Pricing Congestion to Increase Traffic: The Case of Bogotá (preliminary and incomplete, please do not circulate)

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

First Draft: December 10, 2021
This Draft: April 11, 2022

Abstract

In September of 2020, the city of Bogotá introduced a significant market-based reform to its odd-even driving restriction, better known as Pico y Placa. Drivers now have the option to pay a daily fee to be exempted from the restriction. Despite the increase in traffic, we find substantial efficiency gains from the reform. An important fraction of these gains—around 40%—comes from simply abolishing the restriction (i.e., setting the daily fee equal to zero); the rest from setting a strictly positive fee. Interestingly, and after accounting for the increase in remote work relative to its pre-covid level, our model suggests that the existing exemption fee is nearly optimal. We also discuss distributional and air-quality impacts of the reform. Letting the exemption fee to vary with the vehicle's value provides limited distributional alleviation, not nearly as much as using the fee collection to reduce public-transport fares. On the other hand, letting the daily fee to vary with the vehicle's emission rate provides some pollution containment, but not enough to offset the increase in emissions from the reform.

Key words: congestion, local pollution, driving restrictions, road pricing

*Montero (jmontero@uc.cl): Department of Economics, Pontificia Universidad Católica de Chile, and Sistemas Complejos de Ingeniería (ISCI); Sepúlveda (fsepulveda@fkeconomics.com): FK Economics; and Basso (lbasso@ing.uchile.cl): Civil Engineering Department, Universidad Chile, and ISCI. This paper was commissioned by the Inter-American Development Bank. Comments from Agustina Calatayud, Germán Escovar, Andrés Gómez-Lobo, and Tomás Serebrisky are much appreciated. Also thanks to Santiago Sanchez for access to the Waze database.

1 Introduction

Congestion remains a serious problem in many cities around the world. According to the INRIX Global Traffic Scorecard,[1] the city of Bogota lead the pre-covid ranking of the most congested cities around the world, with 192 hours per capita lost in heavy traffic in 2019. It was followed closely by Rio de Janeiro (191 hours), Mexico City (158), Istanbul (153) and Sao Paulo (152). Chicago closes the top-10 list with 145 hours lost in congested roads. Unfortunately, when authorities have decided to deal with this externality, they have rarely turned to pricing schemes.[2] Instead, they are increasing relying on rationing schemes, better known as driving restrictions or license-plate bans. One of the most stringent of these programs is precisely found in Bogota.[3] Since 2012, Bogota's restriction program, known as Pico y Placa, bans most drivers from using their vehicles every other day of the week (excluding weekends). 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.

These type of restrictions—which treat all cars the same—have been widely critized 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).[4] In response not only to this "second-car" concern but also to the need to raise extra funds to pay for a better public transport, Bogota's transport authority introduced a major reform to its Pico y Placa program in September of 2020: since then drivers have the option to pay a daily fee (or congestion charge) to be exempted from the restriction, with the entire fee collection earmarked by law to go to public transport.[5]

The goal of this paper is to study the efficiency and distributional implications of Bogota's 2020 market-based reform. We initiate our analysis in Section 2 with an evaluation of the change in traffic due to the reform, for which we use a large Waze database including not only data of Bogota but also of two other large cities in Colombia—Cali and Medellin—that serve as control. Our dif-in-dif estimates suggest a statistically significant increase in traffic of 3% due to the reform (**preliminary**). Since at the time of the reform many drivers were expected to pay the daily fee, this increase in traffic should come as no surprise. Some may interpret it as the inevitable short-run sacrifice needed to finance a better public-transport system in the medium and long-run. Only then, drivers could be persuaded to give up their cars in favor of public transport and attain a substantial reduction in congestion. Complementing the latter,

[1] See https://inrix.com/scorecard/

[2] Notable exceptions include London, Stockholm, Singapore, Milan, and Gothenburg. Cite IADB's recent working paper.

[3] There are many othere programs, including Mexico City and Sao Paulo. Complete list.

[4] An important exception are the vintage-specific restrictions. See Barahona et al (2020).

[5] The program's name has also been updated to Pico y Placa Solidario, in reference to drivers' "solidarity" to help funding the public-transport system.

the introduction of an exemption fee can also be seen as the first step toward a full-blown congestion-pricing scheme over the long-run.[6]

Without questioning the potential long-run benefits of the reform, the main contribution of this paper is to show that such two-phase interpretation of the reform (a sacrifice phase followed by a benign phase) is wrong. We find no evidence of a short-run sacrifice, quite the contrary. One piece of evidence comes from a simple model of homogeneous drivers (Section 3). If, as in Barahona et al (2020), a driving restriction without an exemption fee resembles a proportional-rationing scheme,[7] where any trip is equally likely to be restricted (after controlling for remote work), then an impossibility result emerges (Proposition 2): no matter the extent of the restriction, drivers always end up worse off with the restriction. The gains from faster travel in days of no restriction is not enough to compensate for the lost of valuable trips.

There is a fix to this impossibility result, however. Following what Bogotá did in September of 2020, the fix is to allow drivers to pay a fee (i.e., congestion charge) that exempts them from the restriction. Despite the increase in traffic, the fix always delivers welfare gains (Proposition 3). And if the exemption fee is properly set, welfare is strictly higher than in the no-intervention scenario.

Motivated by these theoretical results, in Section 3 we extend the model to consider heterogenous commuters coming from different income groups and calibrate it to more closely capture Bogota's transport reality in 2019, i.e., before the reform and the covid crisis. The calibrated model is then used in Section 4 to evaluate the impact of the reform. Consistent with the theory, we find substantial gains from the reform, 40% of which comes simply from abolishing the restriction (i.e., setting the daily fee equal to zero) and the remaining 60 % from setting a strictly positive fee. Interestingly, and after accounting for the increase in remote work relative to its pre-covid level, our model suggests that the existing exemption fee is nearly optimal.[8]

We also discuss the distributional impacts of the reform over the different income groups. In particular, we find that letting the exemption fee to vary with the vehicle's value, as Bogota later did in 2021, provides limited distributional alleviation, not nearly as much as using the fee collection to reduce public-transport fares. Finally, in Section 5, we discuss air-quality impacts of the reform. We find that letting the daily fee to vary with the vehicle's emission rate, as Bogota also did in 2021, provides some pollution containment, but not enough to offset the increase in emissions from the reform. This increase in emissions, however, is not enough to offset the welfare gains reported in Section 4.

We are certainly not the first to study the impact of driving restrictions (including low

[6]Personal communication with Nicolás Estupiñán, Chief of Bogotá's Transport Authority, May 20, 2021. See also Secretaría Distrital de Movilidad de Bogotá (2021), Segunda fase del permiso especial de acceso a áreas de restricción vehicular", August 2021.

[7]As opposed to an efficient-rationing scheme, which is the way a congestion charge works. See Tirole (1989) for more on different rationing rules.

[8]There are still substantial gains to make if the authority were to implement a full-blown congestion pricing policy. If optimally done, the extra gain is comparable to the 60 % gain mentioned above, which is accompanied by a significant drop in congestion relative to pre-reform levels.

emission zones) whether on traffic, air pollution or fleet size and composition (see, e.g., Eskeland and Feyzioglu, 1997; Davis, 2006; Gallego et al, 2013, Wolff, 2014; Viard and Fu, 2015; Barahona et al 2020). We are the first, however, to look at the impact of introducing some form of congestion pricing within an existing restriction program. Despite the large number of existing restriction programs, Bogotá is one of the only two programs where exemption fees have been used (the other is Cali, also in Colombia).

This unusually low use of exemption fees is unfortunate, but perhaps not surprising. It may be in part explained by a fear of an increase in traffic. Here is where Bogota's reform provides such a valuable policy lesson: the introduction of an exemption fee is welfare enhancing no matter its value and the extent of the restriction. This is in addition to other benefits such as the possibility of raising extra funds for financing a better public transport or setting the start for a full-blown congestion-pricing policy in the future. Natural candidates for the evaluation (and eventual introduction) of these exemptions fees are, for example, the existing programs in Mexico City and Sao Paulo. In a similar vein, Bogota's reform should also serve to call the attention of any authority considering the introduction of a restriction policy to fight traffic, as seems to be the case in Lima and Santiago. Absent of an exemption fee, no restriction is better than any restriction.[9]

The rest of the paper is organized as follows. Section 2 contains the empirical analysis. Theory results are in Section 3. Our evaluation of the reform is in Section 4. Air-pollution implications of the reform are covered in Section 5. We conclude in Section 6.

2 Evidence: Bogota's Market-Based Reform

In this section we first provide some details about the evolution of Bogota's Pico y Placa and then an empirical evaluation of the impact of the 2020 reform on traffic.

2.1 Bogota's Pico y Placa

Bogota, Colombia's capital and home to more than 7 million people, has long suffered congestion problems. In 1995, Bogotá introduced its first vehicle restriction (District Agreement N-3-1995), when 20% of the fleet was restricted each day. The restriction was based on the last digit of the car's license plate, restricting on Monday the license plates ending in 1 and 2, on Tuesday the license plates ending in 3 and 4, and following this criterion until reaching Friday with the license plates finished at 9 and 0. In this case, the same patent would always be restricted on the same day of the week, so people could easily organize themselves. This policy excluded vehicles arriving from other cities from the restriction.

Since then, the policy suffered minor changes, but it kept its spirit intact until 2003. That year, through District Agreement N-212-2003, the restriction began to apply to 40% of cars per

[9] If the aim is primarily to fight air pollution, the restriction may abstract from the exemption fee but follow a vintage-specific design, as recommended by Barahona et al (2020).

day, based on the last digit of the license plate. Thus, for example, on Monday cars with license plates ending in 7, 8, 9 and 0 were restricted, and the following days the restriction rotated to the following numbers. In this way, the same patent could be restricted on different days of the week, so that people could not easily organize themselves to avoid the restriction.

The policy remained without major changes until 2012. In that year, through District Agreement N-271-2012, the authority restricted the mobility of 50% of city's cars. To do this, the restricted vehicles depended on whether their license plate ended in an odd or even number. In addition, the authority incorporated an exception for electric vehicles (that remains until today).

This setting was maintained for several years and being ratified by the different mayors of the city. In this context, since 2016 (CONFIRM) the city began to discuss the possibility of implementing a policy that allows people to pay a certain fee, which allows them to be exempted from the restriction. However, this measure was not implemented at that time.

2.1.1 Pico y Placa Solidario: The 2020 Reform

In March 2020, due to the Covid-19 pandemic, the authorities decided to suspend the Pico y Placa vehicle restriction measure (District Agreement No. 121-2020) until September 2020.

In September 2020, under District Agreement No. 200-2020, the BogotÃ¡ authority decided to resume the Pico and Plaza, but this time they allowed that those who pay a new congestion fee could circulate, (as had been discussed in previous years). This measure was called Pico y Placa Solidario.

It is important to notice that, at that time, these rates were the same for all vehicles, regardless of their size, emissions, or other characteristics.

The only additional exemption in this setting were electric or hybrid vehicles, medical services or health personnel vehicles (in the context of the Covid-19 pandemic) and vehicles with 3 or more passengers -including the driver-.

These were the congestion fees: COP 51,700 (Colombian pesos) per day, COP 413,200 per month and COP 2,066,200 per semester. The above is equivalent to \$13 per day; \$104 per month (equivalent to \$5 per week); and \$523 per semester (equivalent to \$4 per day). This means that the policy incorporates an effective discount of 62% if it is paid monthly and 69% if it is paid once every 6 months.

2.1.2 Pico y Placa Solidario with varying fees

Almost a year later, in August 2021, the authority implemented an important change to the Pico y Placa Solidario. From that moment, the congestion fees began to be vary according a vehicle's characteristic, more precisely, its value and pollution rate. The objective was to promote the use of modes of transportation with less impact on city's air quality and to create a distributional impact.

Firstly, two thresholds were defined according to the commercial value (CV) of the vehicle:

- CM <$12,500 (represents 91.7% of cars)
- $12,500 <Value < $27,500 (represents 7.7% of cars)
- Value> $27,500 (represents 0.6% of cars).

Secondly, two thresholds were defined based on vehicle emissions (VE):

- $VE < 0.15$ (represents 61.5% of cars)
- $0.15 < VE < 0.3$ (represents 25.6% of cars)
- $VE > 0.3$ (represents 12.9% of cars).

It is important to note that the vehicle emissions factors consider both local and global pollutants, such as PM, NOx, HC, CO and CO2. To define the weighs of each pollutant, the authority of Bogota consulted an experts committee.

Then, based on both thresholds, the authority defined a matrix of multipliers to increase the daily fee for those vehicles that had high emissions or a high commercial value.

The multipliers were constructed as follows: low emission, low value vehicles have a multiplier of 1; medium emission vehicles have a multiplier of 1.1; vehicles of average commercial value have a multiplier of 1.25; high emission vehicles have a multiplier of 1.2; and high trade value vehicles have a multiplier of 1.5.

With this scheme, the policy has operated since 2021.

2.2 A Price Reform that Increases Traffic

In this section we shows that, not surprisingly, the introduction of Pico y Placa Solidario increases congestion in Bogota.

To analyze this phenomena, we use data from Waze App [10] that incorporates hourly average speed in three cities: Bogota, Cali and Medellin.

The idea, is to use Cali and Medellin as control groups for Bogota's traffic context. Even though there could be problems of endogeneity, we do not want to estimate the causal effect of the policy, nor the exact increase of Bogota's congestion, but to show that there is an increase in traffic in the city that is not followed by the rest of the country and that is not normal considering the historical data of the city.

To define the average speed of each city we defined an area of relevance that is affected by the Pico y Placa. Then, the average speed is calculated as the time (hours) of all of the trips that were made in an specific hour in those cities, divided by the total distance that those trips reach (kilometers).

[10] The data was provided by the Interamerican Bank of Development in the context of an Agreement with the authors.

Having these variables built, we could estimate a simple difference in difference model in which the dependent variable is the average speed observed in Bogota and the control groups are the speeds observed in Cali and Medellin.

The model to be estimated is the following:

$$Speed_{it} = \beta_0 + \beta_1 Treat_t * Bogota_i + \beta_2 September2020_t + \mathbf{X}_{it} + \alpha_i + \gamma_t + u_{it} \tag{1}$$

Where $Speed_{it}$ is the speed of each city i on each hour t, $Treat$ is a binary variable that indicates if each city implemented a Pico y Placa Solidario in t, $Bogota$ is an indicator Bogota, $\mathbf{X}_{it}$ is a vector of controls, which includes for example the evolution of the pandemic, city quarantines, among others. Then, α_i are cities fixed-effects, and $\gamma_t + u_{it}$ are time fixed effects (week and hour). Finally, u_{it} is a normally distributed error.

The results of the estimation are not available yet for review, but we can advance that there is an increase in Bogota's traffic compared with the other cities. The magnitude of the increase is under revision.

3 Theory: An Impossibility Result

Consider a unit mass of a continuum of homogeneous drivers. The surplus that driver i obtains from x_i kms of driving in a given period, say a week, is given by the difference between the benefit of driving minus its cost

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

where x_{-i} represents the total amount of driving excluding i, $\nu > 0$, $\alpha < 1$, $\beta > 0$, and $\gamma > 0$ is the free-flow cost of driving per km (i.e., the unit cost in the absence of congestion; here, when $x_{-i} = 0$). This functional form, which follows Barahona et al (2020), captures in a most tractable way 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

$$\frac{\partial}{\partial x_i} S_i(x_i, x_{-i}) = \alpha \nu x_i^{\alpha-1} - \gamma x_{-i}^{\beta} = 0 \tag{2}$$

And since all drivers are the same, we have that in equilibrium $x_i = x_{-i}$, which plugged into (2) yields the no-intervention amount of driving

$$x^{ni} = (\alpha\nu/\gamma)^{1/(1-\alpha+\beta)}$$

and corresponding consumer welfare $S^{ni} = S(x^{ni}, x^{ni})$. Given the congestion externality, this amount of driving is obviously above the socially efficient (or first-best) level, which is given by

$$x^{fb} = (\alpha\nu/(1+\beta)\gamma)^{1/(1-\alpha+\beta)} = \arg\max_{x}\{\nu x^{\alpha} - \gamma x^{\beta} x\}$$

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

Proof. Faced with a congestion fee, i solves $\max_{x_i}\{S_i = vx_i^{\alpha} - (\gamma x_{-i}^{\beta} + \tau^{fb})x_i\}$, 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. As discussed in the Introduction, however, in many instances the authority does not have this market-based instrument at her disposal, so 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. Obviously, they would cancel only those that report net benefits below the congestion fee, which is socially efficient. Under a driving restriction, in contrast, drivers do not have that choice. At times, they would be forced to cancel highly valuable trips (and work from home or take the public transit) and at others allowed to make 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. While one can entertain different views about the extent of this inefficiency, the view we adopt in this paper, following Barahona et al (2020), is that of a driving restriction as a proportional rationing scheme, where all trips are equally likely to be canceled until to comply with the restriction. If so, we have the following "imposibility" result.

Proposition 2 *Under the assumption that a driving restriction works as a proportional rationing scheme, any driving restriction leads to welfare losses, no matter its extent.*

Proof. Let $R \in [0,1]$ denote the extent of the driving restriction, with $R = 1$ the case of no restriction and $R = 0$ the case of full restriction. If x_{-i}^r is the amount of actual driving prompted by some R level of restriction, the surplus that i actually obtains under a proportional-rationing rule is given by

$$S_i^r(x_i^*, x_{-i}^r; R) = R(\nu[x_i^u]^{\alpha} - \gamma[x_{-i}^r]^{\beta}x_i^u) \tag{3}$$

where $x_i^u \equiv x_i^u(x_{-i}^r)$ is the unrestricted amount of driving that i would pursue when the total driving from the remaining drivers adds to x_{-i}^r, i.e., $x_i^u(x_{-i}^r)$ solves (2) for $x_{-i} = x_{-i}^r$.

Taking the derivative of (3) with respect to R and applying the envelope theorem leads to

$$\frac{\partial}{\partial R}S_i^r(x_i, x_{-i}^r; R) = (\nu[x_i^u]^\alpha - \gamma[x_{-i}^r]^\beta x_i^u) - R\gamma\beta[x_{-i}^r]^{\beta-1}x_i^u\frac{\partial x_{-i}^r}{\partial R}$$

which captures i's welfare gain/loss from changing the extent of the restriction. Increasing R has two effects on i's welfare. Captured by the terms in parenthesis, 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 $\partial x_{-i}^r/\partial R > 0$ (see (4) below), 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 $x_i^r = Rx_i^u(x_{-i}^r)$ and that $x_i^r = x_{-i}^r$, from (2) we obtain

$$x_i^r = x_{-i}^r = x^r = (\alpha\nu/\gamma)^{1/(1-\alpha+\beta)}\, R^{(1-\alpha)/(1-\alpha+\beta)} \tag{4}$$

that plugged into (3) yields

$$S^r = (1-\alpha)\,(\alpha\nu/\gamma)^{\alpha/(1-\alpha+\beta)}\, R^{1-\alpha\beta/(1-\alpha+\beta)}$$

Since $\alpha\beta < 1-\alpha+\beta$ for all $\alpha < 1$ and $\beta \geq 0$, it follows that $\partial S^r/\partial R > 0$ for all R and, hence, $S^r < S^{ni}$ for all $R < 1$. ■

Proposition 2 indicates that no matter R, the lower travel cost from less traffic ($x^r < x^{ni}$) is never enough to compensate the cancelation of socially valuable trips. There is a fix to this impossibility result, however. Following what Bogotá did in September of 2020, the fix is to allow drivers to pay a fee that exempts them from the restriction.

Proposition 3 *Despite the increase in congestion, the authority can improve upon a driving restriction $R < 1$ with the introduction of an exemption fee $p \geq 0$ that allows drivers to use their cars in times of restriction: $x^{rp} > x^r$ and $S^{rp} > S^r$ for any $p \geq 0$, where x^{rp} and S^{rp} are, respectively, the amount of driving and consumer welfare under a restriction design that considers $R < 1$ and $p \geq 0$. Furthermore, if $\bar{p}$ denotes the optimal exemption fee for a given R and $S^{r\bar{p}}$ the corresponding consumer welfare, then $S^{r\bar{p}} > S^{ni} > S^r$.*

Proof. The proof that $x^{rp} > x^r$ can be omitted since it is intuitively obvious (drivers face fewer restrictions). Let x_i^p denotes i's amount 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

$$\alpha\nu[x_i^p]^{\alpha-1} - \gamma[x_{-i}^{rp}]^\beta - p = 0 \tag{5}$$

Thus, i's welfare can be written as

$$S_i^{rp}(x_i^p, x_i^u, x_{-i}^{rp}; R, p) = R\left(\nu[x_i^u]^\alpha - \gamma[x_{-i}^{rp}]^\beta x_i^u\right) + (1-R)\left(\nu[x_i^p]^\alpha - \gamma[x_{-i}^{rp}]^\beta x_i^p\right) \tag{6}$$

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 (6) 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 (6) with respect to p, and applying the envelope theorem (twice) yields

$$\frac{\partial S_i^{rp}(\cdot)}{\partial p} = -R\gamma\beta[x_{-i}^{rp}]^{\beta-1}x_i^u\frac{\partial x_{-i}^{rp}}{\partial p} - (1-R)\gamma\beta[x_{-i}^{rp}]^{\beta-1}x_i^p\frac{\partial x_{-i}^{rp}}{\partial p} + (1-R)p\frac{\partial x_i^p}{\partial p} \tag{7}$$

where $x_{-i}^{rp} = x_{-i}^p + R(x_{-i}^u - x_{-i}^p)$. An expression for $\partial x_{-i}^{rp}/\partial p$ can be obtained by combining $\partial x_{-i}^{rp}/\partial p = (1-R)\partial x_{-i}^p/\partial p + R\partial x_{-i}^u/\partial p$, (2), (5), and the fact that in equilibrium $x_i = x_{-i} = x$ as follows

$$\frac{\partial x_{-i}^{rp}}{\partial p} = \Upsilon(R)\frac{\partial x_{-i}^p}{\partial p} = \Upsilon(R)\left(\alpha(\alpha-1)\nu[x_{-i}^p]^{\alpha-2} - \Upsilon(R)\gamma\beta[x_{-i}^{rp}]^{\beta-1}\right)^{-1} < 0 \tag{8}$$

where $\Upsilon(R) = (1-R)/(1-R\beta x^u/(\alpha-1)x^{rp}) > 0$. Although it is intuitively obvious, expression (8) serves to formally see that $x^{rp} > x^r$. Now, evaluating (7) at $p=0$ (where $x_i^u = x_i^p = x_i^{rp} = x^{ni}$ also) leads to

$$\left.\frac{\partial}{\partial p}S^{rp}(R,p)\right|_{p=0} = -\gamma\beta[x^{ni}]^\beta\frac{\partial x^{rp}}{\partial p} > 0$$

Based on the latter, the fact that $S^{rp}(R, p=0) = S^{ni}$ and $S^{rp}(R, p\to\infty) = S^r$ by construction, and that $S^{ni} > S^r$ from Proposition 2, it is clear, by continuity, that there will be some price $\bar{p} \in (0,\infty)$ where $S^{rp}(R,p)$ is maximized and $S^{rp}(R,\bar{p}) \equiv S^{r\bar{p}} > S^{ni} > S^r$ (the proof that $\partial S^{rp}(R,p)/\partial p = 0$ has a unique solution is in the Appendix). ■

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 Bogota in the 1990s, can only lead to welfare losses (Proposition 2). And the second is that despite the increase in congestion, these uniform restrictions can be fixed, as Bogota did in 2020, with the introduction of an exemption fee (Proposition 3).[11] There is an additional result, connecting Propositions 1 and 3, with a more political-economy flavor: the possibility of implementing the first-best by adjusting R and p accordingly, i.e., $R=0$ and $p = \bar{p}(R=0) = \tau^{fb}$.[12] In practice, this adjustment may take place gradually, as politics permits it, turning an existing driving restriction without an exemption fee into a full-blown congestion-pricing policy over the long run. This seems to be Bogota's ultimate goal.[13]

[11]See also Daganzo (2000) and Basso et al (2021) for a discussion about the benefits of restriction policies with exemption fees.

[12]From here, it is not difficult to see that $\bar{p}(R>0) > \tau^{fb}$ in an effort to compensate for the fact that a fraction R of trips remain unpriced.

[13]Personal communication with Nicolás Estupiñán, Chief of Bogotá's Transport Authority, May 20, 2021. See

Having already confirmed the first of our predictions—that the introduction of Bogota's exemption fee did lead to an increase in traffic ($x^{rp} > x^r$)—we now turn our attention to the second prediction—-that despite this increase in traffic, Bogota's reform did report a welfare gain ($S^{rp} > S^r$). In exploring this potential gain, many questions arise. How much of the gain 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} and S^{ni})? How far is the existing p from $\bar{p}$? 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? To answer these and other questions we need to extend our model to more closely capture Bogota's transport reality.

4 A Driving-Restriction Model for Bogotá

Our homogeneous-driver setting certainly abstracts from elements that may prove relevant in a practical application, affecting some of our results. One of these elements 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 ever-more-frequent episodes of local air pollution, Bogota's policy has mostly responded to congestion concerns. It is easy to see that our result in Proposition 3 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 (driving) surplus ($S^{rp} > S^r$). In Section 6, we attend these air pollution considerations and show that the increase in vehicle emissions have had a rather limited welfare effect, not affecting our main result.

A second element absent in our simple 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 2005). In policies without an exemption fee (i.e., where $p \to \infty$), this possibilty 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). By contrast, in restriction policies that include an exemption fee, the incentive to buy a second car is significantly reduced, if not completely eliminated (see, e.g., Basso et al 2021). For this reason, in what follows we abstract from this "second-car" possibility.

A final aspect that will prove particularly relevant for what follows, not only for its efficiency but also distributional implications, is the presence of heterogenous commuters with different preferences over available transportation modes (e.g., car, public transport, etc.), including the

also Secretaría Distrital de Movilidad de Bogotá (2021), Segunda fase del permiso especial de acceso a áreas de restricción vehicular", August 2021.

possibility to work remotely. In this heterogenous setting, it is possible to imagine a driving restriction without an exemption fee that could eventually increase overall welfare, going against the result in Proposition 2. For that we would need, on the one hand, that the actual extent of the restriction R to vary by income group, with high-income groups being less affected by the restriction than middle and low-income groups, as actually documented by Gallego et al (2013) for Mexico City. We would also need high-income drivers to value travel time more than their lower-income counterparts, as widely documented in the transportation literature (see, e.g., Basso et al 2021). In that case, it is trivial to extend our simple setting to generate an increase in overall welfare.[14]

If interpreted correctly, however, this more favorable result does not entirely invalidate the imposibility result of Proposition 2. It still remains impossible for the restriction to leave *everyone*, high and low-income drivers, better off. To formally study this and other efficiency and distributional implications, we use the rest of this section to explain first, how our simple setting is extended to accomodate for commuter heterogeneity and then, how this extended model is calibrated to Bogota's transport reality.

4.1 Heterogeneous Commuters, Public Transport and Remote Work

We build upon the standard origin-destination transport model of Basso and Silva (2014) and Basso et al (2021). 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. Let d_i be the number of days of the week (excluding weekends) that $i = 1, ..., n$ commutes by car, h_i the number of days that works from home, and $(5 - d_i - h_i)$ the number of days that uses public transport. Since the majority of public-transport rides in Bogota, involve some combination of subway and buses, we treat public transport in a "reduced form" combining both modes. The (transport) surplus that individual $i = 1, ..., n$ obtains after a week of travel is given by

$$S_i = \Omega_i(d_i, h_i) - C_i(d_i, h_i, r_i) - T_i(d_i, h_i) \tag{9}$$

where $r_i = 0, ..., 5$ measures the extent of the restriction, i.e., the number of days her car, provided she owns one, is restricted from entering the city center during the week,[15] $\Omega_i(d_i, h_i)$ captures the gross benefit of travel, $C_i(d_i, h_i, r_i)$ is the financial cost of travel, and $T_i(d_i, h_i)$ is the time cost of travel. Note that $\Omega_i(d_i, h_i)$, $C_i(d_i, h_i, r_i)$, and $T_i(d_i, h_i)$ are all measured in dollars and vary across individuals according to their income levels, which we divide in five income groups: (1) low, (2) middle-low, (3) middle, (4) middle-high and (5) high. We use $g = 1, ..., 5$ to denote an income group.

[14] One can make the case, for instance, if drivers are divided in two income groups, high (h) and low (l), with $R_l < R_h \approx 1$, $\beta_h > \beta_l$, and $\nu_h > \nu_l$.

[15] 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.

The gross benefit of travel for $i \in g$ depends on her intrinsic (relative) preferences for each transport mode as follows

$$\Omega_{i\in g}(d_i, h_i) = \lambda_i^{-1}\phi_0[d_i + (5 - d_i - h_i)\theta_{i\in g} + H_{i\in g}(h_i)]$$

where λ_i corresponds to i's marginal utility of income, ϕ_0 is a constant, $\theta_{i\in g}$ captures i's taste for public relative to private transport and is drawn from a normal distribution with mean $\bar{\theta}_g$ and standard deviation σ_g^θ, and $H_{i\in g}(h_i)$ captures the benefit of remote work relative to private transport. According to ProBogota GSD+ (2021), the demand for remote work has showned to be increasing and more elastic with income,[16] a trend that we capture with a linear function of the form $H'_{i\in g}(h_i) = \vartheta_{i\in g} - \zeta(6-g)h_i$, where ζ is a constant and $\vartheta_{i\in g}$ is drawn from a normal distribution with mean $\bar{\vartheta}_g$ and standard deviation σ_g^ϑ. Below, in the calibration section, we explain how we obtain values for the different parameters.

On the other hand, i's weekly financial travel cost is given by

$$C_i(d_i, r_i) = d_i c + p \max\{0, d_i + r_i - 5\} + (5 - d_i - h_i)f \tag{10}$$

where c is the daily cost of using a car (set to infinity for those individuals that do not own one), including expenses on fuel, parking, lubricants, tires, and so on, p is the exemption fee, and f is the daily expense on public transit (i.e., the product of single-ride fare and the average number of daily rides). Two observations regarding how the restriction enters into the model are in order. The first is that we allow the extent of the restriction to vary accross 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.[17] The second is that individuals have ample flexibility to accomodate to the restriction. For example, an individual that faces two days of restriction ($r_i = 2$) would need to spend nothing on exemption fees if she is planning to use the car three days ($d_i = 3$); the days of restriction would be those in which she either worked from home or took public transit.

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

$$T_i(d_i, h_i) = \lambda_i^{-1}\left[d_i\gamma_i^c t^c l + (5 - d_i - h_i)(\gamma_i^p t^p l + \gamma_i^w w^p)\right] \tag{11}$$

where γ_i^m is i's marginal utility of time when using transport mode $m \in \{c, p\}$, $t^m \equiv t^m(n_c, n_p)$ is time (in hours per kilometer) spent on transport mode m on any given day, which is a Bureau of Public Roads (BPR) function (see, e.g., Basso and Silva, 2014) of the number of individuals

[16]In fact, ProBogota GSD+ (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 Bogota (Rodriguez et al. 2021), indicating that 81% of individuals in the lowest-income group believe they won't be teleworking once the pandemic is over, in contrast to 40% in the highest-income group.

[17]In the case of Bogotá, such inviduals 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.

that either commute by car (n_c), use public transport (n_p) or stay home ($n_h = n - n_c - n_p$), l is the average travel-distance (in kilometers) involved 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/train station, and w^p is the average waiting time at the station. Values for the parameters that enter into both $C_i(d_i, r_i)$ and $T_i(d_i, r_i)$ come for different sources, independent of the calibration. See the Appendix for details.

The decision problem of individual i is to chose d_i and h_i so as to that maximize (9), while taken as given the equilibrium choice of the remaining individuals, that is, taken as given n_c, n_p 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 Calibration

In this section we explain the calibration of the model, and how its predictions fit with Bogota's observed data.

Before explaining the parameters of the model, it is important to delve into the structure of the different income groups that we consider.

We divide commuters in five income groups based on the strata definition of Bogota's 2019 Moibility Survey (MS-2019)[18]. As shown in Table 1, groups are of different sizes (they are not quintiles), but similar with the sizes considered in BMS(2021) for Santiago de Chile.

Group No.	Strata level	Fraction of total	Average monthly income per household	Car ownership	Marginal utility of time ($/hr)
1	Low (1)	12%	$184	11%	0.60
2	Middle-low (2)	40%	$288	21%	1.36
3	Middle (3)	34%	$502	39%	2.59
4	Middle-high (4)	10%	$1,027	66%	4.60
5	High (5 & 6)	5%	$1,564	82%	12.38

Note: This table contains household characteristics for five income groups based on information provided by the BMDS, Bogota's 2019 Moibility Survey (MS-2019), and our own model calibration.

Table 1: Income-group characteristics and preferences

On the other hand, the table shows that cars are heavily used only by higher income groups.

[18] Even though Bogota has 6 strata groups, we colapsed stratas 5 and 6 into one group that we called the high-income group.

Also, the Marginal Utility of Time vary a lot between groups. In the next section we explain the sources of this parameters and the rest of the parameters of the model.

4.2.1 Inputs

The model is parametrized to capture Bogota's traffic and air pollution context with the most recent available data. In the case of traffic, we use Bogota's 2019 Mobility Survey (MS-2019) and ProBogota GSD+ (2021). Also, some parameters are taken from the literature or available information. In the case of vehicle and pollution parameters, we use data provided by the Bogota's Mobility District Secretary (BMDS), which indicates the emission factors for vehicles of each district of the city.

The parameters could be divided into two types, the ones that are assumed by external evidence, and the ones that have to be calibrated.

The first group of parameters can be taken -not calibrated- from external sources because of assumptions that are standard in the literature or because of the availability of useful information in public or reliable information (for example, that is provided by Bogota's authorities). In that sense, some of the parameters need to represent the context of Bogota, but others are more general. Those parameters that could not be calibrated are divided in four groups: income-level; transportation; car ownership; and vehicle and pollution parameters.

Firstly, in relation to income-group parameters we have the marginal utility of time, the marginal utility of income and the intrinsic preference parameters that we have explained in the calibration process. On one hand, BMDS has estimated a general value for the marginal utility of time γ_i^c, which is supported by evidence of a recent survey made in Bogota by reseachers of UC Berkeley (To introduce some heterogeneity between groups we use the dispersion observed in Santiago for each income-group. Also, to generate heterogeneity within each group we let γ_i^c be normally distributed (and truncated at zero) with a mean value equal to BMDS's numbers and a standard deviation of 20%. On the other hand, the commuter's marginal utility of income, λ_i, is obtained in two steps: firstly, we let the income distribution of our sample of n commuters replicate the actual income distribution observed in the MS-2019; secondly, if Y_i is household i's income, we let $\lambda_i = \lambda_0/Y_i$, where λ_0 is a scaling factor to be obtained from the model calibration, as we mentioned before. Finally, while intrinsic preferences for public transport and remote working are obtained from the model calibration, we assume $\theta_{i\in g}$ and $h_{i\in g}$ to be uniformly distributed.

Secondly, in relation to transport parameters, some of them were provided by the BMDS, such as the average distance (l), the daily cost of using a car to c \$/day, the duration of the peak period to q, the car occupancy to a, the network length to K, the free-flow speed of cars and the free-flow speed of buses (v). On the other hand, the public-transit (daily) fare is set at its current value. In addition, other transport parameters were selected because of the literature, such as the equivalence factor (κ). Values of α_s and β_s comes from BMS(2021) and are very standard in the transportation literature. Finally, the capacity of a lane to C is adjusted to match observed and predicted speeds in the reference scenario.

Thirdly, in relation to car ownership, Table 1 shows that the higher is the income-level the higher is the car ownership. Consistently, Table 2 shows there is an important heterogeneity in the number of cars per household. While, the lowest income group has only 1% of the population with two or more cars, the highest income group (stratas 5 and 6) has 36% of the group with two or more cars.

Income Level	Percentage of people who own more than 1 car
Low	1%
Middle-low	2%
Middle	6%
Middle-high	16%
High	36%

Note: This table contains car possession per household for five income groups based on information from Bogota's 2019 Moibility Survey (MS-2019).

Table 2: Percentage of people who own more than 1 car

We consider that this heterogeneity must affect the success of the weekly restriction. In this context, we assume that all individuals who have more than 1 car, have the capability to reduce the impact of the restriction in one day.

Fourthly, in relation to vehicle and pollution parameters, BMDS provide us data of the car fleet, which they used to generate the recent multiple tolls based on cars emission and commercial value. On the one hand, the data has the environmental factors of the car fleet, which weighs local and global pollutants[19]. Based on the actual policy, there are three groups of emissions: low, middle, and high emission vehicles. The data consolidate the number of vehicles that are in each emission category, by income-group. Therefore, we have a pollution index, and we can measure the air pollution change when applying different types of policies. In addition, Bogota's data also incorporates commercial value categories for each income group (the thresholds are $12,500 and $27,500). In this way, there are three categories based on cars value.

This information is useful to understand the effect of multiple tolls depending on car's value and emissions. Table 3 shows the percentage of cars in each category (value-emission). Surprisingly, there isn't much heterogeneity between the percentage of cars in each category among different income groups. In addition, there is an important concentration of cars in the categories of high emissions and low commercial value for all income-groups.

The table suggest us that there could be possibility to increase the efficiency of the policy by changing the thresholds of the car's value and emissions the are defined by the authority.

[19]To define the weights for each pollutant, the SDC consulted a committee of experts. For more information see Online Appendix

Income group	Low-value			Middle-value			High-value		
Emissions:	H	M	L	H	M	L	H	M	L
1	47,8%	32,8%	17,1%	1,5%	0,5%	0,1%	0,1%	0,1%	0,0%
2	54,3%	28,0%	14,4%	2,2%	0,6%	0,2%	0,1%	0,2%	0,0%
3	56,6%	23,9%	13,1%	4,6%	1,1%	0,3%	0,2%	0,3%	0,0%
4	57,8%	20,8%	10,1%	8,5%	1,7%	0,4%	0,3%	0,3%	0,0%
5	49,9%	21,8%	10,6%	12,2%	3,4%	0,8%	0,7%	0,6%	0,0%

Table 3: Car fleet distribution by cars' emissions and values

As we explain in Air Quality Implications section, a next version of this paper we will include the damage estimation of vehicle emissions in Bogota, in order to compare air quality changes with transportation benefits.

On the other hand, there are parameters that must be calibrated. In this context, we calibrate preference parameters, such as $\theta_{i \in g}$, which represents i's taste of public transport relative to private transport, and $h_{i \in g}$, which represents i's taste of remote working relative to private transport. Also ϕ_0 must be calibrated, because adjusts the gross benefit of commute $(\Omega_i(d_i, h_i))$ in relation to the financial travel cost and the time travel cost. Finally, λ_0 is calibrated to adjust the relevance of financial travel cost in relation to the rest of the surplus function.

These 12 parameters are specific from Bogota's traffic context and do not have any reliable source to be estimated, so they must be calibrated.

We obtain the value of the parameters by calibrating the model's predictions to match the modal shares (i.e., fraction of individuals using public transport according to MS-2019, and fraction of individuals doing remote work according to ProBota GSD+). The loss function we minimize is

$$L = (MP_0 - \hat{MP}_0)^2 + \sum_{g=1}^{5}(MP_g - \hat{MP}_g)^2 + (MH_0 - \hat{MH}_0)^2 + \sum_{g=1}^{5}(MH_g - \hat{MH}_g)^2$$

where MP_0 is the observed modal share of using public transport for the entire population in the MS-2019 and $\hat{MP}_0$ is the corresponding model's prediction, MH_0 is the observed remote working pre-pandemic for the entire population according to ProBogota GSD+ and $\hat{MH}_0$ is the corresponding model's prediction, and finally the sub-indexes represents the modal shares for specific income groups ($g = 1, ..., 5$).

4.2.2 Model calibration and fit

The calibration proceeds as follows. We work with a representative sample of $n = 2000$ individuals. Thus, we first take 2000 random draws from MS-2019's income distribution. Second, we get random draws for γ_i^c. Third, we assign a value to ϕ_0 and get random draws for λ_i, θ_i, and h_i after assigning values to λ_0 and to the boundaries of the θ's and h_i uniform distributions of each income group. Fourth, we compute the equilibrium based on these parameters and compare the equilibrium modal shares to actual shares. Steps three and four are repeated until L is minimized.

As shown in Table 4, calibrated parameters provide a close match between actual modal shares and those predicted by the model for each group and overall.

	PT modal share		Remote working	
Income groups	Observed	Model prediction	Observed	Model prediction
Low	85%	85%	0,0%	2%
Middle-low	72%	69%	0,1%	4%
Middle	61%	62%	5%	12%
Middle-high	44%	42%	15%	15%
High	25%	20%	25%	30%
Overall	55%	59%	10%	12%

Note: This table shows how our model calibration matches observed data. The first and second columns contrast the observed modal shares of the public transportation to the predictions of our model. The third and the fourth columns report estimations of remote working pre-pandemic with predictions of our model.

Table 4: Model fit

It is important to notice that we consider remote working before pandemic (10%) because we are calibrating Bogota's traffic context in 2019. However, in some of the simulations presented in the section 5.1.2 we study the effect of considering an increase in the remote working, which would be consist with the post-pandemic perspectives. In fact, recent studies shows that 20% of the workers will do remote working in Bogota in 2021 (ProBogota GSD+).

5 Policy Evaluation

After presenting the calibration and the fit of the model, in this section we run different simulations in order to analyze the effects of Bogota's 2020 reform.

Firstly, we analyze the efficiency implications of a driving restriction with an exemption fee relative to a standard driving restriction without an exemption fee. Then, we analyze distributional implications, considering heterogeneous effects between income-groups and the impact of using the available tools to handle equity problems, such as recycling the fee collection into the public transit system, or considering fess that vary with the car's characteristics. Finally, we analyze air quality implications by measuring the change of the environmental pollution under different policy scenarios, and we study if varying fees by emission factors are a reasonable way to handle air quality concerns.

5.1 Efficiency Implications

As a benchmark case, our first set of simulations considers a distributionally neutral recycling policy: all the revenue collected from toll payments is returned in a lump-sum fashion to individuals while preventing any transfers between income groups and from individuals that own a car to those that do not. In other words, the entire toll collection coming from individuals that own a car in group $g = 1, ..., 5$ is returned to that same subset of individuals. Using this neutral revenue-recycling criteria, we exclusively concentrate on efficiency and distributional implications of allocating road capacity among heterogeneous users. It is as if a perfectly informed (surplus-maximizing) planner could directly inform each individual that owns a car whether she can use it or not.

As the policy in Bogota considers between 2 and 3 days of restriction (because affects exactly the 50% of the license plates each day), we firstly study only the surplus changes of moving from a policy of 2 days of driving restriction without exemption to a similar policy but with toll exemptions. Then we study how an increase in remote working could affect the results of the model, specially considering that it is expected that remote working will increase post-pandemic with respect to the situation in 2019 (ProBogota GSD+). Thirdly, we comment the efficiency gains of moving towards a full congestion pricing scheme. Finally, we show the change in the consumer surplus among different income groups, in order to understand who are the winners and the losers with these types of policies, which motivates the distributional analysis of the next section.

5.1.1 Driving restriction with an exemption fee

In this section we analyse surplus changes of moving from a driving restriction of 2-days scheme without exemption to a similar policy but with toll exemptions. We considers tolls from \$0 to a value that is as high as we want[20].

In this context, is possible for us to separate the pure effect of eliminating driving restriction (Proposition 2) with the effect of putting the optimum price (Proposition 3). Figure 1 shows the benefits from passing from a pure driving restriction policy -Bogota's scenario in 2019- to different policy scenarios, starting from no-policy scenario (i.e. driving restriction with toll exemption of \$0) and ending in a driving restriction policy with an extremely high daily toll.

Consistent with Proposition 3, the figure shows that there is an optimal price that maximize the aggregate transportation surplus, which is around \$20. Additionally, the figure shows that moving from a full-driving restriction policy to a no-policy scenario is beneficial for Bogota's drivers, consistent with Proposition 2. The explanation is that in a full driving restriction scheme there are many valuable trips that are banned, and there are low-value trips that are made, generating a very inefficient result in terms of traffic.

In this context, eliminating the full-driving restriction policy generates a 42% of the total

[20] For example, the case of eliminating the driving restriction is considered in the simulations when the toll is \$0, and a driving restriction without toll exemption is equivalent to giving a toll an infinite value

benefits obtained in a driving restriction with an optimal toll exemption, so setting the price in the optimal value provides 58% of the total benefit.

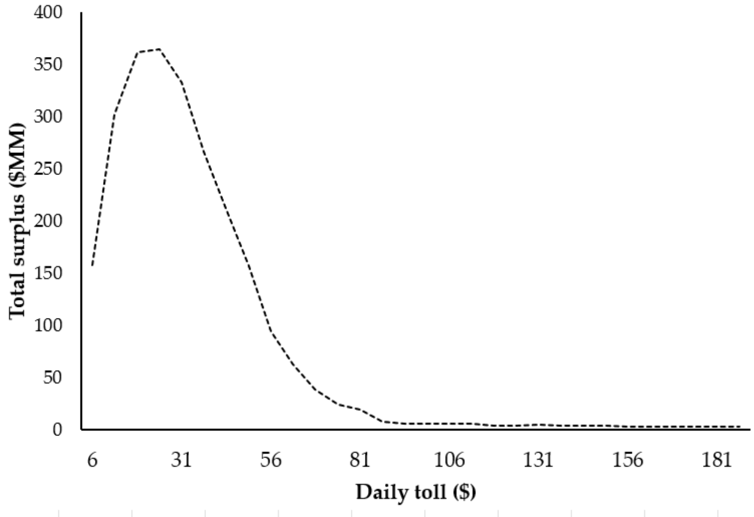


Figure 1: Efficiency gains for implementing exemption tolls

This result is extremely important, because it shows that the policy could have welfare improves even driving restriction is completely eliminated and traffic increases in the city. The literature usually have criticize driving restrictions for other factor, such as the incentive of buying a second car (BGM, 2021), but this paper shows that full-driving restrictions could be bad policies even without taking into account dynamic effects.

We test the welfare changes among different days of restriction policies (1 to 5), always reaching the same conclusion. However the benefits of moving from a driving restriction scenario to a no-policy scenario are increasing in terms of the days of restriction. Obviously the valuable trips that are banned are higher when days of restriction increases. In the extreme case of 5-days of restriction, there are no cars commuting so the welfare loss is maximum.

Nevertheless, this is a result that only represents Bogota's traffic context. In this sense, to understand how the results could vary depending on city's structure, we analyzed these dynamics in the context of Santiago de Chile (following BMS-2021). Our results shows that Santiago has welfare losses when moving from 1 and 2 days of pure driving restriction scenario to no-policy scenario. Therefore, the benefit of allowing valuable-trips are not sufficient to compensate the costs of a higher traffic context. On the other hand, when moving from a 3, 4 and 5 days of restriction (R) to a no-policy scenario there are always welfare gains, which are increasing in R in the same way as Bogota. In conclusion, even though there is always an optimal price that allows to maximize aggregate transportation surplus, the benefits or losses given from passing from a pure driving restriction policy to a no-policy scenario depends on the city traffic context.

In the Online Appendix we show that there are three main factors that explain the difference between cities: the traffic context (speed) in the reference scenario, the inequality between individuals, and the modal shares.

Having discussed the welfare changes generated by a toll-exemption policy we now review some of the changes observed in transportation variables. Our simulations shows that a surplus-maximizing policy leads to a reduction in the transportation speed: -6% in the case of cars (from the benchmark speed of 20 km/hr to 19 km/hr) and -2% in the case of buses (from the benchmark speed of 17 km/hr to 16.6 km/hr). Even though, the policy generates substantial surplus gain from setting the toll at its optimal level: $364 million per year, or 0.134% the country's GDP. On the other hand, we estimate a toll revenue of $181 million that could be used to improve the public transport system (which now return into the same income-groups), and increases in the use of public transport mode (from 59% to 61%) and remote working (from 12% to 14%).

5.1.2 An increase in remote working

As we show before, the optimal price under a 2-days restriction policy is approximately $20. However the price that Bogota's authority imposed since September 2020 is $5.3. This gap could be used to suggest an increase in the exemption toll set by the authority, however there is a factor that could overestimate the optimal congestion toll, the percentage of remote working in the city.

In the reference scenario, we predict a remote working of 10% based on Bogota's information in 2019. In this context, a price of $20 could be the optimal policy based on our estimations. However, there is evidence that Bogota's traffic context changed significantly due to Covid-19 pandemic. In fact, recent studies indicates that in 2021 approximately 20% is doing remote working.

Based on this new information, we simulate the effect of an exogenous increase in the aggregate remote working to 20%. We do this by multiplying the h_i by a constant factor which increase the value of remote working.

Our simulation suggest that the optimal toll drops 50% respect to the previous scenario, so the the optimal price is about $10. In that way, the gap between the price set by Bogota's authority and the optimal price predicted by the model reduces notoriously.

Someone could put in doubt the accuracy of this new estimation, but the idea is not to have a precise estimation of the optimal price in a 2-days of restriction setting, but to understand how the optimal price could change if we consider changes in remote working levels in the future.

In this context, we recall that the objective of this paper is to understand complex economic dynamics that comes from Bogota's unique policy and the propositions of the theoretical model, not to generate precise estimates of the changes in transportation parameters.

5.1.3 Moving towards a full congestion pricing policy

An important efficiency implication of a policy with this design is that in the long-term it could be used to reach full-congestion pricing, which is represented by a driving restriction of 5 days a week ($R = 5$) with toll exemption.

With the calibrated model we estimate which are the benefits from moving from a pure driving restriction scheme of 2 days to a full-congestion pricing. The Figure 2 compares the benefits obtained in both 2-days restriction and 5-days restriction scenario with toll exemption. The figure shows that there are important gains of moving to a full-congestion pricing scheme, which in the optimal price ($15) represents an increase of 32% respect to the 2-days restriction scheme ($480 million compared to $364 million).

Another way to analyze this result, is that a 5-days restriction policy with a toll of $5 could reach similar welfare gains than a 2-days restriction policy with optimal price ($20). Similarly, the welfare gains generated by actual policy set by the authority (2 days with a price if $5.3) could be reach with a 5-days restriction policy with a toll exemption of about $3.

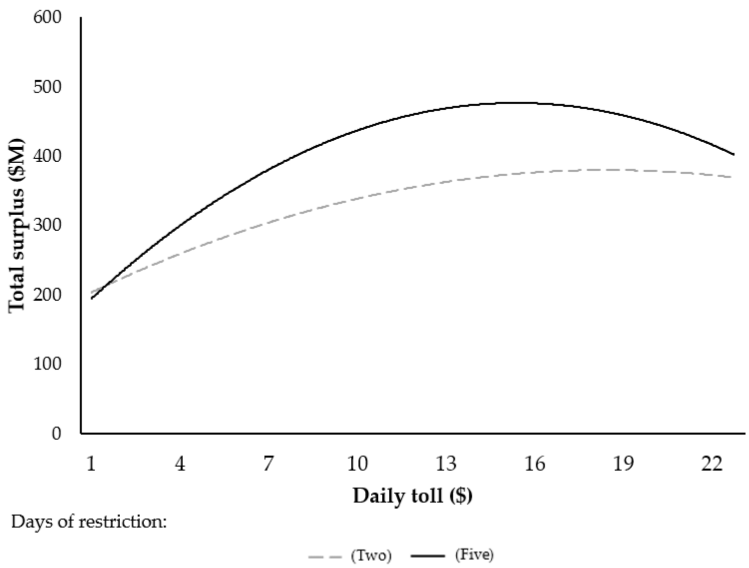


Figure 2: Efficiency gains for moving towards a full congestion pricing

In addition, simulations shows that a 5-days surplus-maximizing policy leads to a increase in the transportation speed: 25% in the case of cars (from the benchmark speed of 20 km/hr to 25 km/hr) and 18% in the case of buses (from the benchmark speed of 17 km/hr to 20 km/hr). The policy generates substantial surplus gain from setting the toll at its optimal level: $480 million per year, or 0.17% the country's GDP. [21] On the other hand, we estimate a total

[21] Interestingly 0.4% of GDP (or $87 billion) is the congestion cost (i.e., hours lost in traffic) estimated by INRIX (2019) for the entire United States in 2018. Similar estimations are in De Palma and Lindsey (2011). In addition. Finally, BMS-2021 estimates surplus gains of about 0.37% in Santiago de Chile, but do not consider

revenue of \$ 820 million that could be used to improve the public transport system (which now return into the same income-groups). Finally, we see important increases in the use of public transport (from 59% to 72%) and increases in remote working (from 12% to 19%).

In conclusion, there are important benefits of moving towards a full congestion pricing policy, which we know is the objective of the transportation authority of Bogota (according to BMDS-2021). However, this analysis only considers surplus gains for all the individuals in the aggregate. As we mention in section 1, the problems of these types of politics -that charges vehicle use- is that they are seen as very unequal measures. In the following section we show the surplus effect of the policy among different income-groups.

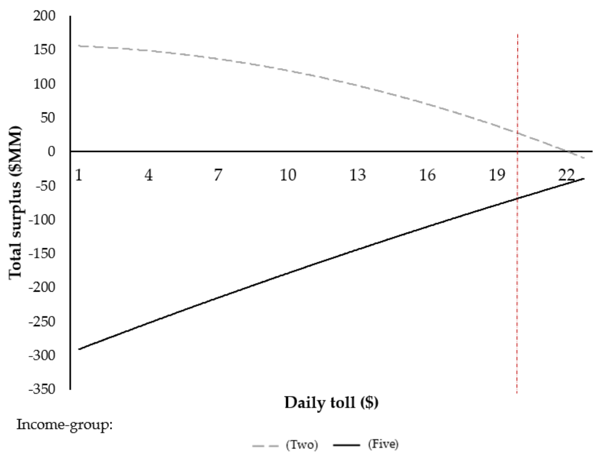
5.1.4 Heterogeneous effects by income-group

One can imagine a situation in which a regulator is willing to sacrifice efficiency in favor of more equitable outcomes that leave individuals in all income groups -or some relevant groups- better off. If these more equitable designs have a much better chance at being implemented in practice, this is a compromise worth considering.

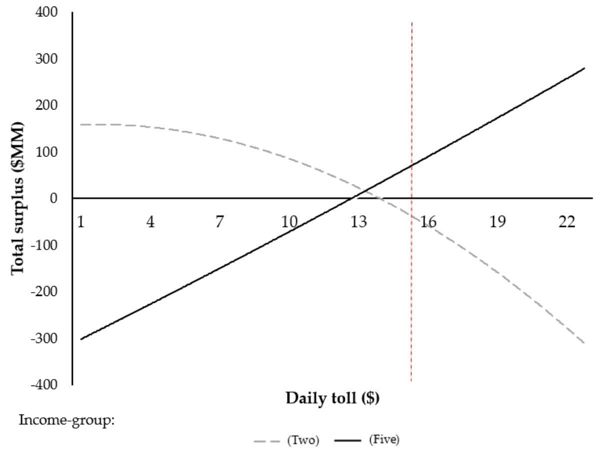
Figure 3 shows the heterogeneous impact of a driving restriction with toll exemption policy. For simplification in the figure we analyze 2-days and 5-days restriction schemes for groups 2 and 5. The rest of the analysis for different scheme options and income-groups is included in the Online Appendix.

---

the existence of remote working, which affects the transportation surplus.



(a) 2-days restriction scheme



(b) 5-days restriction scheme

Note: This figure depicts changes in group 5's transport surplus for different restriction formats (2 and 5 days of restriction), and toll levels. All toll revenues collected from this group are given back to car owners in that group in a lump-sum fashion.

Figure 3: Groups 1 and 5's surplus change under neutral recycling

As shown in Figure 3(a), 2-day restriction with optimal price ($20) scheme ends up leaving the low-income (i.e., group 2) individuals better off, while leaving the highest-income (i.e., group 5) individuals worse off. The reason is that that there is an important amount of trips that were restricted in a the driving restriction policy that now could be done, which benefits both groups. In addition, as the toll increases the surplus change of the low-income group it reduces, and it turns to negative after $22 -after the optimal price-, because of two factors: the toll increases the monetary cost of private mode and the higher number of individuals in the public transport mode generates overcrowding. At the same time, as we explain in the previous section, with toll exemptions traffic increases, affecting the highest-income group more than the low-income group. Therefore, considering all the effects, in 2-day restriction scheme the high-income group experience a loss of transport surplus, while low-income group are benefited. In other words, high-income group prefers to commute in a car 3-days a week and then combining bus and remote working the rest of the week rather than commute 5-days a week with more traffic and paying the toll.

On the other side, as shown in Figure 3(b), 5-days restriction with optimal price ($15) scheme ends up in the opposite direction, leaving high-income individuals better off, while leaving low-income individuals worse off. The difference here is that 5-days restriction scheme prohibit much more trips than 2-days restriction scheme, reducing a significant amount of surplus. Also, the weekly cost of transport is much higher, reducing the probability that low-income group could pay the toll. Finally, this policy reduces the traffic (increases car and public transport speed), which have a bigger impact in high-income group than low-income group. Here, high-income group prefers to commute in a car 5-days a week paying the exemption toll rather than 3 days a week combining public transport and remote working the rest of the week.

Something relevant to add in this analysis is that people with no car is always worse of under this neutral recycling scheme, which suggest that public transportation is significantly affected by the overcrowding and remote working is not the solution even for people in the highest income-groups. Results are summarized in Figure 4.

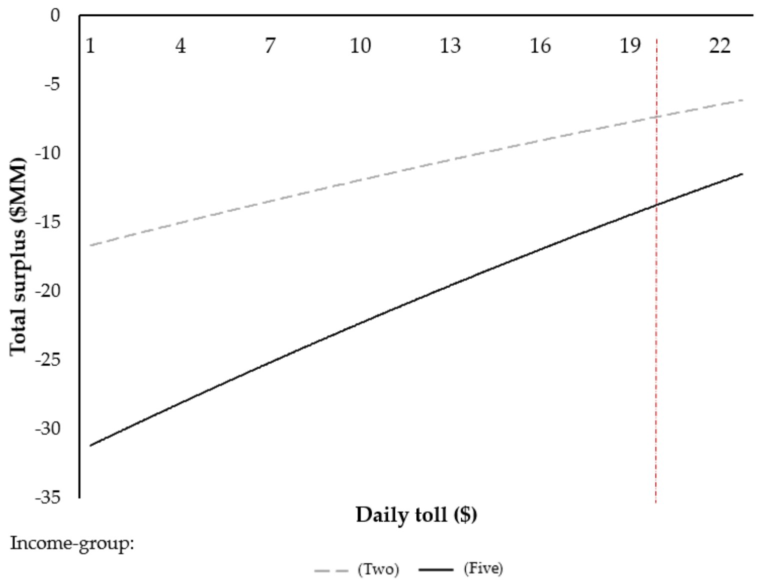


Figure 4: Surplus change of people who do not have a car in Groups 2 and 5 under a neutral recycling scheme with 2 days of restriction

In this case, we see that surplus change is negative, but the loss is reduced when the toll is higher. This implies that the benefits of commute at a higher speed in public transport somewhere intersects the costs of overcrowding, but not before the optimal price. Furthermore, if authorities wants to implement these types of policies they have to take into account people who do not have a car that is worse off. In they do not, is highly possible that the policy is rejected by some groups.

The analysis for the other income-groups is trivial and is included in the Online Appendix. The lowest-income group (i.e., group 1) -not analyzed in this section- behaves similar to group 2, but has a faster reduction of its transportation surplus. In fact, the surplus change it turns to negative at \$12 in a 2-day restriction scheme and at \$4 in a 5-day restriction scheme.

The conclusion of this analysis is that there are always winners and losers under the neutral recycling scheme. While in the 2-days restriction format (Bogota's actual scenario) group 2 is better off, it is important taking into account that in a transition to a full-congestion pricing scheme some individuals could affected. Also, group 1 is strictly worse off in these exemption toll schemes.

Considering these results in the next section we delve into the available tools that could be used to address distributional concerns.

5.2 Distributional Implications

Given that is not possible to leave everyone better off— we need to find ways to more evenly distribute the aggregate benefits of any of the policy designs described above. In theory, this should help reach wider public support for their implementation.

With that aim, in this section we consider two alternative uses of the toll collection. In the first case, the totality of the collection is directed to reduce the public-transport fare (P_b). In the second case, the authority use multiple tolls to affect people of higher income more than people of lower income.

5.2.1 Public transport fare reduction

As it is defined in the District Agreement N. 200-2020, Bogota's transportation authority must recycle the entire toll collection into the public transport system. Even though, there are many possibilities to use the revenue, we modeled this regulation as the possibility to make reductions in the public transport fare. Figure 5 summarize the results.

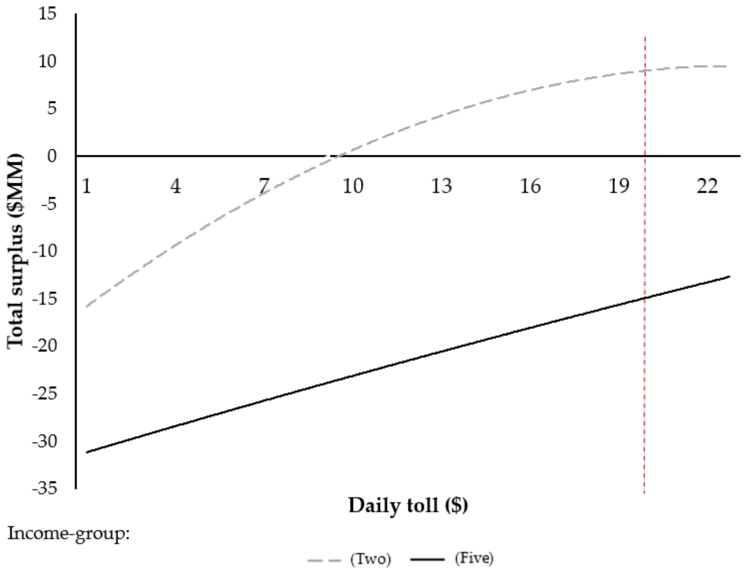


Figure 5: Surplus change of people who do not have a car in Groups 2 and 5 under a fare-reduction recycling scheme with 2 days of restriction

As it shows in the figure, in the optimal price policy with 2 days of restriction, a reduction in the public transport fare have an important impact in individuals with lower income, and specifically in individuals who do not have a car. Individuals of group 2 (low-income) who do not own a car benefit greatly from the fare reduction whereas those of the group 5 (highest-income) have not an important impact in relation to the neutral recycling scheme. This is because monetary cost affects less to people of higher income, because they have a lower utility marginal of income (λ_i). For these high-income individuals, the fare reduction is not enough to compensate for a more crowded public-transit system. The Online Appendix contains equivalent figures for groups 1, 3 and 4), but the dynamic is monotonous: group 1 have an even major effect than group 2 in terms of surplus benefit, and the effect of groups 3 and 4 are between the effect of groups 2 and 5.

In conclusion, this is an important instrument that could be used to balance the surplus gains from different income-groups in order to design a policy that could be implemented.

5.2.2 Varying exemption fees by vehicle's characteristics

Since August 2021, Bogota's authority could have different fees varying on vehicle's characteristics.

As we show in section 2.1.2, Bogota's authority defined two thresholds based on cars' emissions and two thresholds based on cars' value. Then, we simulate the effect of implementing this policy on total surplus and among different income-groups.

Firstly, the results shows a total surplus reduction of about 4% compared to the homogeneous toll scheme. This is because there is an obvious cost of charging more the cars that are more expensive, because trips that are equally valuable have different tolls. Additionally, as the authority charges more the vehicles with higher emission rate, there is a relative-benefit for people of higher income that have cleaner cars.

In sum, it is not clear to find which are the income-groups that more benefited or affected by the multipliers defined by the authority, so the entire effect of the policy could even generate a worse welfare distribution.

In fact, comparing the transportation surplus of groups 2 and 5 in a 2-days driving restriction scheme with toll exemption in both schemes (fixed and multiple toll), we find that, under exactly the same effective toll paid, group 2 is 13% worse in the multiple toll scheme than in the fixed toll scheme, while group 5 is 1.3% worse in the multiple-toll scheme.

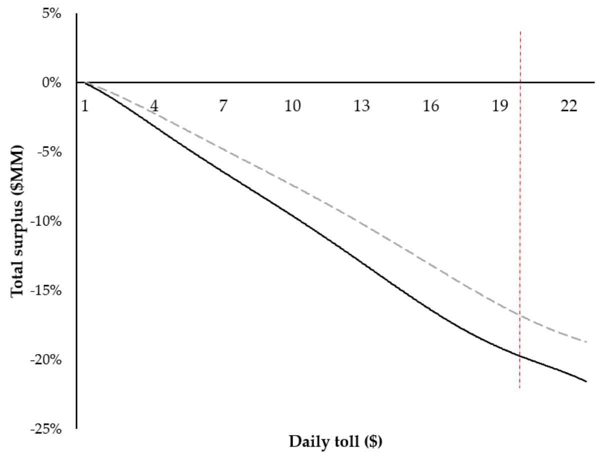
In conclusion, we find that the multiple-toll scheme implemented by the authority n August 2021 have important problems from an efficiency perspective (the air quality perspective is analyzed in the next section). A possible policy recommendation for Bogota's authorities is to use this instrument (multiple-toll) only to address air quality concerns, and focus the distributional concerns only in the public fare reduction instrument that is available by law. Finally, is important to notice that a better analysis should incorporate changes in the thresholds that were defines by the authority. At least for now, we only reviewed the cutoff defined by the authority, but in future versions of this paper we will analysis how the results could change under different thresholds.

6 Implications for Air Quality

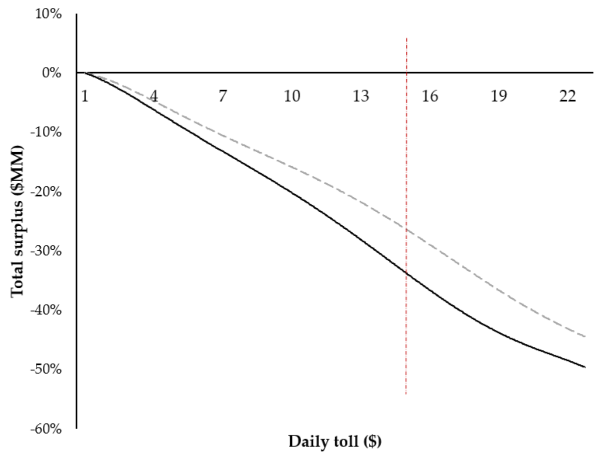
So far we have been silent about air pollution concerns. In this paper, we want to make two points about this. Firstly, it is known that nowadays pollution effects are more important than before. Bogota's context is a good example of that, since in the 90's when driving restriction policies were implemented the only concern was congestion (see, for example, District Agreement N. 626-1998), nevertheless the introduction of multiple-tolls based on cars' emissions implies that Bogota's authorities are considering air pollution when implementing transportation policies.

Considering that the city has the multiple-toll instrument available, in this section we simulate how it could be working under the thresholds defined by the authority.

As it shown in Figure 6, multiple-toll policies (black continuous lines) have a bigger impact on air pollution than fixed-toll schemes (dash lines). In fact, Figure 6(a) shows a reduction of about 20% in the optimal 2-days of restriction scheme and Figure 6(b) shows a reduction of about 35% in the 5-days restriction scheme in the multiple-toll policy. Relative to this policy, a fixed-toll scheme have a 4% less impact on a 2-days restriction scheme and a 9.6% less impact on a 5-days restriction scheme.



(a) 2-days restriction scheme



(b) 5-days restriction scheme

Note: This figure depicts changes in vehicle emissions among different driving restriction schemes (2 and 5 days of restriction), and toll levels.

Figure 6

As we show in the previous section, the aggregate surplus is affected by the structure of the multiple toll, which includes thresholds for cars's value. Therefore, a better knowledge of the efficiency implications of this scheme (i.e., the trade-off of reducing air pollution and losing transportation surplus) should incorporate a multiple-toll designed exclusively on vehicle emissions. In a future version of this paper we will include an analysis on that direction. Also we will incorporate an estimation of damages of the air pollution.

7 Final remarks

to be written

References

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

[2] 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.

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

[4] Bento, A., Goulder, L., Jacobsen, M., and von Haefen, R. (2009). Distributional and efficiency impacts of increased US gasoline taxes, *American Economic Review* 99, 667-699.

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

[6] Currie, J. and M. Neidell (2005), Air pollution and infant health: What can we learn from California's recent experience? *Quarterly Journal of Economics* 120, 1003-1030.

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

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

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

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

[11] 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.

[12] Harsman, B. and J. Quigley (2010), Political and public acceptability of congestion pricing, *Journal of Policy Analysis and Management* 29, 854-874.

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

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

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

[16] ProBogota GSD+ (2021), Gestión de Demanda en un Escenario POST COVID –Teletrabajo y Cambios en la Demanda de Viajes. June 2021.

[17] 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. January, 2021.

[18] Viard, V. B. and Fu, S. (2015), The Effect of Beijing's Driving Restrictions on Pollution and Economic Activity, *Journal of Public Economics* 125, 98-115.

[19] Wolff, H. (2014), Keep Your Clunker in the Suburb: Low-emission Zones and Adoption of Green Vehicles, *The Economic Journal* 124, F481-F512.

.