MIT CEEPR
Center for Energy and Environmental Policy Research

Working Paper Series

Pricing Congestion to Increase Traffic: The Case of Bogotá

Juan-Pablo Montero, Felipe Sepúlveda, and Leonardo J. Basso

FEBRUARY 2025
CEEPR WP 2025-04

Working Paper Series.

Since 1977, the Center for Energy and Environmental Policy Research (CEEPR) has been a focal point for research on energy and environmental policy at MIT. CEEPR promotes rigorous, objective research for improved decision making in government and the private sector, and secures the relevance of its work through close cooperation with industry partners from around the globe. Drawing on the unparalleled resources available at MIT, affiliated faculty and research staff as well as international research associates contribute to the empirical study of a wide range of policy issues related to energy supply, energy demand, and the environment.

An important dissemination channel for these research efforts is the MIT CEEPR Working Paper series. CEEPR releases Working Papers written by researchers from MIT and other academic institutions in order to enable timely consideration and reaction to energy and environmental policy research, but does not conduct a selection process or peer review prior to posting. CEEPR's posting of a Working Paper, therefore, does not constitute an endorsement of the accuracy or merit of the Working Paper. If you have questions about a particular Working Paper, please contact the authors or their home institutions.

Pricing Congestion to Increase Traffic: The Case of Bogotá

Juan-Pablo Montero, Felipe Sepúlveda, and Leonardo J. Basso*

Abstract

In September 2021, the city of Bogotá introduced a major market-based reform to its odd-even driving restriction, known as *Pico y Placa*. Since then, drivers have had the option to pay a daily congestion fee to be exempt from the restriction. We find that while the reform increased traffic, it brought overall benefits. The welfare gains of middle-income individuals, who now use their cars more often, far outweighed the losses experienced by their higher-income counterparts, who now spend more time in traffic. Additionally, we show that these overall benefits could quintuple if Bogotá were to expand the reform into a comprehensive road-pricing scheme, not to mention the extra gains in air quality.

Keywords: congestion, driving restrictions, road pricing, air quality

JEL Classification: H23, L62, R41, R48, Q53.

*Montero (jmontero@uc.cl): Pontificia Universidad Católica de Chile, Aalto University, Sistemas Complejos de Ingeniería (ISCI), and MIT-CEEPR; Sepúlveda (fasepu@ubc.ca): University of British Columbia; and Basso (lbasso@ing.uchile.cl): Universidad de Chile, and ISCI. We are grateful to four anonymous referees, Laura Alvarez, Patrick Behrer, Jorge Bonilla, Marcelo Caffera, Agustina Calatayud, Victor Cantillo, Pierre Dubois (the Editor), Germán Escovar, Carolyn Fischer, Andrés Gómez-Lobo, Josh Graff Zivin, Jeanne Lafortune, Tomás Rau, Tomás Serebrisky, as well as seminar participants at different venues for many useful comments, to Dominga Selman and Sebastián Figari for superb research assistance, and to the Inter-American Development Bank (IADB) for financial and technical support, in particular for giving us access to the Waze database used in the empirical analysis. The opinions expressed in the paper, however, are not necessarily shared by the IADB. Montero also thanks financial support from Fondecyt and the Chair Energy and Prosperity at Fondation du Risque.

1 Introduction

Congestion remains a serious problem in many cities around the world. According to the INRIX Global Traffic Scorecard, the city of Bogotá led the pre-covid-19 ranking of the most congested cities in the world, with 192 hours per capita lost in heavy traffic in 2019. It was followed closely by Rio de Janeiro, Mexico City, Istanbul, Sao Paulo and Rome.[1] Unfortunately, when authorities have decided to deal with this externality, they have rarely turned to pricing schemes.[2] Instead, they increasingly rely on rationing schemes, better known as driving restrictions or license-plate bans.

One of the most stringent driving restriction today is precisely found in Bogotá, where restrictions were first introduced in 1998.[3] Since 2012 Bogota's driving restriction, better known as *Pico y Placa*, bans from circulation the vast majority of residential and commercial vehicles every other day of the week (excluding weekends) from 6:00 a.m. to 8:30 a.m. and then from 3:00 p.m. to 7:30 p.m. Buses, police cars, ambulances, fire trucks, government and diplomatic vehicles, school buses and vans, and electric and hybrid vehicles are exempt. To decide which half of the fleet is restricted in any given day, the program follows an odd-even schedule based on the last digit of the vehicle's license plate.[4] Compliance with the restriction is effectively enforced with a combination of different measures.[5]

These type of restrictions—which treat all cars the same—have been widely criticized for the perverse incentives they create on drivers to buy additional (often older and more polluting) vehicles, not only increasing the fleet size but also moving its composition toward higher-emitting vehicles, resulting in more congestion and pollution. The best documented evidence supporting this claim comes from Mexico City's *Hoy No Circula* program, as implemented in 1989 (e.g., Eskeland and Feyioglu 1997, Davis 2008, Gallego et al. 2013).

In response not only to this "second-car" concern but also to help finance the public transport system, Bogotá's transport authority introduced a major reform to its *Pico y Placa* program in September 2021: since then, drivers of passenger cars, including SUVs and pickup trucks, have had the option to pay a daily fee to be exempted from the restriction, with the entire fee

[1] See https://inrix.com/scorecard/ for details on the rankings construction. In its November 23th 2021's edition, La Republica (https://www.larepublica.co/), Colombia's main business newspaper, also reports Bogotá as the "most congested city in the world."

[2] Notable exceptions include London, Stockholm, Singapore, Milan, Gothenburg and, starting in 2025, New York City. For more on the political economy of congestion pricing see Baranzini et al (2021) and Calatayud et al (2021).

[3] Other restriction programs include, for example, Athens (where restrictions were first introduced in 1982), Santiago (1986), Mexico City (1989), Teheran (1991), São Paulo (1996), Manila (1996), Cali (2002), La Paz (2002), Medellín (2005), Beijing (2008), Tianjin (2008), several German cities (2008), Quito (2010), Hangzhou (2011), Chengdu (2012), Paris (2016), and Madrid (2019).

[4] Although our analysis covers up to December 2021, it is important to mention that the program has suffered some modifications after that. See footnote #15 for details.

[5] It includes a ground force of more than 1000 agents (divided between police patrols and city officials), a network of more than 200 traffic cameras, high penalties, and confiscation of the vehicle when caught in non-compliance by a ground agent. More details can be found in a series of district decrees (decretos distritales in Spanish), in particular, Decreto 575 (2013), Decreto 515 (2016), Decreto 846 (2019) and Decreto 208 (2020). These decrees are avaible at https://bogota.gov.co.

collection going to the public-transport system.[6]

Of all the possible variations on a driving restriction policy one might think of, the introduction of an exemption fee represents a radical departure from early designs. By allowing drivers to bypass the restriction not by purchasing a second car but by paying an exemption fee, driving restrictions with exemption fees have the potential not only to restore many socially valuable trips that were inefficiently rationed by the restriction in the first place but also to make drivers face the external cost of at least some of their trips.[7]

The restoration of valuable trips, however, comes at the cost of increasing traffic. Our difference-in-difference estimates, based on a large Waze database with vehicle speed records from both Bogotá and Medellín—Colombia's second largest city, serving as control—confirm that the reform reduced city-level speed in Bogotá during peak hours by about 5%.[8] Given this increase in traffic, the introduction of an exemption fee presents authorities with a clear tradeoff: increasing traffic versus restoring valuable trips. Using Bogotá's 2021 reform as motivation and evidence, the objective of this paper is to study this tradeoff, both overall and for each income group in particular.

The main contribution of this paper is to show that such a tradeoff can always be resolved favorably—at least at the overall level. According to our theory (Proposition 4 and Appendix B), there is always a range of exemption fees that leads to overall welfare gains. In the application of the theory to Bogotá, we find that the existing exemption fee—an average of \$8.8 per day (all currency in this paper is in 2020 U.S. dollars)—falls well within this range, resulting in annual gains of \$42 million. In fact, the current fee is close to its optimal average level of \$10.2. These annual benefits decrease slightly, by \$9 million, when we account for the increase in pollution prompted by the reform.[9]

However, consistent with the theory, when evaluating the impact of the reform across different income groups, with varying preferences and options for transportation modes and remote work, we find major differences. The big winners of the reform are middle-income individuals, who now use their cars more often, restoring many of their socially valuable trips that were previously inefficiently rationed. Their gains amount to \$133 million per year.[10]

By contrast, the big losers of the reform are higher-income individuals, who now spend more

[6] Bogotá initiated its reform in September 2020 with a lump-sum exemption fee, when drivers had only the option to purchase a six-month pass, and then, in September 2021, switched closer to a per-trip exemption fee, when drivers were also offered the option to purchase a daily pass. Although our empirical analysis focuses on the impact of the reform from September 2021 onward, our motivating theory also serves to show that a per-trip fee (Proposition 4) is highly superior to a lump-sum exemption fee (Proposition 3). In Section 6, we also (numerically) evaluate this alternative exemption-fee design.

[7] In the limit, when the introduction of the exemption fee is accompanied by an extension of the restriction to every hour of the day and day of the week, we converge to a full-fledged road pricing scheme.

[8] Salgado and Mitnik (2024) is another attempt at using Waze data to study traffic. There are also similar studies using Google Maps, for example, Hanna et al (2017), Akbar and Duranton (2017), and Akbar et al (2023).

[9] Note that our evaluation only considers short-run effects over commuters affected by the restriction (i.e, during peak hours), about 2.06 million individuals. We neglect any effects on fleet composition, other individuals (e.g., those traveling off-peak), or commercial activities.

[10] Low-income individuals also benefit, though less, as the entire fee collection is used to finance the public transport system, which in our application takes the form of a fare reduction.

time in traffic, with losses amounting to $117 million. Two reasons explain these losses. One is that many high-income individuals have access to more than one car, so they have more easily accommodated to the restrictions before the reform. And a second, closely related reason is that these individuals have greater access to remote work. For them, the reform only brought heavier traffic.[11]

We are certainly not the first to study the impact of driving restrictions (including low emission zones) whether on traffic, air pollution, crime activity, fleet size and composition, consumer spending, or commuter welfare (see, e.g., Eskeland and Feyzioglu 1997, Davis 2006, Gallego et al 2013, Wolff 2014, Viard and Fu 2015, Zhang et al 2017, Blackman et al 2018, Carillo et al 2018, Bonilla 2019, Barahona et al 2020, Galdon-Sanchez et al 2023, Salgado and Mitnik 2024). We are the first, however, to look at the impact of introducing an exemption fee into an existing restriction program.[12]

Despite the large number of existing restriction programs, Bogotá is one of the only two programs where exemption fees have been introduced. The other is Cali, also in Colombia.[13] This unusually low use of exemption fees is unfortunate but perhaps not surprising. It may be in part explained by the resulting increase in traffic and the opposition of higher-income individuals.

Here is where Bogotá's reform provides such a valuable policy lesson: despite the increase in traffic, exemption fees can always be made (overall) welfare-enhancing. This is in addition to other benefits such as the possibility of raising extra funds for the public transport system or paving the way toward a full-fledged road-pricing scheme in the future. In fact, if Bogotá were to expand the reform into such an scheme, overall benefits could quintuple, reaching $214 million, not to mention the additional gains in air quality ($17 million).

Natural candidates for the evaluation (and eventual introduction) of these exemption fees include a long list of existing programs, notably *Hoy no Circula* in Mexico City and *Rodízio* in São Paulo. In a similar vein, Bogotá's reform should also serve to call the attention of any authority considering the introduction/expansion of a restriction policy to fight traffic, as it is currently the case in Lima and Santiago. Absent of an exemption fee, no restriction may better than any restriction (Proposition 2).[14]

The rest of the paper is organized as follows. Section 2 contains the empirical analysis.

[11] As discussed in Section 6, our results are qualitatively unchanged to changes in the value of key parameters, such as the elasticity of the congestion function, the level of non-compliance, the post-reform (i.e., late 2021) level of remote work and the value of time. For example, if we assume that the late 2021 level of remote work was 30% higher than its 2019 level (as opposed to 50% as in our baseline), then overall welfare gains from the reform would increase to $55 million, with the same winners and losers. Similarly, if we assume that the (average) value of time is 30% of the hourly wage (as opposed to 50% as in our baseline), then overall gains from the reform would increase to $166 million. This time higher-income individuals would lose a lot less.

[12] Daganzo (2000) and Basso et al (2021) also discuss the potential benefits of exemption fees.

[13] Unlike Bogotá, which, as explained in footnote #6, switched to a per-trip format in September 2021, Cali maintains its lump-sum format. For more on the Cali program, see Soto et al (2023).

[14] If the increase in pollution is also a concern, one could follow the vintage exemptions in Barahona et al. (2020) and make the exemption fee available only to cars with pollution rates below a certain threshold or, alternatively, increase its price accordingly for the more polluting cars.

Theory results are in Section 3. The extension of the theory model to capture Bogotá's transport reality is in Section 4. The impact of the reform on different dimensions—traffic, welfare, and air quality—are studied in Section 5. We provide some policy-design extensions and sensitivity analysis in Section 6. We conclude in Section 7. Some proofs and additional results are collected in the Appendix.

2 Bogotá's Market-Based Reform

In this section we first briefly explain the evolution of Bogota's *Pico y Placa* and then offer an empirical evaluation of the impact of the reform on traffic.

2.1 Bogotá's *Pico y Placa*

Bogotá, Colombia's capital and home to more than 7 million people, has long suffered congestion problems. In response, it introduced in August 1998 a restriction program, better known as *Pico y Placa*, that placed a circulation ban on 20% of the fleet each day of the week (excluding weekends) from 7:00 a.m. to 9:00 a.m. and then from 5:30 p.m. to 7:30 p.m. Over the years *Pico y Placa* has gone through some modifications looking to extend its scope, in particular, with regard to the number of cars restricted on a single day. Since July 2012, *Pico y Placa* affects the vast mayority of residential and commercial vehicles every other day of the week (excluding weekends) from 6:00 a.m. to 8:30 a.m. and then from 3:00 p.m. to 7:30 p.m. Buses, police cars, ambulances, fire trucks, government and diplomatic vehicles, school buses and vans, and electric and hybrid vehicles are exempt. To decide which half of the fleet is restricted in any given day, the program follows an odd-even schedule based on the last digit of the vehicle's license plate.[15]

The 2012 design remained in place until March 19th 2020 when the authority ordered its complete suspension in response to the covid-19 pandemic. As the covid-19 crisis begun to recede, the program was reinstated in September 1st 2020 according to its 2012 design except for a major provision: the possibility for drivers of passenger vehicles, including SUVs and pickup trucks, to pay a congestion fee to exempt their cars from the restriction. At the time, the exemption fee made no distinction between different type of cars and, most importantly, was only available as a six-month pass. Both aspects of the 2020 reform were revised in September 1st 2021. Since then, exemption fees vary according to cars' characteristics—commercial value and pollution rate—and drivers have the flexibility to also pay them on a daily and monthly basis.

Probably, the six-month format as opposed to the daily format does not make much of a difference for drivers who are prepared to pay the exemption fee every time their cars are

[15] Although our analysis covers up to December 2021, it is important to mention that the program has suffered some modifications after that. Since January 2022 the restriction runs uninterrupted from 6:00 a.m. to 9:00 p.m. and since January 2023 it no longer follows an odd-even schedule but a sequential one (e.g., plates ending in 1, 2, 3, 4 or 5 are restricted one day and those ending in 6, 7, 8, 9 or 0 are restricted the next, and so on).

restricted. But for many others, those who are prepared to pay the fee only sporadically, say, once week, it does make a big difference. The six-month format comes closer a lump-sum fee while the daily format comes closer to a per-trip fee (it would be exactly a per-trip fee if cars were used exclusively for commuting purposes). Once the six-month pass is paid, it becomes a sunk investment that does not affect a driver's decision at the margin, i.e., as to whether use her or his car in a particular day. The distinction between these two formats has profound welfare implications. As formally shown in Section 3, a per-trip exemption fee is highly superior to a lump-sum exemption fee, so much that the latter may render useless in some contexts, even under homogeneous drivers.

Since September 2021, the exemption fee that applies to a particular car is the product of a base value, of $8 per day, and a factor that increases with the commercial value of the car and its pollution rate, which weighs both local and global pollutants. Although this factor can be as high as 1.8 for some cars—for 0.1% of the fleet—the relevant factor for 92% of the fleet is 1.2 or less, leading to an average exemption fee of $8.8 per day. By April 2022, the first and only month for which we obtained detailed data from Bogotá's Mobility District Secretary, the total number of exemption fees issued by day in its different formats was anywhere between 25,291 and 60,692.[16]

2.2 The impact of the reform on traffic

For most cities, if not all, traffic after covid-19 did not returned to its pre-covid-19 level, even in the absence of any policy change. This is particularly true for the initial months following the crisis as cities gradually returned to their usual day-to-day activities. For this reason, we evaluate the impact of Bogota's reform on traffic following a Difference-in-differences approach that uses the city of Medellín as control.

Medellín, home to 2.6 million people, is the second largest city in Colombia. Despite their distance—a driving distance of 425 km—Medellín and Bogotá share similar trends in many aspects of economic activity, most importantly for our work, in traffic congestion. In fact, in February 2005 Medellín introduced its own *Pico y Placa* program, placing a circulation ban on 20% of the fleet each day of the week (excluding weekends) from 6:30 to 8:30 a.m. and then from 5:30 to 7:30 p.m. In August 2013, Medellín decided to extend its circulation ban to 40% of the fleet, while delaying its morning start in 30 min, to 7:00 am.

Medellín's program remained unaltered until March 19th 2020 when it was completely suspended in response to the covid-19 crisis. But unlike Bogotá, Medellín reinstated its *Pico y Placa* program not only a year later, in September 6th 2021, but more importantly without giving drivers the option to pay a congestion fee to be exempted from the restriction. The only

[16] The available information only tells us that a particular license plate is associated to the payment of at least one exemption fee during the month, when in fact it could be associated to multiple payments during the month, up to 10 or 11 payments (the number of weekdays a car is restricted during a month). Given the range in the number of exemption fees paid per day, the revenue generated in a year is anywhere between $59 and $143 millions, that is, between 10 and 23% of what the city of Bogotá spent on subsidies to public transport fares in 2021.

difference with its 2013 design is that now the circulation ban applies to only 10% of the fleet, although during the entire working day, from 5:00 a.m. to 8:00 p.m. If anything, the 2021 design appears less stringent than the pre-covid-19 design.

In the rest of the paper we focus on the impact of the reform as of September 2021 onward. We do this not only because it captures the latest changes in Bogotá's reform but also because this is when the *Pico y Placa* programs in both cities were in operation once again, which is essential for our diff-in-diff estimation. In the rest of the section we first describe the data used in the analysis, then offer a justification for using Medellín as control, and finally present the empirical strategy and results.

2.2.1 Data

The data we use in our analysis comes from the Waze application, which collects speed data via the GPS signal of a driver's mobile device on which the application is installed.[17] We use data comprising the urban areas of Bogotá and Medellín from January 2019 through December 2021, omitting the period when only Bogotá had its *Pico y Placa* active, September 1st 2020 through August 31st 2021.

Each city is divided into ZATs (*Zona de Análisis de Transporte* or Zone for Transport Analysis in English) and each ZAT includes several segments (e.g., streets, drives, avenues, etc) for which vehicle speeds are recorded. The city of Bogotá (or Bogotá D.C.) is made of 898 ZATs scattered in 20 counties and the city of Medellín is made of 342 ZATs scattered in 16 counties.[18]

Our unit of observation is the average speed at the ZAT level every 15-minute intervals. Because segments in each ZAT vary by length and whether they exhibit high levels of congestion at a given time interval, our unit of observation comes in four different formats: v_1 is the average velocity or speed considering only highly-congested segments within the ZAT, and v_2 is the average speed recorded on highly-congested segments but weighted by each segment's length. Analogously, v_3 is the average speed of all segments, and v_4 is the average speed of all segments but weighted by each segment's length. We are particularly interested in the results obtained from v_3 and v_4 since our analytical framework does not make any distinction between highly congested and less congested roads. It takes an aggregate view at the city level.

2.2.2 Medellín as control

Figure 1 shows trends in travel speed for both Bogotá and Medellín during the morning peak. The figure is constructed by averaging at the weekly level the natural logarithm of all ZAT records of speed format v_4 for the morning hours during which *Pico y Placa* were active in both cities, that is, weekdays from 6:30 am to 8:30 am (trends for other times of the day are

[17] According to the 2019 Bogotá's Mobility Survey (BMS 2019), Waze is by far the most popular navigation application, with a market share growing by income-group, from 32% to 58%.

[18] Figure A1 in Appendix A depicts ZATs in Bogotá.

discussed below).[19] The vertical line (Week -1 or last week of August 2020) marks the last month in which Bogotá and Medellín shared comparable traffic policies and Week 0 (first week of September 2021) marks the time in which the two cities separated under alternative *Pico y Placa* models, with and without exemption fees.

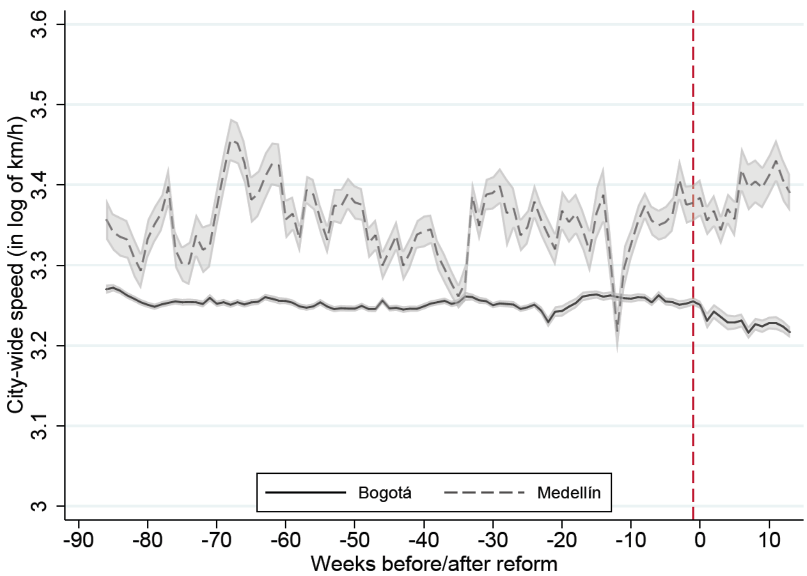

Figure 1: Trends in travel speed during morning hours (6:30-8:30 a.m.) in Bogotá and Medellín

Notes. The figure shows trends in travel speed for both Bogotá and Medellín during the morning peak. It averages the natural logarithm of all ZAT speed v_4 records at the weekly level (95% confidence intervals are also included). The vertical (dashed) line at Week -1 (last week of August 2020) marks the last week the two cities shared comparable traffic policies. Week 0 (first week of September 2021) marks the first week under alternative *Pico y Placa* models, with and without exemption fees, respectively.

Figure 1 suggests a parallel trend lasting until the cities separated with different *Pico y Placa* designs. This is confirmed by the estimation of the following event study for the data in the figure:

$$\ln(v_4^{it}) = \alpha + \sum_{j=1}^{19} \beta_j Lead_{jit} + \sum_{k=0}^{3} \gamma_k Lag_{kit} + \mu_i + \lambda_t + \varepsilon_{it} \tag{1}$$

where v_4^{it} is the (weighted average) speed in ZAT i during the 15-minute time interval t (similar trends are observed for the other three speed formats), $Lead_{jit}$ is a dummy variable that takes the value of 1 for observations coming from the jth month before the first "treated" month, which is September 2021, and Lag_{kit} is a dummy variable that takes the value of 1 for observations coming from the kth month after the first treated month. Thus, we have 19 "leads" (from January 2019 to August 2020) and 4 "lags" (from September 2021 to December 2021). The specification also controls for city fixed effects, μ_i, and time (day of the week and year) fixed

[19] Note that morning restrictions start one hour apart, 6:00 am and 7:00 am, respectively, so we adopt something in between, 6:30 am. Numbers barely change for different windows.

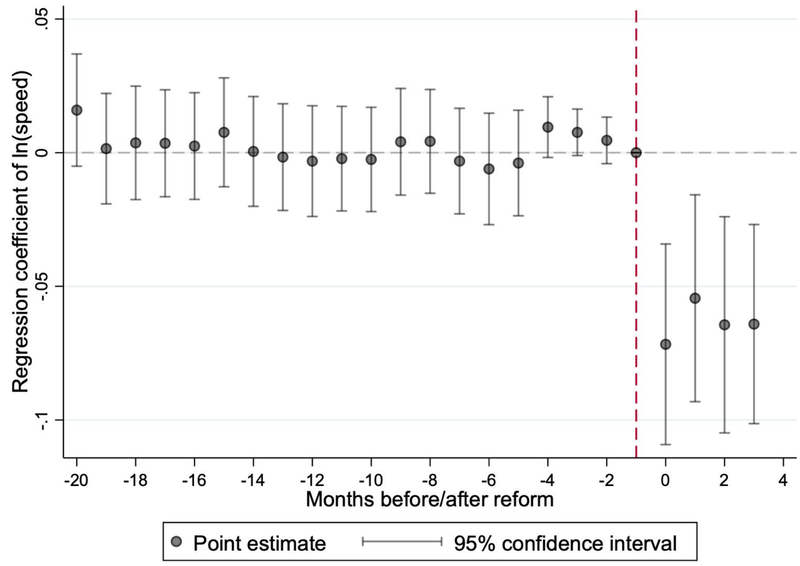

Figure 2: Results from event study comparing Bogotá and Medellín

Notes. The figure presents estimations of the *Lead* and *Lag* coefficients in equation (1) using all ZAT speed v_4 records during morning hours at the monthly level. The vertical (dashed) line at Month -1 (August 2020) marks the last month Bogotá and Medellín shared comparable traffic policies. Month 0 (September 2021) marks the first month under alternative *Pico y Placa* models, with and without exemption fees, respectively. Standard errors, presented in the figure in 95% confidence intervals, are clustered at UTAM level (see footnote #20).

effects, λ_t. The error term is denoted by ε_{it}.[20]

We are interested in the value of the *Lead* (i.e., β_j) and *Lag* (i.e., γ_k) coefficients, which capture the difference between the speed levels in Bogotá and Medellín relative to the omitted base month, August 2020. As shown in Figure 2, estimations of the *Lead* coefficients suggest that traffic patterns in Medellín and Bogotá have followed, for the most part, parallel trends while the two cities shared similar traffic policies. Estimations of the *Lag* coefficients, on the other hand, indicate that traffic patterns in Medellín and Bogotá have moved apart under their different *Pico y Placa* designs. Bogotá exhibits a relative drop in speed of around 7%, a number that is in line with the difference-in-differences estimations that we describe next.

2.2.3 Empirical strategy and results

We estimate the following diff-in-diff equation for the morning peak, weekdays from 6:30 am to 8:30 am:

$$\ln(v_f^{it}) = \beta_0 + \beta_1 Post_t + \beta_2 Bogota_i + \beta_3 Post_t \times Bogota_i + \boldsymbol{\beta}_4' X_t \times Bogota_i + \mu_{it} \qquad (2)$$

[20] We cluster standard errors at the UTAM (*Unidades Territoriales de Análisis de Mobilidad* or Territorial Units for Mobility Analysis in English) level in Bogotá. UTAMs are geographic areas larger than ZATs but smaller than counties. There are 115 of them in Bogotá.

where v_f^{it} is the average speed corresponding to speed format $f = 1, ..., 4$ in ZAT i during 15-min time interval t, $Post_t$ is a dummy variable equal to 1 for time intervals following the reactivation of both *Pico y Placa* programs, $Bogota_i$ is a dummy variable equal to 1 for ZATs located in Bogotá, X_t is a vector of time fixed effects (i.e., the day of the week and month of the year), which we allow to differ across cities, and μ_{it} is the error term. We are interested in the sign and magnitude of β_3, the impact of the exemption fee on travel speed.

As shown in Table 1, we estimate equation (2) for different specifications and samples. Panel A shows results when we use all ZAT records available while Panel B shows results for a 20%-smaller sample that excludes ZATs in Bogotá with significant missing records. As shown in Figure A1 of Appendix A, these ZATs include semi-rural areas in the city's periphery and large green areas within its urban perimeter (e.g., parks, playing fields, cemeteries, golf courses, etc.). Finally, Panel C revisits Panel A with a larger set of fixed effects that control for location (i.e., UTAM; see footnote #20) and its interaction with the time fixed effects, increasing the number of fixed effects from 34 (in Panels A and B) to 2087 (note that Panels B and C only report the coefficients of interest).

Specifications (1)–(4) adopt September 6th 2021 as the time when Medellín reintroduced its *Pico y Placa* program. But since Medellín considered a "pedagogic/trial" period from September 6th to September 20th in which offenders to the restriction were offered the option to engage in a driving education course instead of paying a fine, specification (5) drops the data from this pedagogic/trial period. Finally, specification (6) restricts the pre-treatment period to before the arrival of covid-19 and the suspension of the *Pico y Placa* programs, that is, before March 19th 2020.

Results are consistent across specifications and samples. The numbers in the first row of columns (3)–(6) of either panel indicate that the impact of the exemption fee was a reduction in average speed during morning hours, somewhere between 6 and 9%. In highly-congested segments, columns (1) and (2), this reduction was much larger, consistent with a strictly convex congestion function that is common to congested road systems (see, e.g., Small and Verhoef, 2007).

2.2.4 Additional empirical results

We extend our empirical analysis in different directions (more details are in Appendix A). In our first extension we ignore Medellín as control and simply run a before-and-after regression for Bogotá, which is essentially running (2) without data from Medellín. Results in columns (3)—(6) of Panel A in Table A1 (also for the morning-peak hours) show much smaller reductions in speed than the diff-in-diff estimations of Table 1, of about 2.5%. The problem with these before-and-after estimations is that they do not control for changes in trends unrelated to the treatment, here in particular for traffic changes due to covid-19.

In fact, Figure 1 suggests a slight increase in traffic in Bogotá before treatment, from April 2020 through August 2020. If we run a before-and-after estimation using a shorter pre-treatment

Table 1: Difference-in-differences estimations (6:30-8:30 a.m.)

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln(v_1)$	$\ln(v_2)$	$\ln(v_3)$	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$
Panel A: it includes all ZATs						
$Post \times Bogota$	-0.292***	-0.294***	-0.090***	-0.091***	-0.089***	-0.086***
	(0.044)	(0.044)	(0.015)	(0.016)	(0.016)	(0.016)
$Post$	0.071***	0.064***	0.065***	0.066***	0.065***	0.063***
	(0.021)	(0.021)	(0.003)	(0.003)	(0.003)	(0.003)
$Bogota$	0.418***	0.420***	-0.023	-0.079***	-0.079***	-0.079***
	(0.032)	(0.032)	(0.014)	(0.015)	(0.015)	(0.015)
Observations	1,463,535	1,463,534	1,669,357	1,669,354	1,660,892	1,325,420
Adjusted R^2	0.009	0.009	0.006	0.016	0.015	0.018
Panel B: it excludes "green and rural" ZATs						
$Post \times Bogota$	-0.296***	-0.298***	-0.074***	-0.073***	-0.071***	-0.069***
	(0.044)	(0.044)	(0.011)	(0.011)	(0.012)	(0.012)
Observations	1,217,803	1,217,804	1,399,052	1,399,049	1,392,064	1,108,567
Adjusted R^2	0.010	0.010	0.020	0.040	0.039	0.044
Panel C: it includes UTAM fixed effects						
$Post \times Bogota$	-0.264***	-0.266***	-0.062***	-0.059***	-0.057***	-0.054***
	(0.043)	(0.043)	(0.010)	(0.011)	(0.011)	(0.011)
Observations	1,463,521	1,463,521	1,669,356	1,669,353	1,660,891	1,325,420
Adjusted R^2	0.174	0.173	0.418	0.414	0.414	0.404

Notes. The table reports results from running equation (2) for different speed formats and time windows. All regressions in Panels A and B include time fixed effects (i.e., day of the week and month of the year). All regressions in Panel C include UTAM fixed effects (see footnote #20), both directly and interacting with the time fixed effects. Standard errors, in all regressions, are clustered at the UTAM level in Bogotá and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

window, covering just the first five months of the covid-19 crisis, the drop in speed almost doubles, as shown in columns (3)—(6) in Panel B of Table A1. The covid-19 effects on traffic (i.e., less traffic than otherwise) are likely to be present post-treatment as well, from September 2021 through December 2021. The diff-in-diff estimation is supposed to capture that, obviously, under the assumption that these covid-19 effects are similar in both cities. We have no means to test for that, but it seems a reasonable assumption since, as shown in Figure A2 in the appendix, both cities have followed very similar contagion dynamics.

Our second extension considers different time windows. For example, column (1) of Table A2 reports estimates of running (2) for the evening peak when both *Pico y Placa* programs were active, that is, from 4:15 pm to 7:30 pm.[21] We fail to see any effects, which is surprising. Figure A3 in the appendix, which plots trends in travel speed for both Bogotá and Medellín during the evening peak, offers an explanation. It is evident that the parallel-trend assumption

[21] Evening restrictions also start at different times, 3:00 p.m. and 5:30 p.m. respectively, so again we adopt something in between, 4:15 p.m.

for the entire pre-treatment period does not hold.

Inspired by the synthetic diff-in-diff approach of Arkhangelsky et al (2021), in columns (2) and (3) of the table we report the results of running (2) for a shorter pre-treatment period where the parallel-trend assumption holds, one that places no weight in observations from five or more months away from the treatment.[22] Consistent with the post-treatment trend of Figure A2, we find the reform to have also reduced the speed in Bogotá during the evening peak, anywhere between 2 and 3%.[23] In column (4) we report the drop in speed during the morning peak but for the same shorter pre-treatment period, finding no difference with the results of Table 1. As shown in column (5), pooling observations from columns (3) and (4) into a single equation reports a lower speed fall (3.7%) than the simple average of morning and evening estimates. The latter is more representative of the overall impact of the reform on commuting time, however, as column (5) includes four times more observations for the evening peak than for the morning peak.

Table A2 also reports the effect of the Bogotá reform on hours not affected by the *Pico y Placa* restriction, for instance, between 10:00 a.m. and 12:00 p.m. (noon). As seen in columns (6) and (7), there is an increase in travel speed, between 1 and 2%. We believe this increase in speed is entirely consistent with the decrease during the morning peak: restricted cars covered by exemption fees are on the road now at morning-peak hours and, hence, not longer available to other household members at later (non-restricted) hours.

In a third extension we look at the effect of the reform by income group. We do this by labeling each ZAT in Bogotá to one of the income groups in BMS's (2019) survey. Table 3 below provides a summary of the survey for the different income groups. The analysis that follows collapses the two lowest-income groups into one group (Groups 1&2). We run separate diff-in-diff regressions for each group, using all the ZATs in Medellín as control (all four regressions satisfy the parallel-trend assumption). Results for the morning peak are reported in Table 2 below, with and without expansion factors (which are used to expand individual responses up to an estimate for the entire population).

Either case shows a non-monotonic (U-shaped) impact of the Bogotá reform, with middle-income ZATs (i.e., groups 3 and 4) suffering the largest reductions in speed and ZATs at the two extreme of the income distribution suffering the least. Despite drivers may cross ZATs of different income levels as they complete their trips, these numbers are entirely consistent with our results in Section 5. Unlike middle-income individuals, low and high-income individuals show less interest in buying an exemption fee, but for very different reasons; the former because few of them own a car or can afford the fee while the latter because many of them have already

[22] Alternative approaches (e.g., Borusyak et al. 2024, Callaway and Sant'Anna 2021, Sant'Anna and Zhao 2020) do not seem to work any better, partly because of our high-frequency data, unbalanced panel, and the non-staggered nature of the treatment.

[23] A potential explanation for this smaller impact is that the evening restriction lasts longer than the morning restriction—4.5 vs 2.5 hours. This extended duration spreads additional traffic over a longer period, diluting its effect. Salgado and Mitnik (2024) also find for Lima's driving restrictions that effects at evening (peak) hours are much smaller than at morning (peak) hours.

Table 2: Difference-in-difference estimations by income group (6:30-8:30 a.m.)

	(1) all groups	(2) Groups 1&2	(3) Group 3	(4) Group 4	(5) Group 5
Panel A: with expansion factors					
Post × *Bogota*	-0.091***	-0.055*	-0.103***	-0.052**	-0.048**
	(0.0161)	(0.030)	(0.020)	(0.023)	(0.019)
Observations	1,669,354	347,336	930,587	534,748	147,468
Adjusted R^2	0.016	0.080	0.031	0.024	0.027
Panel B: without expansion factors					
Post × *Bogota*	-0.091***	-0.044	-0.111***	-0.062*	-0.020
	(0.016)	(0.029)	(0.017)	(0.032)	(0.021)
Observations	1,669,354	342,228	906,908	551,338	159,665
Adjusted R^2	0.016	0.087	0.031	0.025	0.028

Notes. The table reports results from running equation (2) for speed format $f = 4$ and different income groups, from the lowest-income groups (groups 1 and 2) to the highest income groups (groups 5 and 6). Only the most relevant information is reported. Panel A reports results when individual responses are expanded to capture their weights in the entire population while Panel B reports results without such adjustment. To facilitate comparison, column (1) of the table replicates column (4) of Panel A in Table 1. All regressions include time fixed effects (i.e., day of the week and month of the year). Standard errors are clustered at the UTAM level in Bogotá (see footnote #20) and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

accommodated to the restriction, with either an additional car or more remote work.[24]

In our final extension we investigate whether nothing else happened at the time when both *Pico y Placa* programs were re-instated. One concern is that in response to the covid-19 crisis many cities around the world reduced their roadways to give more space to pedestrians, cyclists and amenities, notably restaurants and cafes. While it would not necessarily pose a problem for our difference-in-differences estimations if Bogotá and Medellín also adopted similar measures, it would be problematic if they implemented them differently, in ways that could have had differential impacts on travel speeds.

We address this concern by collecting data from Google Maps to rank ZATs in Bogotá and Medellín based on their density of restaurants and cafés per square kilometer within each city (see Figure A4 of the Appendix). As depicted in Table A3 of the Appendix, our diff-in-diff estimations barely change if, for example, we exclude observations from Bogotá's ZATs within the top 25% of the "restaurant/cafe density" ranking. This holds true whether we use all ZATs in Medellín as control or only those within the bottom 75% of the Medellín's ranking. More interesting are the results from conducting separate diff-in-diff regressions for the distinct quartiles in Bogotá, while employing the corresponding quartile in Medellín as control. As illustrated in Figure A5, reductions in speed exhibit a U-shaped pattern. The most significant reductions occur in areas with a moderate concentration of restaurants and cafés, while areas

[24]A larger response from middle-income individuals is also found in other restriction programs, notably, in Mexico City's 1989 *Hoy No Circula* (Gallego et al 2013) and in Santiago's 2017 *Restricción Vehicular* (Fardella et al. 2023).

with high concentrations show virtually no reductions. Since commuting routes are less likely to pass through areas with numerous cafés and restaurants, this reinforces that the reduction in speed on these routes is probably smaller than the 9% reported in panel A of Table 1 and closer to the 7 or 6% reported in panels B or C of the table.

3 Motivating Theory

We now illustrate with a simple model the different forces acting behind the introduction of an exemption fee. For this, consider a unit mass of a continuum of homogeneous drivers (to ease the exposition we relegate the case of heterogeneous drivers to Appendix B.1 and refer to its results as we introduce the application to Bogotá in the next section). Driver i's net surplus from driving can be written as

$$S(x_i, x_{-i}) = B(x_i) - C(x_i) - T(x_i, x_{-i}) \tag{3}$$

where x_i is i's amount of driving in a given period, say a week, and x_{-i} is the overall amount of driving by all the other drivers. The amount of driving can be measured by the number of trips made or kms traveled during the period.

With the goal of illustrating a fundamental tradeoff that motivates much of our work, in this section we adopt very simple forms for the different elements in (3). In the next section (as well as in Appendix B) we extend these forms in different directions to better capture Bogotá's transport reality, most importantly, drivers' heterogeneity.

The benefit of driving is captured by a quadratic (concave) function, $B(x_i) = x_i - x_i^2/2$, so i's inverse demand for driving is the linear $B'(x_i) = 1 - x_i$. Given that a driver always has the option to take the bus or work from home, $B'(x_i)$ must be interpreted as the net benefit of an extra car trip relative to the best alternative option, which could be either complete that trip by bus or cancel it and work from home.[25]

The cost of driving has two components. One is the financial cost of travel, $C(x_i)$, which includes expenses on fuel, parking, lubricants, tires, repairs, and so on. This cost is captured by the linear function $C(x_i) = cx_i$, with $c < 1$. The other component is the time cost of travel, $T(x_i, x_{-i})$, measured in monetary terms. This cost, which is increasing in traffic x_{-i}, is also captured by a linear function, $T(x_i, x_{-i}) = \gamma x_{-i} x_i$.[26] Since x_{-i} can be regarded as the time per km it takes for an individual to travel through the city, γ can be interpreted as her value of time.[27]

[25] For tractability, we assume these outside options to be invariant to any policy intervention. In the application to Bogotá, the option to take the bus is somewhat affected by the policy either by altering its speed, crowding level, or fare. Nevertheless, assuming fixed outside options does not change anything fundamental in the analysis that follows.

[26] We do not make any attempt, neither here nor in the application to Bogotá, to let $C(x_i)$ be affected by the amount of traffic. Fuel consumption may go up at lower speed levels but the probability of having an accident may go down.

[27] More details about this value and its estimation can be found in the next section, in the application to Bogotá.

In the absence of any government intervention, the amount of driving in equilibrium is given by the first-order condition

$$B'(x_i) - C'(x_i) - \partial T(x_i, x_{-i})/\partial x_i = 0 \tag{4}$$

Since all drivers are alike, in equilibrium $x_i = x_{-i}$. Plugging the latter into (4), our simple functional forms yield the following no-intervention amount of driving

$$x^{ni} = \frac{1-c}{1+\gamma}$$

and corresponding consumer welfare $S^{ni} \equiv S(x^{ni}, x^{ni}) = (1-c)^2/2(1+\gamma)^2$.

Given the congestion externality, the no-intervention amount of driving is obviously above the socially efficient (or first-best) level, which is given by

$$x^{fb} = \arg\max_x S(x, x) = \frac{1-c}{1+2\gamma}$$

Proposition 1 *The authority can restore the first-best amount of driving with a congestion fee* τ *per trip equal to* $\tau^{fb} = \gamma x^{fb}$.

Proof. Faced with such congestion fee, i solves $\max_{x_i}\{B(x_i) - C(x_i) - \tau^{fb} x_i - T(x_i, x_{-i})\}$, which yields $x_i = x_{-i} = x^{fb}$. ■

As well known, the reason the first-best is restored is because τ^{fb} is exactly equal to the externality that i imposes upon the remaining drivers evaluated at the optimal level of driving. Depending on the value of γ, restoring the first-best may call for a significant reduction in traffic, $\gamma/(1+2\gamma)$ or 33% when $\gamma = 1$.

As discussed in the Introduction, however, in many instances authorities do not have this market-based instrument at their disposal, so they must rely on alternative instruments. Among these, one that have received much support in practice is the rationing of driving according to the last digit of a vehicle's license plate, a so-called driving restriction. While a congestion fee is also intended to ration the amount of driving, it does it quite differently than a driving restriction. Under a congestion fee, drivers have a choice as to which trips to make and which to cancel (and take the bus or work from home). Obviously, they would cancel only those that report net benefits below the congestion fee, which is socially efficient provided the fee is set at its socially optimal level. Under a driving restriction, in contrast, drivers do not have that choice. At times, they would be forced to cancel highly valuable trips and at others allowed to make car trips of negative social value.

Thus, the main difference between a congestion fee and a driving restriction—leaving aside fiscal considerations—is that the former works as an efficient rationing scheme and the latter does not. One can certainly entertain different views about the extent of this inefficiency. If, following Barahona et al (2020), one adopts the view that a driving restriction works as a proportional rationing scheme—where all trips are equally likely to be canceled—then an "unpleasant" result may follow.[28]

Proposition 2 *A driving restriction that works as a proportional rationing scheme leads to welfare losses unless the congestion externality (i.e., γ) is sufficiently large.*

Proof. Let $R \in (0,1)$ denote the extent of the driving restriction, with $R \to 1$ the case of no restriction and $R \to 0$ the case of full restriction. If x^r_{-i} is everybody else's amount of driving for a given level of restriction R, then the surplus that i actually obtains under proportional rationing is equal to

$$S^r(x^u_i, x^r_{-i}; R) = R(B(x^u_i) - C(x^u_i) - T(x^u_i, x^r_{-i})) \tag{5}$$

where $x^u_i \equiv x^u_i(x^r_{-i})$ is the unrestricted amount of driving that i would pursue when the rest is driving x^r_{-i}, i.e., $x^u_i(x^r_{-i})$ solves (4) for $x_{-i} = x^r_{-i}$. Taking the derivative of (5) with respect to R and applying the envelope theorem leads to

$$\frac{\partial S^r(\cdot)}{\partial R} = (B(x^u_i) - C(x^u_i) - T(x^u_i, x^r_{-i})) - R\frac{\partial T(\cdot)}{\partial x^r_{-i}}\frac{\partial x^r_{-i}}{\partial R} \tag{6}$$

Using the fact that in equilibrium $x^r_i = Rx^u_i = x^r_{-i} = x^r$, our simple functional forms yield

$$x^r = \frac{(1-c)R}{1+\gamma R} < x^{ni}$$

and

$$\frac{\partial S^r(\cdot)}{\partial R} = \frac{(1-c)^2(1-\gamma R)}{2(1+\gamma R)^3} \tag{7}$$

It follows that a necessary condition for a driving restriction to increase welfare is $\gamma > 1$; otherwise is optimal to set $R = 1$, i.e., to have no restriction. ■

Expression (6) helps convey the intuition. Increasing R (i.e., relaxing the restriction) has two effects on i's welfare. Captured by the terms in parentheses, one effect is the direct effect, which is positive. It amounts to the net benefit of marginally increasing i's driving while keeping congestion unchanged. Working in the opposite direction is the indirect or congestion effect. Since $\partial x^r_{-i}/\partial R > 0$ (and $\partial T(\cdot)/\partial x^r_{-i} > 0$), increasing R leads to more congestion and, hence, to higher travel costs. According to expression (7), for the congestion effect to dominate the direct effect, we need $\gamma > 1$; otherwise, the restriction policy will lead to welfare losses, no matter R. Whether $\gamma > 1$ is a demanding condition is ultimately an empirical question to which we will come back in the next section. In our simple model $\gamma > 1$ calls for a first-best reduction of traffic of more than 33%.[29]

Propositions 1 and 2 show not only that restrictions are a poor alternative to congestion fees but also that they can potentially reduce welfare. Does this imply that authorities should

[28] For more on different rationing rules, see Tirole (1988).

[29] In Appendix B.2 we provide an "impossibility result", of an isoelastic demand for driving (with no choke price), in which a restriction always lead to welfare losses for all γ and R. This result holds for heterogeneous drivers as well.

abandon driving restrictions as a tool to curb traffic, even though in many cases they appear to be the only available tool (other than improving public transport)? The answer is no, but subject to a fix. Following what Bogotá did, the fix is precisely to allow drivers to pay a fee that exempts them from the restriction.

As explained earlier, exemption fees can come in different formats, from lump-sum to per-trip based (and anything in between). Bogotá initiated its reform in September 2020 with a lump-sum fee, when drivers had only the option to purchase a six-month pass, and then, in September 2021, switched closer to a per-trip fee, when drivers were also offered the option to purchase a daily pass. Commuters in our application to Bogotá make on average 1.03 round trips per day according to Bogotá's 2019 Mobility Survey (BMS 2019), so a daily fee comes very close to a per-trip fee.[30]

Given their use in practice, we will study both types of exemption fees here, but in the application to Bogotá we will primarily consider the per-trip fee, which is today's relevant case. Although intuitive, the next two propositions show that a per-trip fee is highly superior to a lump-sum fee, so much that the latter may render useless in some contexts, as the following proposition indicates.

Proposition 3 *Consider a driving restriction $R \in (0,1)$ that allows drivers to use their cars in times of restriction upon payment of a lump-sum or fixed fee $F \geq 0$, independent of how much they drive. Assume that the entire fee collection is returned to drivers in a lump-sum fashion. If conditions (i) $\gamma > 1$ and (ii) $\gamma R < 1$ hold, then it is optimal to set the fee at*

$$F^* = (1-R)(1-c)^2/8 \tag{8}$$

so a fraction

$$z^* = \frac{1-\gamma R}{\gamma(1-R)} \in (0,1) \tag{9}$$

of individuals pay the fee. If, on the other hand, condition (i) holds but (ii) does not, then it is optimal to leave the restriction as it is, that is, to set the fee at

$$F^* \geq \bar{F} \equiv (1-R)(1-c)^2/2(1+\gamma R)^2$$

so that nobody pays it ($z^* = 0$). Finally, if condition (i) does not hold and, hence, (ii) does, then it is optimal to terminate the restriction, that is, to set the fee at

$$F^* \leq \underline{F} \equiv (1-R)(1-c)^2/2(1+\gamma)^2$$

so that everybody pays it ($z^* = 1$).

Proof. See Appendix B.3. ■

[30] According to Basso et al (2021), this number is 1.35 in Santiago.

The proposition shows that the conditions under which the introduction of a fixed fee can improve upon a plain restriction are quite limited. Congestion (i.e., γ) must be neither too high nor too low for the fee to be of any help. The reason is that a fixed fee does not have the ability to sort out socially valuable trips from socially non-valuable trips. When congestion is too high, the (traffic) cost of adding non-valuable trips to the road is higher than the benefit of restoring valuable trips, so it is optimal to keep the restriction as it is. On the other hand, when congestion is not that high, the benefit of restoring valuable trips is higher than the cost of adding non-valuable trips to the road, so it is optimal to get rid of the restriction altogether.

A per-trip exemption fee works quite differently. It has the ability to sort out valuable from non-valuable trips. For this reason, it can always be designed in a way to improve overall welfare (as discussed in Appendix B.2, this result extends to the case of heterogeneous drivers).

Proposition 4 *Consider a driving restriction $R \in (0, 1)$ that allows drivers to use their cars in times of restriction upon payment of a per-trip fee $p \geq 0$. Assume that the entire fee collection is returned to drivers in a lump-sum fashion. Let x^{rp} and S^{rp} denote, respectively, the amount of driving and consumer welfare under this (R, p) restriction. Despite the increase in traffic (i.e., $x^{rp} > x^r$), the introduction of a per-trip fee leads to welfare gains (i.e., $S^{rp} > S^r \equiv S^r(x^r, x^r; R)$) for any $p \in (\underline{p}, \bar{p})$, where $\underline{p} \geq 0$ and $\bar{p}$ is the choke price that eliminates the demand for exemptions.*

Proof. See Appendix B.4. ■

That the per-trip fee p of Proposition 4 can always be designed to increase welfare indicates already its superiority over the lump-sum fee F of Proposition 3. The case when is optimal to set F either very low, so everyone pays it, or very high, so nobody does, is straightforward, as p can always be chosen to replicate that outcome. The remaining case, when there is an interior optimal F, where a fraction $z < 1$ of individuals pays it in equilibrium, is less obvious. In any event, such F leads to a highly asymmetric outcome, with a fraction z of individuals facing no restriction, while the remaining fraction continue facing the original restriction. We can always find a value of p that results in the same amount of traffic as that F. However, such p would achieve this by "replacing" low-value trips from individuals who paid F with high-value trips from those who did not. This replacement is necessarily welfare-enhancing.

The work of the exemption fee p is illustrated in Figure 3 for two scenarios, A and B, that differ in the value of time, $\gamma_A > \gamma_B$. Consistent with Proposition 2, a restriction $R < 1$ with no exemption fee is better than no restriction in scenario A ($S^r_A > S^{ni}_A$) while worse in scenario B ($S^r_B < S^{ni}_B$). In either case, and consistent with Proposition 4, $S^{rp} > S^r$ as long as $p \in (\underline{p}, \bar{p})$.

An interesting aspect of Figure 3 is the level of the optimal exemption fee for a given level of restriction, say, $p^*(R)$. It is not obvious how this value compares to the Pigouvian level τ^{fb} (see Proposition 1), which corresponds to the optimal price under full restriction, i.e., $p^*(R = 0) = \tau^{fb}$. The reason is because there are two forces at work. When only a fraction trips can be priced, the regulator would like to set the exemption fee above the first-best level in order to bring the overall level of congestion closer to the first-best level. But since only a fraction of

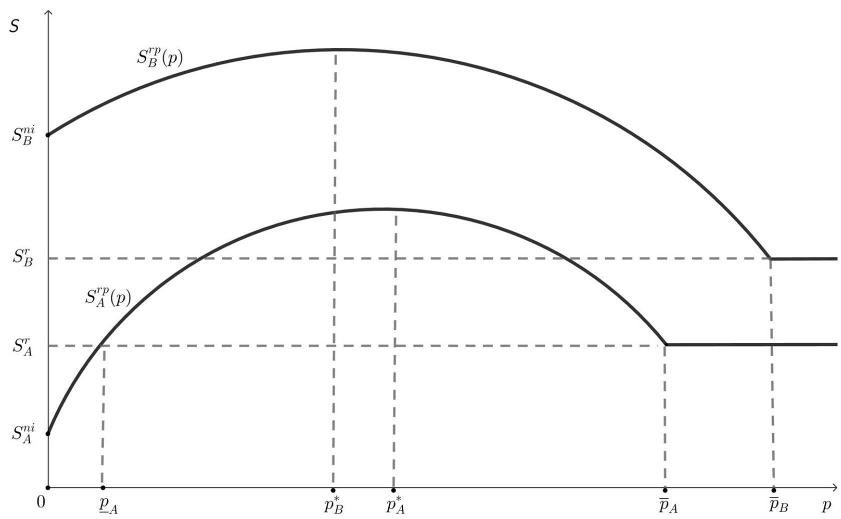

Figure 3: Welfare gains from introducing a per-trip exemption fee

Note: The figure depicts welfare gains for two scenarios, A and B, as a function of the exemption fee, ranging from zero (equivalent to abolishing the existing restriction) to a value high enough that nobody pays it (equivalent to maintaining the restriction in its original form). Scenario A considers a higher value of time than B (i.e., $\gamma_A > \gamma_B$).

cars face a price, this lower level of congestion would encourage drivers of unrestricted cars to increase their trips, some of which are non-valuable from a social point of view. In our simple (homogeneous) setting, the second force dominates so the optimal exemption fee is below the Pigouvian level, as the next lemma indicates.[31]

Lemma 1 *The optimal exemption fee in a (R, p) restriction is given by*

$$p^*(R) = \frac{\gamma(1-c)}{(1+2\gamma+\gamma^2 R)} < \tau^{fb}$$

for any $R \in (0,1)$.

Proof. See Appendix B.5. ■

This homogeneous-driver setting provides us with two important results that motivate much of our analysis. The first is that uniform restrictions, like the one introduced in Bogotá in the late 1990s, can potentially lead to welfare losses (Proposition 2). And the second is that despite the increase in congestion, these uniform restrictions can be fixed, as Bogotá did in 2021, with the introduction of an exemption fee, ideally per-trip based (Proposition 4) as opposed to lump-sum based (Proposition 3).[32]

[31] It is outside the scope of this paper to prove the generality of the lemma, particularly in a context of heterogeneous individuals. In the context of homogeneous individuals, however, it seems to hold quite generally, from highly convex (e.g., $B'(x) = (1-x)/x$ with $x \in [0,1]$) to highly concave (e.g., $B'(x) = \ln(2-x)$ with $x \in [0,1]$) demands.

[32] It is not difficult to see that Proposition 4 extends beyond the linear-quadratic setting. An exemption fee p equal or above the existing traffic externality—γx^{rp} in our case—can only increase welfare.

Having established the superiority of a per-trip fee over a lump-sum fee, for the rest of the paper we will concentrate exclusively on the former, which is today's revelant case in Bogotá. It may sound paradoxical but the welfare gain from introducing an exemption fee is likely to be decreasing in the value of time γ: higher in scenario B than in scenario A of Figure 3. This can be shown formally for the optimal exemption fee.

Lemma 2 *The welfare gain from introducing an optimal exemption fee into a restriction policy $R < 1$, $S^{rp}(R, p^*(R)) - S^r(R)$, is decreasing in the value of time γ.*

Proof. Replace $p^*(R)$ into $S^{rp}(R, p^*(R))$ to obtain that $\partial(S^{rp}(R, p^*(R)) - S^r(R))/\partial\gamma < 0$ for all γ and $R < 1$. ■

This result will prove useful for some sensitivity analyses in the application to Bogotá. Some intuition can be grasped from Proposition 2. An important reason for introducing an exemption fee into an existing (plain) restriction is to fix it. According to Proposition 2, the fix is more desirable the lower the value of γ. When γ is low, as in scenario B of the figure, introducing an exemption fee always carries benefits, even if it is set to zero. In contrast, when γ is high, as in scenario A of the figure, introducing an exemption fee carries benefits only if it is set sufficiently high, above $\underline{p}_A$.

In exploring how much of the potential welfare gain from introducing the exemption fee of Proposition 4 applies to Bogotá, many questions arise. How much of the gain, if any, is due to moving from $p \to \infty$ to $p = 0$ (the difference between S^{ni} and S^r) and how much to moving from $p = 0$ to $p > 0$ (the difference between $S^{rp}(p)$ and S^{ni})? How far is the existing p from $p^*(R)$? How is the welfare gain allocated among different individuals, many of whom may not even own a car? Has the reform left everyone better off? What are the implications of the increase in traffic for air quality? What are the efficiency and distributional implications of letting p to vary according to a car's value and pollution rate? What is the additional gain of deepening Bogotá's reform so as to replicate a full congestion pricing scheme, that is, of moving toward a "restriction scheme" where $R = 0$ and, ideally, $p = p^*(R = 0) = \tau^{fb}$. To answer these and other questions we need to extend our model to more closely capture Bogotá's transport reality.

4 Application to Bogotá

Our homogeneous-driver setting certainly abstracts from elements that may prove relevant in any practical application. The most important is the presence of heterogeneous commuters. One source of heterogeneity is that the demand for driving depends on preferences over and availability of different transportation modes (e.g., car, public transport, etc.) and also on the possibility to work remotely. Other sources of heterogeneity include the extent of the restriction, R, and the cost of being stuck in traffic, γ. As documented by Gallego et al (2013) for Mexico City, high-income households tend to be less affected by the restriction (they perceive a higher R) than middle and low-income households given their access to more than one car. Similarly, high-income drivers tend to value travel time more than their lower-income counterparts, as widely documented in the transportation literature (see, e.g., Small and Verhoef 2007, Basso et al 2021).

In Appendix B.1 we extend our theory to formally study these two sources of heterogeneity, in γ and R. New results emerge, which will certainly help explain some of our findings for Bogotá. One is that high-type drivers (those with high values of γ and R) may benefit from a plain restriction (i.e., a restriction without exemption fees) while low types suffer. And closely related to this is that the introduction of an exemption fee to an existing (plain) restriction, while positive from an overall welfare perspective, can lead to winners (low types) and losers (high types). Furthermore, there may be cases in which high types are always worse off with the introduction of an exemption fee, except when the fee is set so high that nobody pays it.

A second element absent in our simple setting is air pollution associated to vehicle travel, whether at the local or global level. Unlike the restriction policies introduced in Santiago and Mexico City in the late 1980s, which were mainly triggered by increasingly frequent episodes of local air pollution, Bogotá's policy has primarily responded to congestion concerns. It is easy to see that our result in Proposition 4 may not look as favorable in the presence of air pollution (for the same reason that the result in Proposition 2 may not look as negative). The increase in traffic prompted by the the exemption fee ($x^{rp} > x^r$) may lead to higher levels of air pollution that can dissipate, at least partially, some of the gains in consumer surplus ($S^{rp} > S^r$). At the end of Section 5, we attend these air pollution considerations and show that the increase in vehicle emissions have had a rather modest effect on welfare, not affecting our results.[33]

We use the rest of this section to explain first, how our simple setting can be extended to accommodate for commuter heterogeneity and then, how this extended model is calibrated to Bogota's transport reality. We leave the evaluation of the Bogotá's reform for the following section.

4.1 Heterogeneous commuters, public transport and remote work

We consider a standard origin-destination transport model with income and time constraints (see, e.g., Small and Verhoef 2007). On a daily basis, a large number of individuals, say n, must decide whether to commute to the city center to work/study either by car or public transport, or to work/study from home.

Since car owners will transition between weeks with two and three days of restriction, we consider the week to be the relevant planning horizon.[34] Call d_i the number of days of the week

[33] Yet another element absent in our setting is the possibility of buying a second (often older and more polluting) car to bypass the restriction, something that has been documented in other restriction programs (see, e.g., Davis 2008). In policies without an exemption fee (i.e., where $p \to \infty$), this possibility does not change the result in Proposition 2. It basically amounts to a costly investment that only affects the extent of the restriction (higher R), certainly undoing some of the initial gain in traffic (and pollution) reduction (see, e.g., Gallego et al 2013). In what follows, however, we can safely abstract from this "second-car" possibility since we are studying the introduction of an exemption fee into an existing driving restriction. If anything, this should prompt some individuals to sell their "second cars", something we do not evaluate given our short-run focus.

[34] This weekly horizon separates our work from the existing literature and at the same time imposes limits

(excluding weekends) that $i = 1, ..., n$ commutes by car, h_i the number of days that works from home, and $b_i = 5 - d_i - h_i$ the number of days that i uses public transport, i.e., buses; since all public-transport in Bogotá runs on buses, whether as part of the Bus Rapid Transit (BRT) system or zonal buses.

In a model where individuals face income and time constraints, the net surplus that individual $i = 1, ..., n$ obtains after a week of travel can be written as

$$S_i(d_i, h_i, b_i) = B_i(d_i, h_i, b_i) - C_i(d_i, b_i; r_i) - T_i(d_i, b_i; n_c, n_b) \tag{10}$$

where $r_i = 0, ..., 5$ measures the extent of the restriction, i.e., the number of days i's car, provided she owns one, is restricted from circulation during the week,[35] n_c is number of individuals that commute by car in any given day and n_b is the number of individuals that commute by bus, so $n_h = n - n_c - n_b$ is the number of individuals that work from home. Given the large number of individuals, the partition (n_c, n_b, n_h) is invariant to the day of the week. Unlike the previous section, the functions $B_i(\cdot)$, $C_i(\cdot)$ and $T_i(\cdot)$ now vary across individuals.

The benefit of travel depends on i's intrinsic (relative) preferences for each transport mode and remote work as follows

$$B_i(d_i, h_i, b_i) = \lambda_i^{-1}[d_i + \theta_i b_i + H_i(h_i)]$$

where λ_i corresponds to i's marginal utility of income (i.e., the Lagrangian multiplier for the budget/income constraint),[36] θ_i captures i's preference for public transport relative to private transport, and $H_i(h_i)$ corresponds to the benefit of remote work relative to private transport, which we capture with the linear demand $H_i'(h_i) = \vartheta_i - \xi_i h_i$. In the next section we explain how to obtain values for the parameters λ_i, θ_i, ϑ_i and ξ_i.

In turn, i's weekly financial travel cost is given by

$$C_i(d_i, b_i; r_i) = c_i d_i + p_i \max\{0, d_i + r_i - 5\} + f b_i \tag{11}$$

where c_i is the daily cost of using a car (set to infinity for those individuals who do not own one), p_i is the exemption fee (set to infinity before the reform) and f is the daily expense on

as to how much structure and geographic granularity we can include in the model. For instance, it would be hard to follow the structural approaches of Durrmeyer and Martinez (2022) and Barwick et al (2024). Our weekly horizon is open to dozens of bundle options—one cannot rule out individuals who find it optimal to combine public transport with cars and remote work in a certain way this week, and in a different way next week—making it difficult to estimate preferences for bundles, let alone running counterfactual scenarios, for data-availability and computational reasons (origin-destination surveys typically ask individuals for their travel decisions in a given day, not in a given week). Moreover, even under our simpler structure is not trivial to add geographic differentiation. This would require to determine how these bundle options change for each possible origin-destination pair (for example, access to public transport is likely to vary, sometimes greatly, from pair to pair; information we do not have).

[35] In an odd-even restriction, half of the cars will face two days of restriction in a given week and the other half three days of restriction in that week.

[36] Note that by including i's marginal utility of income we are assuming that transport-related expenditures have non-trivial income effects. This is well documented, particularly for lower-income individuals (see, e.g., Small and Verhoef 2007).

public transit (i.e., the product of single-ride fare and the average number of daily rides), which is the same across individuals. In contrast, we let individuals to face different exemption fees to account for the fact that they may vary by vehicle type. Values for all these financial-cost parameters are obtained from external sources.

Two observations regarding how the driving restriction enters into (11) are in order. The first is that we allow the extent of the restriction to vary across individuals with different access to cars. In particular, and following the evidence documented by Gallego et al (2013), we let individuals in households with two or more cars to face a milder restriction, more precisely, one less day of restriction a week than the nominal level.[37]

The second observation is that individuals are expected to accommodate to the restriction. For example, an individual that faces a week with two days of restriction ($r_i = 2$) would not need to spend on exemption fees if she is planning to use the car only three days ($d_i = 3$); the days of restriction would be those in which she either works from home or takes public transit. Note that this flexibility, if anything, would work against the result in Proposition 2 that a restriction without an exemption fee may be welfare decreasing.

Finally, i's time cost of travel per week is expressed as follows

$$T_i(d_i, b_i; n_c, n_b) = \lambda_i^{-1} \left[\gamma_i^c t^c(n_c) l d_i + \left(\gamma_i^b(n_b) t^b(n_c) l + \gamma_i^w w^p \right) b_i \right] \tag{12}$$

where γ_i^m is i's marginal utility of time (i.e., the Lagrangian multiplier for the time constraint) when using transport mode $m \in \{c, b\}$, $t^m(n_c)$ is the time per unit of distance spent on transport mode m on any given day, l is the average distance traveled in a round trip from home to work including any shorter trips during the day, γ_i^w is the marginal utility of time when waiting at the bus station, and w^p is the average waiting time at the station. Following Basso and Silva (2014), we assume that that $\gamma_i^w = 2\gamma_i^c$.

We allow γ_i^c and γ_i^b to differ and also to control for any inconvenience that may result from increasing public-transport use without the corresponding adjustment in service frequency. Following Tirachini et al (2017) we let

$$\gamma_i^b(n_b) = \gamma_i^c \left(1 + \zeta \frac{n_b l}{ysqL} \right) \tag{13}$$

where ζ is a crowding penalty, y is the bus frequency, s is the average bus size, q is the duration

[37] An individual with two cars with license plates ending in even and odd numbers, respectively, could completely bypass the restriction by alternating cars. At the other extreme, two drivers in a household with two cars and in need of them at restricted hours will be as restricted as a driver in a single-driver household with one car. The reality for most households with two or more cars lies in between. We have two options to accommodate for this: either allow one day less of restriction (equivalent to reduce from 5 to 3 restricted days in two weeks) or allow two days less of restriction (equivalent to reducing from 5 to 1 restricted days in two weeks). We think the reality is closer to the first option. Thus, we assume that individuals in households with 2 or more cars would alternate between weeks of 1 and 2 days of restrictions, while the rest of individuals between weeks of 2 and 3 days of restriction.

of the peak period,[38] and L is length of the road network.[39]

To model travel times t^c and t^b we adopt a standard Bureau of Public Roads (BPR) function (see, e.g., Small and Verhoef, 2007, p.76)

$$t^m(n_c) = t_f^m \left(1 + \alpha_m \left(\frac{y\kappa + n_c l/qL}{K}\right)^{\beta_m}\right) \tag{14}$$

where $t_f^m = 1/v_f^m$ is the free-flow travel time of mode $m \in \{c, b\}$, v_f^m is the free-flow speed of mode $m \in \{c, b\}$, κ is an equivalence factor between buses and cars, K is the capacity of a road lane (maximum number of cars per hour a road lane can absorb without affecting travel time and taking into account traffic signals), and α_m and β_m are positive parameters. With the exception of K and β_c, which are estimated jointly with preference parameters, values for all the other travel-time-cost parameters, including marginal utilities of time, are obtained from external sources.

The decision problem of individual i is to chose d_i and h_i or b_i (recall that $b_i = 5 - d_i - h_i$) so as to that maximize (10), while taken as given the equilibrium choice of the remaining individuals, that is, taken as given n_c, n_b and n_h. According to David and Fourcat (2014), a game like ours, with network externalities, may accept multiple equilibria. There are two reasons, however, this potential multiplicity is less of a problem here than in David and Fourcat (2014). One is the fact that public-transit quality is exogenous (i.e., determined outside the game), so Morhing's (1972) positive externality from public-transit use is absent in our setting. And the second reason is that in our model public transit become less attractive (i.e., more crowded) as more people switch to it. We only share with David and Fourcat (2014) the fact that buses run faster as more people switch to public transport, leaving behind less congested roads. Whether this network externality alone is enough to generate multiplicity is something that none of our simulations supports.

4.2 Parameter values and calibration

The model is parameterized to capture Bogotá's traffic and air pollution reality, using information from both before and after the 2021 reform. Most importantly, this reality accounts for the fact that in any given week, half of Bogotá's commuters face two days of restriction and the other half three days of restriction.

We are interested in commuting trips at peak hours, which is when restrictions apply. Bogotá's 2019 Mobility Survey (BMS 2019) registers a total of 523,766 car trips during peak hours, 400,877 or 76.5% are classified as essential, which we interpret as commuting trips to work or school. These essential trips differ from the rest in terms of their duration—they take 32% longer than the non-essential trips— and their geographic scope—only 15% of the

[38] Since l is the round-trip average distance, q includes duration of both morning and evening peaks.

[39] The difference between γ_i^b and γ_i^c is similar to the difference in Basso and Silva (2014), i.e., about two times larger.

essential trips stay within the same UTAM (*Unidades Territoriales de Análisis de Mobilidad* or Territorial Units for Mobility Analysis in English) while 60% of the non-essential trips do.[40]

Furthermore, BMS (2019) indicates that the modal split at peak hours between public transport and cars is 78.4% and 21.6%, respectively (this split varies substantially across income groups, as seen in Table 8 below). PBGGSD (2021), on the other hand, documents that by 2019, 10% of otherwise commuting trips are replaced by work from home (these numbers also vary significantly across income groups, as seen in the same Table 8). Putting these statistics together we arrive at 2.06 million commuters at peak hours—the ratio between 400,877 and the product of 0.9 and 0.216.[41]

Our application, including calibration and policy analysis, focuses exclusively on these 2.06 million commuters, which we scale down to $n = 10,000$ for computational reasons, while keeping the rest of the transportation variables unchanged. This involves two assumptions. One is that (peak) commuters never consider moving some of their trips to off-peak hours in response to a policy intervention (this is already in our model).[42] This seems reasonable given the commuting nature of the trips we are considering. It is also consistent with Basso and Silva (2014), who find low substitution possibilities between peak and off-peak travel while making no distinction between essential and non-essential trips. The second assumption is that non-essential trips at peak hours are not affected by the policy. Again, this seems reasonable since many of these non-essential trips are sporadic and local in nature, occurring on roads outside the main network of commuting roads and avenues.

Most of the relevant information for our application (including car ownership, use of private vs. public transport, amount of remote work, value of time, etc.) is available at the income-group level, so we follow the characterization in BMS (2019) and cluster our individuals according to their income levels in five income groups: (1) low, (2) middle-low, (3) middle, (4) middle-high and (5) high.[43] We use $g = 1, ..., 5$ to denote the income group.

As shown in Table 3, groups are of different sizes (they are not quintiles). Not surprisingly, the table shows substantial heterogeneity in several dimensions. For instance, cars are significantly used only by the higher income groups, while the majority of individuals in the lower income groups rely heavily on public transport.

Individuals vary not only in whether they own a car but also in the type of cars they own. This impacts the exemption fee, p_i, as it varies with car characteristics, namely the value of the car and its pollution rate. For each of these two dimensions, authorities have classified all

[40]BMS (2019) divides the city of Bogotá in 115 UTAMs, which serve as basis to gather and process information.

[41]Note that this computation completely neglects, as done, for example, by Durrmeyer and Martinez (2022), the fact that a small fraction of car trips, less than 10% according to Bogotá's Mobility District Secretary (BMDS 2021), involve two or more passengers. Besides it is a small fraction, there is no simple way to incorporate this car-sharing decision into our model. For instance, it would require to endogenously find the most suitable individuals willing to coordinate their weekly schedules.

[42]Likewise, essential off-peak trips, if any, are never moved to peak hours in response to the policy.

[43]The only difference with BMS (2019) is that we collapse its high-income groups 5 and 6 into a single high-income group.

Table 3: Individual characteristics by income group

Income group	Fraction of total	Income per-capita	Car ownership	More than one car	Marginal utility of time ($/hr)
1. Low	11%	100	11%	1%	0.71
2. Middle-low	40%	157	21%	2%	1.82
3. Middle	34%	273	39%	6%	3.57
4. Middle-high	10%	588	66%	16%	7.48
5. High	5%	850	82%	36%	14.13

Notes: The table is elaborated with information from different sources: BMS (2019), Bogota's Mobility District Secretary (BMDS 2021), and Colombia's 2022 Great Integrated Household Survey (GIHS 2022). Values shown under Income per-capita and Marginal utility of time correspond to group average (to facilitate the comparison, the average income per-capita of the low-income group was normalized to 100).

cars registered in Bogotá into three ranges: low, medium, and high.[44] Cars with a commercial value up to $12,500 are classified in the low-value range while cars with a commercial value of $27,500 and above are classified in the high-value range. Similarly, cars with a pollution rate up to 0.25 are classified in the low-pollution range while cars with a pollution rate of 0.4 and above are classified in the high-pollution range.

Based on these classifications, the exemption fee p_i corresponding to each car in the fleet is the product of a baseline exemption fee of $8 and the factor in Table 4. Thus, exemption fees vary from $8, for the cleanest and cheapest cars, to $15, for the most polluting and expensive cars. As shown in Table 5, however, there are very few drivers that face such high exemption fees. The large majority of drivers face exemption fees of $9.6 or less, which results in an average exemption fee of $8.8.

Table 4: Exemption-fee factors

Commercial value \ Pollution rate	Low	Medium	High
Low	1.00	1.10	1.20
Medium	1.25	1.38	1.50
High	1.50	1.65	1.80

Note: A car's exemption fee p_i is the product of a baseline value and the corresponding factor shown in the table, which is increasing in its commercial value and pollution rate. Each characteristic is classified as either low, medium, or high.

To determine the types of cars owned by individuals in different income groups, we use information from BMDS (2021). The result is shown in Table 6, which displays the fraction of each type of car by income group. Perhaps surprisingly, these fractions are not that different across income groups, with a great concentration of cars in the low-value, high-pollution range. We allocate cars randomly within each income group while maintaining the fractions from the

[44]The information used by authorities is in BMDS (2021).

Table 5: Fraction of cars in each value-pollution category

Commercial value \ Pollution rate	Low	Medium	High
Low	55.31%	23.93%	12.48%
Medium	5.96%	1.41%	0.36%
High	0.25%	0.30%	0.01%

table. For the purposes of our (short-run) analysis, we assume that these fractions are invariant to both the policy intervention and any effects the covid-19 crisis crisis might have had on fleet composition.

Table 6: Car portfolio by income group

Commercial value	Low (L)			Medium (M)			High (H)		
Pollution rate	L	M	H	L	M	H	L	M	H
Group 1	17.1%	32.8%	47.8%	0.1%	0.5%	1.5%	0.0%	0.1%	0.1%
Group 2	14.4%	28.0%	54.3%	0.2%	0.6%	2.2%	0.0%	0.2%	0.1%
Group 3	13.1%	23.9%	56.6%	0.3%	1.1%	4.6%	0.0%	0.3%	0.2%
Group 4	10.1%	20.8%	57.8%	0.4%	1.7%	8.5%	0.0%	0.3%	0.3%
Group 5	10.6%	21.8%	49.9%	0.8%	3.4%	12.2%	0.0%	0.6%	0.7%

An important policy parameter that the model in the previous section is silent about is the use of the exemption-fee collection. We assume that the entire fee collection goes to the public transport system, as Bogotá currently considers.[45] There are certainly different forms to allocate these resources into the system. In our application, we assume that all of them are used to reduce the existing public-transport fare f.

Individuals also vary greatly in their valuation of time, as shown in the the last column of Table 3. Following the literature (e.g., Small and Verhoef 2007, Yang et al 2020, Durrmeyer and Martinez 2022), we assume that individuals' marginal utility of time vary around half of their hourly market wage.[46] Market wages were obtained from Colombia's 2022 Great Integrated Household Survey (GIHS 2022) and adjusted to their 2019 values using Colombia's general wage index. The values in the last column of Table 3 correspond to the group-average marginal utility of time, say, $\bar{\gamma}_g$. We let $\gamma_i \equiv \gamma_{i\in g}$ to be drawn independently from a (truncated) normal

[45]In the Extension section we discuss alternative ways to recycle the fee collection, in particular, give it back to individuals as lump-sum transfers.

[46]As recently stressed by Barwick et al (2024), the marginal utility of time is perhaps the most important preference parameter for transportation decisions, explaining different attempts at measuring it. Results vary, with estimates often ranging from 30% to 70% of the hourly wage, including the 67% recent estimate of Durrmeyer and Martinez (2022) for Paris and the 65% of Almagro et al. (2024) for Chicago. Unfortunately, there are no equivalent studies for Bogotá. In Section 6, we conduct sensitive analysis around our baseline (average) value of 50%, which also happens to be the value recommended by many transportation authorities, including the US Department of Transportation. Results do not qualitatively change, but nevertheless some interesting inter-group dynamics emerge.

distribution with mean $\bar{\gamma}_g$ and standard deviation $\bar{\gamma}_g/5$.[47]

Values for the remaining financial- and travel-time-cost parameters of the model are summarized in Table 7. There are two parameters in the table that deserves further explanation. One is K, the capacity of the road lane, and the other is β_c, the curvature of the BPR function (14) for cars. These parameters are estimated jointly with the remaining parameters of the model—marginal utilities of income and preferences for transport modes and remote work—using a a two-stage iterative approach that uses information from both before and after the 2021 reform.[48]

Table 7: Summary of financial- and travel-time-cost parameters

Parameter (units)	Symbol	Value	Source
Trip length (km)	l	27.8	BMS (2019)$^{(a)}$
Network length (km)	L	2,171	Transmilenio$^{(b)}$
Passenger car equivalence factor for buses	κ	2.06	Basso and Silva (2014)
Public transport fare ($/day)	f	1.5	BMDS (2021)
Average waiting time at station (min)	w^p	2	Basso and Silva (2014)
Car operating cost ($/day)	c	16.4	BMDS (2021)$^{(c)}$
Lane capacity (car/h)	K	515 (10.9)	Own estimation$^{(d)}$
Free-flow speed – cars (km/h)	v_f^c	43	BMDS (2021)
Free-flow speed – buses (km/h)	v_f^b	30	BMDS (2021)
Bus frequency (bus/h)	y	13.4	BMDS (2021)
Bus average size (m^2)	s	26.4	BMDS (2021)
Crowding penalty	ζ	0.2	Basso et al (2021)
Parameters of BPR function – cars	α_c	0.15	Basso and Silva (2014)
	β_c	3.41 (0.08)	Own estimation$^{(d)}$
Parameters of BPR functions – buses	α_b	0.23	Basso and Silva (2014)
	β_b	1.05	Basso and Silva (2014)

Notes: (a) The value considers two trips per day of approximately 12.5 km each. (b) Transmilenio 2021: Estadísticas de oferta y demanda del Sistema Interconectado de Transporte Público (SITP). (c) This is the operating cost of a car in the middle-value range. The costs in the low- and high-value ranges are 10% lower and higher, respectively. (d) Standard errors in parentheses. See text for details on the estimation.

The first stage of our two-stage calibration exploits pre-reform (and pre-covid-19) information and proceeds according to the following steps. First, we take guess values for K and β_c. Second, we let the income distribution of our simulation sample of $n = 10,000$ commuters—half of which face a week with two days of restriction and the other half with three days of restriction—replicate the actual income distribution observed in BMS (2019). Third, we let $\lambda_i = \lambda_0/Y_i$, where Y_i is i's income and λ_0 is a scaling factor be estimated together with the preference parameters. Fourth, we let car ownership in each group replicate the distributions

[47] Distributions, including those below, are truncated at the 1% and 99% levels to prevent any negative values.

[48] Note that including α_c, α_b, and β_b in the estimation does not make much difference since we have only two observables—speed change and the number of exemption fees paid in response to the reform—to estimate K and β_c. See the text below for more details.

observed in columns 4 and 5 of Table 3.

Fifth, we let $\theta_i \equiv \theta_{i\in g}$ be drawn independently from a (truncated) normal distribution with mean $\bar{\theta}_g$ and standard deviation σ_g^θ. Sixth, based on PBGSD (2021), which documents that the demand for remote work has shown to be increasing in income,[49] we let $\xi_{i\in g} = \xi_0(6-g)$, where ξ_0 is a constant to be estimated. In addition, we let $\vartheta_i \equiv \vartheta_{i\in g}$ to be drawn independently from a (truncated) normal distribution with mean $\bar{\vartheta}_g$ and standard deviation σ_g^ϑ.

Seventh, we reduce the number of preference parameters to be estimated following Basso et al (2021) in that the variance of the distribution of these parameters is assumed to be related to the number of people owning a car in the group, directly in one case (for the remote-work preferences) and inversely in the other (for the transport preferences). Otherwise, it would be hard to explain, for example, why some individuals in low-income groups are so keen to use their cars. Thus, we let $\sigma_g^\theta = \omega^\theta/\pi_g^c$ and $\sigma_g^\vartheta = \pi_g^c/\omega^\vartheta$, where π_g^c is the fraction of individuals owning a car in group g—as indicated in the fourth column of Table 3. This reduces the number of parameters to be estimated in the first stage to fourteen: λ_0, ξ_0, $\bar{\theta}_1, ..., \bar{\theta}_5$, $\bar{\vartheta}_1, ..., \bar{\vartheta}_5$, ω^θ, and ω^ϑ.

As the last step of the first stage, we assign commuters to the different income groups according to the proportions and characteristics of Table 3 and their corresponding distribution functions. The estimation of these 14 parameters is done by minimizing the sum of the square of the difference between what the model predicts and the pre-covid-19 observation of both public vs. private transport use (modal share) and remote work at the income-group level and overall. The second column in Table shows group-level observations on modal shares, sourced from BMS (2019), while the fourth column shows group-level observations on remote work from PBGSD (2021).

Table 8: Model calibration — first stage

Income group	Public transport use: Observed	Public transport use: Model Prediction	Remote work: Observed	Remote work: Model Prediction
1. Low	95%	97%	1%	2%
2. Middle-low	80%	84%	3%	5%
3. Middle	66%	67%	13%	15%
4. Middle-high	42%	42%	26%	26%
5. High	22%	21%	35%	34%
Overall	70%	72%	10%	12%

Note: The table shows how our model matches observed (i.e., surveyed) data for the calibrated parameters. The first and second columns contrast (pre-covid-19) observed modal shares of public transport to the predictions of our model. The third and the fourth columns do the same for remote working.

[49] In fact, PBGSD (2021) shows that approximately 35% of workers in the IT and financial sectors often telework, in contrast to only 10% of workers in the manufacturing sector. These numbers are consistent with those obtained in a survey conducted by the UC Berkeley in Bogotá (Rodriguez et al 2021), indicating that 81% of lower-income individuals believe they will not be teleworking once the covid-19 pandemic is over, in contrast to the 40% of higher-income individuals.

We now move to the second-stage, i.e., to the estimation of K and β_c using the 14 parameters estimated in the first stage. This second stage exploits post-reform (and post-covid-19) information, particularly changes in traffic speed and the number of exemption fees paid. Recall from Section 2 that our post-reform information includes speed data for the period from September 2021 to December 2021 and exemption fee data for April 2022. Therefore, for the purposes of this paper, the post-reform period is supposed to cover from September 2021 to April 2022.

Based on the empirical analysis of Section 2, we assume the 2021 reform reduced traffic speed during peak hours by about 5.5%—the average between an 8% reduction during morning hours and a 3% reduction during evening hours. As explained in that same section, the available information on the number of exemption fees paid in any given day is not as precise, suggesting it can range from 25,291 to 60,692 exemptions.[50] Under the assumption that most, if not all, exemption fees paid go to cover (otherwise restricted) essential trips during peak hours, we adopt 55,000 exemptions in our calibration.[51]

Another piece of information that enters into the second stage is the level of remote working during the post-reform period, from September 2021 to April 2022. PBGSD (2021) suggests that the overall amount of remote work doubled by early 2021 due to covid-19, increasing from 10% to 20%. Since then, it has been in constant decline, as observed in other places (see, e.g., Delventhal and Parkhomenko, 2024). According to a survey commissioned by the country's National Department of Statistics, 14.6% of Bogotá's workers operated remotely by early 2022.[52] We assume that remote working experienced an overall increase of 50% during our post-reform period, rising from 10 to 15%.[53] We capture this shift in our model by increasing the constant ξ_0 accordingly, that is, to exactly yield an overall level of 15% of remote work while estimating K and β_c. As in the first stage, the latter is done by minimizing the sum of the square of the difference between what is observed and what the model predicts for speed change and exemption fees paid, properly normalized.

With the estimations of K and β_c we return to the first stage and repeat the sequence until convergence is reached. Note that our two-stage sequence assumes that preferences for public vs. private transport, as well as the fleet composition, remain invariant at their pre-reform (i.e., 2019) levels. While we cannot rule out that covid-19 might have altered some of that, we lack the information to incorporate it.

The first-stage estimated parameters are in Table 9 and how they fit the model predictions to the actual (i.e., survey) data is in columns three and five of Table 8. The second-stage

[50] See footnote #16 for an explanation regarding this range.

[51] Note that lower values lead to estimates for β_c that are clearly off from values often found in the literature. Nonetheless, adopting a number of exemptions larger than the actual number is consistent with accepting some level of non-compliance with the reform, which is reasonable. We elaborate further on this non-compliance possibility in the Extensions section.

[52] See https://nearshoreamericas.com/colombians-remote-working-office/

[53] In Section 6, we conduct sensitive analysis around our baseline value of 15%. Results do not qualitatively change.

estimated parameters are in Table 7.[54] They predict a drop in speed and exemption fees paid—6.2% and 62,983, respectively—above the numbers used in the calibration—5.5% and 55,000 respectively.

A better fit would result if we started from a higher number of exemption fees paid. Strictly speaking, our model assumes full compliance with the restriction policy while in reality there may be some level of non-compliance, despite its relatively high non-compliance fine, of almost \$100.[55] From conversations with Bogotá's Mobility District Secretary, full compliance is a reasonable assumption for the pre-covid-19 period but perhaps less so for the post-covid-19 period. Not only detecting non-compliance has become more demanding, as enforcement agents must also verify the validity of the exemption, but also drivers are acting less socially responsibly.[56]

Table 9: Preference parameters

Parameters	Preference for car		Remote work	
Income group	$\bar{\theta}_g$	σ_g^θ	$\bar{\vartheta}_g$	σ_g^ϑ
1. Low	-4.21 (0.079)	2.27	-8.41 (0.078)	0.01
2. Middle-low	-3.30 (0.081)	1.19	-3.09 (0.070)	0.02
3. Middle	-1.78 (0.073)	0.64	-0.70 (0.075)	0.04
4. Middle-high	-0.15 (0.039)	0.37	0.35 (0.028)	0.07
5. High	-0.09 (0.059)	0.30	0.05 (0.021)	0.08

Notes: Standard errors in parentheses. Those of σ_g^θ and σ_g^ϑ have been omited, as they only depend on the standard errors of ω^θ and ω^ϑ, respectively, which have been quite precisely estimated. The estimation also includes values for the scaling factor for the marginal utility of income, $\lambda_0 = 0.074$ (0.0014), and the slope of remote working demand, $\xi_0 = 0.04$ (0.0051).

Before we move to evaluating the 2021 reform, it is interesting to observe in Table 9 that while higher-income individuals have, on average, stronger preferences for cars, the estimations for lower-income individuals exhibit a much larger standard deviation. This is an indication that some lower-income individuals value their cars more than their higher-income counterparts. It is also interesting to observe that the demand for remote work is highly non-linear with respect to income. For example, individuals in Group 5 value the first unit of remote work, on average, less than individuals in the next highest group (Group 4), despite having 45% more income.

[54] Note that the estimation of β_c is well within values often found in the literature (see, e.g., Small and Verhoef 2007).

[55] See https://www.valoraanalitik.com/2022/12/26/pico-y-placa-estas-son-las-sanciones-por-infringir-en-bogota-en-2023/. Compliance with the program would nevertheless be relatively high according to our model. For instance, our model predicts 456,362 essential car trips when the exemption fee is set to zero and 369,012 when it is at its current level of \$8.8. The difference, 87,350, corresponds to the number of drivers in compliance with *Pico y Placa*: 62,983 by paying the exemption fee and 24,367 by leaving their cars at home. Suppose the number of exemption fees actually paid is lower, say, 45,500, leading to 17,483 non-compliant drivers (the difference between 62,983 and 45,500). This results in a non-compliance rate of 20% (the ratio between 17,483 and 87,350). Given this rate and the current non-compliance fine of \$100, our model suggests that one in five (risk-neutral) drivers assign a probability of being caught in non-compliance of 8.8% or less. For the remaining four drivers, that probability would be higher than 8.8%.

[56] We have also seen a surge in evasion in several public-transport systems.

Yet, individuals in Group 5 work more from home, on average, than individuals in Group 4, 35% more. The reason is that they care twice as much about spending time in traffic.

5 Policy Evaluation

In this section, we evaluate the impact of the September-2021 reform on different economic variables, including traffic, welfare, and air pollution. Our evaluation focuses exclusively on short-time effects—the first few months—so we do not evaluate any potential impact on fleet composition.

There are two natural counterfactuals one can think of when evaluating the impact of the reform. One is to assume that Bogotá could have followed Medellín and reintroduced its *Pico y Placa* in September 2021 without the option to pay for an exemption fee. The alternative counterfactual is to assume that Bogotá could have postponed the reintroduction of its *Pico y Placa* indefinitely or for at least another year (in our model, this would be equivalent to having introduced a free exemption). The evaluation that follows uses the first counterfactual, which is also the one used in the second stage of the calibration.

5.1 Impact on traffic

Our model predicts that the reform caused a drop in city-level speed of 6.2% during peak hours—from a pre-reform level of 21.81 km/h to a post-reform level of 20.46 km/h—and a daily purchase of 62,983 exemption fees. This response resulted in 45,337 more essential car trips during peak hours, a 14.0% increase over the 323,675 essential trips that would have occurred in the absence of the reform, according to our model. Note that the latter figure is significantly less than the 400,887 essential trips used in the first stage of our calibration. There are two reasons for this. First, the amount of remote working in the first stage of the calibration is much smaller than its post-reform level—10 vs. 15%. Second, as seen from Table 8, our calibration fits a slightly larger share of buses over cars than what we actually observe, 72% instead of 70%.

These changes in speed and essential car trips reflect an elasticity of speed with respect to the number (or density) of car riders of -0.44. This number is not different from values found in the literature (see, e.g., Ardekani and Herman 1987, Small and Verhoef 2007, Geroliminis and Daganzo 2008, and Yang et al. 2020),[57] but it is certainly higher, in absolute terms, than the estimates from Akbar and Duranton (2017) for Bogotá, ranging from -0.06 to -0.20.

More interestingly, the actual increase in traffic is only a fraction of the number of exemption fees paid, 72%. Figure 4, which illustrates the car-use response to the reform as a function of income, offers an explanation (to avoid cluttering, the figures plots a (random) sample of 700 car owners). It shows that the increase in traffic has prompted many individuals in higher-income

[57] For example, Panel A in Figure 3 of Yang et al. (2020) suggests a speed-density elasticity of -0.55 for Beijing. Barwick et al (2024) estimate an even larger (city-wide) elasticity, of -1.10, also for Beijing.

groups, particularly in group 4, to abandon their cars in favor of more remote work or public transport. For example, the individual identified as C in the figure reduced her driving to work by four days.

Individuals in these higher-income groups are still purchasing some exemption fees, but not as many as individuals in lower-income groups, particularly in group 3 (for example, individuals A and B in the figure increased their driving to work by two and three days, respectively). In fact, our model predicts that car owners in group 3 purchased almost four times as many exemption fees as car owners in group 4, 13.3 vs. 3.5%, once normalized by the total number of cars owned in each group.

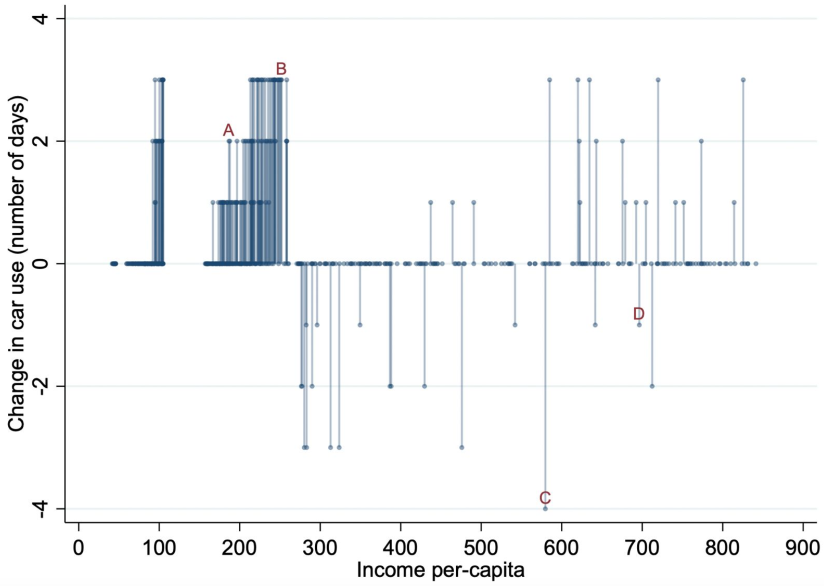

Figure 4: Car-use response to the 2021 reform as a function of income

Notes: The picture plots car-use responses to the reform for a random sample of 700 car owners, ordered by income. Recall from Table 3 that the per-capita income of the lowest-income group is normalized to 100. Capital letters identify car-use patterns followed by specific individuals.

Unfortunately, we cannot contrast these numbers with the numbers of exemption fees actually paid by drivers from different income groups. While we have information on the number of exemptions fees actually paid under the different factors of Table 4—see Table 10 below—there is not much we can infer from these numbers given the symmetric allocation of cars across the different income groups that we observe in Table 6.

Table 10: Fraction of exemption fees actually paid

Commercial value \ Pollution rate	Low	Medium	High
Low	60.45%	11.34%	3.03%
Medium	17.66%	3.31%	0.57%
High	1.85%	1.78%	0.02%

5.2 Impact on welfare

Figure 4 summarizes much of the impact of the 2021 reform across individuals of different income levels, suggesting winners and losers. The first column of Table 11 documents the impacts under the existing exemption fee, \$8.8 on average. It is clear that the big winners of the reform are middle-income individuals (groups 2 and 3) who now use their cars more often, restoring many of their socially valuable trips that before were inefficiently rationed. Their welfare gains amount to \$133 million a year. Individuals in group 1 have also benefited, but from a lower fare.

Table 11: Welfare impact of the reform and alternative policies$^{(a)}$

	Exemption fee$^{(b)}$			Road pricing$^{(c)}$	
Income group	Existing	Free	Optimal	Public Transport	Neutral
1. Low	25.78	-0.07	24.70	78.01	-0.08
2. Middle-low	73.87	111.25	65.04	176.44	-36.95
3. Middle	59.04	157.05	41.77	27.12	148.29
4. Middle-high	-36.07	-88.10	-29.64	-62.82	-25.41
5. High	-80.57	-244.34	-53.76	-5.08	124.22
Total	42.05	-64.21	48.11	213.67	210.08

Notes: (a) Numbers in the table are in million of dollars. (b) The first three columns show welfare impacts for each income group and overall for three exemption fee levels, respectively: the existing level (p_i = \$8.8 on average), free ($p_i = 0$ for all, equivalent to abolishing the restriction), and the optimal level (p_i = \$10.2 on average). (c) Columns 4 and 5 consider two full-fledged road pricing schemes with fees set at their optimal levels, \$7.5 and \$7.4 respectively. In column 4 the collection fee is recycled back in the public transport system in the form of lower fares while in column 5 is returned lump-sum to individuals in a neutral way, i.e., preventing transfers across groups.

By contrast, the big losers of the reform are higher-income individuals (groups 4 and 5) with losses that amount to \$117 million a year. There are two reasons that explain their losses. One is that many of these individuals have access to more than one car (see Table 3), so they have more easily accommodated to the pre-reform restrictions. And a second, closely related reason is that these individuals have greater access to remote work. Imagine an individual who faces a week with two days of restriction. He or she could completely prevent the destruction of valuable car trips by combining the use of a second car during one of the days of restriction and work from home during the other. For these individuals the reform has only made their car trips longer.

For these reasons, and consistent with the theory results with heterogeneous drivers, high-income individuals can only lose from the introduction of an exemption fee. In fact, the second column of Table 11 shows their losses could triple under a free exemption fee, which is equivalent to abolishing the restriction. Consistent with Scenario A in Figure 3 of the theory section, abolishing the restriction is a bad idea overall, not to mention that low-income individuals (i.e., group 1) would also be worse off, from both slower buses and fares that remain at their pre-reform levels. The only ones who would gain from abolishing the restriction are car owners in groups 2 and 3 who would then use their cars freely.

Overall, the best course of action would be to increase the exemption fee to \$10.2, its (average) optimal level. However, the only individuals who would support such an increase are those in groups 4 and 5, particularly car owners, as they would see an improvement in traffic relative to the existing situation (speed would increase by 1.3 percentage points). Individuals in all the other groups would be worse off, although for different reasons: car owners would face more expensive exemption fees, and bus riders would experience higher fares. In fact, the fee collection reaches its maximum at \$7.5. Considering all these factors, including the relatively small overall gain, it is difficult to support an increase of the exemption fee from its current level.

More promising is to leverage the existing reform to advance a full-fledged road pricing scheme. The last two columns of Table 11 show the welfare gains from such an initiative, provided congestion fees are set at their optimal levels, around \$7.5. From an overall perspective, it does not matter much whether the fee collection is used to reduce public transport fares, as is currently done (fourth column), or returned lump-sum to individuals in a neutral way, i.e., preventing transfers across groups (fifth column). In either case, overall benefits could more than quintuple existing benefits. What is clear from these two distinct recycling options, however, is that leaving all consumers better off is unlikely, unless the authority has access to targeted transfers, which is rarely the case.

5.3 Impact on air pollution

So far, we have omitted from the analysis any impact of the 2021 reform on air pollution, whether at the local or global level. There is a reason for this. Unlike the restriction policies introduced in Santiago and Mexico City in the late 1980s, which were mainly triggered by increasingly frequent episodes of critically high local air pollution, Bogotá's policy has mostly responded to congestion concerns. The numbers that follow confirm this.

The pollution rate of a car is important not only to determine its exemption-fee factor but also to estimate its contribution to the air pollution costs borne by society before and after the reform. To estimate these pollution costs we use the same pollution rates used by the authority to classify cars in Tables 4 and 5. These pollution rates are based on a composite of local and global pollutants weighted by their pollution harm according to the responses of a group of 10 experts consulted by the authority.[58] In this composite, fine particulate matter accounts for 50.4%, while carbon dioxide accounts for 18.5%; the remaining 31.1% corresponds to other local pollutants such as carbon monoxide and nitrogen oxides.

In our policy evaluation, we assign to each car the average pollution rate of the pollution category to which it belongs, rather than its actual pollution rate, which is information we do not have. Specifically, these category rates are 0.1, 0.3, or 0.5.[59] Base on these composite pollution rates, our model predicts that the reform has increased emissions (relative to the counterfactual of a *Pico y Placa* without an exemption fee) by 6.5%.

According to SDG (2018) the pre-reform social cost of fine particulate matter in Bogotá from all passenger vehicles (including SUVs and pickup trucks but excluding commercial and industrial trucks) amounted to \$68.4 million a year. Since the weight of particular matter in this composite index is 50.4%, our estimate is that the reform increased pollution damages by \$8.8 million per year, calculated as the product of 68.4 and 0.065, divided by 0.504. These losses could turn into \$16.7 million in benefits if Bogotá were to expand the 2021 reform into a comprehensive congestion-pricing scheme, like the one described in the fourth column of Table 11. According to our model, such a scheme would lead to a drop in (composite) emissions of 12.3%.

6 Extensions

In this section, we explore some additional policy-design questions and how our results might change with variations in some key parameters, namely, the shape of the congestion (BPR) function, the level of compliance with the reform, the post-reform level of remote work, and the value of time. Let us discuss the latter first. Table 12 reports results when we let the value of these parameters move away from their baseline levels used to construct Table 11. Qualitatively speaking, these sensitivity analyses confirm our main results: the 2021 reform led to important overall gains but also resulted in winners (middle-income individuals) and losers (high-income individuals).

Quantitatively speaking, the numbers in Table 12 also serve to confirm that our results change as expected; for instance, that the (overall) welfare gains from introducing an exemption fee into an existing restriction increase as we assume a more elastic BPR (congestion) function, more compliance, lower levels of remote work and lower values of time. The first column of the table examines a more elastic BPR (congestion) function by increasing α_c in equation (14) fourfold.[60] As a result, the calibrated value of β_c decreases by 32%, to 2.32, while the calibrated

[58] More detail can be found in BMS (2021).

[59] The fact that cars in the high-pollution range are five times more polluting than cars in the low-pollution range is amply consistent with the evidence in Kahn (1996), Barahona et al (2020) and Jacobsen et al (2023), for example. They document that this wide range is mostly explained by the high pollution rates of older vehicles.

[60] Our baseline case considers $\alpha_c = 0.15$, commonly found in the literature (Small and Verhoef 2007), and calibrates $\beta_c = 3.41$, not far either from what is commonly found in the literature either, a value of 4.0. Some

Table 12: Sensitivity analysis: BPR, non-compliance, remote work and value of time[(a)]

	BPR[(b)]	N.C.[(c)]	Remote work[(d)]		Value of time[(e)]			
		Pico y Placa			*Pico y Placa*		Road Pricing	
Income group	×4	11%	30%	70%	30%	70%	30%	70%
1. Low	23.99	3.72	28.85	21.56	24.88	22.30	64.82	63.29
2. Low-middle	70.30	34.41	83.05	60.96	74.19	82.22	137.85	71.04
3. Middle	61.44	79.51	87.59	45.75	70.83	55.10	86.41	-8.44
4. Middle-high	-23.21	-26.78	-41.34	-28.78	-13.92	-34.51	-39.42	-34.63
5. High	-59.21	-69.27	-102.84	-57.96	10.45	-94.63	23.84	72.38
Total	73.31	21.60	55.31	41.54	166.43	30.47	273.50	169.72

Notes: (a) Numbers in the table are in million of dollars. For variations in the BPR congestion function, level of non-compliance (N.C.) and remote work, we only report welfare impacts for the existing *Pico y Placa* reform, so their numbers should be compared to the ones in the first column of Table 11. (b) The first column considers a fourfold increase in the value of α_c in the BPR congestion function compared to our baseline value of 0.15. (c) The second column allows for an 11% of non-compliance with the reform, corresponding to individuals circulating with randomly allocated "free" exemption fees. (d) The two columns under the heading remote work consider two post-reform levels of remote work around our baseline level: 30% and 70% increase from the pre-reform level of 10%. (e) The last four columns consider two average values of the marginal utility of time, 30% and 70%, around our baseline level of 50% of the hourly wage for two policy formats, the existing *Pico y Placa* and an optimal road-pricing scheme. Numbers from the last two columns should be contrasted with numbers in the fourth column of Table 11.

value of K (road capacity) increases by 40%, to 690. This more elastic BPR function leads to a smaller drop in speed due to the reform, of only 4.5% (as opposed to 6.2%), leading to a 74% increase in overall welfare, mainly explained by smaller losses suffered by higher-income individuals.

The second column of the table introduces non-compliance in a "reduced form", by randomly allocating a number of free exemption fees across the entire commuter population. A 5% random allocation, for example, leads to 11% of non-compliance: of the 55,779 restricted cars on the street, 49,604 are covered by costly exemption fees while the rest, 6,175 or 11%, by free exemption fees.

Table 12 also show the effects of varying the post-amount of remote work and of values of time (γ) in either direction. That higher values of γ lead to smaller overall welfare gains may seem less intuitive but is entirely consistent with Proposition 2 and Lemma 2. The reason is that welfare under a plain driving restriction (i.e., without an exemption fee) is increasing in γ: it may deliver benefits under a high γ but losses under a low γ. As seen from last two columns of the table, this logic extends if we were to replace the existing restriction with a road-pricing scheme.[61]

recent studies consider higher values of α_c and correspondingly lower values of β_c. Pan et al (2023, Table 5), for example, uses $\alpha_c = 0.56$ for Los Angeles and 0.81 for Beijing together with $\beta_c = 3.26$ and 2.16, respectively.

[61] Note that the welfare gain from moving from the no-intervention benchmark to a restriction with an exemption fee is nevertheless increasing in the value of time, as expected. This trend may revert only for very high values of γ, when higher-income individuals massively start adopting remote work.

Recall that Bogotá initiated its market-based reform in September 2020 with non-discriminatory lump-sum exemption fees, when drivers had only the option to purchase a six-month pass. The September 2021 reform, which is the focus of our study, introduced two important changes. First, it let fees vary with car characteristics—with its commercial value and pollution rate. And second, it switched closer to a per-trip exemption fee by giving drivers the option to purchase a daily pass. We study the impact of each change separately. While there is no congestion-based reason to discriminate across cars (although there are pollution- and safety-based reasons),[62] discrimination appears to have a minor welfare impact when we compare the first column of Table 13 to the first column of Table 11.[63] If anything, it helped higher-income groups, but this was not easy to anticipate given the fleet symmetry across income groups observed in Table 6.

This minor impact extends if the authority were to implement a road-pricing scheme, as shown by the second and fourth columns of Tables 13 and 11, respectively. An important policy lesson from these exercises is that the use of varying fees can facilitate their introduction without compromising their welfare goals. Communicating that more expensive and polluting cars will face higher exemption fees, and that the entire fee collection will be allocated to the public transport system, may help persuade the public to more easily support these fees.

Table 13: Alternative policy designs$^{(a)}$

	Per-trip fees$^{(b)}$		Lump-sum fees$^{(c)}$	
Income group	*Pico y Placa*	Road Pricing	*Pico y Placa*	Road Pricing
1. Low	26.48	80.39	18.97	91.66
2. Low-middle	80.86	177.69	60.20	239.45
3. Middle	69.73	10.20	35.22	65.72
4. Middle-high	-40.58	-61.11	-26.54	-129.24
5. High	-98.24	17.16	-57.88	-104.85
Total	38.25	224.33	29.96	162.75

Notes: (a) Numbers in the table are in million of dollars. The table compares the performance of daily (or per-trip) fees vs lump-sum fees for two different policy formats: the existing *Pico y Placa* and a road pricing scheme. Both formats consider uniform pricing, so all drivers face the same fees. (b) Daily fees are, respectively, \$8.8, the existing *Pico y Placa* average level, and \$7.5, the optimal road-pricing level. (c) Lump-sum fees are, respectively, the product of the daily fees and ten days of restrictions, the number of restricted days in a two-week period, the relevant planning horizon in our model for an odd-even restriction with lump-sum fees.

The minor impact of varying fees contrasts sharply with the major impact of using lump-sum fees as opposed to per-trip fees. As anticipated by Propositions 3 and 4, switching to a per-trip fee results in an overall welfare gain of 27.7%, as shown by the first and third columns of Table 13. The second and fourth columns of the same table indicate that this gain is even larger, 37.8%, in a road-pricing scheme, where drivers pay every time they use their cars, explaining

[62] The congestion-pricing scheme in London is a good example where congestion fees vary with the car's pollution rate. According to a recent article in *The Economist*—America's killer cars, September 7th 2024—-these fees should probably also vary with the vehicle's weight.

[63] This welfare impact does not include the 29.2% increase in pollution damages, however, from \$8.8 to \$11.4 million.

why lump-sum fees hit higher-income groups particularly hard then.

7 Final Remarks

In September 2021, the city of Bogotá introduced a major market-based reform to its *Pico y Placa* driving restriction. Since then, drivers have had the option to pay a daily congestion fee to be exempt from the restriction. This reform offers valuable efficiency and equity lessons. The introduction of an exemption fee inevitably presents authorities with an efficiency tradeoff: more traffic in exchange for more socially valuable car trips that were previously rationed.

One of this paper's key contributions, however, is showing that this tradeoff can always be resolved favorably, as demonstrated in Bogotá, resulting in significant overall efficiency gains. This finding holds across a wide range of—and possibly all—parameter values for key variables, including the elasticity of congestion, the level of compliance, the post-reform amount of remote work, and the value of time. But even if overall efficiency gains happens to be modest, a second key contribution of this paper is demonstrating that the introduction of an exemption fee can result in significant wealth transfers from higher-income individuals to their middle-income counterparts. As observed in Bogotá, the restoration of valuable social trips primarily benefits the latter, while the increase in traffic disproportionately impacts the former.

Appendix A

A.1 Figures

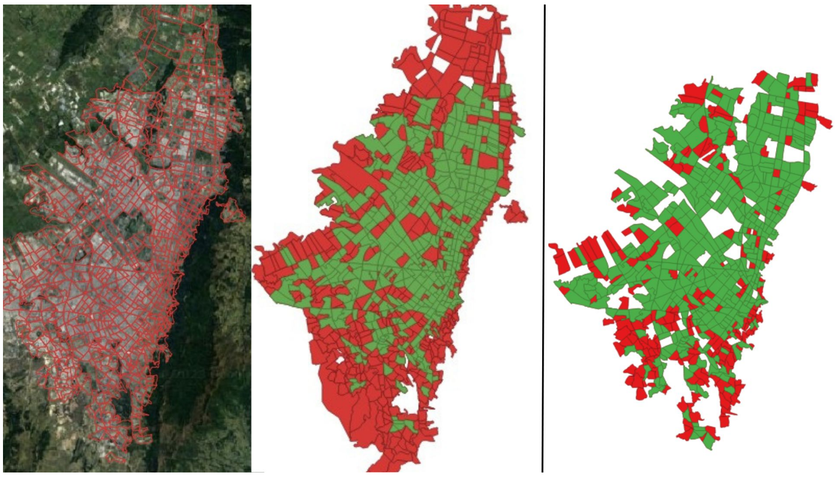

Figure A1: Zones of Transport Analysis (ZATs) in Bogotá

Notes: The picture on the left shows the 898 ZATs that make the city of Bogotá. The picture in the middle shows the number of ZATs, pictured in green (or light gray in a black-and-white display), with available data at a given 15-min interval, in this case at 7:30 a.m. on July 15th, 2019. By looking at these two pictures, we can see that many of the ZATs with missing data correspond to rural areas in the city's periphery and urban green spaces (e.g., parks, playing fields, cemeteries, golf courses, etc.). Discarding these "rural/green" ZATs, we are left with the ZATs depicted in the picture on the right, which reduces the sample in almost 20%.

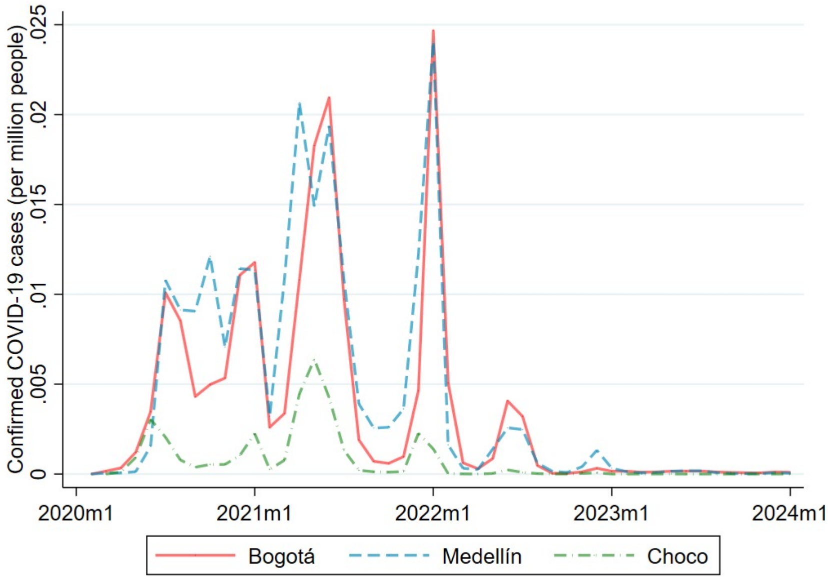

Figure A2: Evolution of covid-19 cases in different Colombian cities

Notes: The figure shows the monthly evolution of covid-19 cases in three Colombian cities—Bogotá, Medellín and Choco—based on the country's official figures (https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia-/gt2j-8ykr/about_data). The figure includes the Choco region, located west of Bogotá and Medellín, solely to highlight that Bogotá and Medellín have exhibited remarkably similar contagion dynamics, with a correlation exceeding 0.9.

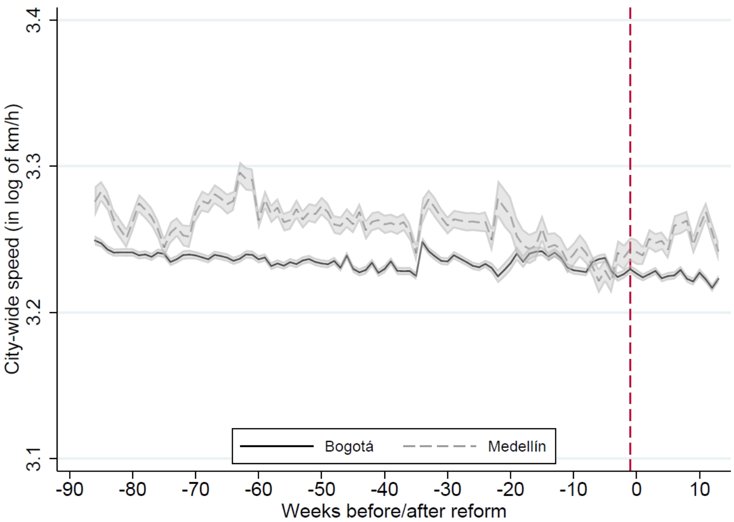

Figure A3: Trends in travel speed during evening hours (4:15-7:30 p.m.) in Bogotá and Medellín

Notes: The figure shows trends in travel speed for both Bogotá and Medellín during the evening peak. It averages the natural logarithm of all ZAT speed v_4 records at the weekly level (95% confidence intervals are also included). The vertical (dashed) line at Week -1 (last week of August 2020) marks the last week the two cities shared comparable traffic policies. Week 0 (first week of September 2021) marks the first week under alternative *Pico y Placa* models, with and without exemption fees, respectively.

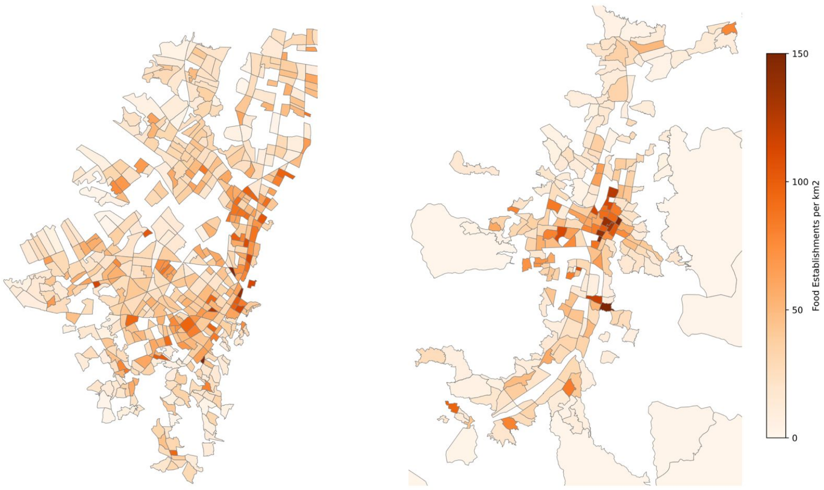

Figure A4: Concentration of cafés and restaurants in Bogotá and Medellín

Note: The figure shows ZAT-level density of food establishments—cafés and restaurants—in Bogotá and Medellín based on a sample of 12,945 establishments, 66% of which are located in Bogotá, collected from Google Maps.

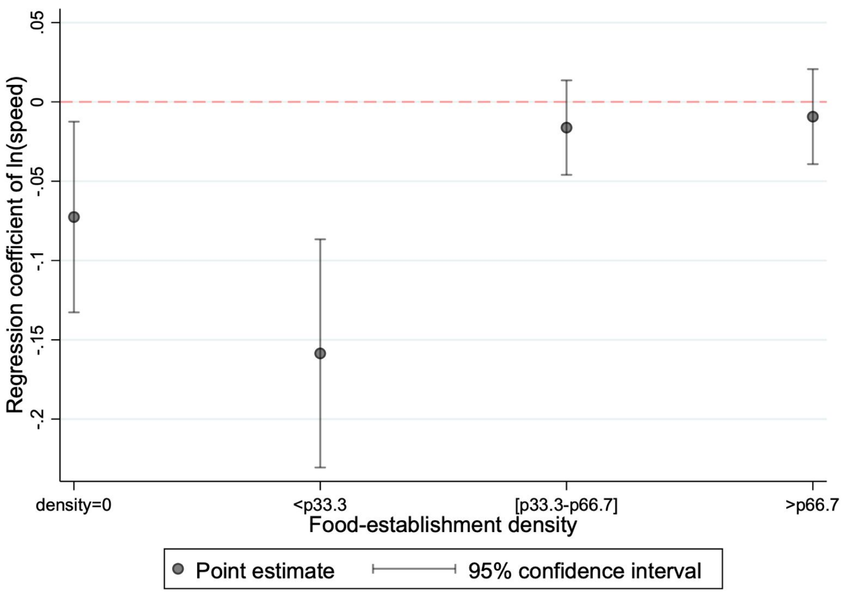

Figure A5: Diff-in-diff estimations by "food" density in Bogotá and Medellín

Notes: This figure shows point estimates and 95% confidence intervals for the $Post \times Bogota$ coefficient (i.e., β_3) after running equation (2) for different quantiles of food-establishment density. The first quantile includes ZATs with zero density. The remaining three quantiles equally split the ZATs with positive densities.

A.2 Tables

Table A1: Before-and-after estimations (6:30-8:30 a.m.)

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln(v_1)$	$\ln(v_2)$	$\ln(v_3)$	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$
Panel A: Long pre-treatment period (20 months)						
Post	-0.2194***	-0.2274***	-0.0253***	-0.0246***	-0.0238***	-0.0235***
	(0.0066)	(0.0067)	(0.0008)	(0.0009)	(0.0010)	(0.0010)
Observations	1,439,324	1,439,324	1,565,126	1,565,123	1,558,238	1,237,718
Adjusted R^2	0.009	0.009	0.001	0.001	0.001	0.001
Panel B: Short pre-treatment period (4 months)						
Post	-0.4452***	-0.4432***	-0.0407***	-0.0429***	-0.0430***	.
	(0.0320)	(0.0318)	(0.0039)	(0.0043)	(0.0043)	.
Observations	235,417	235,417	286,095	286,095	279,210	.
Adjusted R^2	0.037	0.037	0.005	0.004	0.004	.

Notes: The table reports results from running equation (2) without Medellín. All regressions include time fixed effects (i.e., day of the week and month of the year). Standard errors are clustered at the UTAM level in Bogotá (see footnote #20) and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Diff-in-Diff estimations for different time windows and pre-treatment periods

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$
Time window	4:15-7:30 p.m.	4:15-7:30 p.m.	4:15-7:30 p.m.	6:30-8:30 a.m.	pooling a.m.-p.m.	10-12:00 a.m.	10-12:00 a.m.
Pre-treatment	Long	Short	Short	Short	Short	Long	Short
Bogota $\times$ *Post*	0.005	-0.024***	-0.028***	-0.089***	-0.037***	0.016**	0.019**
	(0.003)	(0.007)	(0.006)	(0.018)	(0.006)	(0.007)	(0.008)
Observations	4,659,824	1,532,237	1,343,152	316,070	1,611,295	2,269,464	627,210
Adjusted R^2	0.003	0.002	0.011	0.030	0.015	0.014	0.018

Notes: The table reports results from running equation (2) for different time windows and pre-treatment periods, long (20 months) and short (4 months). All regressions include time fixed effects (i.e., day of the week and month of the year). Standard errors are clustered at the UTAM level in Bogotá (see footnote #20) and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Diff-in-Diff estimations after excluding ZATs with restaurants and cafés (6:30-8:30 a.m.)

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln(v_1)$	$\ln(v_2)$	$\ln(v_3)$	$\ln(v_4)$	$\ln(v_4)$	$\ln(v_4)$
Panel A: it excludes ZATs in Bogotá in top 25% of food-density ranking						
Bogota $\times$ *Post*	-0.2866***	-0.2896***	-0.0942***	-0.0925***	-0.0909***	-0.0880***
	(0.0462)	(0.0468)	(0.0175)	(0.0188)	(0.0187)	(0.0185)
Observations	1,159,458	1,159,457	1,344,616	1,344,614	1,337,434	1,064,570
Adjusted R^2	0.010	0.010	0.004	0.012	0.011	0.014
Panel B: it excludes ZATs in Bogotá and Medellín in top 25% of food-density ranking						
Bogota $\times$ *Post*	-0.2306***	-0.2338***	-0.0920***	-0.0861***	-0.0840***	-0.0809***
	(0.0462)	(0.0468)	(0.0175)	(0.0188)	(0.0187)	(0.0185)
Observations	1,154,653	1,154,652	1,318,482	1,318,480	1,311,557	1,042,375
Adjusted R^2	0.008	0.008	0.006	0.015	0.015	0.018

Notes: Following Table 1, the table reports results from running equation (2) for different speed formats and samples during morning hours. Panel A drops all ZATs in Bogotá ranked in the top 25% of the "food density" ranking. Panel B drops in addition all ZATs in Medellín ranked in the top 25% of the "food density" ranking. All regressions include time fixed effects (i.e., day of the week and month of the year). Standard errors are clustered at the UTAM level in Bogotá (see footnote #20) and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B

B.1 Heterogeneous drivers

We extend the model to allow for heterogenous drivers. They differ in two important dimensions, the value of time (γ) and the extent of the restriction (R). These two variables are highly correlated, as both are increasing in income. Higher-income individuals not only value time more than their lower-income counterpars but also they tend to better accomodate to the restrictions, as they usually own more than one car.

Thus, consider two type of drivers, high (H) and low (L), who exist in proportions $\mu \in (0, 1)$ and $1-\mu$, respectively, and differ in their value of time ($\gamma_L < \gamma_H$) and the extent of the restriction ($R_L < R_H \leq 1$). In the absence of any intervention, it is easy to anticipate that both types will continue driving above their first-best levels, with high-income individuals driving relatively less (and working more remotely):

$$x_L^{ni} = \frac{(1-c)(1+\mu\Delta_\gamma)}{1+\bar{\gamma}} > \frac{(1-c)(1-(1-\mu)\Delta_\gamma)}{1+\bar{\gamma}} = x_H^{ni}$$

where $\Delta_\gamma \equiv \gamma_H - \gamma_L$ and $\bar{\gamma} = (1-\mu)\gamma_L + \mu\gamma_H$. The total amount of driving is given by $(1-\mu)x_L^{ni} + \mu x_H^{ni}$. Note that Δ_γ cannot be too large for high types to drive in equilibrium. In particular we are assuming that $\Delta_\gamma < 1/(1-\mu)$.

Following Proposition 1 in the text, the authority can restore the first-best amount of driving with a per-trip congestion fee equal to $\tau^{fb} = (1-\mu)\gamma_L x_L^{fb} + \mu\gamma_H x_H^{fb}$, where

$$x_L^{fb} = \frac{(1-c)(1+\mu\Delta_\gamma)}{1+2\bar{\gamma}-\mu(1-\mu)\Delta_\gamma^2} > \frac{(1-c)(1-(1-\mu)\Delta_\gamma)}{1+2\bar{\gamma}-\mu(1-\mu)\Delta_\gamma^2} = x_H^{fb}$$

Note that these results converge to the ones in the text as $\Delta_\gamma \to 0$.

It can be shown that both type of individuals gain when moving from the no-intervention to the first best allocation.[64] This Pareto-improving result may not hold as we turn to second-best policies such as driving restrictions. In fact, consider the driving restriction (R_L, R_H). Proceeding as in the main text yields the following equilibrium amount of driving for each group,

$$x_L^r = \frac{(1-c)R_L\,(1+\mu R_H\Delta_\gamma)}{1+(1-\mu)R_L\gamma_L+\mu R_H\gamma_H} \text{ and } x_H^r = \frac{(1-c)R_H(1-(1-\mu)R_L\Delta_\gamma)}{1+(1-\mu)R_L\gamma_L+\mu R_H\gamma_H}$$

respectively.

It is clear from these expressions that the amount of driving of high types is no longer unambiguously lower than that of low types. In fact, if $R_L \to R_H$ (and $\Delta_\gamma > 0$), then $x_H^r < x_L^r$, but if $\Delta_\gamma \to 0$ (and $R_L < R_H$), then $x_H^r > x_L^r$. More interesting is the possibility that $x_H^r > x_H^{ni}$ if R_H is sufficiently large. This opens up the possibility that high types may benefit from the restriction while low types suffer. To see this, let $R_L = R$ and $R_H = \rho R$, with $\rho \in [1, 1/R]$, and denote by S_L^r (resp. S_H^r) the surplus of a low (resp. high) type driver under restriction (R, ρ). We want to show that when $S_L^r/\partial R|_{R=1} \geq 0$, so low types can only suffer if we introduce a restriction $R < 1$, high types may still benefit. In particular, we want to show that $S_H^r/\partial R|_{R=1} > 0$, so high types benefit if we introduce a restriction $R < 1$, when $S_L^r/\partial R|_{R=1} = 0$. This requires

$$R\Delta_\gamma\,(1+\mu(\rho-1))\,(3+R\gamma(1-\mu)+\mu R\rho\Delta_\gamma+\mu R\rho\gamma) > 0 \tag{15}$$

to hold, which is obviously the case.

Even if $\rho = 1$ (i.e., $R_L = R_H$), expression (15) shows that high types may benefit from a restriction while low types do not. The reason is that a proportional rationing of traffic reports greater time-saving benefits to high types than to low types (this can be formally seen from the fact that $x_L^r/x_H^r < x_L^{ni}/x_H^{ni}$ even when $\rho = 1$). These benefits are even higher when $\rho > 1$, as high types see their own driving less restricted.

Finally, consider the introduction of an exemption fee p to an existing restriction (R_L, R_H). Following the steps in the proof of Proposition 4 we can obtain the equilibrium amounts of driving

$$x_L^{rp} = \frac{(1-c-(1-R_L)p)(1+\mu\Delta_\gamma)-\mu\gamma_L(R_H-R_L)p}{1+\bar{\gamma}}$$

[64] It is easy to find a setting in which drivers with low demand may be worse off (provided no transfers across groups are allowed) when moving from the no-intervention to the first-best allocation. An obvious case when this would happen is when low types stop driving under the first-best, for example, when $B_L'(0) < \tau^{fb}$. Other cases can be found in Nie and Liu (2010).

and

$$x_H^{rp} = \frac{(1-c-(1-R_L)p)(1-(1-\mu)\Delta_\gamma) + (1+(1-\mu)\gamma_L)(R_H-R_L)p}{1+\bar{\gamma}}$$

respectively. Note that from these expressions we can arrive at the no-intervention allocation by setting either $p=0$ or $R_L = R_H = 1$.

If an existing restriction is removed by setting $p=0$, there may be winners (the low types) and losers (the high types). This is not surprising given our discussion above, that we could not rule out that $S_H^{rp}(p=0) = S_H^{ni} < S_H^r = S_H^{rp}(p \geq \bar{p}_j)$, where $\bar{p}_j$ is the price that exhausts type $j \in \{L, H\}$'s demand for exemptions. What is somewhat surprising, however, is that it may exist no price $\hat{p} \in (0, \bar{p}_H)$ that leaves high types better off than under both the pure driving restriction and the no-intervention allocation, i.e., that leads to $S_H^{rp}(\hat{p}) > S_H^r > S_H^{ni}$. Following the proof of Proposition 4, for $\hat{p}$ to exist we need both (i) $\partial S_H^{rp}/\partial p\big|_{p=0} > 0$ and (ii) $\partial S_H^{rp}/\partial p\big|_{p=\bar{p}_H} < 0$ to hold. Showing (i) is easy since $(1-(1-\mu)\Delta_\gamma) > 0$ from $x_H^{ni} > 0$, but (ii) requires

$$(1-\mu)(\Delta_\gamma(1-R_L) + \gamma_L(R_H-R_L)) + R_H - 1 < 0$$

which is not clear to hold beyond $\Delta_\gamma = 0$ and $R_L = R_H < 1$.

If (ii) does not hold (and $S_H^{ni} < S_H^r$), high types would prefer not to introduce an exemption fee in the first place.[65] This would only generate more traffic on the road. This does not mean, however, that the introduction of an exemption fee may lead to no gain in overall welfare. In fact, it always does from the fact that (i) $\partial S_L^{rp}/\partial p\big|_{p=0} > 0$, (ii) $\partial S_L^{rp}/\partial p\big|_{p=\bar{p}_L} < 0$, and (iii) $\bar{p}_H < \bar{p}_L$. Any price between $\bar{p}_H$ and $\bar{p}_L$ would report a gain in overall welfare relative to the pure restriction alternative, although the (overall) optimal price is obviously less than $\bar{p}_H$.

B.2 An impossibility result

Consider the setting of heterogeneous drivers of Appendix B.1 but let drivers have isoelastic demands for driving. The surplus that driver type $i \in \{L, H\}$ obtains from x_i kms of driving is given by

$$S_i(x_i, x_{-i}) = x_i^\alpha - \gamma_i x_{-i}^\beta x_i$$

where $\alpha < 1$, $\beta > 0$, $\gamma_H \geq \gamma_L > 0$, and x_{-i} represents the total amount of driving excluding individual type i. Note that $x_{-L} = x_{-H}$. This functional form captures both diminishing returns to driving ($\alpha < 1$) and increasing travel cost of congestion ($\beta > 0$). In the absence of any government intervention, the amount of driving in equilibrium would be given by the first-order-condition

$$\partial S_i(x_i, x_{-i})/\partial x_i = \alpha x_i^{\alpha-1} - \gamma x_{-i}^\beta = 0 \tag{16}$$

and $x_{-i} = \mu x_H + (1-\mu)x_L$.

The main difference with the linear-quadratic formulation is that now the introduction of a driving restriction that works as a proportional rationing scheme always leads to welfare losses.

[65] They would if $S_H^{ni} > S_H^r$.

To see this, consider a driving restriction $R \equiv R_L \le R_H \equiv \rho R \le 1$, with $\rho \ge 1$, and let x^r_{-i} be the equilibrium amount of driving prompted by the restriction, so individual i's surplus is given by

$$S^r_i(x^u_i, x^r_{-i}; R_i) = R_i([x^u_i]^\alpha - \gamma_i [x^r_{-i}]^\beta x^u_i) \tag{17}$$

where $x^u_i \equiv x^u_i(x^r_{-i})$ is the unrestricted amount of driving that individual type i would pursue when the total driving from the remaining drivers add to x^r_{-i}, i.e., $x^u_i(x^r_{-i})$ solves (16) for $x_{-i} = x^r_{-i}$.

Taking the derivative of (17) with respect to R and applying the envelope theorem leads to

$$\frac{\partial S^r_i(x_i, x^r_{-i}; R_i)}{\partial R} = ([x^u_i]^\alpha - \gamma [x^r_{-i}]^\beta x^u_i) - R_i \gamma_i \beta [x^r_{-i}]^{\beta-1} \frac{\partial x^r_{-i}}{\partial R}$$

which captures i's welfare gain/loss from changing the extent of the restriction. Increasing R has two effects in i's welfare. Captured by the terms in parentheses, one effect is the direct effect, which is positive. It amounts to the net benefit of marginally increasing i's driving while keeping congestion unchanged. Working in the opposite direction is the indirect or congestion effect. Since $\partial x^r_{-i}/\partial R > 0$, increasing R leads to more congestion and, hence, to higher travel costs.

We now show that the direct effect always dominate the congestion effect for any level of R. Using the fact that i's actual amount of driving under a proportional-rationing rule is $x^r_i = R_i x^u_i(x^r_{-i})$ and that $x^r_i = x^r_{-i}$, from (16) and $x_{-i} = \mu x_H + (1-\mu) x_L$ we can obtain equilibrium values of x^r_L, x^r_H, and x^r_{-i}. Plugging these values into (17), while letting $\gamma_H = \varrho\gamma$ and $\gamma_L = \gamma$ (with $\varrho \ge 1$), yields

$$S^r_L = \left(1 + \mu(\rho \varrho^{\frac{1}{\alpha-1}} - 1)\right)^{-\frac{\alpha\beta}{1-\alpha+\beta}} (1-\alpha)(\alpha/\gamma)^{\frac{\alpha}{1-\alpha+\beta}} R^{1-\frac{\alpha\beta}{1-\alpha+\beta}}$$

and $S^r_H = \rho \varrho^{\frac{\alpha}{\alpha-1}} S^r_L$. Since $\alpha\beta < 1 - \alpha + \beta$ for all $\alpha < 1$ and $\beta > 0$, it follows that $\partial S^r_i/\partial R > 0$ for all R; hence, $S^r_i < S^{ni}_i$ for all R.

This result indicates that no matter R, the lower travel cost from less traffic ($x^r < x^{ni}$) is never enough to compensate the cancellation of socially valuable trips, some of which are infinitely valuable. Following Proposition 4, however, here the authority can also improve upon a driving restriction R with the introduction of an exemption fee p that allows drivers to use their cars in times of restriction. The only difference with Proposition 4 is that now $S^{rp} > S^r$ for any $p \ge 0$, where S^{rp} is the consumer welfare under a restriction design that considers R and p.

B.3 Proof of Proposition 3

Consider a fixed F under which a fraction $z \in (0, 1)$ of individuals pay the fee in equilibrium. Since in equilibrium individuals must be indifferent between paying the fee and not, F must be equal to the surplus gain from unrestricted driving, that is,

$$F = (1-R)(B(x_i^u(x_{-i}^{rF})) - cx_i^u(x_{-i}^{rF}) - \gamma x_i^u(x_{-i}^{rF})x_{-i}^{rF}) \tag{18}$$

where x_{-i}^{rF} is the total amount of driving under the restriction (R, F) and $x_i^u(x_{-i}^{rF})$ is i's unrestricted amount of driving given x_{-i}^{rF}. In equilibrium it must also hold that the total amount of driving, x^{rF}, is the weighted average between the equilibrium amount of travel of those who pay the fee, x^u, and those who do not, Rx^u:

$$x^{rF} = zx^u + (1-z)Rx^u \tag{19}$$

Using (19) and the first-order condition $x_i^u(x_{-i}^{rF}) = 1-c-\gamma x_{-i}^{rF}$ we can obtain the equilibrium amounts of travel x^u and x^{rF} as a function of R and z that plugged into (18) leads to

$$F = \frac{(1-R)(1-c)^2}{2(1+R\gamma+z\gamma(1-R))^2} \tag{20}$$

Expression (20) allows us to solve for z as a function of F and, hence, write the equilibrium amounts of travel as a function of R and F, that is $x^u(R, F)$ and $x^{rF}(R, F)$.

Plugging $x^u(R, F)$ and $x^{rF}(R, F)$ into the welfare function $S^{rF}(R, F) = B(x^u(R, F)) - cx^u(R, F) - \gamma x^u(R, F)x^{rF}(R, F)$ is easy to see that (8) is the fee that maximizes S^{rF} and leads, according to expression (9) in the text, to $z^\star$ individuals paying the fee, which is valid as long as $z(F^*) \in (0, 1)$, that is, as long as conditions (i) and (ii) hold. If either condition fails to hold the optimum is to set either $z = 0$, which from (20) is done by setting $F = \bar{F}$ or higher, or $z = 1$, which is done by setting $F = \underline{F}$ or lower.

B.4 Proof of Proposition 4

Let x_i^p denotes i's amount of driving with net value above the exemption fee p when the total driving from the remaining drivers adds to x_{-i}^{rp}. This valuable driving is obtained from the first-order condition

$$B'(x_i^p) - C'(x_i^p) - \partial T(x_i^p, x_{-i}^{rp})/\partial x_i^p - p = 0 \tag{21}$$

Thus, i's welfare, $S_i^{rp}(x_i^p, x_i^u, x_{-i}^{rp}; R, p)$, can be written as

$$S_i^{rp}(\cdot) = R(B(x_i^u) - C(x_i^u) - T(x_i^u, x_{-i}^{rp})) + (1-R)(B(x_i^p) - C(x_i^p) - T(x_i^p, x_{-i}^{rp})) \tag{22}$$

where x_i^u is, as in Proposition 2, the unrestricted amount of driving that i would pursue given x_{-i}^{rp}. The second term in (22) is new; it captures the extra surplus from valuable trips (i.e., with net benefit above p) that were previously rationed. Taking the derivative of (22) with respect to p and applying the envelope theorem (twice) yield

$$\frac{\partial S_i^{rp}(R,p)}{\partial p} = -R\frac{\partial T(x_i^u, x_{-i}^{rp})}{\partial x_{-i}^{rp}}\frac{\partial x_{-i}^{rp}}{\partial p} - (1-R)\frac{\partial T(x_i^p, x_{-i}^{rp})}{\partial x_{-i}^{rp}}\frac{\partial x_{-i}^{rp}}{\partial p} + (1-R)p\frac{\partial x_i^p}{\partial p} \tag{23}$$

Using (4) and (21) to obtain, respectively, $x_i^u = 1-c-\gamma x_{-i}^{rp}$ and $x_i^p = 1-c-\gamma x_{-i}^{rp}-p$, expression (23) reduces to

$$\frac{\partial S_i^{rp}(R,p)}{\partial p} = -\gamma x_i^u \frac{\partial x_{-i}^{rp}}{\partial p} - (1-R)p \tag{24}$$

where, after using $x_i^{rp} = Rx_i^u + (1-R)x_i^p$ and the fact that in equilibrium $x_i^{rp} = x_{-i}^{rp}$,

$$x_{-i}^{rp} = x^{rp} = \frac{1-c-(1-R)p}{1+\gamma} \tag{25}$$

for $p \leq (1-c)/(1+\gamma R) \equiv \bar{p}$. At $\bar{p}$, $x_i^p = 0$, so $x^{rp} = x^r$ for any $p \geq \bar{p}$. On the other hand, $\partial x^{rp}/\partial p < 0$ from (25), so $x^{rp} > x^r$ for any $p < \bar{p}$, which concludes the first part of our proof (somewhat obvious because an exemption fee leaves drivers with a milder restriction). For the rest of the proof, that $S^{rp} > S^r$ for any $p \in (\underline{p}, \bar{p})$, note that (i) $S_i^{rp}(R,p)$ is concave in p, i.e., $\partial^2 S_i^{rp}(R,p)/\partial p^2 < 0$ for all $p < \bar{p}$, (ii) $\partial S_i^{rp}(R, p=0)/\partial p > 0$, and (iii) $\partial S_i^{rp}(R, p \to \bar{p})/\partial p = -(1-R)p/(1+\gamma) < 0$. It remains to determine the value of $\underline{p}$. If γ is not high enough so that $S^r < S^{ni}$ (see Proposition 2), then $\underline{p} = 0$. Conversely, if γ is high enough so that $S^r > S^{ni}$, then $\underline{p} > 0$ solves $S^r = S^{rp}(R, \underline{p})$.

B.5 Proof of Lemma 1

Make (24) in the proof of Proposition 4 equal to zero, replace $x_{-i}^{rp} = x_i^{rp} = x^{rp}$ by (25), use $x_i^u = 1-c-\gamma x_{-i}^{rp}$, and then solve.

References

[1] Akbar, P., V. Couture, G. Duranton, and A. Storeygard (2023), Mobility and congestion in Urban India, *American Economic Review* 113, 1083-1111.

[2] Akbar, P., and G. Duranton (2017). Measuring congestion in a highly congested city: Bogotá, CAF, Working paper.

[3] Almagro, M., F. Barbieri, J. Castillo, N. Hickok, and T. Salz (2024), Optimal urban transportation policy: Evidence from Chicago, NBER, Working paper.

[4] Arkhangelsky, D., S. Athey, D. Hirshberg, G. Imbens, and S. Wager (2021), Synthetic difference-in-differences, *American Economic Review* 111, 4088-4118.

[5] Ardekani, S., and R. Herman (1987), Urban network-wide traffic variables and their relations, *Transportation Science* 21, 1-16.

[6] Barahona, N., F. Gallego, and J.-P. Montero (2020). Vintage-specific driving restrictions, *Review of Economic Studies* 87, 1646-1682.

[7] Baranzini, A., S. Carattini, and L. Tesauro (2021). Designing effective and acceptable road pricing schemes: Evidence from the Geneva Congestion Charge, *Environmental and Resource Economics* 79, 417-482.

[8] Barwick, P.J., S. Li, A. Waxman, J. Wu, and T. Xia (2024), Efficiency and equity impacts of urban transportation policies with equilibrium sorting, *American Economic Review*, forthcoming.

[9] Basso, L., J.-P. Montero, and F. Sepúlveda (2021), A practical approach for curbing congestion and air pollution: Driving restrictions with toll and vintage exemptions, *Transportation Research Part A* 148, 330-352.

[10] Basso, L., and H. Silva (2014), Efficiency and substitutability of transit subsidies and other urban transport policies, *American Economic Journal: Economic Policy* 6, 1-33.

[11] Blackman, A., F. Alpízar, F. Carlsson, and M.R. Planter (2018). A contingent valuation approach to estimating regulatory costs: Mexico's Day without driving program. *Journal of the Association of Environmental and Resource Economists* 5, 607-641.

[12] BMS (2019), Bogotá's 2019 Mobility Survey, https://www.movilidadbogota.gov.co/web/sites/default/files/Paginas/22-04-2020/20191216_presentacion_encuesta_v2.pdf (accessed: March 1, 2021).

[13] BMDS (2021), Bogotá's Mobility District Secretary, Segunda fase del permiso especial de acceso a areas de restricción vehicular. August, 2021.

[14] Bonilla, J.A. (2019), The more stringent, the better? Rationing car use in Bogota with moderate and drastic restrictions, *The World Bank Economic Review* 33, 516-534.

[15] Borusyak, K., X. Jaravel, and J. Spiess (2024), Revisiting Event-Study Designs: Robust and Efficient Estimation, *Review of Economic Studies* 91, 3253-3285.

[16] Calatayud, A., S. Sánchez González, F. Bedoya-Maya, F. Giraldez, and J.M. Márquez (2021). Urban road congestion in Latin America and the Caribbean: Characteristics, costs, and mitigation, Inter-American Development Bank, Transportation Division, Washington DC.

[17] Callaway, B., and P. Sant'Anna (2021), Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment, *Journal of Econometrics* 225, 200-230.

[18] Carillo, P., A. Lopez-Luzuriaga, and A. Malik (2018), Pollution or crime: The effect of driving restrictions on criminal activity, *Journal of Public Economics* 164, 50-69.

[19] Daganzo, C. (2000). Taking turns: Rx for congestion, *Access* 17, 14-19.

[20] David, Q. and R. Foucart (2014), Modal choice and optimal congestion, *Regional Science and Urban Economics* 48, 12-20.

[21] Davis, L. (2008), The effect of driving restrictions on air quality in Mexico City, *Journal of Political Economy* 116, 38-81.

[22] Delventhal, M. and A. Parkhomenko (2024), Spatial implications of telecommuting, USC, Working paper.

[23] Durrmeyer, I., and N. Martinez (2022), The welfare consequences of urban traffic regulations, TSE, Working paper.

[24] Eskeland, G. S. and T. Feyzioglu (1997), Rationing can backfire: The day without a car in Mexico City, *The World Bank Economic Review* 11, 383-408.

[25] Fardella, C., N. Barahona, J.-P. Montero, and F. Sepulveda (2023), On the geography of vintage-specific restrictions, *Resource and Energy Economics* 75, 101-122.

[26] Gallego, F., J.-P. Montero, and C. Salas (2013), The effect of transport policies on car use: Evidence from Latin American cities, *Journal of Public Economics* 107, 47-62.

[27] Galdon-Sanchez, J.E., R. Gil, F. Holub, and G. Uriz-Uharte (2023), Social benefits and private costs of driving restriction policies: The impact of Madrid Central on congestion, pollution, and consumer spending, *Journal of the European Economic Association* 21, 1227-1267.

[28] Geroliminis, N., and C. Daganzo (2008), Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological* 42, 759-770.

[29] GIHS (2022), Colombia's 2022 Great Integrated Household Survey.

[30] Hanna, R., G. Kreindler and B. Olken (2017), City-wide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta, *Science* 357 (6346), 89-93.

[31] INRIX (2020), INRIX Global Traffic Scorecard, Kirkland WA, U.S.

[32] Jacobsen, M., J. Sallee, J. Shapiro, and A. Van Benthem (2023), Regulating untaxable externalities: Are vehicle air pollution standards effective and efficient? *Quarterly Journal of Economics* 138, 1907-1976.

[33] Kahn, M. (1996), New evidence on trends in vehicle emissions, *The RAND Journal of Economics* 27, 183-196.

[34] Mohring, H. (1972), Optimization and Scale Economies in Urban Bus Transportation, *American Economic Review* 62, 591-604.

[35] Nie, Y., and Y. Liu (2010), Existence of self-financing and Pareto-improving congestion pricing: Impact of value of time distribution, *Transportation Research Part A* 44, 39-51.

[36] Pan, Y., H. Zheng, J. Guo, and Y. Chen (2023), Modified Volume-Delay Function Based on Traffic Fundamental Diagram: A Practical Calibration Framework for Estimating Congested and Uncongested Conditions, *Journal of Transportation Engineering, Part A: Systems* 149, 04023112.

[37] PBGSD (2021), Gestión de Demanda en un Escenario Post COVID: Teletrabajo y Cambios en la Demanda de Viajes, ProBogota and GSD+, Bogotá, June 2021.

[38] Rodriguez, D., Lazarus, J. and M. Toledo (2021), Consultoría para el desarrollo del modelo de elección modal para evaluar el impacto de las medidas de gestión de la demanda de vehiculos particulares en Bogota. Bogotá, January, 2021.

[39] Salgado, E. and O. Mitnik (2024), Spatial and Time Spillovers of Driving Restrictions: Causal Evidence from Lima's Pico y Placa, Inter-American Development Bank, Working paper.

[40] Salgado, E. (2022), Situación financiera de la movilidad urbana en Bogotá, Project document (LC/TS.2022/151), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).

[41] Sant'Anna, P., and J. Zhao (2020), Doubly Robust Difference-in-Differences Estimators, *Journal of Econometrics* 219, 101-122.

[42] SDG (2018), Estructuración técnica del sistema de cobro por distancia recorrida para vehículos motorizados privados que circulan en Bogotá. Report No. 23180101 prepared by Steer Davies Gleave (SDG) for the Secretaría Distrital de Movilidad de Bogotá.

[43] SECTRA (2013), Plan Maestro de Transportes de Santiago 2025, Santiago, Secretaría Interministerial de Planificación de Transporte (SECTRA), Santiago, Chile.

[44] Small, K. and E. Verhoef (2007), *The Economics of Urban Transportation*. Routledge, London.

[45] Soto, J., L. Macea, and V. Cantillo (2023), Analysing a license plate-based vehicle restriction policy with optional exemption charge: The case in Cali, Colombia, *Transportation Research Part A* 170, 103618.

[46] Tirole, J. (1988), *The Theory of Industrial Organization*, MIT Press, Cambridge, Massachusetts.

[47] Viard, V. B. and S. Fu (2015), The effect of Beijing's driving restrictions on pollution and economic activity, *Journal of Public Economics* 125, 98-115.

[48] Wolff, H. (2014), Keep your clunker in the suburb: Low-emission zones and adoption of green vehicles, *Economic Journal* 124, F481-F512.

[49] Yang, J., A.-O. Purevjav, and S. Li (2020), The marginal cost of traffic congestion and road pricing, *American Economic Journal: Economic Policy* 12, 418-453.

[50] Zhang, W., C.-Y.C. Lin Lawell, and V.I. Umanskaya (2017), The effects of license plate-based driving restrictions on air quality: Theory and empirical evidence, *Journal of Environmental Economics and Management* 82, 181-220.

Contact.

MIT CEEPR Working Paper Series
is published by the MIT Center for Energy and Environmental Policy Research from submissions by affiliated researchers.
For inquiries and/or for permission to reproduce material in this working paper, please contact:

General inquiries: ceepr@mit.edu
Media inquiries: ceepr-media@mit.edu

MIT CEEPR
Center for Energy and
Environmental Policy Research

MIT Center for Energy and
Environmental Policy Research
Massachusetts Institute of Technology
77 Massachusetts Avenue, E19-411
Cambridge, MA 02139-4307
USA

ceepr.mit.edu

MASSACHUSETTS INSTITUTE OF TECHNOLOGY