IDB WORKING PAPER SERIES Nº IDB-WP-1398

Estimating the effect of urban road congestion on air quality in Latin America

Felipe Bedoya-Maya
Agustina Calatayud
Vileydy González-Mejía

Inter-American Development Bank
Transport Division

October 2022

Estimating the effect of urban road congestion on air quality in Latin America

Felipe Bedoya-Maya
Agustina Calatayud
Vileydy González-Mejía

October 2022

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Bedoya-Maya, Felipe.
Estimating the effect of urban road congestion on air quality in Latin America / Felipe Bedoya-Maya, Agustina Calatayud, Vileydy González-Mejía.
p. cm. — (IDB Working Paper Series ; 1398)
Includes bibliographic references.
1. Traffic congestion-Environmental aspects-Latin America. 2. Urban transportation-Latin America-Econometrics models. 3. Air quality-Latin America-Econometric models. I. Calatayud, Agustina. II. González, Vileydy. III. Inter-American Development Bank. Transport Division. IV. Title. V. Series.
IDB-WP-1398

http://www.iadb.org

Abstract[1]

Road congestion and air pollution are key challenges for quality of life in urban settings. This research leverages highly disaggregated crowdsourced data from Latin America to study the effect of road congestion on levels of carbon monoxide, nitrogen dioxide, and particulate matter in four of the most congested cities in developing countries: Bogota, Buenos Aires, Mexico City, and Santiago. Based on a panel data econometric approach with over 4.4 billion records from Waze and hourly data from 54 air monitoring stations for 2019, our two-stage least square model shows a cumulative increase of 0.6% in response to a 1% of road congestion on the three air pollutants. Moreover, we find a nonlinear relationship between road congestion and air quality and estimate the threshold above which the effect decays. This study provides evidence that supports public policies designed to make urban mobility more sustainable by implementing measures to reduce road congestion in developing contexts.

Keywords: road congestion, air quality, urban mobility, sustainability, developing countries, Latin America.

[1] The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.
Declarations of interest: none.

Introduction

Air pollution is a critical global issue affecting physical and mental health, well-being, and mortality (Heft-Neal *et al.*, 2018). Several respiratory illnesses and cerebrovascular and cardiovascular diseases have been linked to high levels of particulate matter in the air (Burnett *et al.*, 2014). Globally, over 3 million premature deaths per year can be attributed to outdoor pollution. Particulate matter from combustion emissions is linked to around 200,000 premature deaths per year, and 53,000 of these deaths are associated with road transportation emissions (Lelieveld *et al.*, 2015).

Indeed, road traffic is one of the most critical contributors to high CO_2 and NO_2 levels in the air. These levels are linked to health problems affecting the lungs, eyes, larynx, and pharynx (Barr Rosso, 2021). This pollution is a major challenge in global urbanization processes, which are driving higher volume and more geographically concentrated road traffic. These urbanization processes bring positive benefits from agglomeration economies, including more accessible employment opportunities and health services, but they also entail negative externalities such as road congestion and pollution (Higgins *et al.*, 2019).

Latin America and the Caribbean is the most urbanized region in the world (IDB, 2021). About 80% of its population lives in cities, and two-thirds of people in the region live in cities with 20,000 people or more. These levels of urbanization are much higher than the average for industrialized countries (ECLAC, 2012). If this exponential urbanization trend continues, a projected 100 million people will live in either Mexico City, Sao Paulo, Buenos Aires, Rio de Janeiro, Bogota, or Lima by 2025 (IDB, 2021). The rapid pace of urbanization brings significant challenges for cities in the region, such as housing shortages, road congestion, decreasing air quality, and lower life expectancy (ECLAC, 2012). Indeed, higher levels of air pollution can shorten life expectancy by an average of 1.8 years (Barr Rosso, 2021). In a study that included 652 cities in Latin America and the Caribbean, the Economic Commission for Latin America and the Caribbean (ECLAC, 2021a) found that an average increase of 10 $\mu g/m^3$ of PM_{10} and $PM_{2.5}$ was associated with increases in mortality of 0.4% and 0.7%, respectively.

Road congestion is a growing issue for many cities in Latin America and the Caribbean. The region holds eight of the world's thirty most congested cities: Bogota (number 3 in the TomTom ranking), Lima (7), Mexico City (13), Recife (15), Rio de Janeiro (20), Sao Paulo (24), Santiago (26), and Salvador (28) (TomTom, 2020). Cities bear direct and indirect costs from this congestion. For example, in 2019 the monetary cost of road congestion in Buenos Aires and Mexico City, measured as the value of time lost in traffic delays, was three times what these cities spent on education (Calatayud *et al.*, 2021). According to Gómez-Lobo et al. (2022), a 5% increase in road congestion would reduce the GDP of cities in the region by 0.5%. Congestion is also linked to road accidents: Sánchez González et al. (2021) estimated that a 10% drop in road congestion could decrease road accidents by 3.4%.

The potential post-pandemic increase in private vehicle usage is raising further concerns about congestion and its negative externalities. In a global survey, the number of respondents who said having a private vehicle was very important to them increased by 12 percentage points from pre-pandemic levels (Berger, 2020). In Mexico and Brazil, for example, 58% of respondents said the pandemic caused them to reconsider the number of vehicles their household needs (Ford, 2021).

In this context and given the pressing need to transition to a more sustainable mobility model and reduce mobility's impact on pollution and quality of life, it is important to ascertain the extent to which congestion is related to deteriorating air quality.

To help close this knowledge gap, this paper leverages over 4.4 billion crowdsourced records and hourly data from 54 air monitoring stations to estimate panel-data econometric models and discuss the effect of road congestion on CO, NO_2, and PM_{10} in four Latin American cities that are among the world's most congested: Bogota (Colombia), Buenos Aires (Argentina), Mexico City (Mexico), and Santiago (Chile). These are among the six major air pollutants regulated in the United States (EPA, 2022), and are generated considerably by circulating vehicles but have not been widely researched regarding its relationship with road congestion. Moreover, available studies have valid results only for specific settings. Thus, the research also explores the temporal dynamics of congestion on each air pollutant, including its autoregressive behavior and cumulative effects. Finally, to provide results for a broader list of air pollutants ($PM_{2.5}$, NO, and NOx) and explore heterogeneities within urban areas, we take a closer look at the case of Bogota and analyze air quality patterns by geographical area of the city.

The study aims to help policymakers in emerging economies — especially in highly urbanized and congested cities in Latin America — design and implement measures to reduce road congestion. In addition, since transportation is one of the biggest contributors to global warming due to its use of fossil fuels, this research aims to help achieve a more sustainable mobility model, in line with the Paris Agreement's goal of limiting global warming to less than 1.5°C.

This paper is organized as follows: Section 1 reviews the available literature on the relationship between road traffic and air quality. Section 2 explains the theoretical framework for the analysis. Section 3 presents the data and methods we applied. Section 4 presents the results from the econometric approach. Section 5 discusses the results and policy implications, and Section 6 contains the study's conclusions.

1. Literature Review

The literature shows that changes in the scale of cities lead to longer travel times and shifts in modal preferences, generally from walking or using public transportation to increased dependency on private vehicles. These changes usually worsen air quality because of higher aggregate fuel consumption (Lu *et al.*, 2021). The mechanism is as follows: higher vehicle density that exceeds road capacity is followed by congestion, which generates travel delays and increases energy demand (Zhang and Kockelman, 2016). This in turn means more vehicle emissions and higher concentrations of pollution. Worse air quality from congestion poses a significant health risk for a city's inhabitants, shortening life expectancy (Zhanga and Batterman, 2013).

Most existing evidence on the relationship between congestion and pollution is for advanced economies. For instance, Zhang and Batterman (2013) conducted two controlled case studies — one on an interstate highway of Michigan and another on an arterial road in Detroit, USA — to assess how emissions due to road congestion affected emergency doctor visits, hospital admissions, and mortality attributed to NO2 exposure. Their findings suggest a differentiated effect: for freeways, a U-shaped risk factor was reported for on-road populations; for arterial roads, the risk increased sharply for both on-road and near-road populations. Requia et al. (2018) implemented a Stochastic User Equilibrium Traffic Assignment Algorithm with data from a travel survey in Toronto, Canada to estimate emissions as related to congestion. They then computed mortality rates linked to PM2.5. Their results indicate a substantial effect on mortality, especially during morning rush hours.

Lack of data has often been a major constraint for conducting such analyses in developing contexts. However, developing contexts are precisely where the highest levels of urban congestion are observed: 29 of the world's 50 most congested cities are located in developing countries (TomTom, 2022). But recently novel data sources such as environmental monitoring stations, satellite images, and digital traffic-monitoring platforms have allowed researchers to start bridging this gap. For example, using data from one environmental monitoring zone, Heger et al. (2019) studied the relationship between car density and PM_{10} concentrations on a road segment in Cairo in 2016 and 2017. Applying a panel model with fixed effects, they found that PM_{10} concentration increased by 8.6 μg/m³ per 100 cars within the range of the monitoring station. Soleimani et al. (2022) leveraged hourly data from sixteen monitoring stations in Isfahan (Iran) to conclude that vehicle traffic was responsible for 14% of changes in $PM_{2.5}$ between 2018 and 2020.

Furthermore, Dasgupta et al. (2021) combined satellite-based traffic and pollution data with meteorological information (temperature, humidity, and wind speed and direction) and socioeconomic data to investigate the temporal dynamics of vehicle congestion and pollution in Tanzania. They found wind speed to be a critical in explaining the intensity of air pollution from vehicle traffic. Mishra et al. (2019) used data from online platforms such as Google Maps to identify the areas with the highest emissions from road congestion in Delhi (India).

Despite these key advances in the literature, evidence on the causal relationship between road congestion and emissions is still very limited. Lu et al. (2017) used PM_{10} data from the daily air pollution index reported by local authorities in Beijing and $PM_{2.5}$ data from a single monitoring station to analyze the relationship between driving to school, congestion, and emissions around schools. Applying a two-stage least squares regression model using school vacations as exogenous shocks in traffic congestion, the authors found that the 20% drop in traffic congestion during school vacations reduced daily average PM_{10} emissions by 12%. Nonetheless, as recognized by the authors, their estimates are constrained by city-level aggregated data and a small sample size.

Evidence is almost nonexistent In Latin America, and available studies have analyzed indirectly the relationship between air quality and road congestion. In Santiago, Chile, the effect of temporary driving bans derived from alerts on the city's air quality between 2000 and 2015 was evaluated. Rivera (2021) implemented a regression discontinuity design, finding evidence that temporary reductions in driving reduced car trips by 6-9% in peak hours and 7-8% in off-peak hours, which translates into a reduction of air pollution in peak hours, a reduction of congestion - given the decrease in car trips, and the increase of the city's mass transit system.

Pachon et al. (2021) analyzed the factors influencing dust loading on the urban center of Bogota. Specifically, the resuspended particulate matter particles of less than 10 micrometers (RPM10) were found to be influenced by four factors: road characteristics, driving speeds, land use, and meteorological conditions. For the development of the study, the use of a road dust sampler was implemented in which 41 samples collected in different points of the city were collected. Results suggest that the lowest RPM10 values are present under conditions of fast driving speed, absent infrastructure constructions, and high vegetation; on the contrary, RPM10 concentrations increased under heavy-duty traffic conditions. For the same city, Zhang et al. (2017) implement a regression discontinuity design (RDD) and found empirical evidence that license plate-based driving restrictions can cause diverse effects on air pollutants, specifically it can generate a significant decrease in NO but an increase in NO_2, NO_x and O_3.

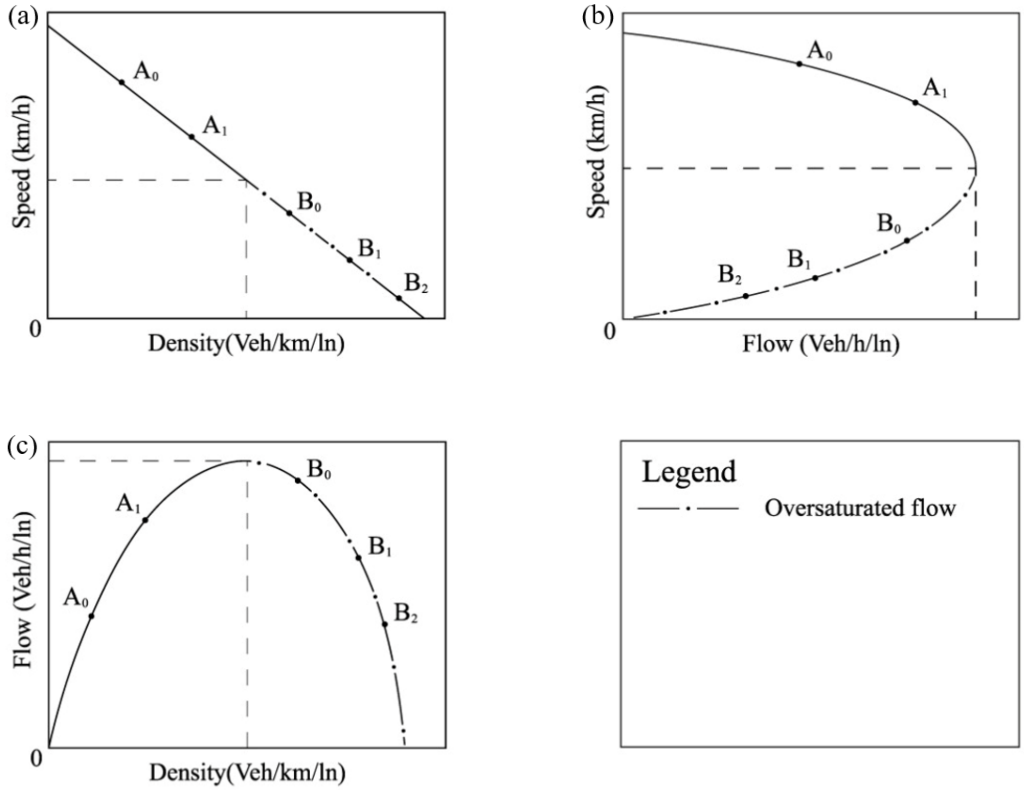
Due to the lack of analysis focused on the causal relationship between road congestion and pollution in developing contexts including Latin America, policymakers have limited ability to properly estimate the negative externalities of congestion, prioritize congestion mitigation in both transportation and health agendas, and implement policies that effectively address this problem in large and highly congested cities. To close this gap, this study uses highly disaggregated big data on road traffic and air quality data from 54 air monitoring stations to estimate the relationship between urban road congestion and air quality in four of the most congested cities in developing countries. It then analyzes the behavior of this relationship across such cities. The granularity of crowdsourced data allows us to rule out fixed and seasonal effects at air monitoring stations across multiple cities and regions, thus yielding results with higher external validity than other available evidence. Furthermore, this research provides evidence on the nonlinear form in which road congestion impacts air quality and explores and accounts for the autoregressive behavior of congestion on air quality.

2. Theoretical framework

According to the Macroscopic Fundamental Diagram of Traffic on an urban scale, the relationship between vehicular density and speed is negative and linear in both non-saturated and oversaturated flow situations (Geroliminis and Daganzo, 2008). However, the relationship between vehicular flow and speed may be nonlinear, reaching a maximum when oversaturated flow begins (**Figure 1**). These dynamics have several repercussions for how congestion can increase vehicles' fuel consumption, and consequently their emissions that affect air quality.

Consider scenario A, where congestion — measured as the number of hours lost in traffic — increases from point A_0 to A_1 in **Figure 1**. According to the Greenshields Macroscopic Fundamental Diagram of Traffic, congestion is directly linked to higher vehicle density (veh/km) and, therefore, to lower speed (km/hour) (Geroliminis and Daganzo, 2008). Because this change occurs in the non-saturated flow section (when there is no road congestion), which is represented by continuous segments in **Figure 1**, the drop in speed occurs while flow (veh/hour) continues to increase. But this increase in flow happens at decreasing rates. In other words, a higher number of vehicles limits the space available for other vehicles to enter a specific lane. With a higher number of vehicles per kilometer per hour, fuel consumption (l/km) is expected to be higher (**Figure 2 [d]**). These dynamics are supported by the findings of Shabihkhani and Gonzales (2015) on the link between density and emissions, as well as the conclusions of the NCHRP (2005) regarding the relationship between flow and emissions.

Figure 1 Traffic Model

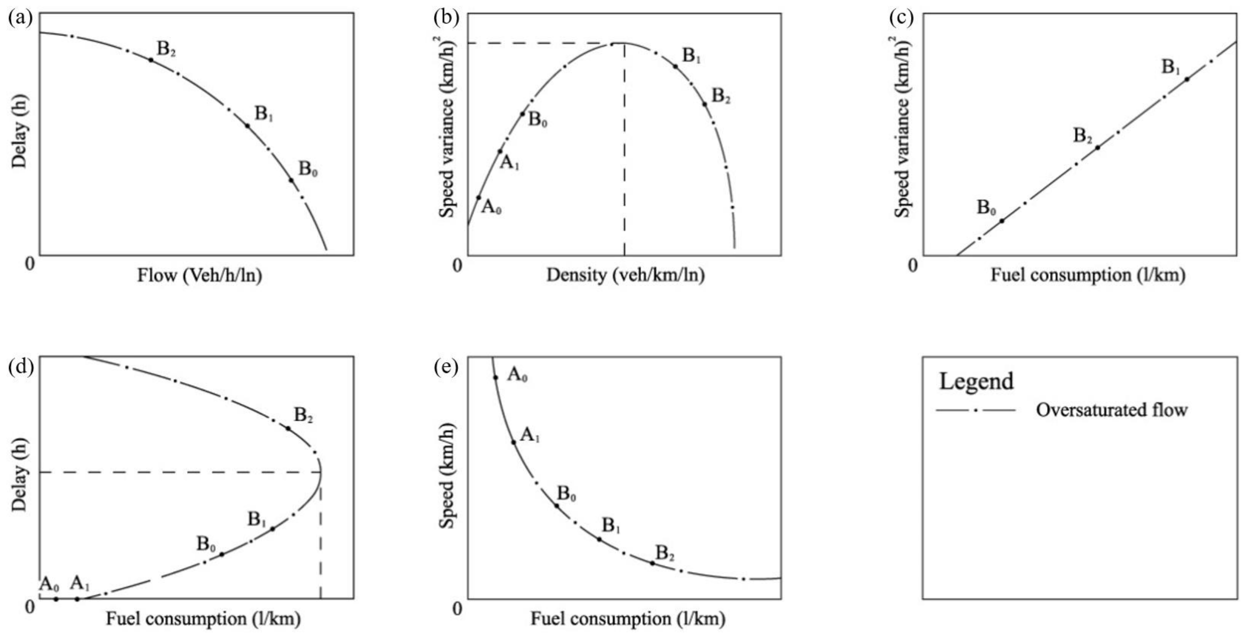


Source: Prepared by the authors based on Geroliminis and Daganzo (2008).

Now, consider scenario B, with an oversaturated flow in which congestion increases from point B_0 to B_1 and B_2 (represented by the discontinuous sections in **Figures 1 and 2**). Although density increases, the drop in speed has a different relationship to vehicular flow. Since the road is already congested, a decrease in speed is accompanied by a drop in vehicular flow. When flow is oversaturated, changes in speed will have considerable influence on the direction of the relationship between congestion and energy demand. Zhou et al. (2023) show that the relationship between speed and fuel consumption is negative, but the magnitude of the change is different in low-speed ranges versus high-speed ranges. Greater changes in fuel consumption are reported during low-speed levels.

Furthermore, speed variability influences the level of fuel consumption (Zhou *et al.*, 2023b). In scenario A and during the initial stages of scenario B, after an increase in vehicle density, speed variance is expected to increase as vehicles move from optimal free-flow speeds, i.e., the maximum point in the relationship between speed and flow, to traffic jam speeds. The advanced stages of the oversaturated flow scenario are different, as vehicles are limited to moving within a considerably lower range of speeds. This situation is represented in the segment after the maximum point is reached in **Figure 2 (b)**. Indeed, speed and flow are both zero in a fully saturated traffic jam. Consequently, in advanced stages of the oversaturated flow scenario, the combination of lower flow and speed variability means there is relatively lower aggregate demand for fuel (**Figure 2 [d]**). This study will empirically assess this nonlinear form by leveraging the highly disaggregated data collected for the four cities being analyzed.

Figure 2 Theoretical relationship between vehicular density, delay, and fuel consumption



Source: Prepared by the authors based on Greenshields (1935), NCHRP (2005), Shabihkhani and Gonzales (2015), and Zhang and Kockelman (2016).

3. Data and Methods

To assess the impact of road congestion on air quality in Latin America and the Caribbean, we collected traffic data for the urban areas of Bogota, Buenos Aires, Mexico City, and Santiago. The boundaries of the study area are outlined by the black polygons in **Figure 3**. TomTom (2020) ranks these cities as 3rd, 66th, 13th and 26th on its list of the most congested cities in the world. We retrieved over 4.4 billion road segment records for these areas from Waze, a mobile navigation app with a high adoption rate in the four cities (Waze, 2019). This data covers all of hours of 2019 (from 00:00 January 1 to 23:59 December 31), which is the year before mobility was severely restricted to contain the COVID-19 pandemic.

Waze has proven to be a valid source of information with enough coverage and precision for conducting statistical inference studies. For instance, Amin-Naseri et al. (2018) compared one year of Waze data with the recorded incidents in Iowa's advanced traffic management system (ATMS) in the same timeframe. The authors indicated that the crowdsourced data stream from Waze is an invaluable source of information with broad coverage of crash and congestion reports, on average 9.8 minutes earlier than a probe-based alternative and achieving a reasonable geographic accuracy. Also, Hoseinzadeh et al., (2020) assessed the quality of Waze data and conducted a case study in Tennessee, USA; the authors replicated speed and jam length remarkably. Furthermore, Goodall & Lee (2019) evaluated the accuracy of Waze data by observing a 2.7-mile corridor on a major urban freeway in Virginia, USA, finding that it replicated the information collected by other means, such as traffic detection cameras and police reports.
Indeed, this mobile navigation application with over 130 million monthly active users has been used to measure the impact of sports megaevents on traffic (Xu and González, 2017), analyze road safety (Goodall and Lee, 2019), repositioning police agents (Fire *et al.*, 2012), and estimate vehicle speed on urban corridors (Nune *et al.*, 2017). Its validity has allowed completing research in developing studies where traditional sensors and cameras are usually not available.

Waze provides two types of data (Calatayud *et al.*, 2021):

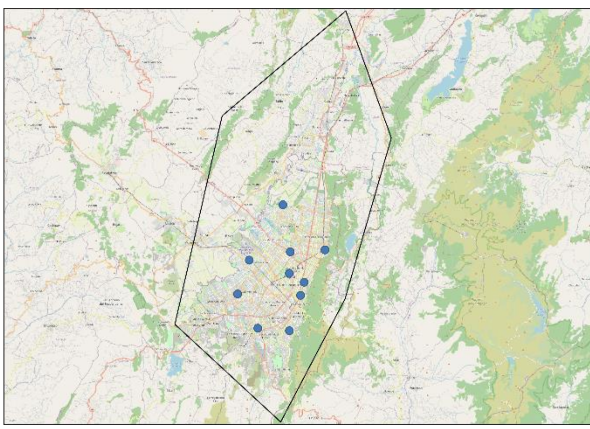
(i) **Alerts:** Waze users report alerts based on what they notice on the road. Once an alert is reported, other users validate it by using the app to report whether the alert still applies. Based on the information received from users, Waze calculates a reliability factor of 1 to 10, where 10 is the most reliable. Users can report three kinds of alerts: *(i)* Accident, which is any type of collision; *(ii)* Hazard, for reporting stranded vehicles or objects on the road, adverse weather conditions, and floods, among other events; and *(iii)* Road Closed, for lane closures due to demonstrations, events, maintenance, or other reasons.

(ii) **Jams:** Waze actively builds this dataset using smartphone GPS signals. When the API identifies a significant group of vehicles moving at an irregularly variable speed, it classifies this as a traffic jam, in contrast to free-flow speed. Waze collects information on average speed, expected delay from the traffic jam compared to free-flow conditions, geographical coordinates, and traffic jam length for each traffic jam. Information on road status is updated every 2 minutes.

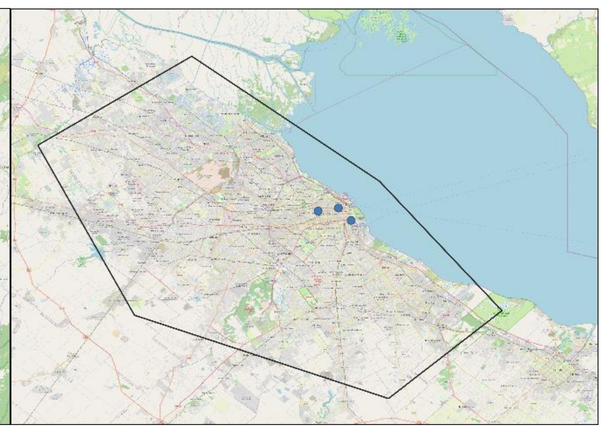
We use the definition of congestion from Goodwin (2004) *"…the impedance vehicles impose on each other, due to the speed-flow relationship, in conditions where the use of a transport system* *approaches its capacity."* Accordingly, we estimate congestion as the extra travel time experienced by road users due to an excess of vehicles on a portion of the road at a specific time that results in slower-than-normal speeds. Put differently, congestion is measured as the total number of hours vehicle occupants lost in 2019 due to traffic delays in the areas within the polygons. We used the methodology developed by Calatayud et al. (2021) to calculate congestion in those areas using Waze data.

As depicted in **Table 1**, Mexico City and Bogota are the cities with the highest number of traffic jams and alerts recorded. They also reported the highest average jam duration in 2019. Consequently, although Buenos Aires had almost half as many records of traffic jams as Bogota, it is only second to Mexico City in number of unique traffic jams. For all variables considered in this study, Mexico City has the highest values of the four cities, while Santiago reports the lowest values, especially in the number of unique jams, but also for average jam duration and number of alerts.

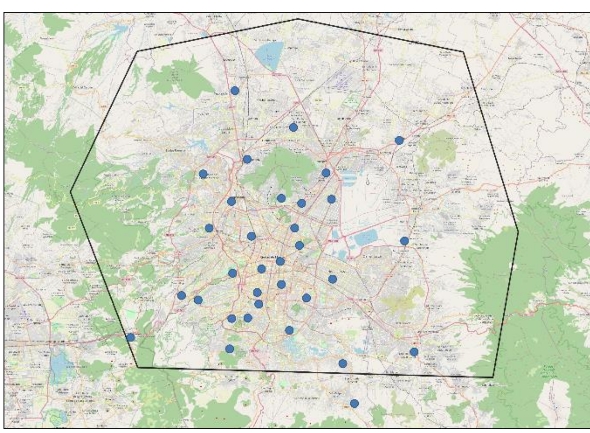
Figure 3 Study areas



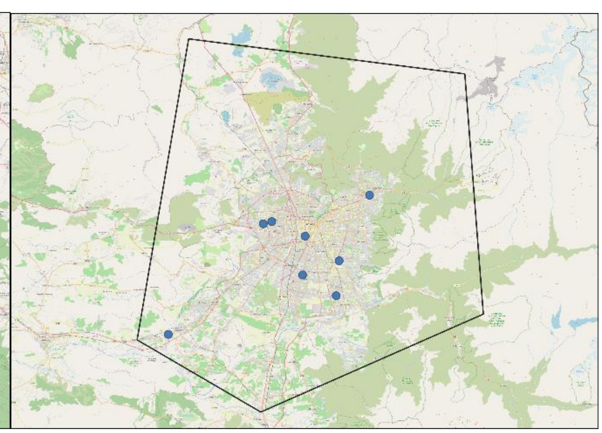
(a) Bogota



(b) Buenos Aires



(c) Mexico City



(d) Santiago

Source: Prepared by the authors.

Note: The black polygons denote the area where road congestion data was collected. The blue dots indicate the location of air quality monitoring stations.

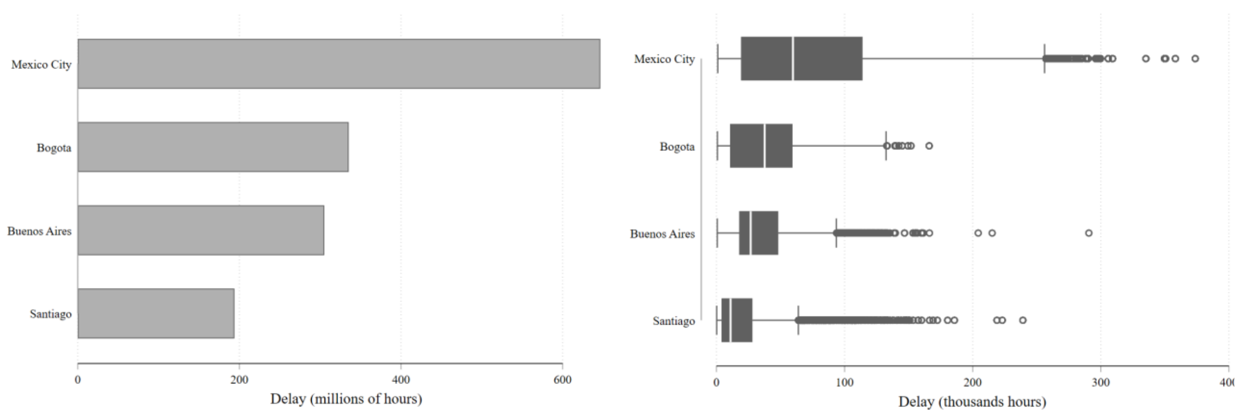
Table 1 Descriptive statistics for road congestion (2019)

City	Records of traffic jams (millions)	Unique jams (millions)	Average jam duration (minutes)	Alerts recorded (millions)
Bogota	1,101	17.8	20.9	13.8
Buenos Aires	634	21.5	15.4	11.0
Mexico City	2,162	37.1	23.1	14.9
Santiago	576	11.8	12.9	7.6

Source: Prepared by the authors from Waze.

In 2019, Mexico City had the highest traffic delays among the four cities analyzed, reaching over 600 million hours lost in traffic, followed by Bogota and Buenos Aires, with over 300 million each. The lowest congestion levels were in Santiago, with under 200 million hours lost in traffic (**Figure 4**). In per-capita terms, during that year each inhabitant lost 31 hours in Bogota, 30 hours in Mexico City, 29 hours in Santiago, and 20 hours in Buenos Aires due to traffic delays. Variability levels are correlated with total delay. The median delay per hour in Mexico City was 59,700 hours, with a standard deviation of 60.5 and a maximum of 373,700 hours. Bogota had fewer outliers in the sample and a median delay of 37,600 hours. While Santiago had the lowest median number of hours lost in traffic (27,200 hours), its standard deviation was higher than that of Bogota and similar to that of Buenos Aires.

Figure 4 Hours lost due to road congestion and its per-hour variability in 2019



(a) Total road congestion

(b) Distribution of road congestion

Source: Prepared by the authors with data from Waze.

To measure air quality, we collected data on carbon monoxide (CO in particles per million, ppm), nitrogen dioxide (NO_2 in particles per billion, ppb), and particulate matter with a diameter of 10 micrometers or less per cubic meter (PM_{10} in µg/m³) from the field monitoring stations in each of the four cities. We did not consider the monitoring stations that do not provide data for prolonged periods and make it inviable to estimate the models. Furthermore, for those short periods of non-reporting data because of station maintenance of related activities, we followed the consideration by Cameron & Trivedi (2005) regarding missing data completely at random (MCAR) to guarantee this type of event does not bias the analysis. The blue dots in **Figure 3** show the geographical distribution of the monitoring stations. Our dataset contains hourly emissions data recorded by each station during 2019. The highly disaggregated data offers an advantage over previous studies, since we were able to account for fixed effects between and within cities, isolating the effect of congestion on emissions while controlling for other factors that may explain different emissions levels.

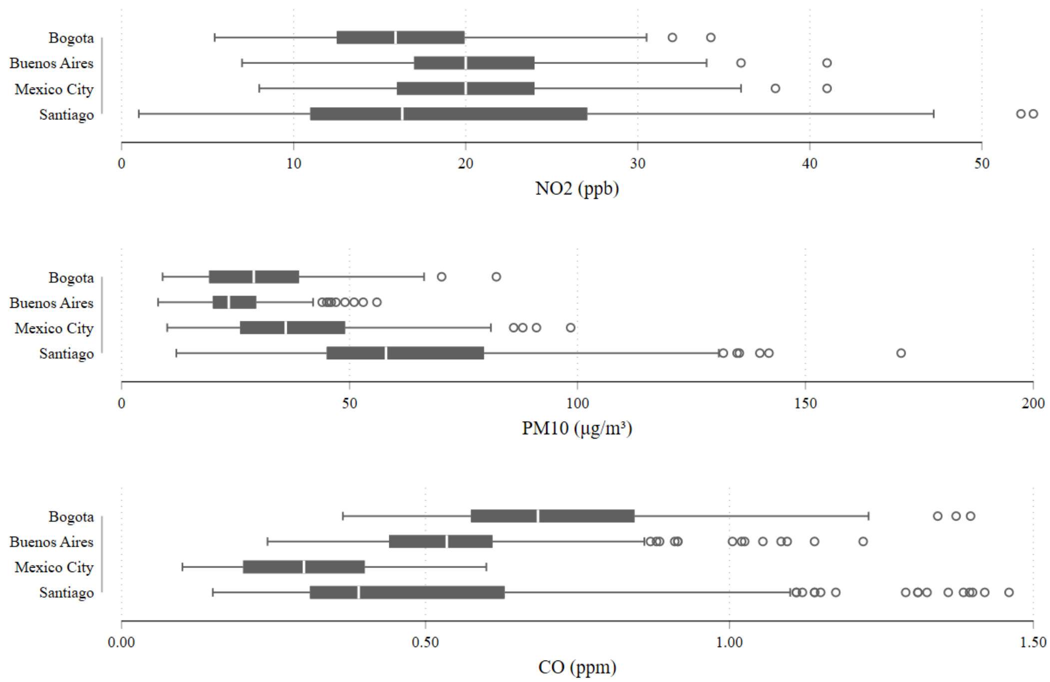
In 2019, Buenos Aires and Mexico City had the highest hourly median level of $\mathrm{NO_2}$, at 20 ppb. Santiago had the highest standard deviation in values of $\mathrm{NO_2}$ at 25 ppm, as well as the highest value, at 565 ppm. When looking at the daily median level in Santiago, the standard deviation was 15 ppm, and the maximum was 50 ppm, as shown in **Figure 5**. For $\mathrm{PM_{10}}$, Santiago reported the highest hourly median value with almost 60 µg/m3, as well as the highest variability and maximum values. Mexico City, Bogota, and Buenos Aires followed, reporting daily median values of under 50 µg/m3. Finally, with nearly 0.70 ppm, Bogota reported the highest hourly median value of CO. However, its standard deviation was lower than Santiago's (0.56 vs. 0.75 ppm, respectively). Mexico City reported the lowest hourly median value of CO among the four cities (0.30 ppm) (**Table 2**).

Table 2 Air quality in study areas in 2019 (in ppm for CO, ppb for NO_2, and µg/m³ for PM_{10})

City	Stations	Contaminant	Mean	S.D.	Min	Median	Max
Bogota	10	$\mathrm{NO_2}$	17.15	9.55	-0.05	16.01	83.39
		CO	0.83	0.56	-0.05	0.69	5.43
		$\mathrm{PM_{10}}$	34.26	23.83	0.00	29.00	385.00
Buenos Aires	3	$\mathrm{NO_2}$	20.76	9.33	0.00	20.00	89.00
		CO	0.58	0.27	0.00	0.53	4.36
		$\mathrm{PM_{10}}$	25.73	9.61	3.00	24.00	86.00
Mexico City	33	$\mathrm{NO_2}$	21.78	12.91	0.00	20.00	109.00
		CO	0.38	0.36	0.00	0.30	5.20
		$\mathrm{PM_{10}}$	43.52	28.20	1.00	38.00	653.00
Santiago	8	$\mathrm{NO_2}$	21.86	24.97	0.00	17.50	565.17
		CO	0.65	0.75	0.05	0.43	10.99
		$\mathrm{PM_{10}}$	69.68	48.92	0.00	58.00	983.00

Source: Prepared by the authors.

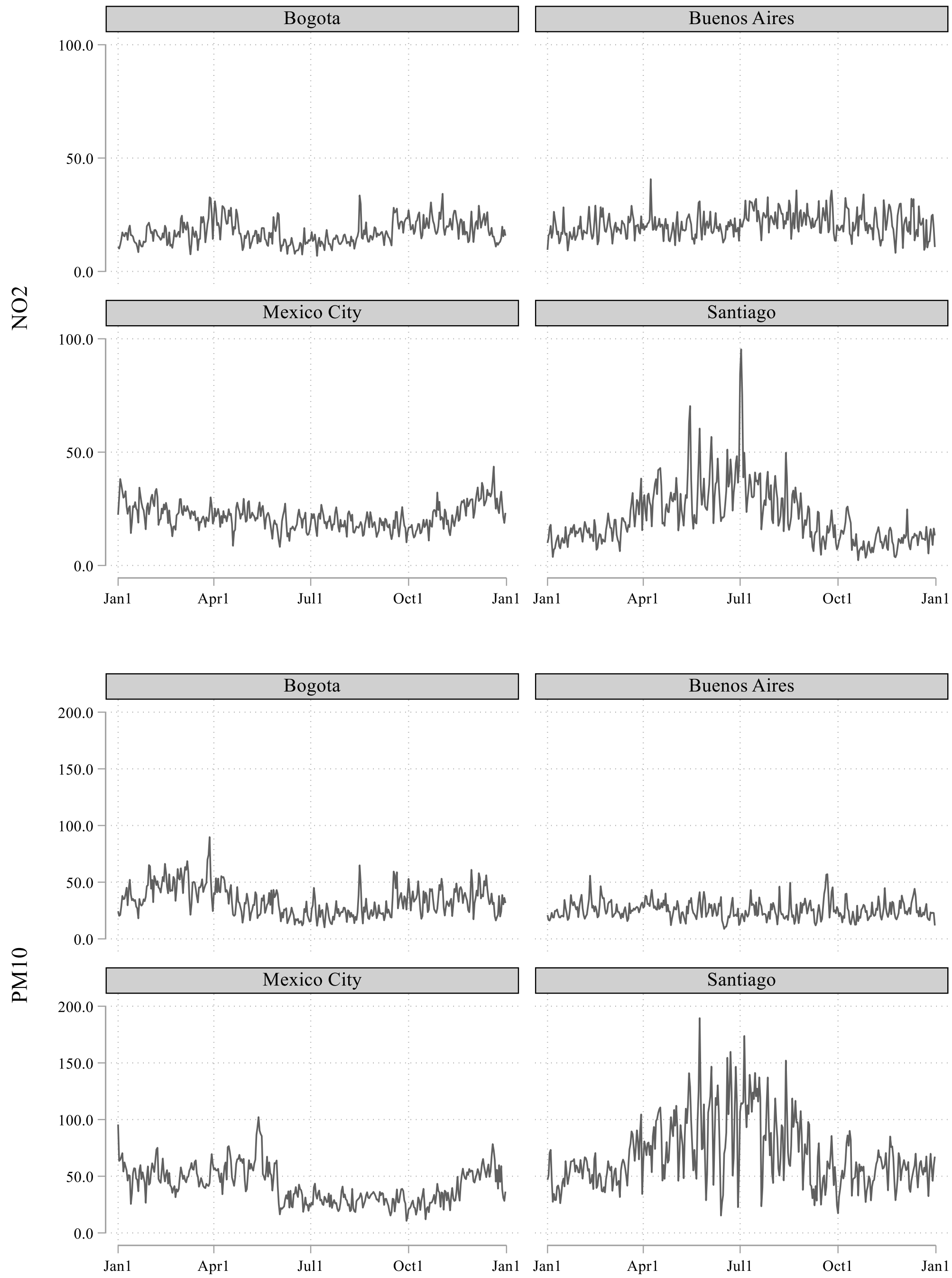
Figure 5 Distribution of daily median air quality, by city (2019)



Source: Prepared by the authors.

Figure 6 shows the temporal dynamics of pollutants in 2019. In Santiago, both the level and variability of pollution, especially CO and PM_{10}, increased considerably from March to October. Moreover, NO_2 and CO values peaked at the beginning of July, and PM_{10} values were relatively higher during this month as well. Meanwhile, the highest CO, NO_2 and PM_{10} levels in Bogota and Mexico City were observed in the first and last months of 2019. In the case of Buenos Aires, NO_2 levels increased from June until the end of the year, and CO also increased at the end of May and the beginning of June but returned to stable levels in October. Unlike the other cities, PM_{10} values in Buenos Aires remained relatively stable throughout the year.

Figure 6 Time series of average air quality (2019)



Source: Prepared by the authors.

Finally, **Figure 7** plots air quality levels against traffic delay due to road congestion for the four cities. To capture the hours of higher variability in road congestion, we differentiate between morning and afternoon rush hours. In all cases there is a positive relationship between pollution levels and traffic delays. Furthermore, in almost all cases the relationship seems to be nonlinear, such as shown in **Figure 2** on the theoretical dynamics between congestion and fuel consumption.

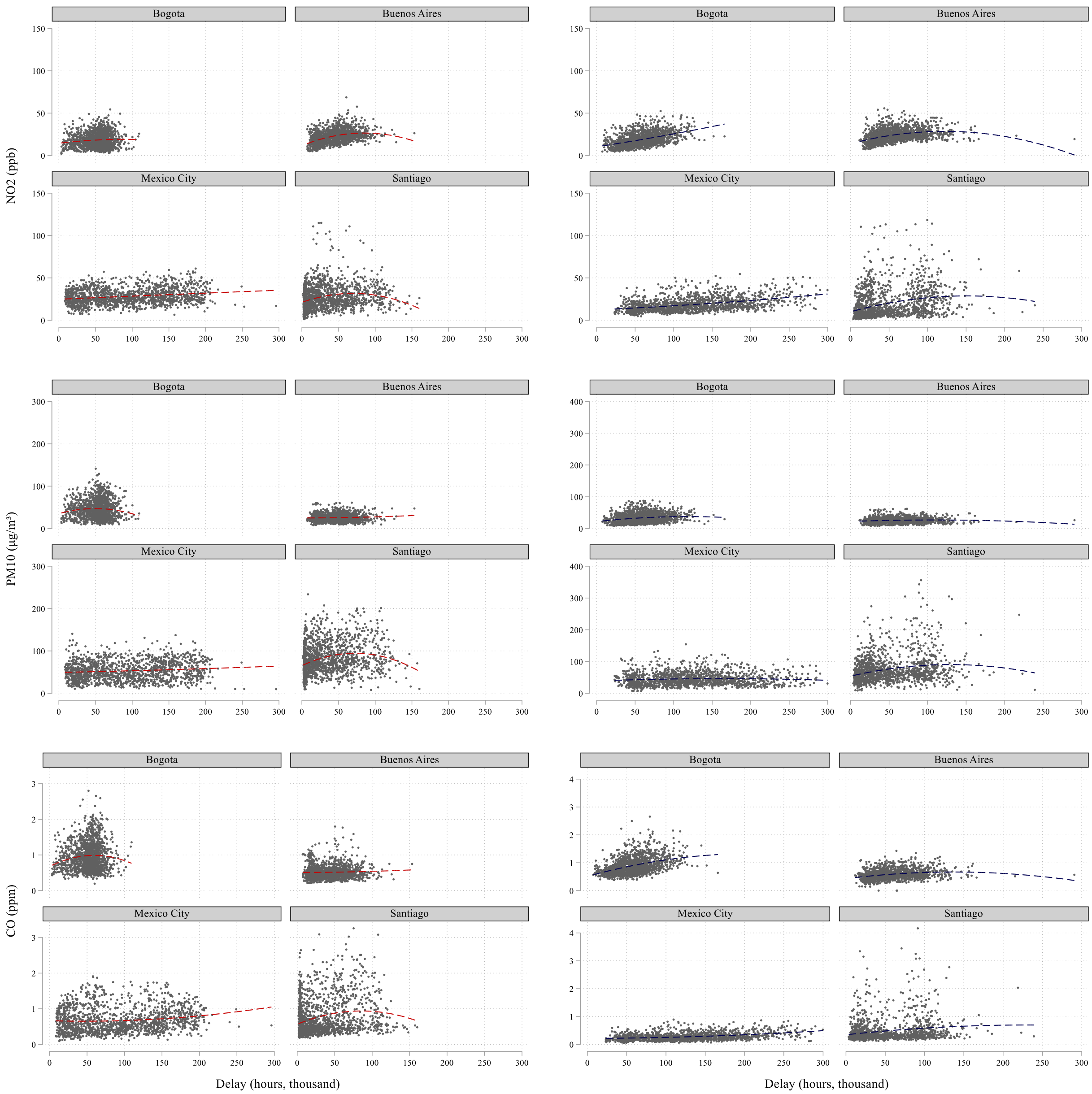
Next, we statistically test whether the effect of road congestion on CO, PM_{10}, and NO_2 levels is indeed nonlinear. If true, then the magnitude of the effect will vary according to the stage of road congestion. According to Brownfield et al. (2003) and Grant-Muller et al. (2007), there are three stages:

1. Pre-congestion stage (PCS): When free-flow conditions break down but complete congestion has not yet occurred. This may happen at the beginning or end of a period of congestion. It can also occur between periods of congestion, due to fluctuations in vehicle movement and speed, depending on the type of traffic jam.

2. Recurrent congestion stage (RCS): Congestion at regular times at fixed locations, for example, during peak hours.

Figure 7 Relationship between congestion and air quality during peak hours (2019)

(a) Morning peak (7-10 am) (b) Afternoon peak (4-7 pm)



Source: Prepared by the authors.
Note: The figure shows city-wide averages for all air monitoring stations.

3. Non-recurring congestion stage (NRCS): Congestion at non-regular times that cannot be predicted by road users. An example is traffic congestion caused by road incidents or weather events.

In addition, we consider the cumulative effect (CE) of road congestion on air quality. Due to the autoregressive behavior of pollutants, a worsening of air quality at a given time is expected to influence air quality during the following hours. Therefore, CE measures the accumulated effect on pollutants from when free-flow conditions break down until the effect disappears.

To test the impact of congestion on air quality in the four scenarios described above, we apply an econometric model that accounts for the serial correlation of contaminants and controls for confounding variables such as weather conditions. We use variability between stations to rule out any systematic fixed differences in contaminants by station and city over time. An example of these differences would be a decrease in air quality caused by fixed industrial sources. Specifically, we follow Cameron and Trivedi (2010) and use a cross-sectional time-series regression model with fixed effects, where the disturbance term is a first-order autoregressive process, represented as follows:

$$y_{itx1} = \alpha + y_{(i(t-1))x1}\lambda_{1x1} + x_{jtx2}\beta_{2x1} + c_{jtxf}\gamma_{fx1} + m_{1xw}\theta_{wx1} + d_{1xe}\omega_{ex1} + h_{1xq}\psi_{qx1} + \upsilon_{ix1} + \varepsilon_{itx1} \quad (1)$$

$$\varepsilon_{itx1} = \varepsilon_{(i(t-1))x1}\rho_{1x1} + \eta_{itx1} \quad (2)$$

Where:

- y_{itx1} is the hourly log level of pollutant NO_2, CO, or PM_{10} captured by station i at hour t. Dependent and main independent variables are considered in logarithms to smooth outliers and interpret the results as elasticities.
- $y_{(i(t-1))x1}$ is the first-order autoregressive term.
- x_{jtx2} includes the log and squared log number of hours lost due to road congestion in city j at hour t.
- c_{jtxf} controls for climatic condition f, including hourly temperature (measured in kelvin), atmospheric pressure (hPa), humidity (%), wind speed (min/sec), precipitation (min), and wind direction as a dummy variable that takes the value of one when the wind is coming from the north. These climatic conditions are captured at the center of city j at hour t.
- m_{1xw}, d_{1xe}, and h_{1xq} control for common variation in air quality for month w of the year, day of the week e, and hour of the day q, respectively.
- υ_{ix1} represents the fixed systematic difference of pollutants between stations over time.
- ε_{itx1} is the error term, which is assumed to follow a first-order autoregressive process, and η_{itx1} follows a normal distribution.

Then, we use the method proposed by Baltagi and Wu (1999), leveraging a Cochrane–Orcutt transformation that allows us to estimate ρ in the first step, to obtain the within estimators of the fixed-effects model (i. e., $\alpha, \beta, \gamma, \theta, \omega,$ and ψ) by applying OLS on the model transformation described in **Equations 3** and **4**:

$$\check{z}_{itx1} = z^{*}_{itx1} - \bar{z}^{*}_{ix1} + \bar{\bar{z}}^{*}; \ \forall\, z: y, x, c, m, d, h, e; \wedge\, t > 1 \quad (3)$$

$$\check{y}_{itx1} = \alpha + \check{y}_{(i(t-1)x1}\lambda_{1x1} + \check{x}_{jtx2}\beta_{2x1} + \check{c}_{jtxf}\gamma_{fx1} + \check{m}_{1xw}\theta_{wx1} + \check{d}_{1xe}\omega_{ex1} + \check{h}_{1xq}\psi_{qx1} + \check{\varepsilon}_{itx1} \quad (4)$$

Under the third scenario, i.e., NRCS, road congestion is also generated by events that are not easily anticipated, such as road closures and hazards that came about during the same hour. These events provide a source of exogenous variation that we use for a robustness check to test the causal validity of the results. Thus, we conduct a two-stage least squares (2SLS) estimation to assess the impact on each of the three contaminants of road congestion from the road closures and hazards reported at a specific hour t. This approach is conducted to confirm that there are no other non-controlled time varying confounders that are misleading the results. The first stage and second stage estimations are represented by **Equations 5** and **6**, respectively.

$$x_{jtx1} = \tau + rc_{jtx1}\vartheta_{1x1} + rh_{jtx1}\varphi_{1x1} + y_{i(t-1)x1}\varsigma_{1x1} + c_{jtxf}\gamma_{fx1} + m_{1xw}\theta_{wx1} + d_{1xe}\omega_{ex1} + h_{1xq}\psi_{qx1} + \xi_{jtx1};\ \forall\, y \tag{5}$$

$$\breve{y}_{itx1} = \alpha + \breve{y}_{(i(t-1)x1}\lambda_{1x1} + \breve{\ddot{x}}_{jtx2}\beta_{2x1} + \check{c}_{jtxf}\gamma_{fx1} + \breve{m}_{1xw}\theta_{wx1} + \check{d}_{1xe}\omega_{ex1} + \check{h}_{1xq}\psi_{qx1} + \breve{\varpi}_{itx1} \tag{6}$$

Where:

- rc_{jtx1} and $rh_{jtx1}$are the number of road closures and hazards that occurred in a city j at hour t.
- $\ddot{x}_{itx2}$ represents the predicted log delay and its squared term in city j at hour t, obtained from the first stage represented by **Equation 5**.
- ξ_{jtx1} and ϖ_{itx1} are the idiosyncratic errors, the first assumed to distribute normal and the second to follow a first-order autoregressive process.
- The remaining group of variables has the same definition as in **Equations 1-4**.

4. Results

Table 3 presents the overall results. The coefficients in the first row provide statistical evidence of a positive effect on all three contaminants during the initial stages of road congestion. We also find consistent evidence that this effect is nonlinear, as indicated by the coefficients in the second row. In other words, regarding the squared term in **Equation 1**, the marginal effect is initially positive but decreases progressively. Thus, in line with the theoretical framework presented in **Section 2**, traffic delays have a higher impact on overall air quality in a city during the initial stages of road congestion. Next, we review these results further by discussing each of the road congestion stages defined in **Section 3**.

4.1. Pre-congestion stage (PCS)

Right after free-flow conditions break down, a 1% increase in road congestion is associated with an immediate rise of 0.13% in CO levels, 0.16% in NO_2, and 0.19% in PM_{10}. This is particularly relevant given that the standard deviation of road congestion is 94% of the hourly mean value and 135% of the hourly median value. Meanwhile, a 1% increase in road congestion raises daily median levels of NO_2 and PM_{10} by 0.39% and 0.66%, respectively. The impact is considerably lower for CO (0.07%). The daily aggregated level of PM_{10} increases by 0.96%, followed by NO_2

(0.84%), while the effect on CO is not statistically significant. These results are influenced by the temporal dynamics of contaminants throughout the day, which will be addressed later in the paper.

Weather conditions also affect air quality. Our results show that an increase in temperature, humidity, wind speed, and rainfall at the city center — where weather conditions are measured in our database — are associated with lower values for all three contaminants, and therefore lower average pollution levels. The effect is higher for PM_{10}, then NO_2, and finally CO. Conversely, an increase of atmospheric pressure in the city center is associated with an increase in NO_2 and PM_{10}. Wind direction is the only climatic variable without a statistically significant effect on air quality at the 5% level for any of the air pollutants. This can be attributed to the fact that we analyze four cities and several air monitoring stations within each city. If the quadrant where the greatest amount of pollution originates differs between cities, wind direction's effect on air quality at a specific location will be mixed. Therefore, variability in the location of air quality monitoring stations explains the mixed and non-statistically significant results regarding wind direction.

Table 3 General results

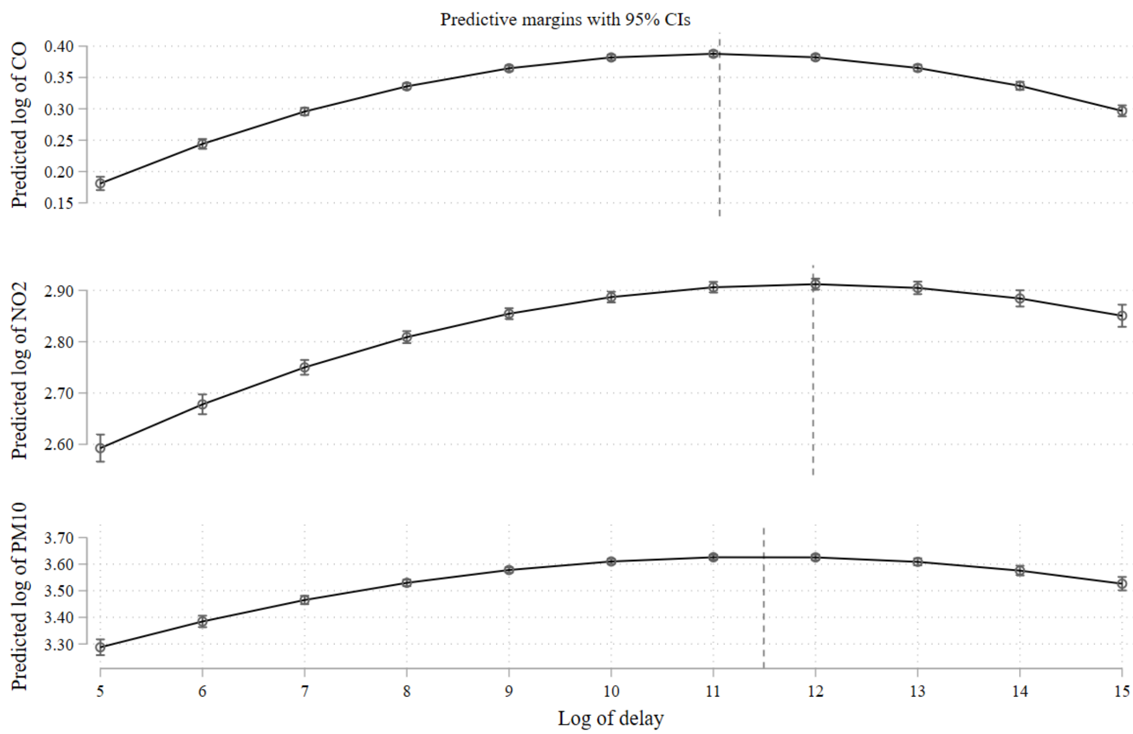
	ln(CO)	ln(NO2)	ln(PM10)
ln(delay)	0.1261***	0.1581***	0.1862***
	(0.0035)	(0.0088)	(0.0095)
ln(delay), sqr	-0.0057***	-0.0066***	-0.0081***
	(0.0002)	(0.0004)	(0.0005)
Temperature	-0.0019***	-0.0032***	-0.0039***
	(0.0001)	(0.0001)	(0.0002)
Pressure	-0.0000	0.0005***	0.0003***
	(0.0000)	(0.0000)	(0.0001)
Humidity	-0.0002***	-0.0001**	-0.0009***
	(0.0000)	(0.0000)	(0.0000)
Wind speed	-0.0028***	-0.0023***	-0.0113***
	(0.0001)	(0.0004)	(0.0004)
Wind direction	-0.0012*	0.0024	-0.0024
	(0.0006)	(0.0016)	(0.0020)
Rainfall	-0.0019***	-0.0044***	-0.0102***
	(0.0005)	(0.0012)	(0.0020)
R2	0.70	0.75	0.73
Obs	329190	317008	314045

Notes: Results from the cross-sectional time-series regression model when the disturbance term is first-order autoregressive. All estimates include monitoring station, month, day-of-the-week, and hour-of-the-day fixed effects. Standard errors in parenthesis. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

The nonlinear effect of road congestion on ir quality described above suggests that while positive, the effect decreases after reaching a threshold and that the value of this threshold can be predicted based on levels of road congestion. To illustrate this finding, we conduct consecutive predictions of each pollutant for different levels of road congestion, while leaving all weather conditions at their average levels (**Figure 8**). The nonlinear nature of the relationship also suggests that PCS alone does not account for road congestion's overall effect on air quality.

In our tests, the first pollutant to reach the maximum predicted threshold level is CO, at nearly 1.5 ppm. This happens when the number of cumulative hours of delay approaches 64,000 at a given calendar hour. In the case of PM_{10}, this point is reached at 37 μg/m³ when road congestion amounts to 98,000 hours. NO_2 is the last contaminant analyzed to reach the maximum prediction (18 ppb) at approximately 160,000 hours of delay. By way of comparison, the median hourly road congestion in all cities during 2019 was almost 40,000 hours, and the 75th percentile was 87,000 hours. These maximum values are above the average hourly level of road congestion in all four cities, consequently, in a *ceteris-paribus* scenario and under common traffic conditions, it is expected that reducing road congestion will most probably end up in air quality improvements. Nonetheless, the magnitude of such improvements is not constant along the levels of road congestion.

Figure 8 Predicted air quality levels based on changes in road congestion



Source: Prepared by the authors.

4.2. Recurrent congestion stage (RCS)

The above results are for the average impact of road congestion on air pollutants throughout all days of the week and all hours of the day during 2019. Given the nonlinear effect we described previously, the impact congestion has on air quality will depend on the baseline level of road congestion when the increase in delays occurs. Therefore, to test the effect of road congestion on air quality, we looked at the times during the day when traffic is higher, which are morning peaks form 7:00 a.m. to 10:00 a.m. and afternoon peaks from 4:00 a.m. to 7:00 p.m. (**Figure 7**). We also assessed the impact considering only weekdays, but the results were virtually unchanged whether weekends are included or not.

In the RCS, we find robust evidence of a higher impact of road congestion on air quality. An initial 1% increase in the level of road congestion is associated with a rise of 0.30% in the level of CO during the same hour. The figures for NO_2 and PM_{10}. are 0.44% and 0.57%. The estimate for the squared term is higher and statistically robust. Notably, the predictive power of congestion is higher for variations in NO_2, less for CO, and almost unchanged for PM_{10}. After controlling for weather conditions and seasonal effects, the predictive power of the econometric model increases by 6 percentage points for NO_2 and decreases by 8 percentage points for CO. This change in predictive power is observed even when using a sample, i.e., peak traffic hours.

Average congestion levels are higher during peak hours than during the rest of the day. In Mexico City, the average delay per hour during the afternoon peak was 123,750 hours (versus an average delay of 74,400 hours for both peak and off-peak hours in 2019). In Bogota, the average peak traffic delay per hour was 59,400 hours (versus 38,600), in Buenos Aires 59,000 hours (versus 35.000), and in Santiago 48,200 hours (versus 22,300). According to the theoretical framework presented in **Section 2**, the baseline effect on air quality should be worse at these times than during subsequent hours.

Table 4 Peak-hours results

	ln(CO)	ln(NO2)	ln(PM10)
ln(delay)	0.2970***	0.4351***	0.5681***
	(0.0091)	(0.0197)	(0.0317)
ln(delay), sqr	-0.0138***	-0.0192***	-0.0258***
	(0.0004)	(0.0009)	(0.0015)
R2	0.62	0.81	0.73
Obs	111694	108052	118288

Notes: Results from the cross-sectional time-series regression model when the disturbance term is first-order autoregressive. All estimates include climatic conditions and stations, month, day-of-the-week, and hour-of-the-day fixed effects. Standard error in parenthesis. * p<0.1, ** p <0.05, *** p <0.01.

4.3. Non-recurring congestion status (NRCS)

According to Brownfield et al. (2003) and Grant-Muller and Laird (2007), non-recurring congestion is caused by events not easily anticipated an hour prior to their occurrence, such as road closures and hazards. The results in **Table 5 (a)** provide evidence of two conditions that validate the use of road closures and hazards as strong instruments of road congestion. Indeed, both closures and hazards significantly raise the level of road congestion, with predictive power reaching around 80%. One additional road closure in the city is associated with a 0.3% increase in road congestion, while one additional hazard raises congestion by 0.26%. These estimated effects are consistent in the first stage of the 2SLS for all three air pollutants.

Table 5. Two-stage least squares estimation

(a) First stage

	ln(CO)	ln(NO2)	ln(PM10)
Road closures	0.0030***	0.0030***	0.0029***
	(0.0000)	(0.0000)	(0.0000)
Hazards	0.0026***	0.0026***	0.0029***
	(0.0000)	(0.0000)	(0.0000)
R2	0.78	0.79	0.78
Obs	325386	313142	311975

Note: OLS model. All controls included. Standard errors in parenthesis. p<0.1, ** p<0.05, *** p<0.01

(b) Second stage

	ln(CO)	ln(NO2)	ln(PM10)
ln(delay), predicted	0.1204***	0.0957***	0.1100***
	(0.0063)	(0.0155)	(0.0171)
ln(delay), sqr, predicted	-0.0049***	-0.0038***	-0.0053***
	(0.0003)	(0.0007)	(0.0007)
R2	0.70	0.75	0.72
Obs	325150	312779	310837

Note: Cross-sectional time-series regression models. All controls included. Standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01

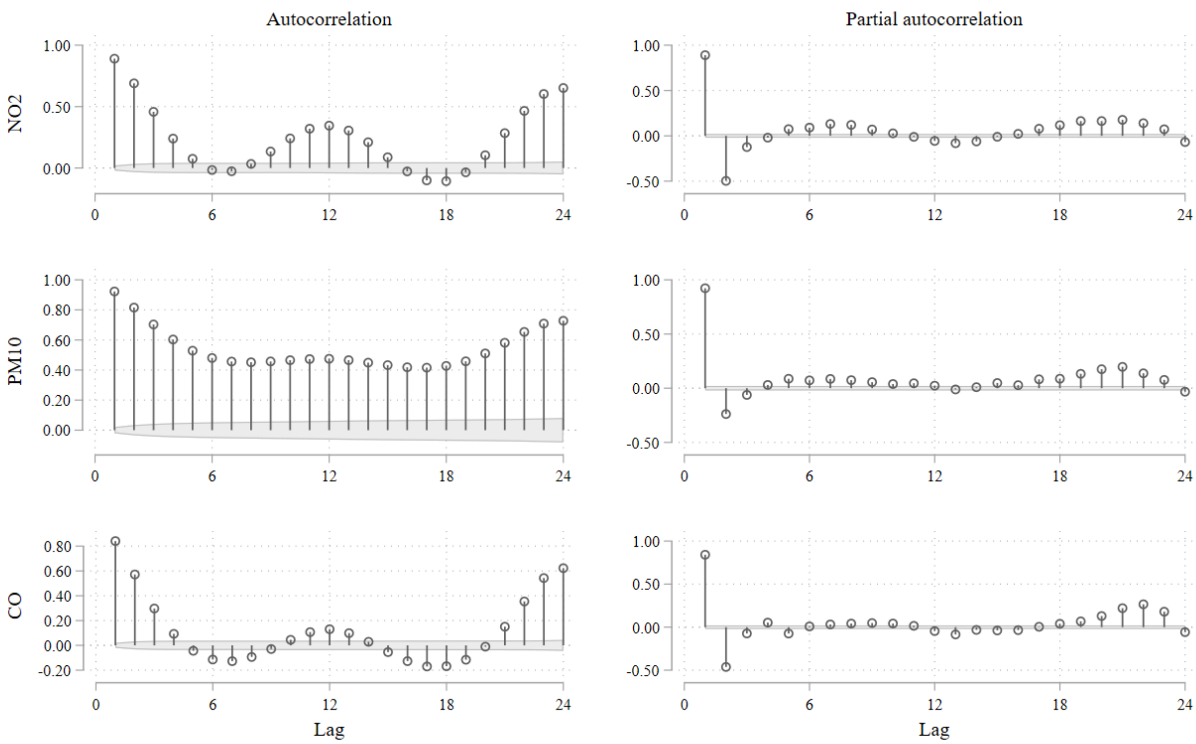
The effect found in the second stage confirms that road congestion has a causal effect on air quality. In this case, additional delays associated with road closures and hazards generate an initial increase of 0.12% in CO, 0.10% in NO_2 and 0.11% in PM_{10}. These initial effects on NO_2 and PM_{10} are less than the general results presented in **Table 3**. The squared term is also lower, which indicates that the effect during more advanced stages of road congestion diminishes at a slower rate.

4.4. Cumulative effect (CE)

An aspect relevant to road congestion's overall impact on air quality is the temporal dynamics of contaminants. **Figure 9** plots the correlogram, or the autocorrelation and partial autocorrelation coefficients, for the 24 hours prior to any time during the day. According to the correlogram, all three contaminants follow a similar autoregressive process and are positively and highly correlated in 12-hour cycles. During the first 6 hours of the cycle, the decay rate is higher for CO levels, followed by NO_2 and PM_{10}. The autocorrelation of PM_{10} is statistically significant during the 24-hour period. Partial autocorrelations suggest that the effect of congestion on air quality at a given point in time starts decreasing after two hours.

From the Cochrane–Orcutt transformation represented by **Equation 6**, we compute the autoregressive term of order one of each contaminant along with the first-order process in the error term (ρ). The temporal dynamics of contaminants indicate that the cumulative effect from an initial 1% increase in road congestion causes an increase of 0.59% in CO, 0.55% in NO_2 and 0.58% in PM_{10}. After that, the cumulative effect depends on the level road congestion reported in the city. According to the correlogram, these effects accumulate faster for CO and NO2 but slower for PM_{10}.

Figure 9 Temporal dynamics of contaminants

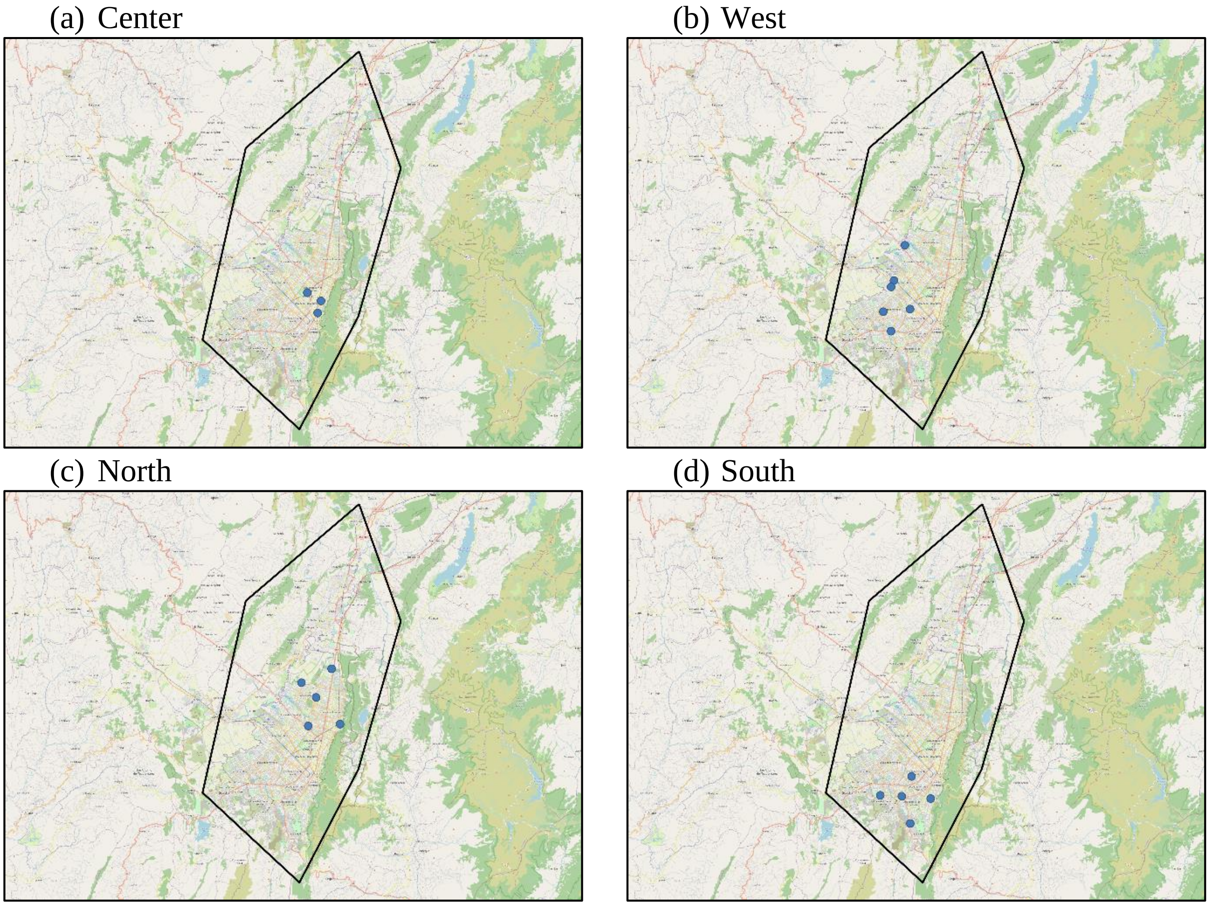


Note: Shaded area denotes 95% confidence intervals.
Source: Prepared by the authors.

4.5. Exploring city heterogeneity: A closer look at Bogota

To analyze whether the relationship between road congestion and air quality changes thoughout the city, we look at different geographic areas in the city of Bogota. As air monitoring stations report hourly climatic conditions at their locations, it is possible to estimate the impact of congestion on air quality using data from a variety of stations geographically distributed across the city (**Figure *10***). Furthermore, the data recorded by the Air Quality Monitoring Network of Bogota (RMCAB, by its acronym in Spanish) enables analyzing the impact of congestion on PM2.5, NO, and NOx.

Figure 10. Air quality monitoring stations according to geographical areas in Bogota.

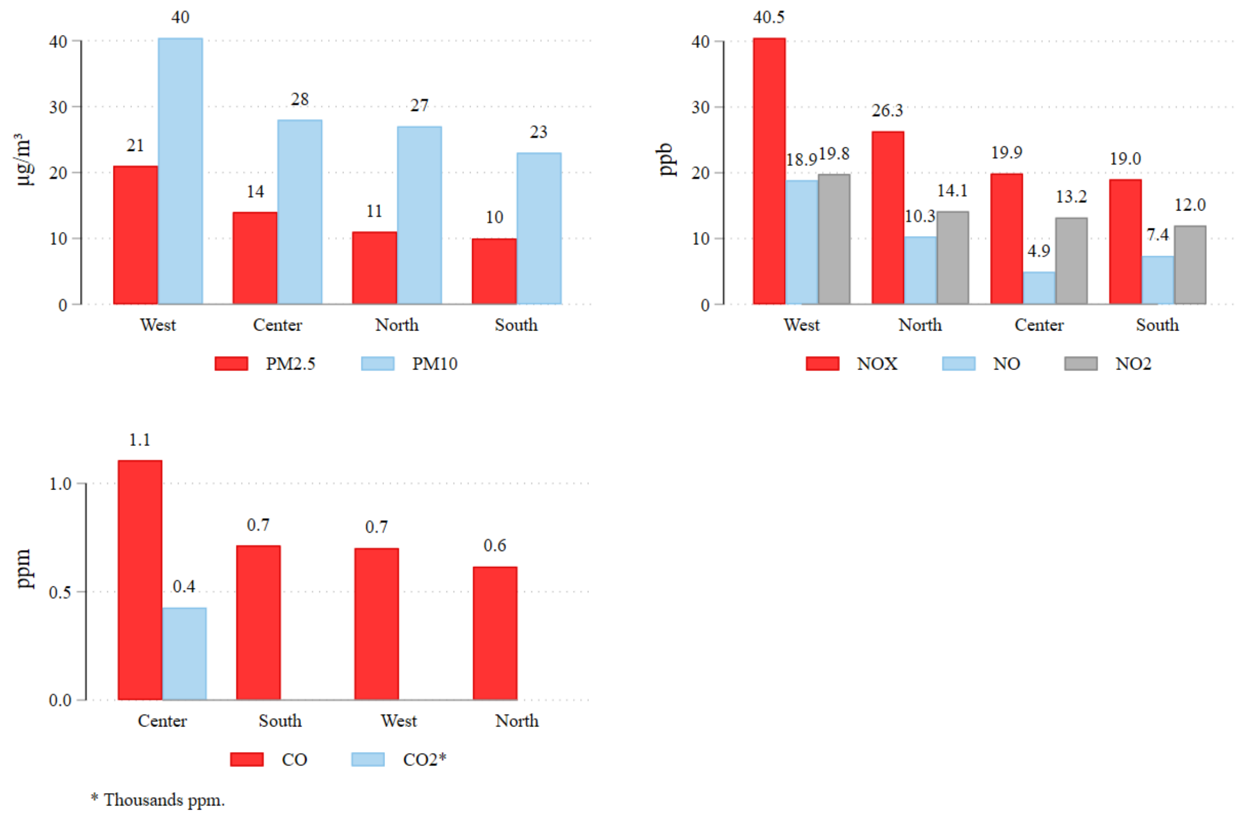


Source: Prepared by the authors.

To avoid the possible bias caused by outliers, **Figure 11** presents the median concentration of particulate matter and contaminant gases during 2019 in each urban area. The highest levels of PM were observed in the western area, with a median of 21 micrograms per cubic meter in PM2.5 and 40 in PM10: 50% and 43% higher, respectively, compared with the second most affected area.

The western area also registered the highest levels of NO, NO_2, and NOx. The median of NOx in this area surpassed 40.5 ppb, while the least affected area, the southern area, registered a median of 12 ppb (a 70% lower). In the case of CO, the central area was the most affected with 1.1 ppm, much above the western, southern, and northern areas, which registered levels between 0.6 and 0.7 ppm.

Figure 11. Median of contaminants concentration according to urban areas in Bogota



Source: Prepared by the authors.

Next, we used a simplified version of the econometric model presented in **Section 3** and developed a sensitivity analysis to compare the behavior of traffic congestion and contaminants in different areas of the city. For this, we considered the logarithm of traffic congestion. **Table 6** presents the results of the analysis. Colors in the table show the degree of sensitivity by contaminant throughout the city, with red meaning higher and green meaning lower sensitivity, respectively. PM had a higher sensitivity to road congestion in the central and northern areas of Bogota. The elasticity for PM2.5 was between 0.05 and 0.07 when considering all four areas. In turn, the range of elasticity for PM10 was broader, with values between 0.02 and 0.06. NOx showed a higher sensitivity to road congestion in the central area of the city. The sensitivity of CO was higher in the western area. NO gases had the highest sensitivity to road congestion when considering all areas. Conversely, PM10 appeared to be the most inelastic among the analyzed contaminants.

Table 6. Comparative sensitivity across urban areas of Bogota

Contaminant	Center	North	West	South
PM2,5	0.067	0.062	0.053	0.063
PM10	0.048	0.063	0.036	0.019
NO	0.153	0.111	0.075	0.074
NO2	0.059	0.046	0.048	0.042
NOX	0.195	0.081	0.062	0.052
CO	0.028	0.030	0.043	0.021

Note: Linear slope of time series model with fixed effects.
Temporal controls included. All results are statistically significant at 1%.

5. Discussion

This paper builds on recent academic literature to provide evidence regarding the causal relationship between congestion and pollution in development contexts, with valid quantitative results for four of the most congested cities globally, located in Latin America, after controlling for all heterogeneity in urban characteristics. We estimate the sensibility of air quality to urban road congestion and provide a policy discussion grounded on the findings.

The results from the econometric approach validate growing concerns about rising congestion levels in cities around the world, particularly in developing countries. Not only does congestion generate a monetary cost and result in lost productivity, it also impacts the environment and quality of life in urban settings. Yet policymakers are facing strong resistance to implementing bolder measures to reduce congestion. Because of a lack of knowledge, traffic mitigation policies aiming to reduce car usage in highly congested areas often gain limited acceptance. It is therefore key to grow the body of available empirical evidence to support more effective actions to revert — or at least contain — the negative trends of rising motorization rates, higher congestion levels, and lower environmental sustainability.

This research helps bridge the knowledge gap by estimating the relationship between congestion and air quality in four of the most populous and congested cities in Latin America and the Caribbean. By processing more than 4.4 billion crowdsourced traffic records from the entire urban areas of Bogota, Buenos Aires, Mexico City, and Santiago, we show that a 1% initial increase in hourly road congestion increases CO by 0.13%, NO_2 by 0.16%, and PM_{10} by 0.19% during the same hour. Moreover, the cumulative effect over time of that initial increase is 0.60% for all three contaminants. To gauge the relevance of this effect, consider the average daily level of road congestion in the four cities, which is approximately 58,000 hours, and its standard deviation of 54,000 hours (94% of the mean). The average hourly levels of the contaminants are 0.52 ppm of CO, 21 ppb of NO_2, and 35 µg/m³ of PM_{10}. Thus, an initial increase in road congestion of one standard deviation would raise CO by 0.29 ppm, NO_2 by 12 ppb, and PM_{10} by 18 µg/m³ in PM_{10} in the same hour.

These results are particularly relevant given the public health impact of worsening air quality (Zhanga and Batterman, 2013). ECLAC (2021) estimated that an average increase of 10 μg/m³ of PM_{10} and $PM_{2.5}$ is associated with increases in mortality of 0.4% and 0.7%, respectively. Analyzing mobility restrictions implemented to contain the COVID-19 pandemic in Bogota, Blackman (2021) calculated that between March 2020 and February 2021, around 115 premature deaths were avoided, a 31% reduction from the 371 deaths that would have occurred in the short term under normal conditions. Our related research provides statistical evidence that, under a *ceteris-paribus* scenario and common traffic conditions, reducing road congestion can potentially decrease the concentrations of CO, NO_2, and PM_{10} in the air, which would help improve public health in Latin America's cities.

It should be carefully acknowledged that not all traffic mitigation policies will enhance air quality if these policies produce counterproductive consequences. For instance, a congestion mitigating policy may increase air pollution when causing increments in vehicle miles traveled, unfavorable travel schedule substitution, or higher vehicle ownership to avoid circulating restrictions (Ortúzar, 2019; Barahona et al., 2020). For instance, Zhang et al. (2017) analyzed the consequences of driving restrictions in Bogota which were designed to control congestion and urban air pollution. The authors developed a theoretical model and implemented an impact evaluation proving evidence that license plate-based driving restrictions can have diverse effects on air pollutants; specifically, it can generate a significant decrease in NO but an increase in NO_2, NO_x and O_3.

Our research also shows that road congestion has lingering effects on air quality. Given the temporal dynamics of this effect and the relationship's nonlinear nature, policymakers should especially focus on policies that mitigate congestion at its initial stages, when free-flow conditions start to break down, since this is the most effective time to prevent deterioration of air quality. In addition to helping make the general case for reducing congestion to improve air quality, our study has implications for how policies and programs could be timed to enhance their efficiency. Of course, reducing congestion requires integrated "pull" and "push" policies (Wang *et al.*, 2022). Pull strategies try to increase public transportation ridership, while push measures involve dissuading people from using private vehicles in congested conditions. This research supports policymakers as they develop push policies that are politically difficult to implement but can provide great social benefits (Ortúzar *et al.*, 2021). In developed countries, new types of policies such as road pricing have successfully reduced excessive use of private cars (Ortúzar, 2019).

Finally, since transportation is one of the most critical contributors to climate change, representing 23% of CO_2 emissions related to energy use worldwide, the sector is among the most critical for achieving the 2030 targets for limiting global warming to 1.5°C (WB, 2016; United Nations, 2019). Furthermore, the main sources of NO_2 are related to transportation (EPA, 2018), and our findings point out that road congestion is critically relevant for reducing concentrations of NO_2 levels in urban settings. This research highlights the previously unexplored impact of road congestion on this contaminant, which is essential to the formation of harmful levels of O3, another greenhouse gas. Understanding the relationship between urban road congestion and NO2 levels is not enough but necessary to provide comprehensive information supporting policy decisions to tackle climate change through actions focused on urban traffic.

6. Conclusions

This study leverages highly disaggregated big data and uses panel-data econometric models to estimate the effect of road congestion on CO, NO_2, and PM_{10} and explore the temporal dynamics of the effect for four of the most congested cities in Latin America. The approach allowed us to rule out fixed confounders between and within cities by using hourly data at the air monitoring station level while controlling for traditional confounders previously reported in the literature. We confirm causality using a 2SLS approach. We find statistical evidence on the impact of road congestion on all three contaminants. Moreover, we find that the effects are lingering and thus accumulate over the following hours. Finally, we report that the relationship is nonlinear and predict the thresholds after which the effect decays. These results are expected to inform and support public policy to develop more sustainable mobility in highly populated, highly congested cities in developing countries. Further research opportunities include applying this methodology to other cities in developing contexts, as well as evaluating the impact of congestion mitigation policies on air pollutants.

Acknowledgements

The authors would like to thank Francisca Giraldez for her constant support and insightful comments on the theoretical framework of the research.

References

Amin-Naseri, M. *et al.* (2018) 'Evaluating the Reliability, Coverage, and Added Value of Crowdsourced Traffic Incident Reports from Waze', *https://doi.org/10.1177/0361198118790619*, 2672(43), pp. 34–43. Available at: https://doi.org/10.1177/0361198118790619.

Baltagi, B.H. and Wu, P.X. (1999) 'Unequally spaced panel data regressions with AR(1) disturbances', *Econometric Theory*, 15, pp. 814–823. Available at: https://doi.org/doi:10.1017/s0266466699156020.

Barahona, N., Gallego, F.A. and Montero, J.P. (2020) 'Vintage-Specific Driving Restrictions', *The Review of Economic Studies*, 87(4), pp. 1646–1682. Available at: https://doi.org/10.1093/RESTUD/RDZ031.

Barr Rosso, K. (2021) *El transporte no sostenible tiene consecuencias: estas son las estrategias para reducir la contaminación ambiental*, *Moviliblog*.

Berger (2020) *How important was it for you to have access to a private vehicle both before and during the COVID-19 pandemic?*

Blackman, A., Bonilla, J.A. and Villalobos, L. (2021) *Quantifying COVID-19's Silver Lining: Avoided Deaths from Air Quality Improvements in Bogotá*. Washington D. C.

Brownfield, J. *et al.* (2003) 'Congestion and accident risk', *Department for Transport. Road Safety Research Report* , 44.

Burnett, R.T. *et al.* (2014) 'An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure', *Environmental Health Perspectives*, 122(4), pp. 397–403. Available at: https://doi.org/10.1289/ehp.1307049.

Calatayud, A. *et al.* (2021) *Congestión urbana en América Latina y el Caribe: Características, costos y mitigación*. Washington, D.C.: Inter-American Development Bank. Available at: https://doi.org/http://dx.doi.org/10.18235/0003149.

Cameron, A. and Trivedi, P. (2005) *Microeconometrics: methods and applications*. Cambridge university press.

Cameron, A. and Trivedi, P.K. (2010) *Microeconometrics using Stata Revised Edition*. Stata Press.

Dasgupta, S., Lall, S. and Wheeler, D. (2021) 'Spatiotemporal analysis of traffic congestion, air pollution, and exposure vulnerability in Tanzania', *Science of the Total Environment*, 778, p. 147114. Available at: https://doi.org/10.1016/j.scitotenv.2021.147114.

ECLAC (2012) *Urbanization Represents Opportunities and Challenges in Terms of Progressing towards Sustainable Development*.

ECLAC (2021a) *Salud y cambio climático: metodologías y políticas públicas*, *Salud y cambio climático: metodologías y políticas públicas*. Ciudad de México.

ECLAC (2021b) *Salud y cambio climático: metodologías y políticas públicas*, *Salud y cambio climático: metodologías y políticas públicas*. Ciudad de México.

EPA (2022) *Regulatory and Guidance Information by Topic: Air* . Available at: https://www.epa.gov/regulatory-information-topic/regulatory-and-guidance-information-topic-air (Accessed: 14 September 2022).

Fire, M. *et al.* (2012) 'Data mining opportunities in geosocial networks for improving road safety', *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2012* [Preprint]. Available at: https://doi.org/10.1109/EEEI.2012.6377049.

Ford (2021) *'The pandemic has caused me to reconsider the number of vehicles I need for my household'.*

Geroliminis, N. and Daganzo, C.F. (2008) 'Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings', *Transportation Research Part B: Methodological*, 42(9), pp. 759–770. Available at: https://doi.org/10.1016/J.TRB.2008.02.002.

Gómez-Lobo, A. *et al.* (2022) *Agglomeration and Congestion in Latin America*. Washington D. C.

Goodall, N. and Lee, E. (2019) 'Comparison of Waze crash and disabled vehicle records with video ground truth', *Transportation Research Interdisciplinary Perspectives*, 1, p. 100019. Available at: https://doi.org/10.1016/J.TRIP.2019.100019.

Goodwin, P. (2004) 'The Economic Costs of Road Traffic Congestion', *University College London*, (May), pp. 1–26.

Grant-Muller, Susan. *et al.* (2007) *Costs of Congestion: Literature Based Review of Methodologies and Analytical Approaches*, *Scottish Executive, Edinburgh*. Scottish Executive, Edinburgh.

Greenshields, B.D. (1935) 'A Study of Traffic Capacity', *Highway Research Board Proceeding*, 14, pp. 448–477.

Heft-Neal, S. *et al.* (2018) 'Robust relationship between air quality and infant mortality in Africa', *Nature*, 559(254–258).

Heger, M. *et al.* (2019) *Motor Vehicle Density and Air Pollution in Greater Cairo*, *Motor Vehicle Density and Air Pollution in Greater Cairo*. Available at: https://doi.org/10.1596/32512.

Higgins, C.D. *et al.* (2019) 'Accessibility, air pollution, and congestion: Capturing spatial trade-offs from agglomeration in the property market', *Land Use Policy*, 84, pp. 177–191. Available at: https://doi.org/10.1016/j.landusepol.2019.03.002.

Hoseinzadeh, N. *et al.* (2020) 'Quality of location-based crowdsourced speed data on surface streets: A case study of Waze and Bluetooth speed data in Sevierville, TN', *Computers, Environment and Urban Systems*, 83. Available at: https://doi.org/10.1016/J.COMPENVURBSYS.2020.101518.

IDB (2021) *Improving the sustainability and quality of life of the cities of Latin America and the Caribbean, Housing and Urban Development Division*. Available at: https://www.iadb.org/en/urban-development-and-housing/housing-and-urban-development.

Lelieveld, J. *et al.* (2015) 'The contribution of outdoor air pollution sources to premature mortality on a global scale', *Nature*, 525, pp. 367–371. Available at: https://doi.org/https://doi.org/10.1038/nature15371.

Lu, J. *et al.* (2021) 'Expansion of city scale, traffic modes, traffic congestion, and air pollution', *Cities*, 108(October 2020). Available at: https://doi.org/10.1016/j.cities.2020.102974.

Lu, M., Sun, C. and Zheng, S. (2017) 'Congestion and pollution consequences of driving-to-school trips: A case study in Beijing', *Transportation Research Part D: Transport and Environment*, 50, pp. 280–291. Available at: https://doi.org/10.1016/j.trd.2016.10.023.

Mishra, S. *et al.* (2019) 'Comparative emission study by real-time congestion monitoring for stable pollution policy on temporal and meso-spatial regions in Delhi', *Journal of Cleaner Production*, 224, pp. 465–478. Available at: https://doi.org/10.1016/j.jclepro.2019.03.122.

NCHRP (2005a) *Predicting Air Quality Effects of Traffic-Flow Improvements: Final Report and User's Guide*.

NCHRP (2005b) *Predicting Air Quality Effects of Traffic-Flow Improvements: Final Report and User's Guide*. Available at: https://nap.nationalacademies.org/catalog/13797/predicting-air-quality-effects-of-traffic-flow-improvements-final-report-and-users-guide.

Nune, R. *et al.* (2017) 'Exploring Social Traffic Data for Evaluating Urban Arterial Congestion', *Transportation Research Board* [Preprint]. Available at: https://trid.trb.org/view/1438955 (Accessed: 14 September 2022).

Ortúzar, J. de D. (2019) 'Sustainable urban mobility: what can be done to achieve it?', *Journal of the Indian Institute of Science*, 99, pp. 683–693.

Ortúzar, J. de D. *et al.* (2021) 'Assessing the potential acceptability of road pricing in Santiago', *Transportation Research Part A: Policy and Practice*, 144, pp. 153–169. Available at: https://doi.org/10.1016/j.tra.2020.12.007.

Pachon, J.E. *et al.* (2021) 'Evaluation of factors influencing road dust loadings in a Latin American urban center', *Journal of the Air & Waste Management Association*, 71(2), pp. 268–280. Available at: https://doi.org/10.1080/10962247.2020.1806946.

Requia, W.J. *et al.* (2018) 'The health impacts of weekday traffic: A health risk assessment of PM2.5 emissions during congested periods', *Environment International*, 111, pp. 164–176. Available at: https://doi.org/10.1016/J.ENVINT.2017.11.025.

Rivera, N.M. (2021) 'Air quality warnings and temporary driving bans: Evidence from air pollution, car trips, and mass-transit ridership in Santiago', *Journal of Environmental Economics and Management*, 108(March). Available at: https://doi.org/10.1016/j.jeem.2021.102454.

Sánchez González, S., Bedoya-Maya, F. and Calatayud, A. (2021) 'Understanding the effect of traffic congestion on accidents using big data', *Sustainability (Switzerland)*, 13(13), p. 7500. Available at: https://doi.org/10.3390/su13137500.

Shabihkhani, R. and Gonzales, E. (2015a) 'Macroscopic relationship between network-wide traffic emissions and fundamental properties of the network', *Transportation Research Circular*, pp. 1–15.

Shabihkhani, R. and Gonzales, E. (2015b) 'Macroscopic relationship between network-wide traffic emissions and fundamental properties of the network', *Transportation Research Circular*, pp. 1–15. Available at: http://tft.eng.usf.edu/tft50/tft50_papers/P31_Shabihkhani.pdf.

Soleimani, M. *et al.* (2022) 'Health effect assessment of PM2.5 pollution due to vehicular traffic (case study: Isfahan)', *Journal of Transport and Health*, 24, p. 101329. Available at: https://doi.org/10.1016/j.jth.2022.101329.

TomTom (2020) *TomTom, Ranking 2019*.

United Nations (2019) *Objetivos de desarrollo sostenible*.

Wang, Y. *et al.* (2022) 'The impact of traffic demand management policy mix on commuter travel choices', *Transport Policy*, 117, pp. 74–87. Available at: https://doi.org/10.1016/J.TRANPOL.2022.01.002.

Waze (2019) *Waze for Cities*. Available at: https://www.waze.com/es-419/wazeforcities/.

WB (2016) *Leaders Call for Global Action to Reduce Transport's Climate Footprint*.

Xu, Y. and González, M.C. (2017) 'Collective benefits in traffic during mega events via the use of information technologies', *Journal of The Royal Society Interface*, 14(129). Available at: https://doi.org/10.1098/RSIF.2016.1041.

Zhang, K. and Batterman, S. (2013) 'Air pollution and health risks due to vehicle traffic', *The Science of the total environment*, 0, p. 307. Available at: https://doi.org/10.1016/J.SCITOTENV.2013.01.074.

Zhang, W. and Kockelman, K.M. (2016a) 'Optimal policies in cities with congestion and agglomeration externalities: Congestion tolls, labor subsidies, and place-based strategies', *Journal of Urban Economics*, 95, pp. 64–86. Available at: https://doi.org/10.1016/j.jue.2016.08.003.

Zhang, W. and Kockelman, K.M. (2016b) 'Optimal policies in cities with congestion and agglomeration externalities: Congestion tolls, labor subsidies, and place-based strategies', *Journal of Urban Economics*, 95, pp. 64–86. Available at: https://doi.org/10.1016/j.jue.2016.08.003.

Zhanga, K. and Batterman, S. (2013) 'Air pollution and health risks due to vehicle traffic', *Sci Total Environ*, 23(1), pp. 1–7. Available at: https://doi.org/10.1016/j.scitotenv.2013.01.074.Air.

Zhou, B. *et al.* (2023a) 'Variability of fuel consumption and CO2 emissions of a gasoline passenger car under multiple in-laboratory and on-road testing conditions', *Journal of Environmental Sciences*, 125, pp. 266–276. Available at: https://doi.org/10.1016/j.jes.2021.12.042.

Zhou, B. *et al.* (2023b) 'Variability of fuel consumption and CO2 emissions of a gasoline passenger car under multiple in-laboratory and on-road testing conditions', *Journal of Environmental Sciences*, 125, pp. 266–276. Available at: https://doi.org/10.1016/j.jes.2021.12.042.