted air pollution across a sample of Indian cities in 2020. Although these restrictions were not intended to mitigate air pollution levels, by accounting for the disparity between social and private costs, they still consist of a governmental intervention which may have affected negative externalities in this setting, i.e. the air pollution species. Based on this theoretical background which puts lockdown restrictions as a governmental intervention to a negative externality, the lockdown restrictions and their effect on air pollution could prove informative for different types of interventions which are intended to address air pollution.

Examples of such interventions include the Indian Air (Prevention and Control of Pollution) Act, 1981. It was the first step in combatting air pollution in India from a legislative standpoint. The act itself was equipped with the main objective to preserve the air quality and control air pollution, which gave the Central and State Boards an additional mandate to regulate air pollution (Mathew & Uppal, 2021; The Air Act, 1981). The Air Act has managed to achieve the imposition of a framework to regulate air pollution. It allows monitoring of air pollution, sets standards for emitters, as well as enforces legal measures such as empowering the State Governments with the authority to denote areas as "air pollution control areas". These areas may be subject to certain standards for pollutants as well as regulating the use of different fuels. Finally, the Air Act was responsible for the development of India's National Ambient Air Quality Standards for parameters including particulate matter ($PM_{2.5}$, PM_{10}) sulphur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and carbon monoxide (CO) (Mathew & Uppal, 2021; National Ambient Air Quality Standards, 2009).

Today the Air Act is subject to criticism partly due to the lack of updates to its content, and its ineffective reliance on expensive prosecution with low conviction rates. Moreover, the lack of incorporation of developments in science, and the inability to account for certain health risks constitute additional flaws to this legislation. There are developments suggesting that future legislation is on the way, which would create a joint act for environmental management to replace the Air Act of 1981 (Mathew & Uppal, 2021; Abraham & Rosencrantz, 1986). In this light, there is an opportunity for this study to contribute to the shape and format of future legislation through its investigation of how lockdown measures, consisting of restrictions on mobility and economic activity, may impact air pollution. Its findings could present policymakers with valuable insight regarding the response of air pollutants to measures which restrict mobility and economic activity, which could be of value for the creation of legal measures, such as the designation of air pollution control areas.

Another example of a policy where the above theoretical background may also be informative is the National Air Monitoring Programme (NAMP), which was executed by the Central Pollution Control Board (CPCB) in 1984 (Centre for Science and Environment, 2018). The motivation behind this program is to observe the status of ambient air quality to identify violations of the set standards and to gain new insights regarding preventive work as well as the process in which air pollutants are dispersed naturally in the environment. There is weekly monitoring of particulate matter, both $PM_{2.5}$ and PM_{10}, oxides of nitrogen including NO_2, and sulphur dioxide SO_2 through approximately 800 stations in 344 cities (CPCB 2021). Since this study is monitoring the responses of several of these air pollutants ($PM_{2.5}$, PM_{10}, and NO_2) to the lockdown restrictions and the heterogeneity in their responses, its findings would be of value both for the preventive work of NAMP as well as for the understanding for the dispersal of the air pollutants in the environment. This would, in turn, make the results of this work useful for the air quality management performed by the CPCB in India (CPCB, 2022).

Based on these possibilities to enhance policies for the prevention and control of air pollution, the theoretical background of this study rests on the theory of externalities and the public goods problem. Viewing air pollution in India as a negative externality which is caused due to clean air being a public good provides a solid background for understanding the emergence of air pollution from a theoretical perspective. In addition, this theory also provides an understanding of how air pollution can be mitigated through governmental intervention, which in turn provides a better insight into the topic regarding the impact of lockdown restrictions on air pollution. Having established the theoretical background for the topic based on the relevant theoretical literature, the next section will explore empirical literature on the relationship between COVID-19 lockdown restrictions and air pollution.

2.3. Empirical literature

The empirical literature on the relationship between COVID-19 restrictions and air pollution is subject to a time constraint, given the relative proximity in time to the outbreak of the COVID-19 pandemic. Despite this, there have been several empirical studies estimating the impact of COVID-19 restrictions on air quality. One segment of the literature has studied the outcomes on a global basis, while another has used a country-by-country basis. In the latter segment, the results are relatively concordant: Dang & Trinh (2020) use a cross-national sample of 164 countries and daily satellite – and station data to study the impact of lockdown restrictions on air quality and find a 5.4 per cent decrease in NO_2, in addition to a 3.1-3.9 per cent decrease in $PM_{2.5}$. Venter, Aunan, Chowdhury & Lelieveld (2020) observed a 60 per cent fall in ground-level NO_2 levels, followed by a 31 per cent decrease in $PM_{2.5}$ for a sample of 34 countries which had been in lockdown for an average of 62 days. Lenzen et al. (2020) include 38 regions and 26 economic sectors in their dataset from the Global Multi-Region Input-Output (MRIO) laboratory. Their findings suggest that global emissions of $PM_{2.5}$, and additional air pollutants such as SO_2 and NO_X, decreased by 3.8 per cent and 2.9 per cent respectively. Although the results are uniform in the negative impact of lockdown restrictions on air pollution, there is a large interval in the magnitude of the impact of the restrictions on air pollution concentrations.

Moreover, the country-based literature has been largely focused on Asian countries, where India and China consist of the two most studied countries in the literature (Addas & Maghrabi, 2021). However, when it comes to India, the country of interest in this study, the usage of microeconometric methods such as the regression discontinuity design is scarce when studying this relationship. Instead, the methods deployed, such as the one of Sikarwar, Rani, and Chattopadhyay (2020), include an Inverse Distance Weighting model to generate spatial interpolation maps of NO_2, $PM_{2.5}$, and PM_{10} pollutants. The authors study Delhi during the lockdown from March 25th and claim to have found a decrease in $PM_{2.5}$, PM_{10}, and NO_2 of 93, 83, and 70 per cent respectively (Sikarwar, Rani & Chattopadhyay 2020).

Additional work from Panda, Satpathy, Das & Ramasamy (2021) provides an overview of the effect of the national lockdown on air pollution across several different Indian cities between different states. Their overall findings suggest air quality fell considerably across their sample of Indian cities for both PM_{10} and $PM_{2.5.}$ They also assert, with referral to Rathore et al. (2021) that the air quality index (AQI), which consists of eight pollutants (PM_{10}, $PM_{2.5}$, NO_2, SO_2, CO, NH_3 and PB) improved significantly due to the lockdown across various cities. The observed maximum decline, - 52 per cent, was in Bengaluru and Lucknow (Panda, Satpathy, Das & Ramasay 2021; Rathore et al. 2021).

Furthermore, Gautam et al. (2021) compare the differences in the AQI between February 17th – May 4th, 2020 and conclude that the air quality in the regions Delhi, Uttar Pradesh, and Haryana improved during the lockdown period with regards to falls in $PM_{2.5}$, PM_{10}, NO_2, NH_3, and CO. After the conditional relaxation of the restrictions was introduced, the concentration of these air pollutants rose again. The sharp decrease in particulate matter was attributed to lower vehicle activity on the roads and lower industrial activity (Gautam et al., 2021).

Additional studies have been investigating China, with varying results in terms of which air pollutants are significantly affected by the lockdown policy. Zhang, Cheng and Jian (2021) use a fixed-effects model and a regression discontinuity design and find generally positive impacts of the restrictions on air quality indicators, excluding O_3 (Zhang, Cheng & Jian, 2021). Furthermore, additional work has made use of the difference-in-differences model to construct a counterfactual for the trend in air quality emissions, without the effects of restrictions due to the pandemic (Ming, Zhou, Ai, Bi & Zhong, 2020; He, Pan & Tanaka, 2020). Ming et al. (2020) discover that $PM_{2.5}$ decreased by -7 $\mu g/m^3$. He et al. (2020) estimate that cities which carried out lockdown policies experienced a decrease of -14.07 $\mu g/m^3$. These findings also suggest that COVID-19 restrictions and air pollutants maintain a negative relationship.

Further work has been conducted on the relationship between pandemic restrictions such as a national lockdown and air quality in Vietnam by Dang and Trinh (2020). They use a regression discontinuity design to obtain their results in both studies. Their results imply that cross-nationally, NO_2 and $PM_{2.5}$ concentrations both significantly decrease because of COVID-19 lockdowns. Similar results appear for the study of the impact on NO_2 levels in Vietnam. They estimate a reduction in NO_2 by 24 per cent from the lockdown over a bandwidth which stretches from two weeks before and after lockdown. Increasing the bandwidth to four and eight weeks before and after lockdown leads to a lower reduction of 18 per cent (Dang & Trinh, 2020).

In general, the findings in the empirical literature suggest that the impact of restrictions intended to mitigate the spread of COVID-19 was negative for air pollution levels. There are some contrasting findings for countries that did not implement a strict national lockdown policy, such as Taiwan, where increasing air pollution during the pandemic was reported. Chang, Meyerhoefer, and Yang (2020) argue that this result may be due to the decision of many people to shift their preferences to use public transport to personal means of transportation (Chang, Meyerhoefer, and Yang, 2020). On the contrary, Lee & Finerman (2021) present evidence from South Korea suggesting that in the absence of a national lockdown, air pollution fell due to lower commuting flows during the COVID-19 outbreak (Lee & Finerman, 2021).

Several studies also suggest that the impact of lockdowns on air pollution may be heterogenous according to certain country characteristics. Examples of these include work on measuring heterogeneity from the impact of lockdown on a global scale, such as another work of Dang & Trinh (2020). The authors assemble a cross-country database on air quality before and after the national lockdowns for 178 countries. Their findings here suggest that there were some heterogeneous impacts across countries with respect to their characteristics regarding shares of trade and manufacturing, initial levels of air pollution and geographical distance to the equator. In the end, the impacts of COVID-19 lockdowns remained negative on air pollution (Dang & Trinh, 2020). Liu et al. (2021) find significant heterogeneity in their difference-in-differences estimates for the lockdown effect in their global city sample. The authors discover that less developed cities (measured by GDP per capita) received a -17.3 unit decrease in $PM_{2.5}$ versus a -4 unit decrease in more developed cities. In addition, cities with larger shares of industries experienced a -11.2 unit decrease in $PM_{2.5}$ versus a -5.6 unit decrease in cities with smaller shares of industries (Liu et al., 2021).

Furthermore, studies based in a country setting have also explored the heterogeneity in air pollution changes from COVID-19 lockdowns. Kumar & Managi (2020) provide a comparison in average levels of $PM_{2.5}$ (µg/m^3) between Indian cities during different phases of lockdown. The city with the highest average level of $PM_{2.5}$, Jodhpur, saw a decrease from 76.66 µg/m^3 to 48.12 µg/m^3 when judging pre-lockdown levels from March 1st – March 24th, 2020, against the level during the first phase of lockdown measured in the paper as from March 25th – April 14th, 2020 (Kumar & Managi, 2020). Ming et al. (2020) test for heterogeneity among Chinese cities due to delayed output from industries due to COVID-19 restrictions by dividing their sample of cities into two categories: larger cities (1) based on whether they cities designated in state plans, capital cities, or municipalities; and ordinary cities (0). Their estimates suggest that there was a significantly larger improvement in $PM_{2.5}$ in the larger cities compared to the ordinary cities. The authors claim that stronger governance, as well as higher production and operation intensity in larger cities, contribute to this finding (Ming et al., 2020).

2.4. Economic motivation from the empirical literature

This study's choice to investigate the relationship between COVID-19 lockdown restrictions and air pollution is based on the economic issues tied to air pollution found in the empirical literature. More exactly, previous empirical work on the relationship between pandemic restrictions and air pollution has been vocal about their contribution to environmental governance and policies regarding air quality, in addition to health outcomes. For example, Dang & Trinh (2020) present their findings for improved air pollution from mobility restrictions as relevant for environmental regulation as well as presenting a back-of-the-envelope cost-benefit analysis of saved health care costs from the dampened air pollution from pandemic restrictions in Vietnam in 2020. Kumar & Managi (2020) suggest their findings on how lockdown restrictions have improved air quality will prove beneficial for Indian policies for hampering the impact of air pollution (Kumar & Managi, 2020). Finally, Liu et al. (2021) argue that their work may contribute to policy interventions, environmental regulations, and the literature on health benefits (Liu et al., 2021).

Overall, environmental and health outcomes appear as the foremost economic issues tied to a study of the relationship between COVID-19 restrictions and air pollution. In this study, the main economic motivation is to contribute to an environmental policy intended to manage air pollution levels. This contribution is made through the testing of hypotheses H1 and H2, which test the effect of lockdown restrictions on air pollution and the heterogeneity of the effect through economic channels such as industries, and through the activity of a segment of the energy sector. Given this study's focus on how restrictions to mobility and economic activity impacted the three air pollutants $PM_{2.5}$, PM_{10}, and NO_2, its findings will be most important for the formation of policies related to mitigating air pollution outcomes. The presentation of how the economic motivation of this study is reflected in the empirical literature ends the chapter of the literature review. The next chapter will present a thorough description of the data sets used to test H1 and H2, in addition to the selection procedure of the estimation samples from the raw data sets.

3. Data

Initially, this chapter will carefully describe the data sources and variables. In addition, motivations behind the use of the selected variables will also be presented. This is followed by a thorough explanation of the formatting of the data into panel data and the selection of the samples used in the estimation.

3.1. Air pollutants

In the research hypothesis H1, this study has set its dependent variable to be represented by three species of air pollutants, particulate matter of a diameter equal to or smaller than 2.5. and 10 μm ($PM_{2.5}$ and PM_{10}), and nitrogen dioxide (NO_2). These air pollutants are imported from the World Air Quality Index Project (2022) as daily data and formatted in terms of their medians and in accordance with the U.S. Environmental Protection Agency's (EPA) standards (World Air Quality Index Project, 2022; EPA, 2022). This study understands there is a difference in the air quality index which is used by the EPA versus India's National Air Quality Index Standard (NAQI) which was adopted by The Central Pollution Control Board in India in April 2015. NAQI differs from the EPA:s index in the sense that it gives higher weight to air pollutants such as particulate matter ($PM_{2.5}$ and PM_{10}) than the U.S. index. This is to account for the prevalence of particulate matter which is the most severe Indian air pollutant. (World Air Quality Index, 2015). However, the NAQI has also been subject to serious criticism regarding its lack of coverage in terms of monitoring stations across Indian cities (Greenpeace, 2015).

Moreover, the selection of these three air pollutants is founded on the fact that $PM_{2.5}$ constitutes an air pollutant for which India occupies one of the top spots globally regarding its concentration. Adding PM_{10} should also yield a rougher estimate of particulate matter concentration to complement the finer measure of $PM_{2.5}$. Another reason for including particulate matter is its contribution to serious health risks, such as heart and lung diseases as well as cancer ($PM_{2.5}$), and respiratory diseases such as asthma and chronic obstructive pulmonary disease (UNEP, 2019; California Air Resources Board, 2022). The strong concentration of $PM_{2.5}$ in India and the health risks from $PM_{2.5}$ and PM_{10} indicate the importance of studying how these air pollutant species react to lockdown restrictions. Furthermore, the reasons for studying NO_2 as an outcome variable partly consist of its role as a precursor to the formation of another dangerous air pollutant, ground-level ozone (O_3). NO_2

responses could therefore be relevant to investigate given that they partly determine the formation of O_3. In addition, NO_2 causes health problems related to the hearts and lungs (UNEP, 2021; Climate & Clean Air Coalition, 2022). Overall, the prevalence of these air pollutant species in India as well as their health implications makes them highly relevant to use when estimating the air pollution responses to the lockdown restrictions on March 25th, 2020.

3.2. Control variables

The dataset imported from the World Air Quality Index Project (2022) also includes daily data for two meteorological control variables: wind speed and temperature medians. These control variables are used extensively in the related literature to control for their respective influences on air pollutants. For example, Dang & Trinh (2020) use temperature alongside precipitation as control variables in their parametric regression discontinuity design (Dang & Trinh, 2020). Zhao, Cheng & Jian (2020) include both daily minimum and daily maximum temperatures as well as windy weather dummies (coded as 1 for the existence of wind on a concerned day and 0 otherwise) in their fixed-effects model and regression discontinuity design for air quality explained by provincial shut-down periods in Chinese provinces (Zhao, Cheng & Jian, 2020) Moreover, Liu, Wang & Zheng (2021) control for humidity, temperature, and wind speed in their fixed-effects OLS estimation and difference-in-differences design of air quality regressed on lockdown measures (Liu, Wang & Zheng, 2021). These control variables will be used in testing both H1 and H2.

3.3. Heterogeneity variables

For the testing of H2, and thus the heterogeneity in the air pollution outcomes to lockdown restrictions, it is important to consider the sources of the three air pollutant species representing the outcome variables in this study ($PM_{2.5}$, PM_{10}, and NO_2). These sources will constitute the foundation of the city characteristics selected to study the heterogeneity in the outcomes of air pollution to lockdown restrictions. According to the Central Pollution Control Board (CPCB) in India, sources of $PM_{2.5}$ and PM_{10} include vehicle emissions, industrial combustion plants, and residential and commercial combustion. Oxides of nitrogen, (NO_X) include the burning of fossil fuels, biomass, and fossil-fuelled power stations (CPCB, 2019). These sources imply that the datasets used for testing the heterogeneity of the impact of national lockdown restrictions on air pollution should be based on city characteristics which correspond to these air pollutant sources. These are presented below in subsections 3.3.1 and 3.3.2.

3.3.1. Industry structure

Based on the sources of the air pollutants, the first set of variables for the heterogeneity analysis consists of the industrial structure measured by the presence of iron ore, steel, and cement industry. The data for the industry structure was collected from the Indian Minerals Yearbook provided by the Bureau of Mines (Bureau of Mines, 2020). The variables are measured in annual production as well as the number of production units per year. The data for these variables will not be included in the estimation of the non-parametric regression discontinuity design but will rather be categorizing cities into different samples for the heterogeneity analysis. This will be explained in more detail in the next section.

Firstly, each city's industrial structure is important to account for in a heterogeneity analysis, not least due to its influence on the level of particulate matter ($PM_{2.5}$ and PM_{10}) at the city level. The selection of industries is based on a classification from the CPCB, where seven industries are listed as critical for the emission of suspended particulate matter (which encompasses $PM_{2.5}$ and PM_{10}) and NO_X (Gurjar, 2021; Gurjar, Ravindra & Nagpure, 2016; European Environment Agency, 2020). These critical industries are constituted by iron and steel, sugar, paper, cement, fertiliser, copper, and aluminium. These findings are supported by Wheeler (1999), stating that iron, steel, and non-metallic products represent particularly polluting industries (Wheeler, 1999). Provided that $PM_{2.5}$ represents the most important air pollutant specie in India, these seven critical industries will constitute the base sample of industries for the test for heterogeneity. Moreover, this study will focus on the iron, steel, and cement industry branches, partly because of their strong contributions to particulate matter emissions. In addition, India is the world's second-largest producer of cement in 2022, the world's second-largest producer of crude steel in 2021, and the fourth largest producer of iron ore in 2022. The sheer magnitude of these industries in India along with their particulate matter pollution makes them particularly relevant to include in the testing of heterogeneity in air pollution outcomes from lockdown (IBEF, 2022; World Population Review, 2022).

3.3.2. Coal power stations

The second set of data for the heterogeneity analysis consisted of the presence of coal power stations. The data for this variable was imported from the Energy Map of India (2021) which is provided by the NITI Aayog and Indian Space Research Organisation (ISRO) (Energy Map of India, 2021).

Coal power stations are the most important supplier of electricity to the grid in India, providing 75 per cent of total power generation in the country. They are also an important source of fine particulate matter and NO_X emissions. In fact, the share of population-weighted ambient $PM_{2.5}$ in India from the coal power stations active in 2018, and including future planned stations, is estimated to be 13 per cent (Cropper, Cui, Guttikunda & Song, 2021; Ministry of Coal, 2022). In 2021, it is reported that 20 per cent of particulate matter emissions in India are explained by coal power stations, accompanied by an explained 30 per cent share of NOx emissions (International Centre for Sustainable Carbon, 2021). Provided the important role of coal power stations in India's domestic energy production and their contribution to the air pollutant species investigated in this study, they constitute a necessary part of the energy sector to account for in the heterogeneity analysis.

3.4. Selection of estimation samples

In the study's pursuit to test hypotheses H1 and H2, the air pollution dataset and its controls were formatted as panel data using the assignment of identities for each city, air pollutant, and control. Given the nature of the non-parametric regression discontinuity design, the data was then cut down into samples which are recentred around the lockdown date, March 25th, i.e. the threshold of the model. The result is a pre-lockdown period and a post-lockdown period numbered from [-26, 26] when the treatment of lockdown restrictions has been conducted. The reason for choosing 26 days is founded in that the lockdown period selected in this study lasted from March 25th to April 20th. This implies that the data should be recentred 26 days before the lockdown, i.e. February 28th – March 24th, and 26 days after the lockdown, i.e. March 26th – April 20th, to yield symmetrical windows around the cut-off point of March 25th.

After the data was recentered around the cut-off, the forcing variable of the non-parametric regression discontinuity design could be constructed as the recentred values ranging from [-26, 26]. In this closed interval, the threshold is represented at 0. Once this procedure had been conducted, the recentred air pollution data was cut down into different intervals, or 'bins', based on the forcing variable, i.e., the recentred variable around the interval [-26, 26]. These bins contain the sample means of the air pollution variables within each interval and will be explained in more detail in the next chapter. This data enabled the graphical analysis from the non-parametric regression discontinuity design.

In addition to the data for the graphical analysis, the panel datasets intended for the local linear regressions were set as the same ones used in the graphical analysis. The exceptions are that no bins are generated, and they also include data for the control variables for wind speed and temperature. Here, these air pollutants and controls, together with the dummy variable assigning treatment, constitute the data for the local linear regressions which will be generated to estimate the impact of the lockdown on each air pollutant.

The steps above were the necessary segments for organizing the data to test H1. Moreover, to test H2, additional data to the one above had to be collected for relevant variables concerning heterogeneity. It is important to explain that the data for these variables were neither included directly in the graphical analysis nor in the local linear regressions but were used to categorize cities into different samples for testing H2. For example, the city sample was divided into two samples for each industry to enable comparisons between samples characterized by a stronger presence of industry versus samples less characterized by industry. For the iron ore industry, this division into two city samples was based on whether there was actual iron industry production located in the state of each city or not. If the iron ore industry was present in the state, the concerned city within that state was assigned to the iron ore industry sample, and if there was no iron ore industry present in the state, the concerned city was assigned to a non-iron industry sample.

The categorization for the steel and cement industries used a different rule for assigning the cities to different samples. If the steel – and cement production at the state level was above the median of the state-level steel – and cement production adhering to all the states represented by the cities in the full sample, then cities in such states with above-median production would be assigned to the above-median steel – and cement production sample. If the opposite was true, i.e., cities were located in states with below-median steel – and cement production at the state level based on all the states represented by the cities in the full sample, then the concerned cities would be located in a below-median steel – and cement production sample.

To clarify the reasons for these categorizations it is vital to understand that iron ore production was not present in some of the states represented in the city sample, while steel and cement production were present in all states represented in the city sample (with the exclusion of Delhi and Punjab).[1] This meant that the categorization based on state-level iron ore production simply could rely on dividing the cities into non-iron ore and iron – ore sample. Since steel and cement

[1] The cities in the sample which are located in the states of Delhi and Punjab are New Delhi and Chandigarh. There was no cement production recorded in either Delhi or Punjab, which in turn means that New Delhi and Chandigarh will not be included in the categorization of city samples based on the presence of cement production at the state level.

production was found in almost all states, the categorizations of the city samples according to cement and steel production used a different rule, in this case, the median for state-level production.

Proceeding to the division into coal power stations, the cities were also categorized into two samples. This categorization was based on the rule of whether a coal power station was located in the same city district or in a neighbouring city district or not. More exactly, the categorization of the two city samples was carried out by using a dummy variable, which was labelled 1 if a coal power station was indeed located in the city district or in a neighbouring city district, and 0 otherwise. Practically, this categorization was made using the containing tracking features for specific coal power stations from the Energy Map of India (Energy Map of India, 2022).

Through the selections of these estimation samples, the methodology used in this study to test H1 and H2 is provided with the data necessary for its analysis. The estimation samples will be useful for graphical analysis which provides a compelling illustration of the impact of lockdown restrictions on air pollution. They will also be important for the local linear regressions generated in accordance with the non-parametric regression discontinuity design used in this study. Both these two parts of the methodology will now be introduced and explained in detail.

4. Methodology

In this chapter, the non-parametric regression discontinuity design will be presented in detail with regard to its idea along with the methods used to estimate it. Firstly, the identifying assumption, intuition, and motivation behind the use of this design are presented. This is followed by a detailed description of the graphical analysis used in this design, alongside various threats against the non-parametric regression discontinuity design. The last part of the chapter introduces the main estimation equations used to test the research hypotheses H1 and H2 and describes them closely.

4.1. The non-parametric regression discontinuity design

This study employs a non-parametric regression discontinuity design as the main model to test hypotheses H1 and H2, i.e., whether the lockdown restrictions had a significant negative impact on air pollution in the sample of Indian cities (H1), and if the impact was significantly heterogeneous depending on city characteristics (H2).

The regression discontinuity design was implemented originally by Thistlethwaite and Campbell (1960) in an article on educational psychology, which studied the effect of treatment from public recognition on near winners in a national competition for scholarships. Some near winners received more public recognition than others which in turn influenced the likelihood of receiving scholarships from other actors (Thistlethwaithe & Campbell, 1960).

The regression discontinuity design may be divided into two categories: sharp and fuzzy. This study relies on a sharp design. The idea behind this design is that treatment is assigned to the observations based on whether they are above or below a certain cut-off. The observations below the cut-off act as a control group and the ones above the cut-off act as a treatment group. The assignment of their position around the cut-off is based on their scores from an assignment or forcing variable. The conditional probability of receiving treatment in sharp regression discontinuity design changes discontinuously at the cut-off, where the conditional probability goes from zero to one. Any "jumps" in the relationship between the outcome variable and the forcing variable at the cut-off are taken as evidence of a treatment effect. The treatment effect is captured by measuring the average outcomes of the treated observations above the cut-off against the non-treated observations below the cut-off (Angrist & Pischke, 2009; Cattaneo, Idrobo & Titiunik, 2019).

Moreover, assuming that the observations were unable to manipulate their placement around the cut-off, this in turn should ensure that the variation in receiving treatment or not receiving treatment around the cut-off is randomized. This will in turn mitigate endogeneity by ensuring that unobservable and observable traits around the cut-off are kept constant. The only difference on both sides of the cut-off is that some observations are treated while others are not (Angrist & Pischke, 2009; Cattaneo, Idrobo & Titiunik, 2019). When measured towards experiments, regression discontinuity designs have indeed proven in several cases to be robust in generating results similar to those of experiments (Cook & Wong, 2008).

This paper is relying on the sharp design, where the treatment assignment is made by a forcing variable representing the days recentred around the cut-off, which in turn represents the date of lockdown, March 25th, 2020. The forcing variable assigns the treatment of lockdown restrictions on March 25th, 2020, and this treatment is considered to remain in place until April 20th, 2020. During the period before lockdown, February 28th – March 24th, no treatment is assigned, which is reasonable since the lockdown restrictions had not been imposed by then.

In the context of this study, evidence of a treatment effect is achieved by comparing the average outcomes, i.e., air pollutants $PM_{2.5}$, PM_{10}, and NO_2 in the city sample, on both sides of the cut-off constituted by the lockdown restrictions on March 25th. Just as stated above, assuming that the cities are unable to affect their placement around the cut-off, the treatment variation should be randomized, which means that hard-measured, unobserved factors influencing air pollution and lockdown restrictions simultaneously in a city can be assumed to be constant around the cut-off, i.e. before and after lockdown. The only difference will be that some cities are treated with lockdown while others are not (Lee & Lemieux 2010; Angrist & Pischke, 2009).

Furthermore, this study uses a non-parametric sharp regression discontinuity design in contrast to a parametric design. The parametric regression discontinuity design requires a correct specification of the forcing variable function, which decides when treatment is assigned. If the forcing variable function is not modelled correctly, any discontinuity in the outcomes may be due to non-linearity in the forcing variable function. To avoid this issue with the misspecification of the forcing variable in the parametric regression discontinuity design, it is possible to use a non-parametric regression discontinuity design. This design focuses on the observations in a small neighbourhood around the cut-off, and it has the benefit that it is not necessary to model the forcing variable function correctly to get the treatment effect of lockdown restrictions on air pollution outcomes (Angrist & Pischke, 2009).

A natural transition from the presentation of the non-parametric regression discontinuity design is to delve into the graphical analysis and regression methods of this design. Thus, the next upcoming section consists of the important graphical analysis, a natural choice for any study using regression discontinuity designs (Lee & Lemieux, 2010). It is followed by the formulation of the estimation equations for the regressions used in the study's non-parametric regression discontinuity design.

4.2. Graphical analysis

The non-parametric regression discontinuity design in this paper to test H1 and H2 is partly based on graphical analysis. To conduct this analysis, several bins (or intervals) were constructed based on the forcing variable, i.e., time formatted as a recentred variable. These bins are then augmented with the average of the outcome variable, air pollution, within each of them, and presented as a scatterplot, with the average outcomes of the air pollution in the city sample explained by the forcing variable. Furthermore, this study also uses two different bin sizes to display the results graphically: one small bin size and one large bin size. The small bin size is set as an interval of three days in the neighbourhood around the threshold, while the large bin size is set to an interval of seven days (Lee & Lemieux, 2010).

In these graphs, outcomes are explained by the forcing variable, without any control variables or covariates. It will however be important for assessing the credibility of the nonparametric regression discontinuity design to conduct graphical analysis where the covariates are explained by the forcing variable. If there are discontinuities in the covariates at the cut-off, then the identification of this nonparametric regression discontinuity design may be invalid (Imbens & Lemieux, 2008).

Another potential threat to the non-parametric RD design is the low number of observations around the selected neighbourhood around the threshold. This could mean that the bandwidth of the neighbourhood needs to be extended to improve efficiency. However, this can lead to more biased estimates (Lee & Lemieux, 2010). The non-parametric model in this study is subject to the time constraint imposed by the lockdown period by the Indian government from March 25th – April 20th. Extending the dataset to a longer period may introduce bias as lockdown restrictions are conditionally relaxed from April 20th. Based on this trade-off between bias and efficiency the study will not extend the bandwidth.

A final note on the graphical analysis is that the bins will be augmented with a local polynomial regression line. Note that this regression is using a local polynomial regression line of order 4, which will differ from the polynomial order in the local linear regressions outlined in the next section. This local polynomial regression line of a higher order provides a good fit to the function of the air pollution outcomes regressed on the forcing variable. It is also more important to have a local linear regression of order 1 when generating the actual local linear regression estimates. The reason behind the latter will now be explained, as the local linear regressions will be introduced in the next section.

4.3. Local linear regression model

Moreover, in addition to the graphical analysis, local linear regressions are also performed for this study's non-parametric regression discontinuity design. The local linear regressions estimate the regression functions above and below the cut-off point by using means of weighted polynomial regressions of order 1. They use a triangular kernel function which generates the weights for each observation based on its distance from the cut-off point. The intuition behind the local linear regression is to concentrate on units within a certain bandwidth on either side of the cut-off (lockdown), discarding the units which fall outside this bandwidth. Then linear regressions are estimated of polynomial order 1 for the remaining units, and an average treatment effect is calculated for the average air pollution outcomes on either side of the cut-off. This treatment effect from the local linear regressions states how much the various air pollutants are affected by the national lockdown in India from March 25th – April 20th, 2020 (Calonico, Cattaneo & Titiunik, 2014; Gelman & Imbens, 2019; Imbens & Lemieux, 2008; Cattaneo, Titiunik & Vazquez-Bare, 2020).

There are several reasons to motivate the usage of local linear regressions in the non-parametric regression discontinuity design. Firstly, Gelman & Imbens (2019) draw out three key arguments for why local linear and quadratic polynomial estimators should be preferred to higher-order polynomial estimators should not be used in regression discontinuity design. Firstly, higher-order polynomial estimators are susceptible to assigning excessive weights to observations with extreme values in the forcing variable and far away from the cut-off. In contrast, local linear estimators are more likely to assign higher weights to observations closer to the cut-off, while assigning zero weight for observations outside the bandwidth. This excessive assignment of weights will lead to estimates which in turn may lead to noisy estimates (Gelman & Imbens, 2019).

Secondly, the estimates of the discontinuity are more sensitive to higher-order polynomial regressions, compared to local linear and local quadratic regression estimates. This is illustrated by a larger difference between the discontinuity estimates from different higher-order polynomials in relation to the difference between local linear and local quadratic estimates, whose difference is much less pronounced (Gelman & Imbens, 2019).

Thirdly, the confidence intervals for the higher-order polynomials have a lower nominal coverage compared to local linear and local quadratic ones. Gelman & Imbens (2019) conduct a test of repeatedly picking a pseudo, or fake, threshold in a pretended regression discontinuity design and then estimating the average treatment effect. Then, for the estimates for the treatment effects from the higher-order polynomial regression discontinuity designs, the 95-per cent confidence intervals include zero for less than 95-per cent of the estimations, which means that they do not cover their assigned nominal coverage. Conversely, the confidence intervals for the local linear and the local quadratic regression discontinuity designs are close to the nominal coverage of 95 per cent. This implies that higher-order polynomial regression discontinuity designs are more susceptible to incorrect inference than their local linear and local quadratic counterparts (Gelman & Imbens, 2019).

Fourthly, another study by Imbens & Kalyanaraman (2009) presents findings suggesting that local linear estimates have attractive bias properties. They assert that the local linear estimator is more likely to have the attractive asymptotic feature of an optimal convergence rate for the estimation of regression discontinuity treatment effect. This property of the local linear estimator is found in the work of Porter (2003) who derive the optimal convergence rate of the estimation of the regression discontinuity treatment effect using local polynomial and partially linear estimators (Imbens & Kalyanaraman, 2009; Porter, 2003).

Provided these attractive properties of the local linear regression model for the non-parametric regression discontinuity design, the next phase is to present the estimation equations for both research hypotheses of the study, H1 and H2. As a reminder, H1 tests whether the national lockdown restrictions imposed on March 25th, 2020, had a significant negative effect on air pollution in the Indian city sample. H2 tests whether there was a significant heterogeneous impact from lockdown on air pollution in the city sample due to city characteristics such as polluting industries and the presence of energy plants. The local linear regression model run in this non-parametric regression discontinuity design will be applicable for both hypotheses since the only differences between the estimations in H1 and H2 are the included cities in the estimation samples. The applied local linear regression model is denoted by the following equation:

$$y_{tc} = \beta_0 + \beta_1 D_t + \beta_2 D_t f(x_t) + \beta_3 f(x_t) + \theta C_c + \varepsilon_{tc} \qquad (1)$$

Here, y_{tc} is the outcome variable, β_0 is an intercept, and $f(x_t)$ is the function of the forcing variable. D_t is the treatment indicator which is a discontinuous function of the continuous forcing variable x_t. D_t adheres to the following treatment rule:

$$D_t = \begin{cases} 0, & if \ x_t < L_0 \\ 1, & if \ x_t \geq L_0 \end{cases}$$

This treatment rule, L_0, refers to the threshold, i.e., the lockdown date. C_c is a set of control variables which are included to account for factors which may influence air pollutants such as temperature and wind speed, and has its coefficient measured by θ. The error term is denoted by ε_{tc}. Finally, the estimate of the treatment effect is measured by β_1. This is the estimate of the change in air pollutants due to the lockdown.

To explain the estimate of the treatment effect in more detail, it is important to explain that in the non-parametric regression discontinuity design, only data in a limited neighbourhood, η, around the discontinuity at the threshold, L_0 will be studied:

$$[L_0 - \eta, L_0 + \eta]$$

From this limited neighbourhood, η, the average outcomes on each side of the threshold are estimated as:

$$E[y_{tc} \mid L_0 - \eta < x_t < L_0] \cong E[y_{0tc} \mid x_t = L_0]$$

$$E[y_{tc} \mid L_0 \leq x_t < L_0 + \eta] \cong E[y_{1tc} \mid x_t = L_0]$$

The final non-parametric RD estimate can be written as these average outcomes of the air pollution outcomes in the limited neighbourhood, η, to the left and right of the threshold L_0:

$$\lim_{\eta \to 0} E[y_{tc} \mid L_0 < x_t < L_0 + \eta] - E[y_{tc} \mid L_0 - \eta < x_t < L_0] = E[y_{1tc} - y_{0tc} \mid x_t = L_0]$$

This estimate will give us the average treatment effect of the lockdown restrictions imposed on March 25th, 2020, on the different air pollutants (Angrist & Pischke, 2009). Again, the average outcomes of air pollution in the limited neighbourhood between February 28th – April 20th are studied, accounting for the control period (February 28th – March 24th) and the treatment period (March 25th – April 20th). The next step will be to test H1 and H2 by generating the average treatment effect from this model. This is executed in the next chapter.

5. Results

In this chapter, the main results from the non-parametric regression discontinuity design are based on the testing of hypotheses H1 and H2. The results for H1 and H2 are reported graphically in *Figures 5.1-3* and in *Figures 5.4-23*, respectively. In addition to the graphical figures, the results from the local linear regressions when testing H1 and H2 are presented in *Tables 5.1* and *Tables 5.2-13*, respectively.

5.1. H1: Air pollution responses to national lockdown

The testing of H1 provides a compelling, graphical view when observing the outcomes of $PM_{2.5}$. It is possible to see a clear discontinuity on each side of the cut-off point. When comparing the graphical outcomes to PM_{10}, it appears as if $PM_{2.5}$ has a sharper discontinuity at the cut-off compared to the one for PM_{10}. Regarding the findings for the outcomes of PM_{10} and NO_2, the discontinuity at the cut-off is pronounced as a positive rise in PM_{10} and NO_2 concentrations.

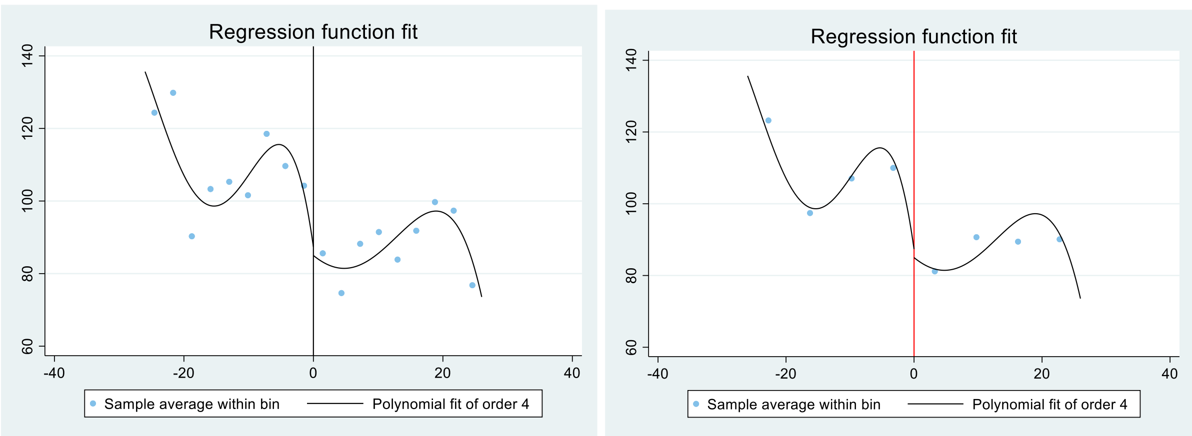

Figure 5.1. $PM_{2.5}$ regression discontinuity design plot using small bins (left) and large bins (right).

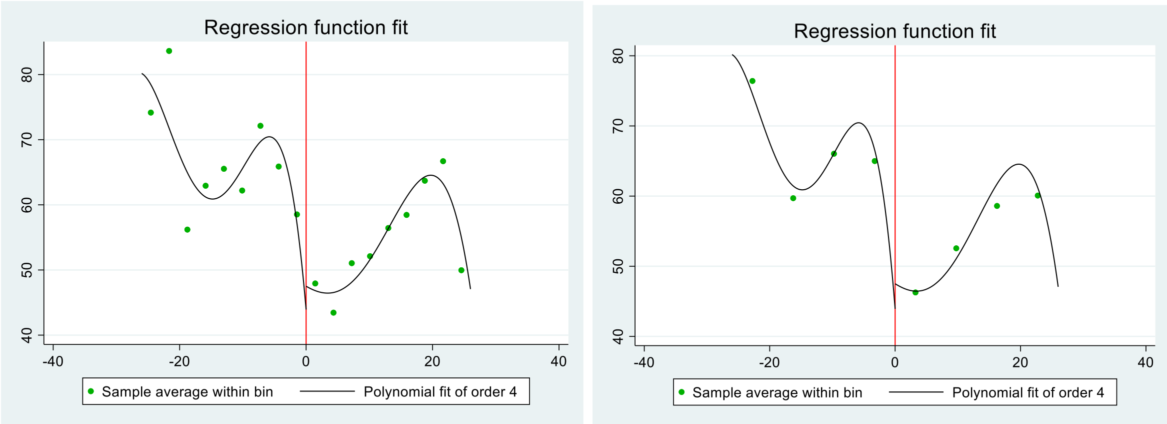

Figure 5.2. PM_{10} regression discontinuity design plot, using small bins (left) and large bins (right).

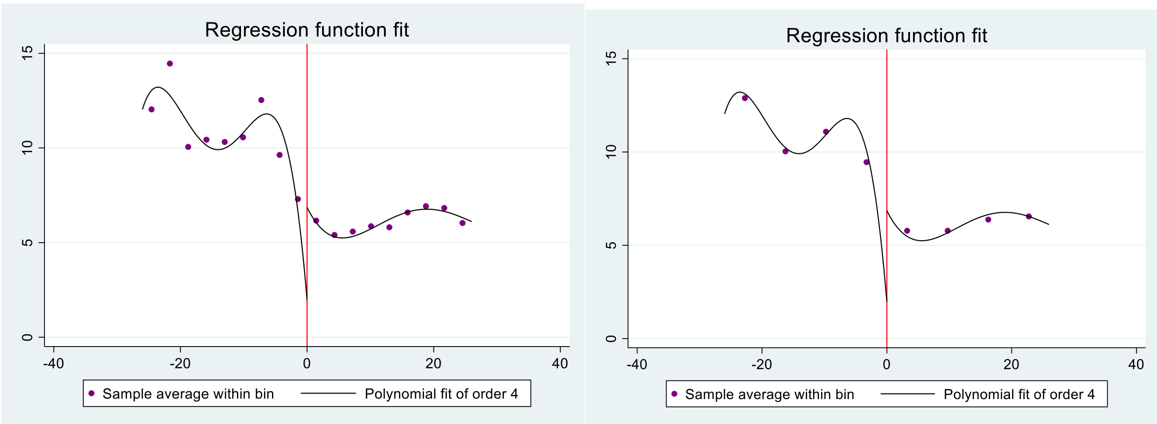

Figure 5.3. NO_2 regression discontinuity design plot, using small bins (left) and large bins (right).

Furthermore, *Table 5.1* presents the results for the local linear regressions for each air pollutant specie. There is not a statistically significant treatment effect of the national lockdown when including and excluding the covariates of temperature and wind speed. Moreover, the economic impact of one day of national lockdown on the air pollutant species, when including covariates, is equivalent to a daily decrease of -11.574 $\mu g/m^3$ in $PM_{2.5}$, -4.047 $\mu g/m^3$ in PM_{10}, and a daily increase of 1.333 ppb (parts per billion) in NO_2. The comparison of the local linear regressions with and without covariates in *Table 5.1* suggests that the economic impact remains highly similar in the comparisons, both from a statistical and economic point of view. The exception is PM_{10} which experiences an approximate 27 per cent reduction in the absolute value of its treatment effect when including covariates. This also yields an initial sensitivity check regarding possible endogeneity, where for $PM_{2.5}$ and NO_2, controlling for wind speed and temperature appears to have a generally minor influence on the treatment effect (Lynch & Brown, 2011).

Table 5.1. $PM_{2.5}$ regression discontinuity estimates with and without covariates. Covariates include accounting for wind speeds and temperature medians.

VARIABLES	(1) $PM_{2.5}$ median	(2) $PM_{2.5}$ median	(3) PM_{10} median	(4) PM_{10} median	(5) NO_2 median	(6) NO_2 median
RD Estimate	-12.159	-11.574	-5.526	-4.047	1.165	1.333
	(10.065)	(9.950)	(5.750)	(5.073)	(1.747)	(1.721)
Covariates	No	Yes	No	Yes	No	Yes
Mean of dependent variable	98.448	98.448	60.520	60.520	8.476	8.476
Observations	1,158	1,158	1,040	1,040	1,106	1,106

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

5.2. H2: Testing for heterogeneity in air pollution responses to national lockdown

Proceeding to test H2, graphical analysis based on the non-parametric regression discontinuity design for the city samples located in states with and without iron, steel, and cement industries are provided in *Figures 5.4-21* below. Local linear regressions for the industrial and non-industrial city samples are in turn presented in *Tables 5.2-12.* The industrial section is then followed by a graphical analysis for city samples with and without coal power stations present in the concerned or a neighbouring city district, presented in *Figure 5.22-27,* alongside local linear regressions in *Table 5.13.*[2]

5.2.1. Iron ore industry

The results from *Figure 5.4* give a compelling impression that the magnitude of the discontinuity for $PM_{2.5}$ at the cut-off is considerably larger for cities without iron ore industry compared to cities located in states with iron ore industry.

[2] The graphical analysis and local linear regressions for PM_{10} and NO_2 when testing H2 are located in the Appendix.

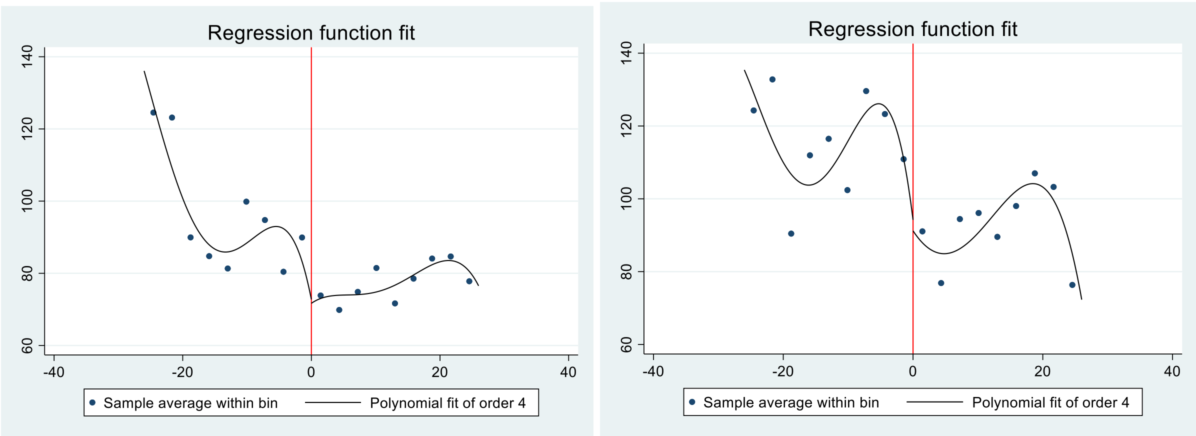

Figure 5.4. $PM_{2.5}$ regression discontinuity design plot for the iron ore industry sample (left panel) versus the non-iron ore industry sample (right panel), using small bins.

The results from the local linear regressions in *Table 5.2* suggest a significant negative treatment effect from lockdown restrictions for the samples with iron industry compared to the ones without iron industry. The city sample located in states with iron ore industry experienced a -29.40 $\mu g/m^3$ decrease in $PM_{2.5}$ when including wind speed and temperature covariates. In contrast, cities located in states without iron industry experienced a non-significant treatment effect of -7.290 $\mu g/m^3$. This finding is to be expected as it was concluded in Chapter 3 that iron ore industry constitutes a chief contributor to particulate matter emissions in India and globally.

Thus, the marginal effect of the lockdown which began on March 25th should be larger in the city sample where iron ore industry was present in the states where the cities were located. This is due to the strictly limited mobility of Indian citizens and the shutdown of several industries (Soni, 2021; Government of India, 2020).

Table 5.2. PM2.5 regression discontinuity estimates with and without covariates in the samples with iron ore industry and without iron ore industry. Covariates include accounting for wind speeds and temperature medians.

VARIABLES	(1) $PM_{2.5}$ median	(2) $PM_{2.5}$ median	(3) $PM_{2.5}$ median	(4) $PM_{2.5}$ median
RD Estimate	-27.252*** (8.250)	-29.402*** (7.244)	-7.716 (14.50)	-7.290 (14.32)
Covariates	No	Yes	No	Yes
Iron ore industry	Yes	Yes	No	No
Mean of dependent variable in sample	86.430	86.430	103.980	103.980
Observations	365	365	793	793

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

5.2.2. Steel industry

In *Figure 5.5.* the graphical representation of the non-parametric regression discontinuity design shows a considerable difference in $PM_{2.5}$ responses between the city sample located in a state with steel production below the median steel production versus the city sample located in a state with steel production above the median. It appears as if there is a positive discontinuity for the cities in states with above-median steel production, while a negative discontinuity is observed for the cities located in states with below-median steel production.

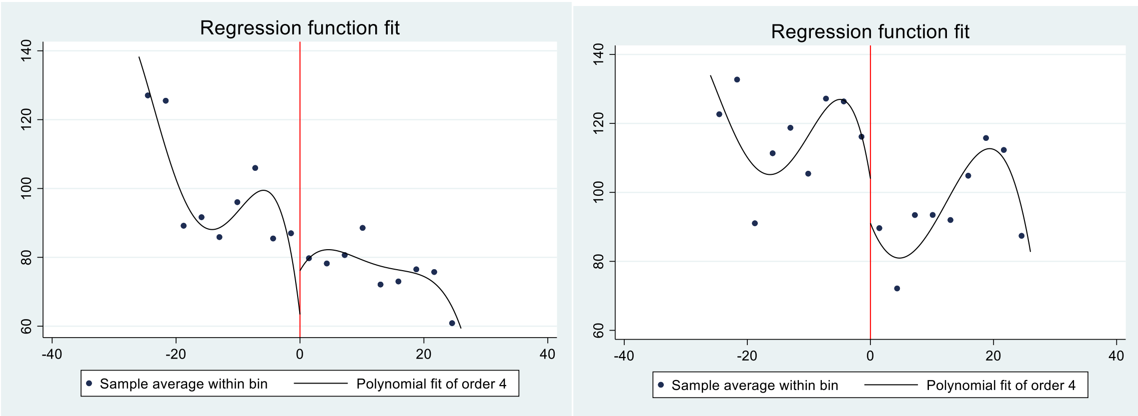

Figure 5.5. $PM_{2.5}$ regression discontinuity design plot for the above-median steel production sample versus below-median steel production sample, using small bins.

Furthermore, the local linear regressions in *Table 5.3* below suggest that the treatment effects of the lockdown restrictions on $PM_{2.5}$ were non-significant in both city samples above and below the steel production median. Including versus excluding covariates show a limited change in the treatment effects, suggesting that endogeneity between the forcing variable and the error term is likely not a problem.

Table 5.3. $PM_{2.5}$ regression discontinuity estimates with and without covariates in the city samples located in regions with steel production above and below the steel production median. Covariates include accounting for wind speeds and temperature medians.

VARIABLES	(1) $PM_{2.5}$ median	(2) $PM_{2.5}$ median	(3) $PM_{2.5}$ median	(4) $PM_{2.5}$ median
RD Estimate	-5.294 (7.463)	-4.121 (6.723)	-14.877 (15.19)	-13.833 (15.04)
Covariates	No	Yes	No	Yes
Median steel industry production	Above	Above	Below	Below
Mean of dependent variable in sample	87.391	87.391	106.029	106.029
Observations	471	471	687	687

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.2.3. Cement industry

In *Figure 5.6* the discontinuity of the $PM_{2.5}$ response to the lockdown is visualized in two city samples whose cities were divided based on whether they were located in states with above-median or below-median cement production. Cities which are situated in states with below-median cement production showed a slightly positive discontinuity while their opposites in the above-median cement production sample experienced no clear discontinuity at all.

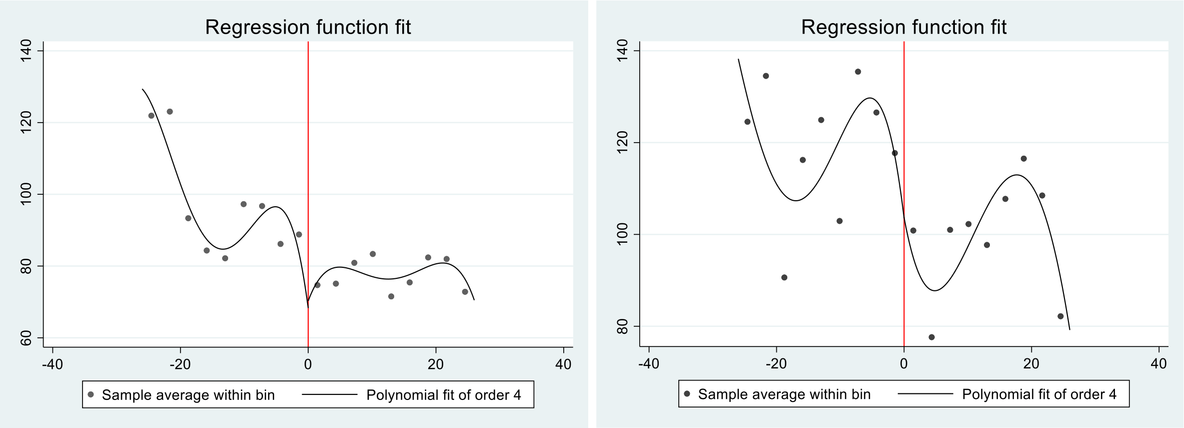

Figure 5.6. $PM_{2.5}$ regression discontinuity design plot for the above-median cement production sample versus below-median cement production sample, using small bins.

Table 5.10 illustrates a significant treatment effect of -11.970 $\mu g/m^3$ on $PM_{2.5}$ in the cities located within states with above-median cement production. For the below-median sample, the treatment effect amounted to an insignificant -3.965 $\mu g/m^3$. Once again, including covariates has a minor impact on the treatment effects' magnitudes.

Table 5.4. $PM_{2.5}$ regression discontinuity estimates with and without covariates in the city samples located in regions with cement production above and below the steel production median. Covariates include accounting for wind speeds and temperature medians.

VARIABLES	(1) $PM_{2.5}$ median	(2) $PM_{2.5}$ median	(3) $PM_{2.5}$ median	(4) $PM_{2.5}$ median
RD Estimate	-13.055** (5.540)	-11.970** (5.211)	0.836 (19.022)	-3.965 (19.354)
Covariates	No	Yes	No	Yes
Median cement industry production	Above	Above	Below	Below
Mean of dependent variable in sample	87.002	87.002	109.135	109.135
Observations	524	524	476	476

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.2.4. Coal power stations

In *Figure 5.7*, the $PM_{2.5}$ responses for cities characterized and not coal power vs non-coal power are introduced. Sharp discontinuities are present in both samples, with the non-coal power sample showing a slightly sharper discontinuity.

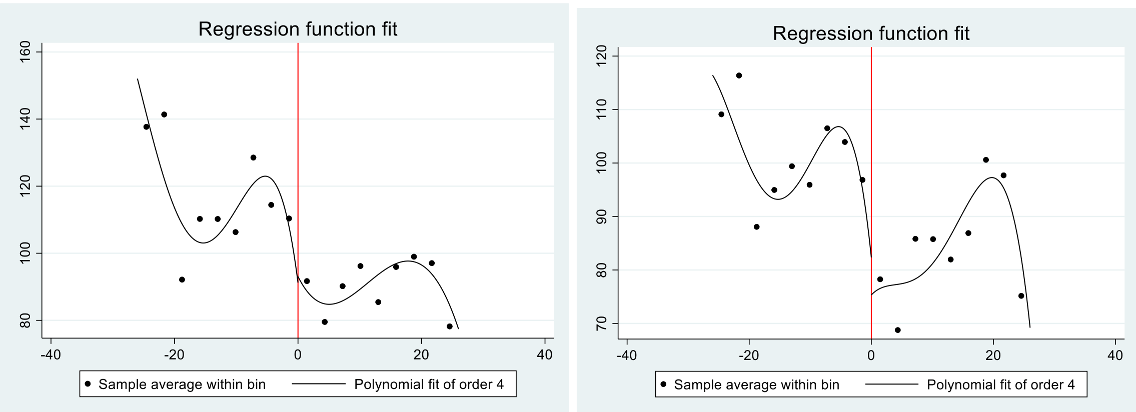

Figure 5.7. $PM_{2.5}$ regression discontinuity design plot for the coal power sample versus non-coal power sample, using small bins.

In *Table 5.5* the local linear estimates suggest that the treatment effects on $PM_{2.5}$ were neither significant in cities characterized by coal power stations nor in cities not characterized by coal power stations. The negative estimates are relatively similar in magnitude between both samples when excluding covariates. However, the difference increases when controlling for covariates, suggesting that there may be some endogeneity between the meteorological covariates and the treatment effect on $PM_{2.5}$ from lockdown in the coal – and non-coal power station samples.

Table 5.5. $PM_{2.5}$ regression discontinuity estimates with and without covariates in the coal power sample and non-coal power sample. Covariates include accounting for wind speeds and temperature medians.

VARIABLES	(1) $PM_{2.5}$ median	(2) $PM_{2.5}$ median	(3) $PM_{2.5}$ median	(4) $PM_{2.5}$ median
RD Estimate	-12.96 (12.59)	-10.54 (11.51)	-12.40 (13.05)	-14.80 (12.98)
Covariates	No	Yes	No	Yes
Coal power	Yes	Yes	No	No
Mean of dependent variable in sample	103.218	103.218	92.758	92.758
Observations	630	630	528	528

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6. Robustness tests

Albeit a powerful tool to study air pollution responses around a national lockdown threshold, the non-parametric regression discontinuity design deployed in this paper faces several threats, where some were outlined in Chapter 4. These threats require robustness tests to be conducted to observe whether they might make the non-parametric regression discontinuity design in this paper invalid. In particular, the estimated treatment effects from the local linear regressions will be subject to scrutiny by applying a linear estimator over two shorter intervals around the cut-off. Each threat and its assigned robustness test are presented with a short description and motivation, followed by the results from the tests.

6.1. Changing the interval and applying a linear regression model

This study performs a robustness test to address the issue of mass points in the forcing variable of the local linear regression estimation above. Mass points denote distinct values shared by many observations. For example, in the non-parametric regression discontinuity design of this study, the forcing variable is discrete by nature and only takes integer values. In the setting of this study, the observations of air pollutants measured in the cities will therefore be sharing many values i.e. there will be mass points. If this number of mass points is very limited, then the local linear regression method above is not applicable. The reason is that the local linear method uses each mass point as a single observation. This would be the equivalent of having the observations aggregated by the values of the forcing variable and then the average outcomes are calculated for all the observations at each value. The effective, total number of observations used will be consisting of these mass points in the forcing variable (Cattaneo, Idrobo & Titiunik, 2018).

This robustness test will investigate whether the main results above from testing H1 through the local linear regression are robust by using an alternative linear estimator in a smaller interval closer to the cut-off, more specifically [-5, 5] and [-10, 10] days before and after lockdown. The linear estimator used here is not specified as the local linear estimator, where the latter would count the mass points as the effective observations. The smaller interval should make the number of polynomial terms needed to model the forcing variable function decrease, based on the logic from the nonparametric regression discontinuity designs explained earlier in the text. The decision to apply this robustness test to just H1 and not H2 is founded on the limited data for the divided samples based on industrial production at the state level and the presence of coal power stations needed for H2. In addition, if the number of mass points is relatively low, then the extrapolation from a linear regression globally would be possible to estimate the treatment effect of the regression discontinuity design. Based on these factors, in the smaller intervals, the usage of the linear estimator globally (not locally as with the local linear estimator) should be suitable for estimating the treatment effect of lockdown restrictions on air pollution (Cattaneo, Idrobo & Titinuik, 2018; Cattaneo, Titiunik & Vasquez-Bare, 2020). Based on the reasons above, cutting down the interval and deploying a linear regression estimator should therefore provide a useful robustness test for the main results which were obtained by the local linear estimator. The results are presented graphically in *Figures 6.1-3* and the linear regression estimates are placed in *Tables 6.1-3.*

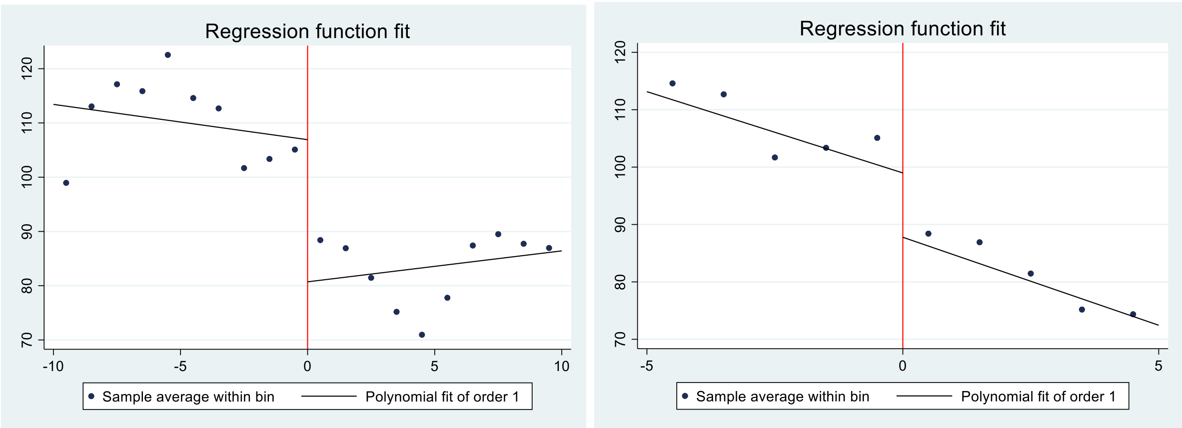

Figure 6.1. Interval of [-10, 10], small bins to the left, and interval of [-5, 5], large bins, to the right.

Table 6.1. Interval changes for $PM_{2.5}$.

VARIABLES	(1) PM25 median	(2) PM25 median	(3) PM25 median	(4) PM25 median
National lockdown	-26.930***	-25.732***	-27.368***	-25.853***
	(3.009)	(3.046)	(4.398)	(4.340)
Wind speed median	-	-2.679	-	-4.931*
		(2.063)		(2.781)
Temperature median	-	-0.640	-	-0.638
		(0.415)		(0.574)
Constant	110.496***	130.655***	107.482***	130.182***
	(2.500)	(11.726)	(3.762)	(16.569)
Interval	[-10, 10]	[-10, 10]	[-5, 5]	[-5, 5]
Observations	462	462	242	242
R-squared	0.152	0.159	0.148	0.161

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

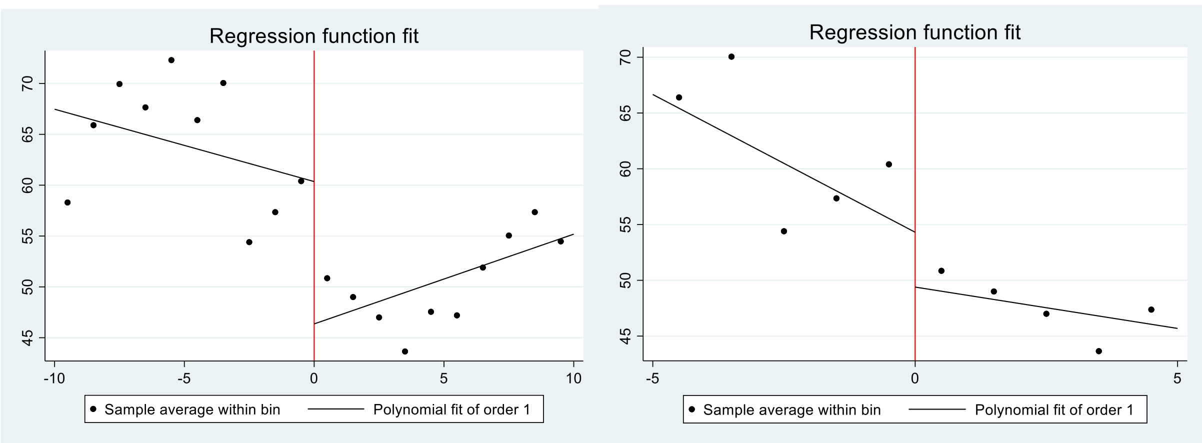

Figure 6.2. Interval of [-10, 10] for PM_{10}, small bins to the left, interval [-5, 5] for PM_{10}, large bins, to the right.

Table 6.2. Interval changes for PM_{10}.

VARIABLES	(1) PM_{10} median	(2) PM_{10} median	(3) PM_{10} median	(4) PM_{10} median
National lockdown	-13.497***	-12.805***	-14.178***	-12.933***
	(2.323)	(2.343)	(3.162)	(3.008)
Wind speed median	-	-8.032***	-	-9.655***
		(1.871)		(2.642)
Temperature median	-	0.443*	-	0.813**
		(0.250)		(0.351)
Constant	64.270***	61.542***	61.720***	50.318***
	(1.761)	(7.474)	(2.601)	(10.736)
Interval	[-10, 10]	[-10, 10]	[-5, 5]	[-5, 5]
Observations	420	420	220	220
R-squared	0.075	0.140	0.088	0.187

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

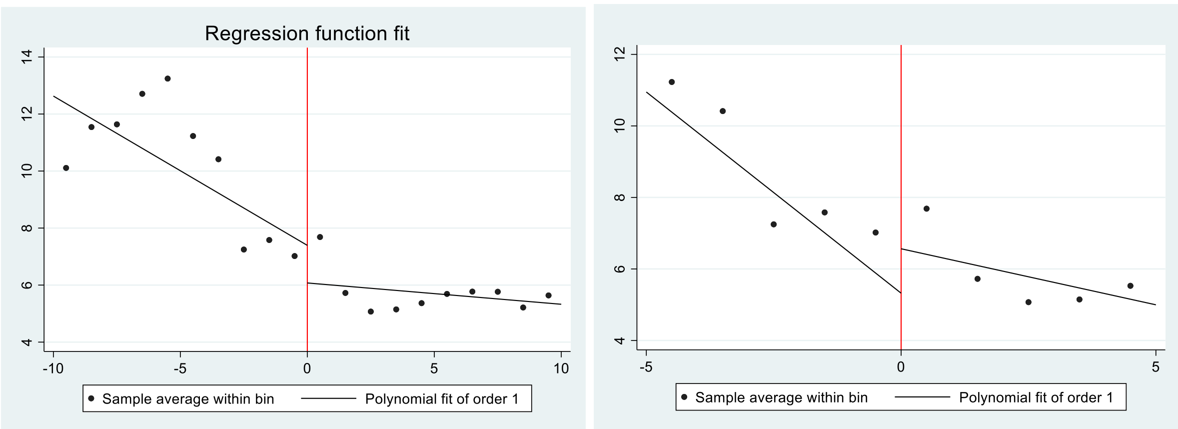

Figure 6.3. Interval [-10, 10] for NO_2, small bins to the left, interval [-5, 5] for NO, large, bins to the right.

Table 6.3. Interval changes for NO_2.

VARIABLES	(1) NO_2 median	(2) NO_2 median	(3) NO_2 median	(4) NO_2 median
National lockdown	-4.572***	-4.584***	-2.917***	-2.737***
	(0.490)	(0.528)	(0.652)	(0.699)
Wind speed median	-	-0.855***	-	-1.175***
		(0.318)		(0.425)
Temperature median	-	0.118	-	0.0746
		(0.0739)		(0.0899)
Constant	10.273***	8.062***	8.698***	7.941***
	(0.410)	(2.005)	(0.492)	(2.471)
Interval	[-10, 10]	[-10, 10]	[-5, 5]	[-5, 5]
Observations	441	441	231	231
R-squared	0.170	0.188	0.081	0.110

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results in *Figures 6.1-3* suggest strongly pronounced discontinuities for the new intervals of [-10, 10] and [-5, 5] days. The discontinuities are generally sharper and larger in magnitude for the [-10, 10] interval. For $PM_{2.5}$, the linear estimator has a better fit to the data when using the [-5, 5] interval based on pure visual inspection. Yet, by comparing the R-squared across the different intervals, it does not differ very much, which implies they have a similar fit to the data. The exceptions are the intervals for NO_2 where the [-10, 10] interval has a better fit than the [-5, 5] interval both with and without covariates. Interestingly, based on *Figure 6.3*, the [-5, 5] interval suggests a positive discontinuity for NO_2 due to lockdown, whereas the [-10, 10] interval suggests a negative discontinuity from lockdown.

The linear regressions presented in *Tables 6.1-3* support the graphical analysis. The treatment effect of the lockdown restrictions is in almost all cases negative and exclusively statistically significant. For the [-10, 10] interval, the treatment effect of lockdown consists of a daily decrease in $PM_{2.5}$ by -25.732 $\mu g/m^3$ when including covariates for temperature and wind speed. This is accompanied by a daily decrease of -14.178 $\mu g/m^3$ in PM_{10} and a fall by -4.584 ppb in NO_2, again including covariates. The linear regressions for the [-5, 5] interval also assert a significant decrease in all three air pollutants. They also indicate approximately similar magnitudes for all air pollutants as for the [-10, 10] intervals, except for NO_2 where there was a 36-40 per cent decrease in the absolute value of the treatment effect when using the [-5, 5] interval compared to [-10, 10] interval.

6.2. Covariates by forcing variable

Another potential threat against the non-parametric regression discontinuity design in this study is if there is a shift in the covariates wind speed and temperature medians at the cut-off. The identification of the regression discontinuity design should not be affected by the presence of covariates, but if a shift in covariates is present at the cut-off, then the assumptions for the regression discontinuity design may be possible to question. If there are other changes which occur at the cut-off, such as changes in the meteorological covariates wind speed and temperature, then they may be influencing the air pollution outcomes and hence included in the treatment effect (Imbens & Lemieux, 2008).

Therefore, to test whether these covariates shift at the threshold, a robustness test recommended by Imbens & Lemieux (2008) which involves graphical analysis, is regressing the covariates on the forcing variable. The test uses bins and the average of the covariates within these bins and measures them on each side of the threshold. The only difference here to the previous graphical analysis in the study is that the outcome variables used will consist of the covariates, i.e. wind speed and temperature (Imbens & Lemieux, 2008).

In *Figures 6.7-8* below, both wind speed and temperature medians are used as the outcome variables, plotted against the forcing variable which again is time recentred around the threshold, 25th of March. The graphs showing wind speed suggest no shift at the threshold. For temperature, no clear discontinuity emerges either in the positive trend of the covariate. The positive trend could be due to temperatures in India rising during the period from February 28th – April 20th.

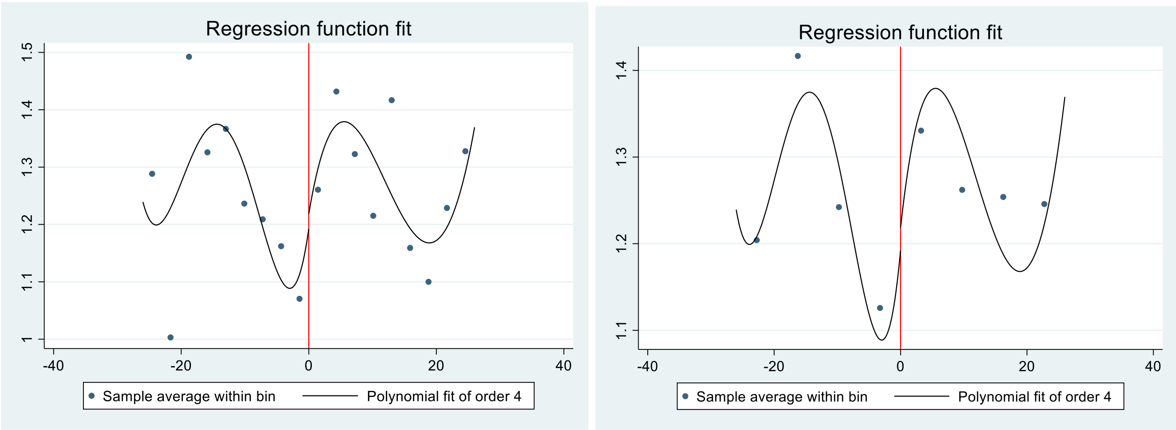

Figure 6.7. Wind speed, small bins (left) and large bins (right).

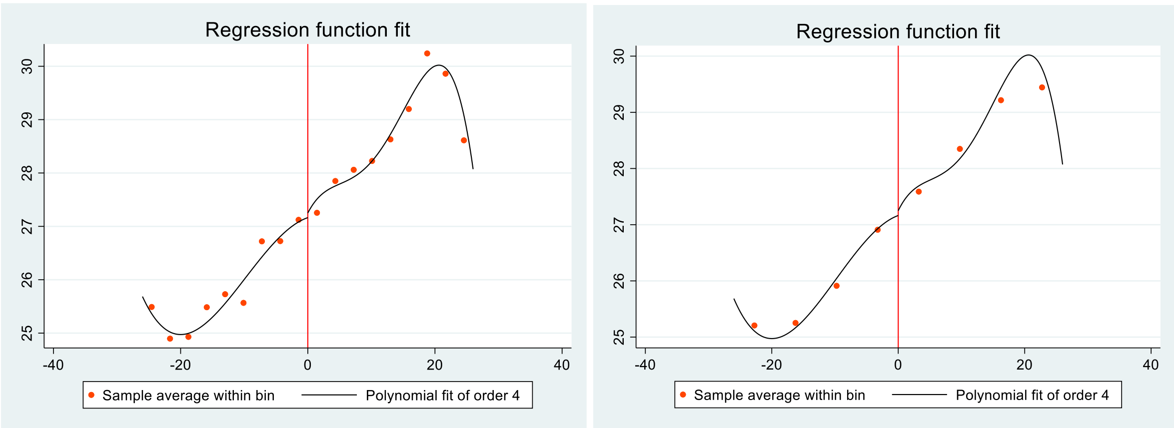

Figure 6.8. Temperature, small bins (left) and large bins (right).

7. Discussion

Based on the results from testing the two research hypotheses H1 and H2, there are several points which are open to discussion regarding their relation to the previous literature and the economic debate. Firstly, it is important to discuss the robustness test with the smaller intervals and its results from using a linear estimator to test H1. These results were jointly statistically significant, in contrast to the results for H1 obtained with the local linear estimator. This provides room for an initial, interesting, discussion about the choice of estimator in this study.

The main differences between the local linear estimator and the linear estimator used in this study are clear in their definitions. The local linear estimator uses weighted polynomial regression of order 1 to estimate regression functions to the left and right of the cut-off. A triangular kernel function assigns more weight to observations closer to the cut-off. The local linear estimator focuses on the observations within a specific bandwidth around the cut-off, and observations outside this bandwidth are discarded. This bandwidth to the left and right of the cut-off selected from the local linear estimator in the testing of H1 ranged from 5.60 – 6.51. When testing H2, the selected bandwidth ranged from 4.03 – 7.55. An average treatment effect is finally calculated from the average outcomes on each side of the cut-off (Imbens & Lemieux, 2008; Calonico, Cattaneo & Titiunik, 2014; Gelman & Imbens, 2019; Cattaneo, Titiunik & Vazquez-Bare, 2020).

Conversely, the linear estimator simply fits an ordinary least-squares (OLS) linear regression to the data. The selected windows around the cut-off consisted of intervals set to [-10, 10] and [-5, 5]. The reason to move closer to the cut-off was to avoid the issues of misspecifications of the forcing variable. In addition, if the number of mass points is low, then extrapolation from a global linear estimator should be used to estimate the treatment effect.

In the context of the choice of estimator, the local linear estimator is not an applicable estimator when the forcing variable is characterized by a small number of mass points. When there is a small number of mass points in the forcing variable, it is more feasible to use linear estimators globally to estimate the treatment effect. (Catteneo, Idrobo & Titiunik, 2018). The large differences in results when performing the robustness test with changed intervals and the linear model to test H1, compared to the main results from the local linear estimator, suggest that the main results are not robust.

The reason this study decided to use the local linear estimator as the model to obtain the main results of this study is based on its positive reputation in the literature. The local linear estimator has the asymptotic property of an optimal convergence rate to the regression discontinuity treatment effect (Imbens & Kalyanraman, 2009). The local linear model also performs better in terms of inference regarding the nominal coverage of confidence intervals compared to higher-order global polynomials. Compared to the global higher-order polynomials, local linear and local quadratic models are also more likely to assign more weights for individuals with values closer to the threshold. Again, this is the context where the average treatment effect estimates are calculated as the difference in weighted averages between treated and controls. In contrast, the weights are sensitive to higher-order polynomials, who may give excessive weights observations far away from the threshold. Finally, the local linear model and the local quadratic model generate less different estimates of discontinuities in regression discontinuity designs compared to global higher-order polynomial models (Gelman & Imbens, 2019).

Despite these good characteristics of the local linear model, the results from it are not robust to a change in the interval around the cut-off and the application of the linear estimator. This may suggest that the local linear estimator is indeed suffering from the presence of mass points in the forcing variable, as the effective number of observations equals the number of mass points. However, if the number of mass points is sufficiently large, the application of the local linear model is reasonable. If a sufficiently large number of mass points are present, then this makes it possible to use the local linear model in a similar vein to the non-parametric regression discontinuity design used in this study (Cattaneo, Idrobo & Titiunik, 2018).

Unfortunately, it is difficult to conclude whether the local linear model is a poor choice compared to the linear model provided the data for this study. When generating the local linear regressions mass points are detected in the forcing variable of the local linear model when attempting to estimate the treatment effect of lockdown restrictions on air pollutants. Whether the number of mass points is too small for the data in this study is difficult to answer. If there was a longer period, it may have been possible to have had a more successful estimation with the local linear model as this possibly could have meant more effective observations due to more mass points. However, while extending the period studied may have improved the efficiency it may have led to more biased estimates, for example, due to new, outside factors arising from changes in the lockdown restrictions during the different phases of lockdown in India in 2020 (Lee & Lemieux, 2010).

Proceeding to discuss the main results from testing H1 in relation to the related literature and economic debate, the findings from the local linear model do not find strong support in the previous literature. The results suggest insignificant treatment effects from lockdown restrictions on both $PM_{2.5}$, PM_{10}, and NO_2 air pollutants. In addition, although the graphical analysis asserted that there were positive discontinuities for both $PM_{2.5}$ and PM_{10} these were modest. Conversely, there was a sharp, positive discontinuity for NO_2 which goes against the formulation of H1.

In contrast to this study's result when testing H1, Dang & Trinh (2020), Zhao et al. (2021), and Liu et al. (2021) find clear evidence of a decline in the air pollutants NO_2, $PM_{2.5}$, and PM_{10}. Dang & Trinh (2020 report a 24-per cent decrease in NO_2 due to lockdown restrictions, using controls for temperature and precipitation, and a bandwidth of two weeks. Zhao et al (2021) report decreases of 17.8, 3.9, and 10.3 per cent of NO_2, $PM_{2.5}$, and PM_{10}, respectively. Finally, Liu et al. (2021) discover a 23-37 per cent decrease in NO_2, a 14-20 per cent decrease in PM_{10}, and a 7-16 percent decrease in $PM_{2.5}$ (Dang & Trinh, 2020; Zhao et al., 2021; Liu et al., 2021).

Moreover, the economic debate surrounding the impact of COVID-19 on air pollution has been reported in a relatively homogeneous manner: the pandemic led to restrictions from governments which in turn proved to reduce air pollution on both the national level and city level. This indication is consistent and appears to be present on a global scale, which is presented through the work of Addas & Maghrabi (2021). Their findings of significant improvements in air quality from lockdowns are supported by Dang & Trinh (2020), Liu et al (2021), and Lenzen et al. (2020).

Furthermore, the main results from the local linear model are interesting to evaluate based on the theoretical background established in Chapter 2, The background was constructed on the notion that the lockdown restrictions would act as a form of governmental intervention, albeit with a different purpose than addressing air pollution. However, as the lockdown restrictions were introduced with mobility and restrictions to economic activity, it could be argued that the individual citizens would have to bear the costs of pollution by being forced to isolate themselves. This could, theoretically, be an alleviating factor to air pollution as an externality caused by private decision-making which does not bear the costs of pollution. The main findings in the study suggest that this intervention in the shape of lockdown was generally not successful at mitigating air pollution outcomes when reasoning about the roots of air pollution as an externality, based on the theoretical literature. On the other hand, various interventions suggested in the theoretical literature, such as Pigouvian taxes and bargaining according to the Coase theorem also have issues with being successful in mitigating an externality such as air pollution. In addition, from the robustness tests, it appears as if the lockdown restrictions may have had a significant impact on air pollution outcomes. The findings should also be relevant for Indian policymakers concerning the implementation of environmental policies intended to manage air pollution, especially since an overhaul to the existing Air Act is potentially on the way.

Overall, the main results of this study suggest that H1 cannot be accepted. There was no significant impact from lockdown restrictions on $PM_{2.5}$, PM_{10}, and NO_2 concentrations in a sample of twenty-two Indian cities. The related empirical and theoretical literature, and the economic debate disagree on this insignificant treatment effect of lockdown restrictions on the air pollutants $PM_{2.5}$, PM_{10}, and NO_2.

Moreover, regarding the second work hypothesis, H2, it is interesting to begin the discussion with the discontinuity estimates in graphical analysis. The main results from the graphical analysis when testing H1 suggested graphical discontinuities in the air pollution outcomes around the cut-off point. The discontinuities were pronounced for $PM_{2.5}$, PM_{10}, and NO_2, where the former two showed negative discontinuities around the cut-off while the latter had a positive discontinuity. When testing H2, negative discontinuities emerged for PM2.5 responses in both the iron ore and non-iron ore industry samples. In the different city samples based on the state-level production of steel and cement, the results were less uniform. In the city sample with above-median state-level steel production, there was a positive discontinuity for $PM_{2.5}$ while there was a negative one for the sample with below-median state-level steel production. For the cement industry samples, there was a slightly positive discontinuity for the cities located in states with above-median cement production. There was no clear discontinuity for cities in states with below-median cement production.

As presented above, the graphical discontinuities for $PM_{2.5}$ for the cement industries were very mild and could prove to be less sensitive to the lockdown restrictions in terms of $PM_{2.5}$ air pollution responses. The steel industry's ambiguous positive and negative discontinuities suggest that the direction of the effect of the lockdown restrictions remains unclear. Conversely, the city samples with and without iron industry both showed a negative discontinuity. A possible, contributing factor to the small and ambiguous discontinuities for the steel and cement industry samples could be that they were selected in a similar way by using the median of state-level production to divide the cities into different samples. On the contrary, the selection of the city samples of the iron ore industry was simply based on whether the cities were located in states with or without iron ore industry. A more detailed selection of the samples for the steel and cement industry made at the city level instead of the state level could potentially have produced different, more clear graphical discontinuities. However, the widespread proliferation of steel and cement industries across India made a more delicate selection of the sample complicated.

An example where the selection indeed was more delicate was the one for the city samples characterized with or without coal power stations in their own district or in a neighbouring district. The graphical discontinuity for the city sample characterized by coal power stations was mild and without any sharp shift in outcomes. The city sample not characterized by coal power stations had a sharper, negative discontinuity.

In general, for $PM_{2.5}$ responses, the city samples which were not characterized by coal power stations, iron ore industry, and below-median steel production, proved to have the stronger, negative discontinuities graphically. This may suggest that there is a heterogeneous response of $PM_{2.5}$ pollutants to the lockdown restrictions based on the presence of polluting industries and power stations, which takes the form of less sensitivity to lockdown restrictions in cities with relatively more polluting industries and power stations.

The above results of the graphical analysis are consistent with the findings of Dang & Trinh (2020) who discover that locations without coal power stations indeed experience larger reductions in NO_2. This is consistent with this study's graphical analysis of NO_2 responses in *Figure 5.26* in the Appendix. As opposed to this work's graphical analysis, Liu et al. (2021) present findings which suggest that cities with more industrial activity experienced larger reductions in $PM_{2.5}$. A more global perspective from Lenzen et al. (2020) suggests that $PM_{2.5}$ emissions were reduced during the pandemic partly due to lower power output in Asia. Based on this finding, one could argue that coal power activity in Indian cities would inhibit their capacity to limit air pollution in the form of $PM_{2.5}$. This would be in favour of the above results from testing H2.

For the results from the local linear regressions when testing H2, the local linear results suggest that $PM_{2.5}$ experienced a significant reduction in cities located in states with iron ore industry. There was also a negative significant treatment effect for $PM_{2.5}$ in cities located in states with above-median cement production. For cities located in states with above-median steel production, there was no significant response in $PM_{2.5}$ to the lockdown restrictions. No significant $PM_{2.5}$ responses were generated for cities with a coal power station in their own or in a neighbouring district.

The results from the local regressions for testing H2 were partly consistent with the related literature. Liu et al. (2021) assert that cities reliant on industrial activity, measured by the share of secondary production and CO2 emissions, experienced a stronger lockdown effect. This finding was consistent in this study for cities located in states with iron ore industry, as this city sample showed a stronger response in $PM_{2.5}$ emissions due to lockdown restrictions. In addition, the city sample located in states with above-median cement production experienced a statistically significant reduction in $PM_{2.5}$ outcomes from the lockdown. However, the $PM_{2.5}$ responses were not significant in the city sample in states with above-median steel production or in the city sample where coal power stations were based in the same city district or in a neighbouring district. Contrary to the findings of the local linear regressions, Dang & Trinh (2020) present findings which indicate a stronger lockdown effect on NO_2 for cities with less industrial production. Their results are not consistent with this study's results for the NO_2 responses of any of the industrial samples or coal power samples, which are located in the Appendix.

The relation of these local linear regression results for testing H2 to the economic debate imply new insights regarding heterogeneous air pollution responses to lockdown restrictions. The heterogeneity in the air pollution responses based on the presence of certain industries suggests that is important to consider such factors when investigating the impact of lockdown restrictions on air pollution. As has been previously explained, industrial activity related to steel, cement, and iron ore is reported as a serious culprit in the pollution of particulate matter. However, given that the presence of iron ore industries and the below-median cement production led to significant reductions in $PM_{2.5}$, one could argue that these industries do not necessarily hamper potential reduction in air pollution from policies such as the lockdown restrictions. Moreover, the lower impact of the lockdown restrictions in cities located in states without iron production and below-median output of cement suggests that less industrial activity measures do not necessarily mean that interventions similar to lockdown restrictions will significantly reduce air pollution. This would suggest that control over air pollution in India is a challenge which cannot necessarily be overcome through policy adhering to one particular source of pollution, in this case, industrial activity. Thus, a future challenge for Indian and other policymakers will be to extend their scope when identifying possible sources of air pollution. Future policy needs to include a larger spectrum of sources contributing to air pollution.

Generally, based on the generated results from the local linear regressions and graphical analysis suggest that H2 can be accepted. There appears to be a significant heterogeneous impact of the lockdown restrictions from March 25th – April 20th, 2020, where cities characterized by state-level industrial activity within the iron ore industry and above-median cement output at the state level experienced significant reductions in $PM_{2.5}$ levels from the lockdown restrictions. Cities with no iron ore industry present at the state level and below-median cement output at the state level did not experience a significant reduction in $PM_{2.5}$. The latter also holds for cities characterized by both above-median and below-median steel output at the state level, and by the presence and absence of coal-power stations in their own or in a neighbouring district.

Finally, it is also important to consider that to this study's knowledge, there has been no previous work using a non-parametric regression discontinuity to estimate the impact of COVID-19 lockdown restrictions on air pollution in India. This provides the study with good scientific relevance to explore this relationship. It also makes it important to consider the context of the lockdown policies and their enforcement in India compared to other countries. Based on the COVID-19 stringency index from Blavatnik School of Government at the University of Oxford, which is developed based on response indicators such as school closures, workplace closures, and travel bans, India scored higher than Italy, the United States, Germany, Canada, France, and the United Kingdom between March 25th – April 20th, 2020 (Our World in Data, 2020). The stringency of India's lockdown was unsurpassed in August 2020 according to Ray & Subramanian (2020) who also suggested that India's lockdown was relatively successful with regard to the enforcement of the lockdown. The authors compare India's lockdown policy model with other developed countries in Europe and North America, suggesting that this model is a good representative of the global restrictions (Ray & Subramanian, 2020). The high stringency of the lockdown restrictions imposed by the Indian government on March 25th, 2020, and this study's findings of their heterogeneous impact on air pollution, are indications of the complexity of the problem of air pollution. Even the strongest of governmental interventions may not achieve the general, uniform results that they set out to obtain. Thus, for Indian policymakers, and policymakers in general, the formulation of policies to control air pollution cannot be formulated according to the doctrine of 'one-size fits all’. It will be crucial to formulate adaptable, flexible future policies for controlling air pollution to cater for the heterogeneous nature of air pollution responses to interventions.

Conclusion

This study investigates the air pollution outcomes in a sample of Indian cities before and after the introduction of COVID-19 lockdown restrictions on March 25th, 2020. Two research hypotheses were formulated, H1 and H2: H1 implied that the lockdown would cause a significant negative response of air pollution concentrations in the city sample; H2 asserted that there would be a significant heterogeneous impact from the lockdown restrictions on air pollution concentration in the city sample due to the presence of polluting industries and energy stations. The results from local linear regressions within the frames of a non-parametric regression discontinuity design suggest that H1 could not be accepted. There was a statistically insignificant impact of the COVID-19 lockdown restrictions imposed on March 25th on the air pollution species $PM_{2.5}$, PM_{10}, and NO_2. The results suggest that H2 can be accepted. The cities characterized by the presence of iron ore industry at the state level as well as above-median cement production at the state level experienced a significant, negative, reduction in $PM_{2.5}$ median concentration levels. There was no significant impact on air pollution in cities with above – and below median steel output at the state level, nor in cities with coal power stations in their own or in a neighbouring district. The main results are robust to changes in the covariates at the cut-off point. However, they are not robust to a change to smaller intervals surrounding the cut-off and the alternative use of a linear regression model to test H1. This finding may be due to the presence of a small number of mass points in the forcing variable.

The findings of this study suggest that pollution for mitigating air pollution need to be aware of the complexity and the extent of the problem. Certain industries may cause heterogeneous outcomes in air pollution concentrations when restrictions on economic activity and mobility are enforced. Future environmental regulations need to be flexible to effectively mitigate air pollution. Moreover, future research should explore the limits of the applicability of local linear models in regression discontinuity designs when using a discrete forcing variable. In addition, future attempts to use non-parametric regression discontinuity designs to analyse the outcomes of broad governmental interventions over a limited period are welcome.

References

A status assessment of National Air Quality Index (NAQI) and pollution level assessment for Indian cities. (2015, December). Retrieved from Greenpeace: https://www.greenpeace.org/static/planet4-india-stateless/2018/05/India-NAQI-PRESS.pdf

About us. (2022, April 20). Retrieved from Central Pollution Control Board: https://cpcb.nic.in/Introduction/

Air quality database. (2022). Retrieved from World Health Organization: https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database

Air Quality Open Data Platform: Worldwide COVID-19 dataset. (2022, August 19). Retrieved from World Air Quality Index Project: https://aqicn.org/data-platform/covid19/

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Annexure to Ministry Home Affairs Order NO . 40-3/2020-D. (2020, March 24). Retrieved from Ministry of Home Affairs: https://www.mha.gov.in/sites/default/files/Guidelines_0.pdf

Basics of COVID-19. (2021, November 4). Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19/basics-covid-19.html

Branson, Z., & Mealli, F. (2019). *The Local Randomization Framework for Regression Discontinuity Designs: A Review and Some Extensions.*

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 2295-2326.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals For Regression-Discontinuity Designs. *Econometrica*, 2295-2326.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2018). *A Practical Introduction to Regression Designs: Volume II.* Cambridge: Cambridge University Press.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations.* Cambridge: Cambridge University Press.

Cattaneo, M. D., Titiunik, R., & Vazquez-Bare, G. (2020). The Regression Discontinuity Design. In L. Curini, & R. Franzese, *Handbook of Research Methods in Political Science and International Relations* (pp. 835-857). Thousand Oaks: Sage Publications.

Coal Burning Responsible for Heavy Air Pollution in India: IEACCC Study. (2021, February 18). Retrieved from International Centre for Sustainable Carbon:

https://www.sustainable-carbon.org/coal-burning-responsible-for-heavy-air-pollution-in-india-ieaccc-study/

Cook, T. D., & Wong, V. C. (2008). Empirical Tests of the Validity of the Regression Discontinuity Design: Implications for its Theory and its Use in Research Practice. *Annals of Economics and Statistics*, 127-150.

December 2021 crude steel production and 2021 global crude steel production totals. (2022, January 25). Retrieved from World Steel Association: https://worldsteel.org/media-centre/press-releases/2022/december-2021-crude-steel-production-and-2021-global-totals/

Dua, A., Mahajan, D., Oyer, L., & Ramaswamy, S. (2020, July 7th). *US small-business recovery after the COVID-19 crisis*. Retrieved from McKinsey & Company: https://www.mckinsey.com/industries/public-and-social-sector/our-insights/us-small-business-recovery-after-the-covid-19-crisis

Evans, S., & Pearce, R. (2020, March 26). *Mapped: The world's coal power plants*. Retrieved from Carbon Brief: https://www.carbonbrief.org/mapped-worlds-coal-power-plants/

Gelman, A., & Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, 447-456.

Gurjar, B., Ravindra, K., & Nagpure, A. S. (2016). Air pollution trends over Indian megacities and their local-to-global implications. *Atmospheric Environment*, 475-495.

He, G., Pan, Y., & Tanaka, T. (2020). The short-term impacts of COVID-19 lockdown on urban air pollution in China. *Nature Sustainability*, 1005-1011.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 615-635.

Imbens, G., & Kalyanaraman, K. (2009). Optimal Bandwidth Choice for the Regression Discontinuity Estimator.

Inhalable Particulate Matter and Health ($PM_{2.5}$ and PM_{10}). (2022). Retrieved from California Air Resources Board: https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health

Iron & Steel Industry in India. (2022, July). Retrieved from India Brand Equity Foundation: https://www.ibef.org/industry/steel

Iron Ore Production by Country 2022. (2022). Retrieved from World Population Review: https://worldpopulationreview.com/country-rankings/iron-ore-production-by-country

Kumar, S., & Managi, S. (2020). Does Stringency of Lockdown Affect Air Quality? *Economics of Disasters and Climate Change*, 481-502.

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 281-355.

Lenzen, M., Li, M., Malik, A. P., Sun, Y.-Y., & Wiedmann, T. e. (2020). Global socio-economic losses and environmental gains from the Coronavirus pandemic. *PLOS ONE*, 1-13.

Liu, F., Wang, M., & Zheng, M. (2021). Effects of COVID-19 lockdown on global air quality and health. *Science of the Total Environment*, 1-13.

Lynch, S. M., & Brown, J. S. (2011). Stratification and Inequality Over the Life Course. In R. H. Binstock, & L. K. George, *Handbook of Aging and the Social Sciences* (pp. 105-117). Cambridge: Academic Press.

Ming, W., Zhou, Z., Ai, H., Bi, H., & Zhong, Y. (2020). COVID-19 and Air Quality: Evidence from China. *Emerging Markets Finance and Trade*, 2422-2442.

NAAQS Table. (2022, 5 2022). Retrieved from United States Environmental Protection Agency: https://www.epa.gov/criteria-air-pollutants/naaqs-table

Nordhaus, W. D. (2016). Revisiting the social cost of carbon. *Earth, Atmospheric, and Planteray Sciences*, 1518-1523.

Overview of India's National Air Quality Index. (2015, May 15). Retrieved from World Air Quality Index Project: https://aqicn.org/faq/2015-05-15/india-national-air-quality-index/

$PM_{2.5}$ Air Pollution. (2017). Retrieved from Our World in Data: https://ourworldindata.org/grapher/pm25-air-pollution

Porter, J. (2003). *Estimation in the Regression Discontinuity Model.* Cambridge.

Raj, D., & Subramanian, S. (2020). India's lockdown: an interim report. *Indian Economic Review*, 31-79.

Rathore, D. S., Nagda, C., Shaktawat, K. T., Chouhan, C. S., Purohit, R., Harish, . . . Jhala, L. S. (2021). COVID-19 lockdown: a boon in boosting the air quality of major Indian Metropolitan Cities. *Aerobiologia (Bologna)*, 79-103.

Ricke, K., Drouet, L., Caldeira, K., & Tavoni, M. (2018). Country-level social cost of carbon. *Nature Climate Change*, 895-900.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., . . . Roser, M. (2020). *COVID-19: Stringency Index*. Retrieved from OurWorldInData: https://ourworldindata.org/covid-stringency-index#learn-more-about-the-data-source-the-oxford-coronavirus-government-response-tracker

Suspended particulates (TSP/SPM). (2020, November 23). Retrieved from European Environment Agency: https://www.eea.europa.eu/publications/2-9167-057-X/page021.html

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 309-317.

Tropospheric ozone. (2022). Retrieved from Climate & Clean Air Coalition: https://www.ccacoalition.org/en/slcps/tropospheric-ozone

Venter, Z. S., Aunan, K., Chowdhur, S., & Lelieveld, J. (2020). COVID-19 lockdowns cause global air pollution declines. *PNAS*, 18984-18990.

Waldhoff, S. T., Anthoff, D., Rose, S., & Tol, R. S. (2011). The Marginal Damage Costs of Different Greenhouse Gases: An Application of FUND. *Social Science Research Network*, 1-33.

Wheeler, D. (1999). *Greening Industry: New Roles for Communities, Markets, and Governments.* Oxford: Oxford University Press.

World Development Report. (2022). Retrieved from World Bank: https://openknowledge.worldbank.org/bitstream/handle/10986/36883/9781464817304_Ch01.pdf

Appendix

Iron ore industry

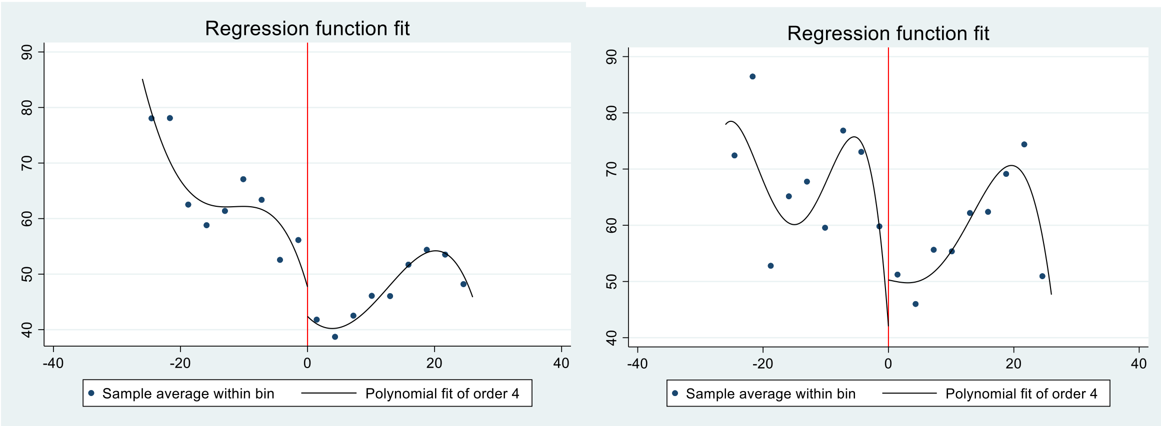

Figure 5.8. PM_{10} regression discontinuity design plot for the iron ore industry sample versus the non-iron ore industry sample, using small bins.

Table 5.6. PM_{10} regression discontinuity estimates with and without covariates in the samples with iron ore industry and without iron ore industry. Covariates include accounting for wind speeds and temperature medians.

	(1)	(2)	(3)	(4)
VARIABLES	PM_{10} median	PM_{10} median	PM_{10} median	PM_{10} median
RD Estimate	-16.99***	-14.33***	-0.992	-0.809
	(6.404)	(5.115)	(10.22)	(8.356)
Covariates	No	Yes	No	Yes
Iron ore industry	Yes	Yes	No	No
Mean of dependent variable in sample	55.318	55.318	63.440	63.440
Observations	365	365	675	675

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

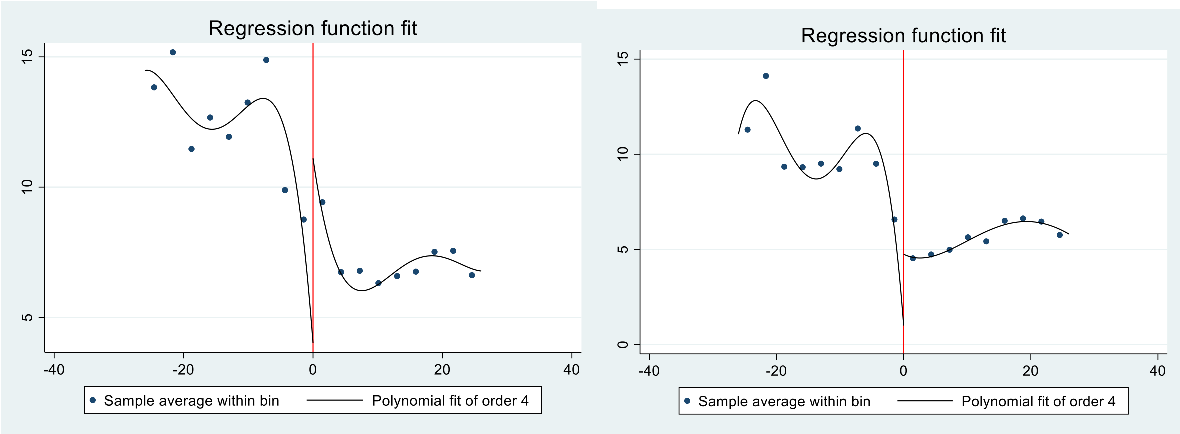

Figure 5.9. NO_2 *regression discontinuity design plot for the iron ore industry sample versus the non-iron ore industry sample, using small bins.*

Table 5.7. NO_2 *regression discontinuity estimates with and without covariates in the samples with iron ore industry and without iron ore industry. Covariates include accounting for wind speeds and temperature medians.*

	(1)	(2)	(3)	(4)
VARIABLES	NO_2 median	NO_2 median	NO_2 median	NO_2 median
RD Estimate	3.550	3.728	0.0692	0.190
	(3.800)	(3.815)	(1.300)	(1.122)
Covariates	No	Yes	No	Yes
Iron ore industry	Yes	Yes	No	No
Mean of dependent variable in sample	9.755	9.755	7.846	7.846
Observations	365	365	741	741

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Steel industry

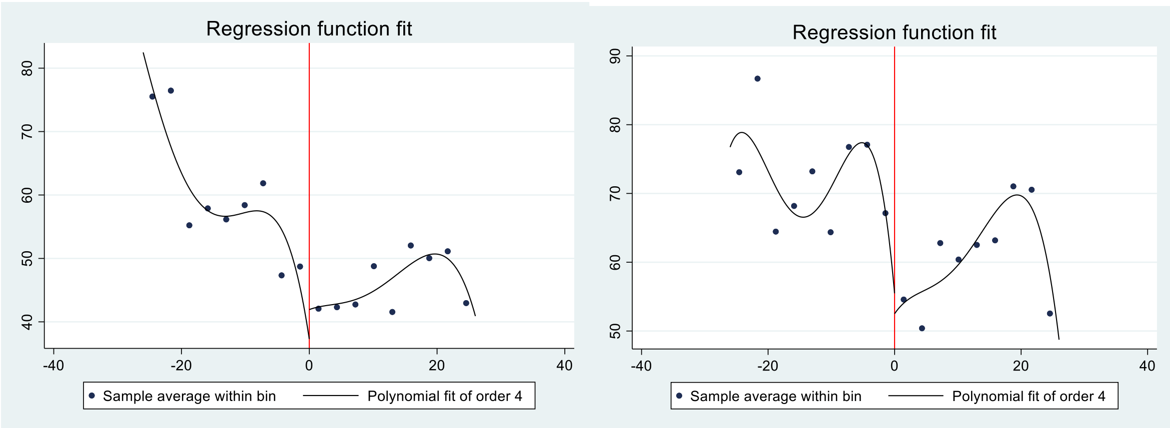

Figure 5.10. PM_{10} regression discontinuity design plot for the above-median steel production sample versus below-median steel production sample, using small bins.

Table 5.8. PM_{10} regression discontinuity estimates with and without covariates in the city samples located in regions with steel production above and below the steel production median. Covariates include accounting for wind speeds and temperature medians.

	(1)	(2)	(3)	(4)
VARIABLES	PM_{10} median	PM_{10} median	PM_{10} median	PM_{10} median
RD Estimate	-3.807	-2.875	-5.803	-6.452
	(7.893)	(6.067)	(8.601)	(8.684)
Covariates	No	Yes	No	Yes
Median steel industry production	Above	Above	Below	Below
Mean of dependent variable in sample	52.810	52.810	69.535	69.535
Observations	458	458	582	582

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

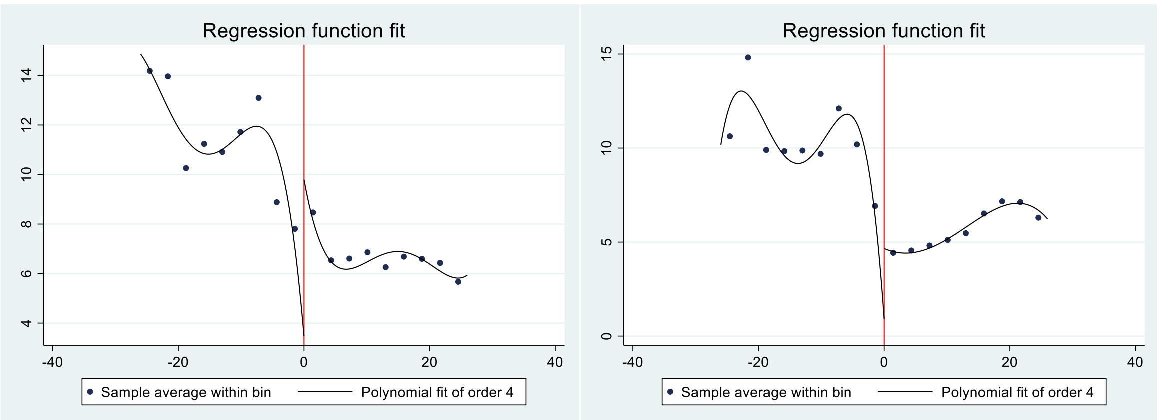

Figure 5.11. NO_2 *regression discontinuity design plot for the above-median steel production sample versus below-median steel production sample, using small bins.*

Table 5.9. NO_2 *regression discontinuity estimates with and without covariates in the city samples located in regions with steel production above and below the steel production median. Covariates include accounting for wind speeds and temperature medians.*

	(1)	(2)	(3)	(4)
VARIABLES	NO_2 median	NO_2 median	NO_2 median	NO_2 median
RD Estimate	3.113	3.454	-0.0833	0.0353
	(3.214)	(3.110)	(1.382)	(1.318)
Covariates	No	Yes	No	Yes
Median steel industry production	Above	Above	Below	Below
Mean of dependent variable in sample	8.985	8.985	9.395	9.395
Observations	471	471	635	635

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Cement industry

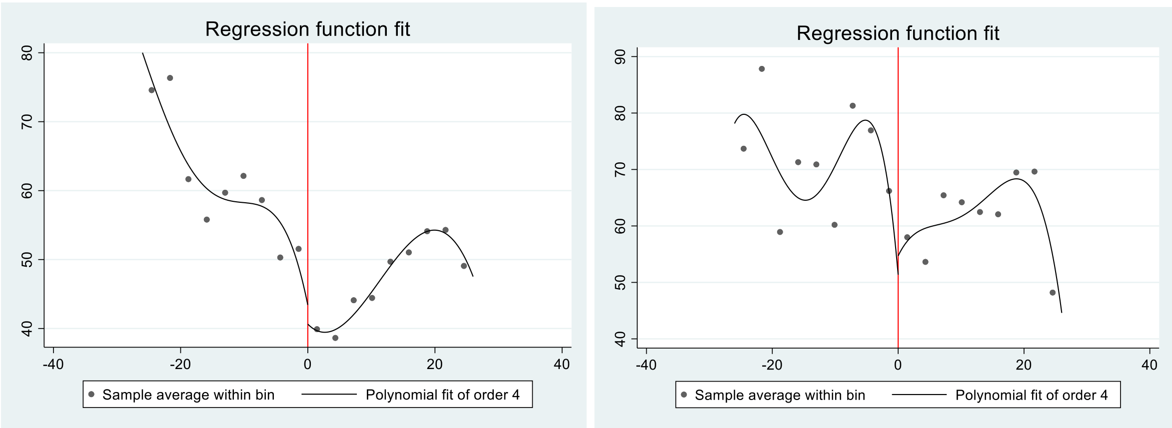

Figure 5.12. PM_{10} regression discontinuity design plot for the above-median cement production sample versus below-median cement production sample, using small bins.

Table 5.10. PM_{10} regression discontinuity estimates with and without covariates in the city samples located in regions with cement production above and below the cement production median. Covariates include accounting for wind speeds and temperature medians.

	(1)	(2)	(3)	(4)
VARIABLES	PM_{10} median	PM_{10} median	PM_{10} median	PM_{10} median
RD Estimate	-10.47	-6.071	-0.701	-3.528
	(7.894)	(5.184)	(10.815)	(10.817)
Covariates	No	Yes	No	Yes
Median cement industry production	Above	Above	Below	Below
Mean of dependent variable in sample				
Observations	511	511	423	423

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

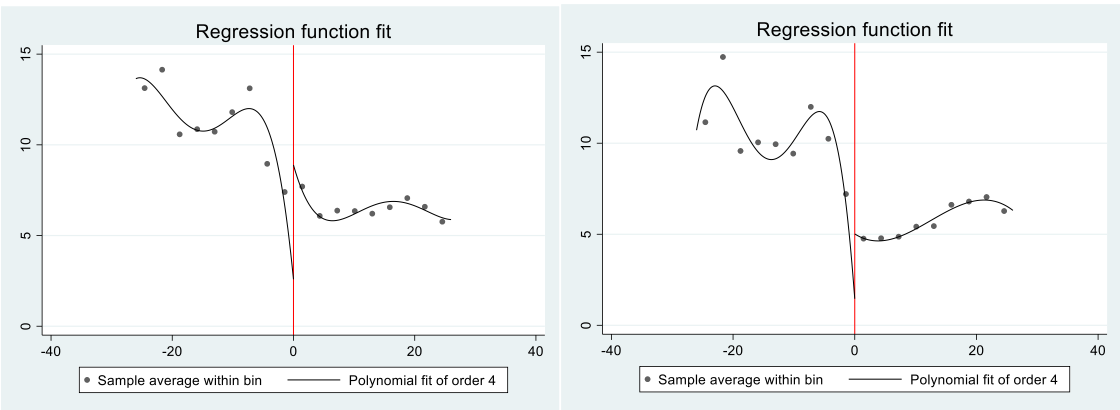

Figure 5.13. NO_2 *regression discontinuity design plot for the above-median cement production sample versus below-median cement production sample, using small bins.*

Table 5.11. NO_2 *regression discontinuity estimates with and without covariates in the city samples located in regions with cement production above and below the cement production median. Covariates include accounting for wind speeds and temperature medians.*

	(1)	(2)	(3)	(4)
VARIABLES	NO_2 median	NO_2 median	NO_2 median	NO_2 median
RD Estimate	2.773	2.983	-0.007	-0.468
	(2.980)	(2.916)	(1.555)	(1.383)
Covariates	No	Yes	No	Yes
Median cement industry production	Above	Above	Below	Below
Mean of dependent variable in sample				
Observations	524	524	476	476

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Coal power stations

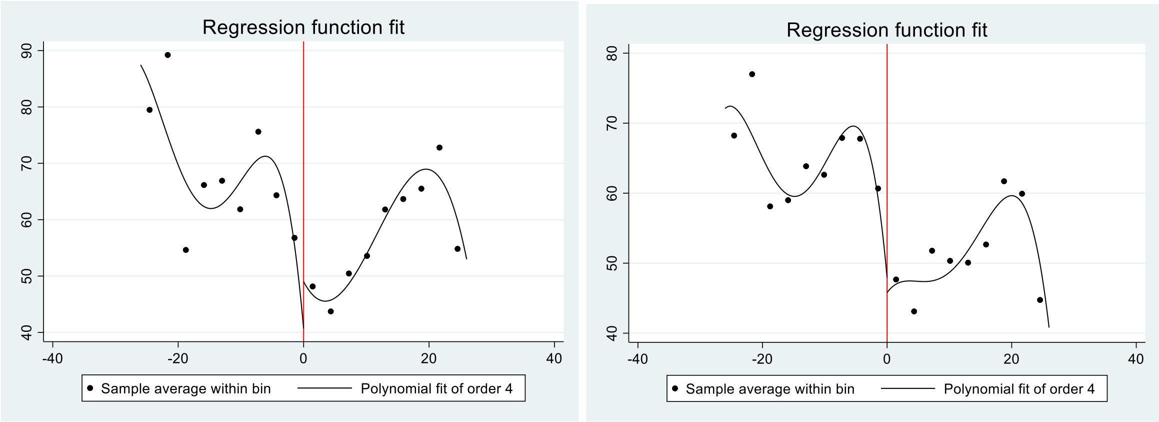

Figure 5.14. PM_{10} regression discontinuity design plot for the coal power sample versus non-coal power sample, using small bins.

Table 5.12. PM_{10} regression discontinuity estimates with and without covariates in the coal power sample and non-coal power sample. Covariates include accounting for wind speeds and temperature medians.

	(1)	(2)	(3)	(4)
VARIABLES	PM_{10} median	PM_{10} median	PM_{10} median	PM_{10} median
RD Estimate	-3.646	-1.539	-5.905	-6.420
	(9.421)	(7.494)	(6.763)	(6.983)
Covariates	No	Yes	No	Yes
Coal power	Yes	Yes	No	No
Mean of dependent variable in sample	62.684	62.684	58.107	58.107
Observations	564	564	476	476

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

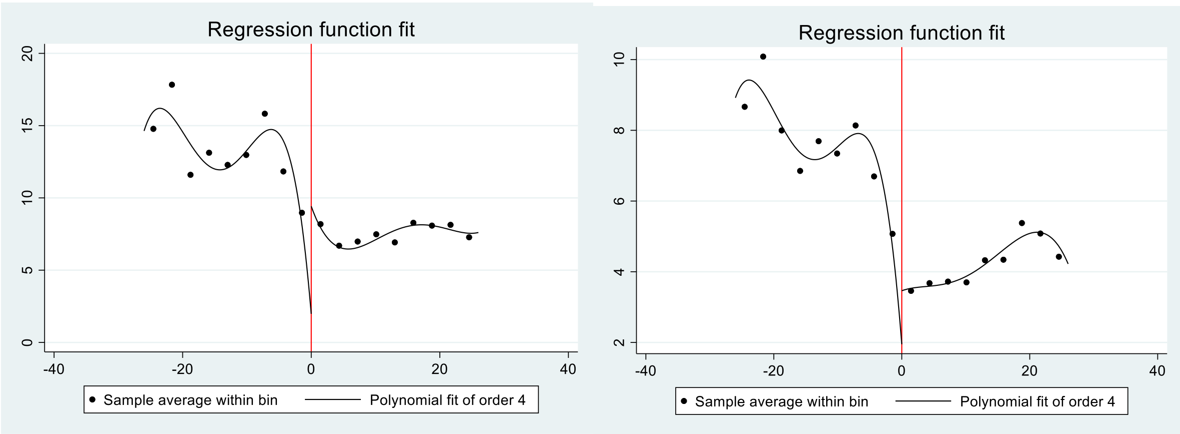

Figure 5.15. NO_2 regression discontinuity design plot for the coal power sample versus non-coal power sample, using small bins.

Table 5.13. NO_2 regression discontinuity estimates with and without covariates in the coal power sample and non-coal power sample. Covariates include accounting for wind speeds and temperature medians.

	(1)	(2)	(3)	(4)
VARIABLES	NO_2 median	NO_2 median	NO_2 median	NO_2 median
RD Estimate	2.329	2.580	-0.361	-0.227
	(2.687)	(2.649)	(0.906)	(0.786)
Covariates	No	Yes	No	Yes
Coal power	Yes	Yes	No	No
Mean of dependent variable in sample	10.396	10.396	5.936	5.936
Observations	630	630	476	476

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1