Vintage-specific driving restrictions

Nano Barahona Francisco Gallego Juan-Pablo Montero*

April 10, 2019

Abstract

Local air pollution has led authorities in many cities around the world to impose limits on car use by means of driving restrictions or license-plate bans. By placing uniform restrictions on all cars, many of these programs have created incentives for drivers to buy additional, more polluting cars. We study vintage-specific restrictions, which place heavy limits on older, polluting vehicles and no limits on newer, cleaner ones. We use a novel model of the car market and results from Santiago's 1992 program, the earliest program to use vintage-specific restrictions, to show that such restrictions should be designed to work exclusively through the extensive margin (type of car driven), never through the intensive margin (number of miles driven). If so, vintage restrictions can yield important welfare gains by moving the fleet composition toward cleaner cars, comparing well to alternative instruments such as scrappage subsidies and pollution-based registration fees.

1. Introduction

Local air pollution continues to be a serious problem in many cities around the world in part because of a steady increase in car use.[1] In an effort to contain such a trend and persuade

*Barahona is with the Economics Department of Stanford University and Gallego and Montero with the Economics Department of the Pontificia Universidad Católica de Chile (PUC-Chile). Emails: hbaraho@stanford.edu; fgallego@uc.cl; and jmontero@uc.cl. We thank Lucas Davis, Liran Einav, Juan Escudero, Claudio Ferraz, Matthew Gentzkow, Marco González-Navarro, Larry Goulder, Christian Hellwig (the editor), Enrique Ide, Gastón Illanes, Kelsey Jack, Shaun McRae, Sebastián Otero, Mar Reguant, James Sallee, Brandon Shaufele, Paulo Somaini, three anonymous referees, and audiences at AARES-Adelaide, Berkeley, Chicago, EAERE, Ecole Polytechnique, Harvard, LSE, Michigan, Norwegian School of Economics, PUC-Chile, PUC-Rio, Sao Paulo School of Economics, Toronto, Toulouse School of Economics, Stanford, Universidad de Chile, Universidad de Montevideo and Universidade de Vigo for helpful comments. Montero acknowledges financial support from Fondecyt (Grant No. 1161802) and the Complex Engineering Systems Institute - ISCI (grant CONICYT FB0816). We also thank José D Salas and Felipe Sepúlveda for excellent research assistance.

[1] Cars are major contributors of carbon monoxide (CO), hydrocarbons (HC), nitrogen oxides (NO_x), and fine particles (PM2.5). HC and NO_x are precursors to ground-level ozone (O_3, also known as smog) and also contribute to the formation of PM2.5. At least in Santiago, vehicles are responsible for 30 and 36% of PM2.5 and O_3 concentrations, respectively (Rizzi and De La Maza, 2017). These local pollutants, unlike global pollutants such as carbon dioxide (CO_2), are characterized as having a local impact, at the city level, that lasts for a short time, sometimes only a few hours. The adverse health effects of these local pollutants are well documented. Currie and Neidell (2005), for example, found a significant effect of CO on infant mortality.

drivers to give up their cars in favor of public transport, authorities increasingly rely on limits to car use, typically implemented on the basis of some combination of the last digit of a vehicle's license plate and colored stickers displayed on its windshield. Good examples of these so-called driving restrictions include Athens (where restrictions were introduced in 1982), Santiago (1986), Mexico City (1989), São Paulo (1996), Manila (1996), Bogotá (1998), Medellín (2005), Beijing (2008), Tianjin (2008), several German cities (2008), Quito (2010), Hangzhou (2011), Chengdu (2012), and Paris (2016).[2]

According to the existing literature, the increasing popularity of these restrictions is problematic. As noted by *The Economist* ("Traffic in megacities," February 27, 2016), the take-away message from this literature is that driving restrictions create perverse incentives for drivers to buy additional vehicles, not only increasing fleet size but also moving its composition toward higher-emitting vehicles. The best documented evidence supporting this claim comes from Mexico City's *Hoy No Circula* (HNC) program, as implemented in 1989 (e.g., Eskeland and Feyzioglu, 1997; Onursal and Gautam, 1997; Molina and Molina, 2002; Davis, 2008; Gallego et al., 2013).[3]

In this paper, we study an aspect of driving restrictions that has been mostly overlooked in the literature yet can be found in some recent programs: namely, vintage-specific restrictions, or more precisely, restrictions that differentiate cars by their pollution rates. In 1992, for example, Santiago reformed its restriction program to exempt all cars equipped with a catalytic converter (a device that transforms toxic pollutants into less toxic gases) from the one-day-a-week restriction. This exemption ended in March 2018 for all cars built before 2012. Mexico City has also introduced several reforms to its restriction program; for example, in today's HNC program, new vehicles are exempt for their first eight years.

Vintage-specific restrictions are also in recent European programs. Authorities in Germany, for instance, have adopted low-emission zones (LEZs) in several cities since 2008. Unlike the partial circulation bans in Santiago and Mexico City, LEZs completely ban certain higher-emitting vehicles from entering city centers (e.g., see Wolff, 2014). This "complete-ban" structure was also in the restriction introduced in Paris in 2016 (where any car built before 1997 is banned permanently from circulation within the city limits weekdays from 8 am to 8 pm) and in recent announcements by several European cities, including London, Paris and Rome, to completely ban diesel vehicles from entering city centers in the coming decade.[4]

[2]Authorities in Santiago, Brussels, London, Madrid, Milano, and Paris, to name a few, have also turned, on occasion, to one-day restrictions (in conjunction with any existing permanent programs) to combat daily episodes of critical air pollution. New Delhi also tried a two-week experiment in January 2016. This paper, however, focuses on permanent restrictions since these have more potential to alter a city's fleet composition.

[3]Zhang et al. (2017) also failed to find air quality improvements from restrictions elsewhere, namely, in Bogotá, São Paulo and Tianjin. They did find effects from the restriction program introduced in Beijing at the time of the 2008 Olympic Games. An initial gain in air quality was confirmed by Viard and Fu (2015) and Liu et al. (2017), but the latter study also showed that the gain disappeared within a year, consistent with the pattern found by Gallego et al. (2013) for HNC.

[4]LEZ programs have also been introduced in China; for example, in Beijing in 2009 and Nanchang in 2013.

Of all the possible variations on a driving restriction policy one might think of, vintage differentiation represents a radical departure from early designs. By allowing drivers to bypass the restriction not by purchasing a second (and possibly older, more polluting) car but by switching to a cleaner car facing lighter or no restrictions, vintage-specific restrictions have the potential to significantly alter the fleet composition towards cleaner vehicles in those places where local pollution is a concern. The objective of this paper is to study such potential.

Our study begins with an illustration of the basic mechanism behind a vintage-specific driving restriction using Santiago's 1992 reform as evidence. Given the sharp discontinuity created by the reform between restricted and nonrestricted vintages, we are able to test for policy effects on fleet composition in restricted and nonrestricted areas by focusing on their fleet differences around the 1992-93 vintage discontinuity.[5] We find that households in areas subject to the restriction (i.e., any municipality in the city of Santiago) own a much smaller fraction of 1992 (and older) models than their counterparts living in nonrestricted areas.[6] In addition, we find significant price effects favoring less polluting models and document a significant impact on vehicle's emissions of local pollutants in Santiago.

While the evidence from these vintage-specific restriction programs is useful to illustrate the fleet-composition effect that vintage differentiation can produce, it still leaves many questions unanswered. For instance, it does not say much about the welfare implications of these policies and says little on how these vintage-specific restrictions ought to be designed and how that compares to alternative policy instruments such as scrappage subsidies, (pollution-based) annual registration fees, and gasoline taxes. With the help of a novel model of the car market, we seek to answer all these questions.

Our model of the car market shares the vertical-differentiation structure of some existing models (e.g. Gavazza et al., 2014) but differs from them in three important respects. First, households in our model decide not only what car to buy, as in Adda and Cooper (2000) and Gavazza et al. (2014), but also how much to drive.[7] Understanding how distinct policy instruments affect extensive and intensive margins differently proves crucial for policy design, as we discuss in detail below. Second, we pay attention to market dynamics following a policy intervention, particularly the dynamics of old, high-emitting cars exiting the market.[8] What explains trade in these vertical-differentiation models is drivers' different willingness to pay for quality. High-willingness-to-pay drivers upgrade to a new car when they decide to sell their used units to medium-willingness-to-pay drivers, who in turn sell their used units to lower-

[5]In Chile, only cars from vintage 1993 onward are equipped with a catalytic converter.

[6]Similar results are in Wolff (2014) for the LEZ programs in Germany.

[7]This latter margin is not in Gavazza et al. (2014), for example, because their focus is on evaluating frictions affecting trade in the second-hand market.

[8]Using a much richer set of vehicle and household characteristics, Bento et al. (2009) also look at how a particular policy intervention (gasoline tax) affects fleet evolution. Our research differs from theirs in scope (by considering a larger set of policy instruments) and modelling assumptions (by letting trade be driven by vertical differentiation and also considering forward-looking agents).

willingness-to-pay drivers, and so on. This trading process over the lifetime of a unit, which can take a long time in developing and emerging economies, ends when a low-willingness-to-pay driver decides to scrap the unit. We are the first to model these long equilibrium transitions and their implications for policy evaluation.[9]

And third, our model is unique in its attention to a variety of policy interventions to curb *local* air pollution. In particular, we study driving restrictions in a wide range of formats, from the uniform restriction introduced in Mexico City in 1989 to the nearly complete-ban structure introduced in Paris in 2016. Since in all these programs the car market affected by the policy intervention extends well beyond the geographic area directly targeted by the policy, our model also considers households in less or unpolluted zones that are affected only by the policy's effect on the car market. This is an important mechanism that can affect the optimality of these restrictions by allowing the flow of older cars to zones where old cars still have value to some drivers. Our model is also flexible enough to allow for temporal variation in pollution harm, which is prevalent in many cities suffering from local air pollution. In this case, it may be optimal to place restrictions only during those hours of the day, days of the week, and months of the year when pollution is of concern.

The main message from our model (and numerical simulations) is that driving restrictions ought to be designed to work exclusively through the extensive margin affecting the type of cars people drive, never through the intensive margin (amount of driving). Hence, restriction designs should follow closely the complete-ban vintage structure seen in Paris and Germany's LEZs, setting a (moving) vintage threshold that separates cars between complete restriction and full exemption. By working exclusively through the extensive margin, a vintage restriction can yield important welfare gains by moving the fleet composition toward cleaner cars; particularly, when emission rates vary widely with (observable and enforceable) car characteristics, most notably vintage.[10]

On the contrary, a driving restriction that aims at the intensive margin, for instance, by placing a uniform restriction on all cars regardless of their emission rate, is sure to result in a significant welfare loss, even without accounting for the "second-car effect" documented for these types of restrictions. Such uniform policy not only fails to remove old cars from the road; it also reduces their prices, extending their lives and dampening sales of new cars. In fact, our simulations show that a uniform one-day-a-week restriction leads to a welfare *loss*

[9] Gavazza et al. (2014) omit any market dynamics by focusing on the steady-state equilibrium. Adda and Cooper (2000) consider dynamics but all agents care equally about quality. This homogeneity assumption leads, among other things, to an immediate adjustment of the equilibrium scrappage age following a policy intervention, which is at odds with our results that exhibit a gradual adjustment as the fleet evolves to its new steady-state equilibrium.

[10] We are not the first to document a large variation in emission rates across vehicles of different vintage (and other observables). See, for example, the work of Kahn (1996) and Knittel and Sandler (2018), based on U.S. data. Our contribution lies in understanding the implication of this variation for policy design and choice.

of 92% of the welfare gain from implementing the first-best. In contrast, a vintage-specific restriction that establishes a complete ban on cars at least 16 years old leads to a welfare *gain* of 51% of the first-best gain.

We also extend the model to study alternative instruments that have either been used or received some attention by policy makers in their quest to curb local air pollution, namely, scrappage subsidies and registration fees (i.e., motor taxes). Much has been written on the use of scrappage subsidies, also known as cash-for-clunker programs (e.g., Hahn, 1995; Adda and Cooper, 2000; Mian and Sufi, 2012; Hoekstra et al., 2017). Our model shows that a scrappage subsidy does not present any advantage over a vintage-specific restriction. One reason is that vintage restrictions allow the flow of older cars to zones where old cars are still socially valuable. Another is implementation constraints. For a scrappage subsidy to reach its full potential, the regulator must prevent old cars from outside the regulated area from being entitled to the subsidy. This can be done, although at the cost of introducing friction in the car market: requiring any scrapping vehicle to have a number of years of registration history in the regulated area. More important, and partly explained by the high fiscal cost incurred by the government, when these subsidies have been used, whether in the U.S. or Europe, they have tended to be short-lived, lasting only a few months.

A growing literature has also been examining the effect of registration fees/subsidies on new car purchase decisions (e.g., d'Haultfoeuille et al., 2014; Adamou et al., 2014).[11] We extend the model to consider annual (pollution-based) registration fees on both new and old units. Because of the temporal variation in pollution harm, the optimal registration design is to offer each year a menu of registration fees that vary by vintage: drivers have the option to pay either a positive fee for unlimited use of the car (approximately equal to the pollution harm that is *expected* from its use) or no fee for use of the car only during times of little or no pollution, say, during weekends and late at night. Like vintage restrictions and scrappage subsidies, registration fees are designed to act exclusively on the extensive margin, but more effectively since they involve a complete set of prices to adjust this margin at each vintage level, which vintage restrictions and subsidies fail to do. In fact, our simulations show that registration fees deliver 79% of the first-best gain. Whether registration fees can be used in practice, and in the menu format that we propose, is an open question, particularly in developing and emerging economies, which, according to Posada et al. (2015), tend to favor quantity instruments (e.g., restrictions) over price instruments (e.g., taxes and subsidies).[12] But even in the absence of these latter, the results of this paper show that well-designed,

[11] So far, these fees/subsidies cover only CO_2 emissions. See Drummond and Ekins (2016) for a proposal to extend them in the UK to also cover NO_x emissions from new diesel cars. London's pollution charge, enacted on October 2017, is another effort to tax cars according to their emission rates of local pollutants.

[12] Introducing pollution-based circulation fees, for instance, seems to be a major policy challenge for any authority, as they imply a complete reversal of existing circulation-fee profiles, under which older cars pay much less than newer cars. This contrasts with existing driving restrictions, under which older cars are already subject to much tougher restrictions than newer cars.

vintage-specific restrictions are a good alternative. In that regard, these restrictions look particularly well suited, for example, to accelerate the introduction of electric vehicles at much lower cost to government than by existing subsidies.[13]

The rest of this paper is organized as follows. Using Santiago's 1992 reform as motivating evidence, Section 2 shows how vintage-specific restrictions can significantly affect fleet composition and vehicle emissions. The model of the car market is developed in Section 3 and estimated in Section 4. Policy exercises for different driving restriction formats and alternative instruments are presented in Section 5. Final remarks are offered in Section 6.

2. Motivating evidence: Santiago's 1992 driving restriction

The city of Santiago, Chile's capital and home to 40% of the country's 17.5 million people, exhibits one of the worst air pollution problems of any urban center in Latin America, due partly to its geography but also to a steady increase in car use. Efforts to control vehicle emissions date back to at least the mid 1980's, first in 1985 with a total prohibition on the import of used cars and then in the winter of 1986, with the introduction of a driving restriction program. At the time, the restriction was intended to operate as an exceptional measure by banning the circulation of 20% of the vehicle fleet only on those days when air pollution was expected to reach critical levels. Over time these restriction episodes were called upon more often, and by 1990 the restriction program applied every weekday from 6:30 am to 8:30 pm from March through September, the time of year when thermal inversions and lack of wind prevent pollutants from dispersing rapidly.

The restriction program experimented an important change in 1992, when the government issued an executive order that, starting in 1993, any new vehicle must be equipped with a catalytic converter in order to circulate in Santiago. In addition, to accelerate the turnover toward these cleaner vehicles, the government decided to exempt from the existing driving restriction all cars equipped with a converter. Given the absence of converters in vehicle models released in 1992 and before,[14] the 1992 reform introduced a sharp discontinuity between the 1992 and 1993 vintages that we exploit here as motivating evidence to illustrate the potential for vintage-specific restrictions to affect fleet turnover and, hence, reduce vehicle emissions.

[13] We also consider gasoline taxes. In the very short-run, when drivers are only allowed to adjust the amount of driving, as in Knittel and Sandler (2018), the gasoline tax delivers 50% of the first-best gain; more than the vintage restriction, and more than the 35% in Knittel and Sandler (2018). This "intensive-margin" advantage fades rapidly as drivers need to be induced to the purchase of cleaner cars, so much that in our simulations the gasoline tax delivers only 16% of the first-best gain. However, since vintage restrictions and gasoline taxes act on different margins, combining the two delivers 61% of the first-best gain, a 20% increase above the gain from the vintage restriction alone.

[14] A negligible number of pre-1993 Honda Accord models were equipped with converters at the time of the reform. We exploit this in an exercise in the online Appendix (Section B.4).

2.1. *Effects on fleet composition*

The database we use to study changes in fleet composition comes from vehicle circulation permits at the municipality level collected by the National Statistics Bureau. In March every year, each car owner is required to obtain a circulation permit upon payment of an annual fee to her home municipality. This data, which is only available for the period 2006-2012 and for 332 of the 346 municipalities in the country,[15] specifies the number of cars of each vintage by municipality and year.

Figure 1 presents evidence suggesting effects of the policy. The figure shows fleet composition in 2006 by vintage for Santiago (the area affected by the driving restriction) and for the rest of the country with darker bars corresponding to pre-1993 models (i.e., 1992 and older), the ones subject to the restriction, and lighter bars corresponding to post-1992 models. Observe first that the fleet in Santiago is indeed cleaner (i.e., it has a larger fraction of post-1992 cars) than the fleet in the rest of the country. Also significant is that while most jumps in the number of cars per vintage are positively correlated between Santiago and the rest of the country, the jumps in 1992 and 1993 vintages (and surrounding vintages) are negatively correlated, suggesting something special regarding these vintages in Santiago relative to the rest of the country. Without controlling for other variables, however, it is not obvious a priori how much of what we see in the figure is due to the 1992 policy and how much is due to characteristics specific to Santiago that might affect car-purchasing decisions (e.g., higher average income in Santiago).

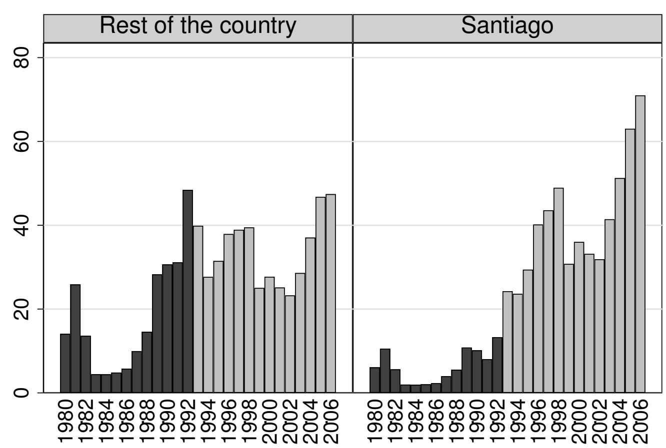


Figure 1: Fleets in Santiago vs the rest of country in 2006

Notes: Each bar represents number of cars (in thousands) of each vintage. Pre-1993 vintages are highlighted as darker bars.

[15]The municipalities missing information are in remote areas with low population density. More details about the database are in the online Appendix (Section B.1).

To isolate the effect of the policy on q_τ^i, the number of vintage-τ cars in municipality i in a given year, we estimate the following regression:

$$\log(q_\tau^i) = \beta_\tau DR_i + \alpha_\tau \log(INCOME_i) + \gamma_\tau \log(POP_i) + \delta_\tau + \varepsilon_\tau^i \tag{1}$$

where DR_i is a dummy that takes the value of 1 if municipality i is affected by the driving restriction (i.e., if it is located in Santiago), $INCOME_i$ is municipality i's income per capita, POP_i is the municipality's total population, and δ_τ is a vintage fixed effect.

Equation (1) describes in a reduced form the relationship between number of cars from each vintage and municipality characteristics. We can use it to estimate causal effects of the policy by focusing on the 92-93 discontinuity. As cars of contiguous vintages share similar characteristics and are freely traded across regions, we should expect the ratio $q_\tau^i/q_{\tau+1}^i$ to be similar in all municipalities except for the 92 and 93 vintages, which is when the attributes of these two contiguous vintages dramatically change with the restriction, i.e., q_{92}^i/q_{93}^i should be much smaller in municipalities located in Santiago than in the rest of the country. Taking the difference of (1) for contiguous vintages, we obtain

$$\log(q_\tau^i/q_{\tau+1}^i) = \Delta\beta_\tau DR_i + \Delta\alpha_\tau \log(INCOME_i) + \Delta\gamma_\tau \log(POP_i) + \Delta\delta_\tau + \Delta\varepsilon_\tau^i \tag{2}$$

where $\Delta\beta_\tau = \beta_\tau - \beta_{\tau+1}$ and so on.[16]

Table 1 presents estimates of $\Delta\beta_\tau$ for different vintages and for the 2006 sample.[17] Columns 1 through 3 present coefficients for the 92-93 dyad using different sets of controls: column 1 includes no controls, column 2 includes the controls specified in (2), and column 3 includes a vector of additional controls as robustness check (quadratic income per-capita, the coefficient of variation of income per capita, urbanization rate, a quadratic function of distance to Santiago, and dummies for municipalities in northern and far away regions). Coefficients, which do not vary much across specifications, are statistically as well as economically significant. According to the point estimate in column 3, for instance, if in a given municipality not affected by the restriction we observe one 93 model for each 92 model (i.e., $q_{93}/q_{92} = 1$), in a similar municipality in Santiago, that ratio would be $\exp(1.241) \approx 3.46$.

The remaining columns in Table 1 present estimates for other contiguous vintages. Under the assumptions of our identification strategy, we should find a zero effect for these dyads: for them, the policy should create no incentives for jumps in car ownership in Santiago relative to the rest of the country. Results in columns 4 and 5 of Table 1 confirm this to be the case for the 91-92 and 93-94 dyads, but it actually extends to other dyads (see the online Appendix,

[16] Note that the left hand side of equation (2) can be re-written as $\log(y_{\tau,\tau+1}^i/(1-y_{\tau,\tau+1}^i))$, where $y_{\tau,\tau+1}^i = q_\tau^i/(q_\tau^i + q_{\tau+1}^i)$, consistent with a logistic model where consumers choose between cars vintage τ and $\tau+1$.

[17] We use 2006 because that is the year closest to policy implementation, but similar results are obtained for more recent years.

Section A.2, Table A.1). In all, we take this as convincing evidence that the driving restriction did have an effect on fleet composition.

Table 1: Effects of the 1992 restriction on share of cars for contiguous vintages

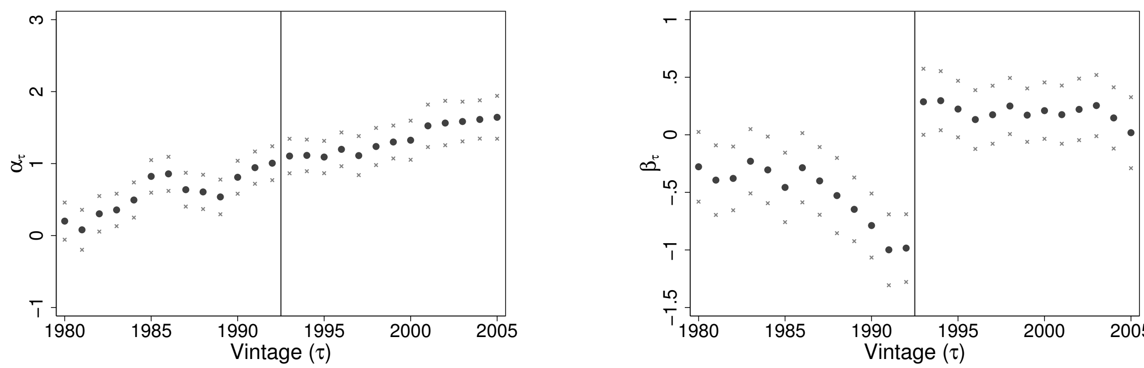
	92-93	92-93	92-93	91-92	93-94
DR_i	-1.283***	-1.257***	-1.184***	-0.0211	0.0882
	(0.062)	(0.072)	(0.106)	(0.070)	(0.058)
Controls	No	Yes	Yes$^+$	Yes$^+$	Yes$^+$
Observations	332	332	332	332	332
R^2	0.465	0.473	0.492	0.079	0.201

Notes: OLS regressions with one observation per municipality. The dependent variable corresponds to $\log(q_\tau)/\log(q_{\tau+1})$, where q^i_τ is the total number of cars of vintage τ found in municipality i in 2006. The first three columns correspond to the case of $\tau = 1992$, while columns 4 and 5 correspond to $\tau = 1991$ and $\tau = 1993$, respectively. Standard errors are calculated via block bootstrap at the province level (53 provinces in total). Municipality controls in column 2 (Yes) include income per capita and population. Municipality controls in columns 3 to 5 (Yes$^+$) include the same controls of column 2 plus quadratic income per-capita, coefficient of variation of income per capita, urbanization ratio, a quadratic function of distance to Santiago, and dummies for municipalities in northern and far away regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Having established a significant policy effect around the 92-93 discontinuity, we return to equation (1) to estimate policy effects away from it. We do so by assuming that, conditional on income and population, unobserved municipality characteristics that may affect q_τ do not differ between Santiago and the rest of the country.[18] Figure 2 presents regression results for year 2006 with estimates for both α_τ and β_τ (to save space, estimates for γ_τ are in the online Appendix, Section A.1, Figure A.1). Results in panel (a) are consistent with the idea that income is a main factor behind purchasing decisions and, therefore, newer models are indeed concentrated in richer municipalities. More importantly for our identification strategy, the relationship is smooth, with no jump around the 92 and 93 vintages (the same happens for the coefficient on population, γ_τ).

In contrast, and consistent with results in Table 1, the 92-93 discontinuity is clear in Panel (b). The point estimate for vintage 92 is -0.984 (statistically significant at the 1% level). This implies that for each 1992 model circulating in a given municipality in Santiago, 2.68 such models will be in a similar municipality not subject to restriction. Conversely, the point estimate for vintage 93 of 0.287 (statistically significant at the 5% level) indicates that for each 1993 model circulating in a given municipality in Santiago, only 0.75 such models will be in a similar municipality not subject to restriction.

[18]In the online Appendix (Section B.2) we also exploit the 1992-93 threshold in a regression discontinuity design. The RDD coefficient we obtain, 1.222, is very close to the values reported in Table 1, confirming our key identification assumption, that DR_i captures only effects of the policy on each vintage's car stock. This and the fact that population and income do not have a discontinuous jump around the 92-93 discontinuity (as shown below) serve to motivate equation (1) in the spirit of the estimator suggested by Angrist and Rokkanen (2015) to identify effects outside a policy discontinuity, which relies on the identification of the policy effect conditional on the included covariates.



(a) *INCOME* coefficients (b) *DR* coefficients

Figure 2: Vintage effects of driving restrictions and income

Notes: This figure presents estimated vintage effects after estimating equation (1) using data at the municipality level for 2006. The panels present the coefficients of (a) income and (b) the driving restriction. Dark dots represent point estimates for each coefficient and light gray dots correspond to 95% confidence intervals using robust standard errors. The vertical line in each panel marks the division between the 1992 and 1993 vintages.

The evolution of the driving restriction's estimated effects as we move away from the 92-93 discontinuity in either direction is also informative. In a market for products that are vertically differentiated and where consumers differ in their willingness to pay for higher quality (i.e., newer models), the null effects of the driving restriction program for the newest models should come as no surprise. Regardless of location, a driver's alternative to, say, a 2004 model is not a model that is ten or more years older but one closer to 2004. In other words, ownership decisions concerning models further from the discontinuity should be independent of the policy. The same logic applies to the fact that the policy (i.e., *DR*) coefficients revert toward zero for very old models, so it would be wrong to interpret this reversal as an indication that some drivers who own a pre-1993 model are bypassing the restriction by purchasing an additional 1980-86 model rather than a 1993 or newer one.[19]

2.2. *Effects on car prices*

In addition to fleet composition effects, documenting price effects is important for several reasons. First, the effect on prices provides an indirect check of whether the policy was enforced or not. If the driving restriction were actually binding, one would expect to find a large impact not only on the allocation of pre- and post-1992 models, but also on market prices given Santiago's large market share (41.8% of the national fleet in 2006). Second, since we have no data on fleet composition before 2006, price effects give a sense of the policy's

[19] In the online Appendix (Section B.3) we test for this "second-car" effect using car ownership information from household surveys. We found no evidence supporting the effect.

effects in years closer to its implementation. And third, estimating the effect on prices is also important, as it provides an estimate of the cost of the restriction to individuals, and in particular, of the (lower) cost of bypassing the restriction not by purchasing a second old, polluting car, but rather by upgrading to a newer, exempt car, preventing the "second-car effect" that is well documented for restriction programs which make no vintage distinctions (e.g., Davis, 2008).

We assembled a dataset of newspaper ads with car offers for new and used cars published in "*El Mercurio*" –Chile's main newspaper– during 1988-2000. Our sample considers price offers for a set of the most traded models on the market covering a wide price range: Fiat Uno, Honda Accord, Honda Civic, Mazda 323, Peugeot 205, Peugeot 505, and Toyota Corolla. Our empirical strategy is motivated by the evident price discontinuity between the 1992 and 1993 vintages that is observed in the data (offers of Toyota Corollas, for example, are displayed in the online Appendix, Section A.1, Figure A.2).

We estimate the following equation that pools observations from the seven models mentioned above:

$$\log(P_{im\tau t}) = \beta DR_\tau + g(\tau) + \delta_a + \delta_t + \delta_m + \varepsilon_{im\tau t} \tag{3}$$

where $P_{im\tau t}$ is the price offer of ad i placed at time t for a vehicle model m of vintage τ, DR_τ is a dummy equal to one for cars equipped with a catalytic converter, i.e., for all $\tau \geq 1993$, $g(\tau)$ is a parametric function of τ,[20] δ_a, δ_t and δ_m are age of the car (where age $a = t - \tau$), date of the ad and model fixed effects, respectively, and $\varepsilon_{im\tau t}$ is the error term. Note that we cannot control for vintage fixed effects, as the DR_τ variable is collinear with them. Identification in this case relies on the assumption that all price differences across vintages unrelated to the driving restriction are captured by age and time fixed effects, and by $g(\tau)$.

Table 2 presents estimates of equation (3) (results for the different models are in the online Appendix, Section A.2, Table A.2). Column 1 presents the estimate of β when controlling only for age and date fixed effects (i.e., $g(\tau) = 0$), while columns 2 to 4 present estimates of β for the alternative specifications of $g(\tau)$. Most of the estimates do not change significantly, suggesting that vintage effects do not affect results (if anything, the catalytic converter estimates increase when controlled for). Overall, we find a 6.5 log point premium for having a catalytic converter installed. As reported in the online Appendix (Section A.2, Table A.2), this premium tends to be larger for more expensive models (e.g., 12 log points for a Honda Accord vs. 5 log points for a Honda Civic), consistent with a situation in which individuals who own more expensive cars have a greater opportunity cost of not driving every day and, therefore, are willing to

[20] As indicated in Table 2, we use four different specifications for $g(\tau)$: (i) in column 1 we assume $g(\tau) = 0$, (ii) in column 2 we use prices of new cars as proxy for a car's intrinsic quality (see online Appendix, Section B.5, for details on how we construct this proxy), (iii) in column 3 we use a linear function of vintage that takes a different slope for before and after 1993, and (iv) in column 4 we use interactions of age dummies with linear trends in time to allow for different depreciation rates.

pay more for cars exempted from the restriction.[21,22]

Table 2: Effects of catalytic converter on the price of used cars

	(1)	(2)	(3)	(4)
DR_τ	0.048***	0.065***	0.065***	0.064***
	(0.006)	(0.004)	(0.004)	(0.003)
Age, Model and Date f.e.	Yes	Yes	Yes	Yes
$g(\tau)$	No	Quality proxy	Flexible line	Flexible age f.e.
Observations	56796	35309	35309	35309

Notes: The table presents results for equation (3) when pooling all models. Results for each model are in the online Appendix (Section A.2, Table A.2). The unit of observation is a car offer published the first Sunday of each month between 1988 and 2000. Standard errors, which are clustered by ad date, are presented in parentheses. More details on $g(\tau)$ are in footnote 20. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.3. *Effects on local pollutant emissions*

We finish the presentation of our motivating evidence with a computation of the potential impact of the 1992 policy on vehicle emissions of local pollutants, and ultimately, on pollution harm. Figure 3 presents our starting point: average smog-check readings of CO and HC as a function of vintage for the year 2008, the first year for which we have data.[23] Results are quite evident: Figure 3 shows the large impact that catalytic converters had on emissions (as evidenced by the big jump between 1992 and 1993 vintages).[24]

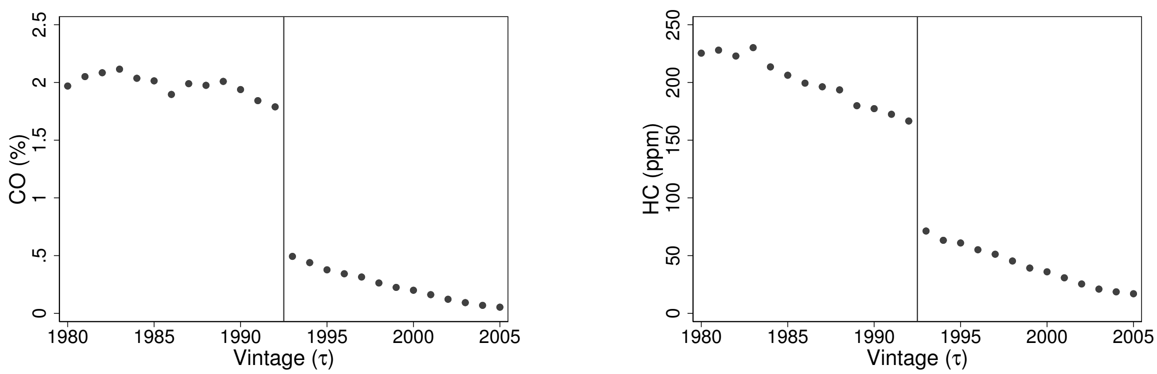
With this evidence in hand, we now make a "back-of-the-envelope" estimation of the policy effects on vehicle emissions using reduced-form evidence. We provide estimates for the two

[21] Notice that the cost of $265 for replacing a catalytic reported in Onursal and Gautam (1997) cannot explain the effects we report in Table 2. First, note that this value is just a small share of the total price of a new car. Second, this cost difference should be captured by our control for the price differences of new cars reported in column 2. Third, if differences were explained by the fixed cost of installing a catalytic converter, we should expect greater percentage differences in prices for less expensive cars, which is exactly the opposite of what we observe. Fourth, notice that converters can only be installed in vehicles with spark-ignition engines (Onursal and Gautam, 1997) , which explains why, at least in Santiago, we did not observe pre-1993 vehicles being retrofitted with converters.

[22] We report similar results from two additional empirical exercises in the online Appendix (Section B.4). One exercise is a regression discontinuity design with τ as the running variable and $\tau \geq 1993$ as treated vintages. The other one exploits the fact that our database contains a few ads for some pre-1993 Honda Accords that were already equipped with a catalytic converter.

[23] Emission rates are obtained from a dataset with information on all smog checks (i.e., vehicle inspections) carried out in the country during the period 2008-2016. With the exception of new vehicles, which are exempt for two years, all vehicles are required each year to pass these inspections before their circulation permit is renewed for the following year. In addition to test results, each observation reports test location and a unique vehicle identification number (i.e., license plate). HC emission rates are given in parts per millions and CO rates are expressed as a percentage of the exhaust, both rates are taken under an engine speed of 2500 rpm.

[24] Figure 3 also shows that emission rates in cars equipped with converters increase with age. As explained in the online Appendix (Section D.3), this is because (i) newer cars enter with cleaner technologies (although much attenuated after 2003) and (ii) pollution-control technologies wear out overtime.



(a) CO emissions

(b) HC emissions

Figure 3: CO and HC emissions

Notes: The figure presents average smog check readings of CO and HC as a function of vintage, based on information taken under an engine speed of 2500 rpm and collected from all inspection stations in the country in 2008.

local pollutants from Figure 3 for both Santiago (the restricted area) and the rest of the country (the non-restricted area).[25]

Denoting by E_τ^k total emissions from vintage-τ cars in area $k \in \{r, nr\}$ in a given year, a first approximation of the policy effects on vehicle emissions that year would be:

$$\Delta E_\tau^k = \left(q_\tau^k x_\tau^k - \hat{q}_\tau^k \hat{x}_\tau^k\right) e_\tau \tag{4}$$

where q_τ^k is the total number of cars of vintage τ that are actually in area k, $\hat{q}_\tau^k$ is the total number of cars of vintage τ that would have been observed in area k in the absence of the policy, x_τ^k is the average number of miles that cars of vintage τ were actually driven in area k, $\hat{x}_\tau^k$ is the average number of miles those cars would have been driven in the absence of the policy, and e_τ is the average amount of pollution emitted per mile by a car of vintage τ. The change in pollution harm due to the policy then would simply be:

$$\Delta H = h_r \sum_\tau \Delta E_\tau^r + h_{nr} \sum_\tau \Delta E_\tau^{nr} \tag{5}$$

where h_k is the externality cost per unit of pollutant emitted in area k.

Because of data availability, we provide only estimates of ΔE_τ^k and ΔH for year 2006. The components that enter in equations (4) and (5) are obtained from different sources. Emission

[25] Unfortunately, we cannot conduct a similar evaluation for NO_x because there are no readings available for pre-1993 models. The (ASM) technology that measures NO_x only works for models with catalytic converters. According to Onursal and Gautam (1997), however, converters can reduce CO and HC emissions by about 90 percent and NO_x emissions by 70 percent, so the policy impact on NO_x emissions should not differ much from that on HC and CO.

rates e_τ for CO and HC are obtained from the values plotted in Figure 3.[26] Miles traveled, x^k_τ and $\hat{x}^k_\tau$, are also obtained from the smog-check database. As odometer readings only began to be collected and reported in recent years, we obtain this information from the more complete readings of 2015 and 2016. We find annual travel to average 12,081 miles for the first year and to decline with age at a constant average rate of 249 miles per year.[27] We take these average estimates to equal $\hat{x}^{nr}_\tau$ and x^{nr}_τ for all vintages, and to equal $\hat{x}^r_\tau$ and x^r_τ for $\tau \geq 1993$. For $\tau \leq 1992$, we let $\hat{x}^r_\tau$ and x^r_τ differ anywhere between 0 and 8.2%.[28]

The more challenging part of this exercise comes from estimating changes in car stocks in response to the policy. While q^k_τ is obtained directly from the circulation-permit data described in Section 2.1, $\hat{q}^k_\tau$ is to be estimated using (1). However, to continue with the exercise, we need to assume at this stage that the policy had no effect on the overall (national) fleet. Taking this latter as given, it considers the policy's impact only on the fleet redistribution between Santiago and the rest of the country. Computing impacts on the national fleet, as well, requires a dynamic model of the car market like the one we develop and estimate in the following sections. Nevertheless, we find this exercise informative, as it provides a benchmark for the policy's effect in a very transparent way.[29]

Thus, taking the overall fleet in 2006 as given implies that:

$$q^r_\tau + q^{nr}_\tau = \hat{q}^r_\tau + \hat{q}^{nr}_\tau \tag{6}$$

and from (1) we have that:

$$q^r_\tau / q^{nr}_\tau = \exp(\beta_\tau)\Psi^r_\tau / \Psi^{nr}_\tau \tag{7}$$

$$\hat{q}^r_\tau / \hat{q}^{nr}_\tau = \Psi^r_\tau / \Psi^{nr}_\tau \tag{8}$$

where $\Psi^k_\tau = \sum_{i \in k} (INCOME_i)^{\alpha_\tau} (POP_i)^{\gamma_\tau} \exp(z_i'\zeta + \delta_\tau + \varepsilon^i_\tau)$ for $k = r, nr$. Dividing (7) by (8) and using (6) we arrive at

$$\hat{q}^r_\tau - q^r_\tau = q^{nr}_\tau - \hat{q}^{nr}_\tau = -\frac{q^{nr}_\tau q^r_\tau (1 - \exp(-\beta_\tau))}{q^{nr}_\tau + q^r_\tau \exp(-\beta_\tau)} \tag{9}$$

[26] Notice that Figure 3 uses emissions rates in 2008 because emission rates in 2006 are not available. This is less of a problem if they are thought to differ by similar percentage levels across vintages. At least, this is what we find when we repeat the exercise using 2009 emission rates. It is also important to note that since our estimates of ΔE^k_τ and ΔH will be presented in percentage terms we need not convert the HC and CO readings into emissions (e.g., grams of pollutant) per mile driven.

[27] This rate is slightly higher than the constant rate of 233 miles per year in Lu (2006), which is based on U.S. data.

[28] Given that the restriction applied only once a week during 30 weeks of the year, 8.2% (the result of dividing 30 by 365) should be seen as the upper limit of policy intensity. In reality the effective restriction is expected to be less than that as some trips in days of restriction may have been moved to days of no restriction. This intertemporal substitution is found in other programs as well (see Gallego et al., 2013).

[29] We provide a more comprehensive policy evaluation using our model in the online Appendix (Section E.1), where we show similar but attenuated results because of a positive policy effect on the entry of new cars.

Using (9) and the estimates of β_τ presented in Figure 2(b), we obtain values for $\hat{q}_\tau^r$ and $\hat{q}_\tau^{nr}$.

Based on this information for the different components in ΔE_τ^k, we estimate that CO emissions in Santiago dropped anywhere between 20 and 27% in 2006 because of the 1992 policy (the reduction in HC emissions is anywhere between 13 and 20%). A good fraction of these reductions, those attributed to changes in fleet composition, must be contrasted with equivalent increases in the rest of the country. Using the numbers in Parry and Strand (2012), who report that vehicle emissions in Santiago are almost 9 times more damaging than in the rest of the country, the net reduction in externality costs, ΔH, is estimated to be anywhere between 15 and 21% for CO and between 10 and 16% for HC. Although these numbers are big by any measure, even if driven exclusively by the fleet-composition effect, they still need to be contrasted with the policy costs incurred by households that had to adjust their purchasing and driving decisions. Doing this requires a model of the car market, which we present next.

3. A model of the car market

The key message that emerges from the Santiago-1992 reform is that vintage-specific restrictions are a potentially useful tool to fight air pollution by helping to accelerate the fleet composition toward lower-emitting vehicles. Yet, the empirical analysis cannot answer many policy-relevant questions: What are these restrictions' overall welfare implications? How does welfare vary as we introduce vintage-specific considerations in the restriction design? How do these vintage-specific designs compare to alternative policy instruments? What mechanisms explain the difference between instruments' performance? We address these questions by developing a model of the car market that is then estimated and used for policy simulation based on data from several local sources (e.g., smog checks, circulation permits). Although the numbers that emerge from the estimated model are specific to Santiago's current pollution problem, their qualitative implications apply more broadly, since nothing specific in the model prevents its application to other cities and contexts. We present the model in this section and leave the estimation and application for the following two sections.

3.1. *Car dealers and households*

There are three agents in the economy: car producers, car dealers and drivers or households. They all discount the future at $\delta \in (0, 1)$. The cost of producing a new car is c, which is also the price at which perfectly competitive producers sell new cars to car dealers.[30] A large number of car dealers buy new cars from car producers and rent them, together with second-

[30] We could change the interpretation of c to represent marginal cost plus a mark up in non competitive markets and conclusions from the model would remain the same. The model's main mechanism is driven by the relationship between car dealers and drivers rather than between car producers and car dealers.

hand cars, to drivers.[31] The (annual) rental price for a car of age $a = \{0, 1, 2, ...\}$ at date t is denoted by $p_{a,t}$ ($a = 0$ corresponds to a new car). Note the change of notation from vintage τ to age $a = t - \tau$. Our model makes no distinction between the two because the car technology is invariant to time (i.e., there is no technological progress), so age is used only to facilitate the exposition, without changing the substance of the results.

Cars exit the market at some exogenous rate due to crashes, fatal malfunctioning, etc. This rate may vary with car age, so the probability that at age a, a car in the market at date t is still in the market at date $t+1$ is $\gamma_a \in (0, 1)$, with $\gamma_a \geq \gamma_{a+1}$. In addition, at any date t, there is an (endogenous) age $T(t)$ at which a fraction of the surviving cars of that age and any older get scrapped for a value v.[32] The remaining fraction of the surviving cars of age $T(t)$ get rented for $p_{T(t),t}$. Since in equilibrium, dealers must be indifferent between scrapping an age $T(t)$ vehicle today and renting it today (and scrapping it tomorrow, provided the vehicle still exists, which happens with probability $\gamma_{T(t)}$), we have that:

$$v = p_{T(t),t} + \gamma_{T(t)} \delta v \tag{10}$$

Furthermore, since car dealers take rental prices as given, their problem in each period t is not only to decide how many old cars of age $T(t)$ to scrap in that period, but also how many new cars to bring to the market in that period so as to satisfy the break-even condition:

$$c = p_{0,t} + \sum_{i=1}^{\Gamma(t)} (\gamma_{i-1}\delta)^i p_{i,t+i} + (\gamma_{i-1}\delta)^{\Gamma(t)+1} v \tag{11}$$

where $\Gamma(t)$ is the (endogenous) age at which a car bought at date t is expected to be retired (or rented for the last time). Since there is no uncertainty, $\Gamma(t) = T(t + \Gamma(t))$.

There is a continuum of households/drivers of mass 1 that vary in their willingness to pay for quality and also in how much they drive. A type-θ consumer who rents an age-a car for p_a and runs it for x miles obtains a utility of (to save on notation henceforth, in many places we will omit the subscript "t" unless it is strictly necessary):

$$u(\theta, a, x) = \frac{\alpha}{\alpha - 1} \theta s_a x^{(\alpha-1)/\alpha} - \psi_a x - p_a \tag{12}$$

where θ is distributed according to the cumulative distribution function $F(\theta)$ over the interval $[0, \bar{\theta}]$. The first term in (12) corresponds to the driver's gross benefit from car travel, which

[31] Note that the renting assumption, which is also in Bento et al. (2009), is equivalent to assuming a frictionless secondary market that clears once per period. Evidence provided in the online Appendix (Section A.2, Table A.3) suggests markets are fairly integrated across the country. Cars in Santiago tend to be 2-3% cheaper than elsewhere in the country, consistent with the costs of moving them from one city to another.

[32] This scrappage value can be interpreted, for example, as the value a dealer gets for selling remaining parts or exporting a car to another country (we assume v to be insensitive to $T(t)$).

exhibits decreasing returns (i.e., $\alpha > 1$). This benefit depends on her type and the car's quality, $s_a > 0$, which is assumed to fall with age according to $s_{a+1} = \varsigma s_a$ with $\varsigma \in (0, 1)$. A declining quality reflects that older cars are more likely to break down and/or that they lack the latest technological advances.[33] The second term in (12) captures monetary (e.g., parking, gasoline, maintenance) and non-monetary (e.g., time) costs of travel.[34] We allow ψ_a to vary with age, with $\psi_{a+1} \geq \psi_a$, to control for the fact that newer models tend to be more fuel efficient than older ones. A driver θ who decides not to rent a car but rather to use public transport obtains an outside utility that we assume equal across households, so we normalize it to zero.

The problem of a type-θ household is to decide what car to rent (extensive margin) and how much to use it (intensive margin) so as to maximize equation (12). If this household happens to rent an a-year-old car, utility maximization leads to:

$$x(\theta) = \left(\frac{\theta s_a}{\psi_a}\right)^{\alpha} \tag{13}$$

miles of driving per period. Anticipating this, it will rent a car of age a that, provided it delivers more utility than using public transport, maximizes (12):

$$a(\theta) \in \arg\max_{a} \{\kappa_a (\theta s_a)^{\alpha} - p_a\} \tag{14}$$

where $\kappa_a = [(\alpha - 1)\psi_a^{\alpha-1}]^{-1}$.

Equations (13) and (14) capture two empirical regularities that we observe in our data (and in Lu, 2006 for U.S. data): that households which value quality more (i.e., higher θ) tend to drive newer cars and that newer cars are, on average, run more often.[35] Central to our analysis is to understand how different policy interventions affect these two decision margins. Some interventions may act exclusively on the extensive margin by only altering rental prices (scrappage subsidies and registration fees), while others may act on both margins by also altering the amount of driving, either with a price (gasoline taxes) or by rationing (driving restrictions).

3.2. *Pollution*

Cars emit all sorts of pollutants, some with global effects (e.g., CO_2) while other with local effects, i.e., effects at the city level (e.g., CO, HC, NO_x). The focus of this paper is on local

[33]A linear quality decay rate is also in Gavazza et al. (2014).

[34]If congestion is a problem, travel costs may also include (socially optimal) congestion charges, which we do not model explicitly.

[35]We abstract from the possibility of households renting multiple cars. Figure B.2 in the online Appendix shows that a very small fraction of households own two or more cars. More importantly, a well-designed vintage-specific restriction eliminates this possibility by construction, as we shall see.

pollutants, so how harmful is a car's emission to society depends not only on how often the car is used but also on where is used.[36] We assume that drivers live and use their cars in two distinct areas: "polluted" and "non-polluted" areas.[37]

Following the motivating section, we also assume that cars emit potentially more local pollution as they age, so we let e_a be an age-a car's emission rate per mile, with $e_{a+1} \geq e_a$. Thus, the harm per mile generated by a car of vintage a is $h_r e_a$ in the polluted area (e.g., Santiago) and $h_{nr} e_a$ in the non-polluted area (e.g., rest of the country), with $h_r \gg h_{nr} \approx 0$. This latter implies that we will consider pollution-control policies in the polluted area only, which nevertheless has implications for the entire car market. For this reason, in the paper we refer to the polluted area as the "restricted" area and the non-polluted area as the "non-restricted" area. Drivers' valuation θ in area $k \in \{r, nr\}$ is distributed according to the cumulative distribution $F_k(\theta)$ over the interval $[0, \bar{\theta}]$, where $\mu F_r(\theta) + (1-\mu)F_{nr}(\theta) = F(\theta)$ for all $\theta \in [0, \bar{\theta}]$ and μ is the fraction of households living in the restricted area.

We abstract from any effect that policy interventions may have on pollution coming from public transport. The main reason for this is that changes in public-transport use predicted by our simulations (from drivers that either give up their cars or reduce their car travel) are too small to prompt any adjustment in a public-transport network with excess capacity.[38]

3.3. *Equilibrium benchmark: No intervention*

We first characterize the market equilibrium in the absence of any policy intervention. We assume that the second-hand market is well integrated, so households observe the same rental prices regardless of where they live. At the beginning of any given period, say, year y, there will be an overall stock of used cars of $\mathbf{S}_y = (q_{1,y}, q_{2,y}, ...)$, where $q_{a,y} = q^r_{a,y} + q^{nr}_{a,y}$ for all $a \geq 1$. As a function of that stock, the market equilibrium must satisfy several conditions. First, it must be true that in equilibrium, drivers of higher-valuation types rent newer cars. A series of cutoff levels $\{\theta_{0,t}, \theta_{1,t}, ...\}_{t=y}^{+\infty}$ precisely determines the prices at which certain drivers rent certain cars at time $t \geq y$.[39] Denote by $\theta_{a,t}$ the driver who, at time t, is indifferent to renting a car of age a at price $p_{a,t}$ versus one of age $a+1$ at a lower price $p_{a+1,t}$, that is:

$$\kappa_a \left(\theta_a s_a\right)^\alpha - p_a = \kappa_a \left(\theta_a s_{a+1}\right)^\alpha - p_{a+1} \tag{15}$$

[36]It also depends on when the car is used (e.g., peak hours, weekends, winter months). We cover this temporal distinction in section 5.4.

[37]The model can be extended to consider households outside the polluted area that commute to it. As discussed in Section 4, we do not pursue this here because it is not empirically relevant for our application.

[38]For a recent evaluation of Santiago's public- trasport system see EMBARQ (2017).

[39]Note that while in the absence of intervention these cutoff levels do not vary across the two areas, this will change as the planner intervenes the market to correct for the pollution externality in the polluted area, even if the second-hand market continued to be well integrated across areas, as we assume throughout.

for all $a = 0, 1, ..., T-1$, where T is the age of the oldest car rented. Consumers of type $\theta \geq \theta_a$ rent age-a vehicles or newer while consumers of type $\theta < \theta_a$ rent older vehicles (or not at all for θ's sufficiently low). As in any vertical differentiation model, an obvious corollary from (15) is that a higher valuation consumer obtains strictly more surplus than a lower valuation consumer.

In equilibrium, the series of cutoff levels $\{\theta_0, \theta_1, ...\}$ must be consistent with the total population of drivers, the existing stock of used cars $\mathbf{S}$, and the new cars coming to the market (q_0) in period t. Hence, it must also hold that

$$q_0 = F(\bar{\theta}) - F(\theta_0) \text{ and } q_a = F(\theta_{a-1}) - F(\theta_a) \tag{16}$$

for all $a = 1, ..., T-1$ (and $t \geq y$) and where $F(\cdot) = \mu F_r(\cdot) + (1-\mu)F_{nr}(\cdot)$.

Since only a fraction of T-year-old vehicles are scrapped in equilibrium (while all older vehicles will be), we also have

$$F(\theta_{T-1}) - F(\theta_T) \leq \gamma_{T-1}q_{T-1} \tag{17}$$

where $\gamma_{T-1}q_{T-1}$ is the number of age T vehicles that survived from the last period.[40]

One last condition must hold in equilibrium: The lowest-valuation household to rent a car today, θ_T, obtains the same utility as using public transport:

$$\kappa_T (\theta_T s_T)^{\alpha} - p_T = 0 \tag{18}$$

If (18) does not hold, a dealer would be better off renting a T-year-old vehicle at a price slightly above p_T instead of scrapping it.

Together with conditions (10) and (11), conditions (15) through (18) determine the unique equilibrium for any given stock of used cars $\mathbf{S}_y$, that is, current and future rental prices of new and used cars, i.e., $\{p_{0,t}, ..., p_{a,t}, ..., p_{T(t),t}\}_{t=y}^{+\infty}$, and current and future sales of new cars, i.e., $\{q_{0,t}\}_{t=y}^{+\infty}$. Unlike some previous work, we are interested not only in the steady state equilibrium, but also in the equilibrium during the transitory states after a policy shock. Transitions can be particularly long in car markets, so they cannot be neglected in policy evaluation and design.

[40]Note that, because quality drops discretely with age, it can happen that in equilibrium, all age $T-1$ vehicles are rented but all age-T vehicles are scrapped; then the relevant scrapping condition is not (10) but

$$p_{T-1} + \delta\gamma v > v > p_T + \delta\gamma v$$

where p_T is the hypothetical rental price for an age-T vehicle.

3.4. *Available policy interventions*

Clearly, in the absence of any policy intervention the market equilibrium above leads to socially inefficient levels of pollution. As prescribed by Pigou almost 100 years ago, one way to restore efficiency is by taxing drivers an amount equal to the externality their driving imposes on the rest of society. Unfortunately, this Pigouvian approach —leaving aside its political resistance— is not technically feasible for handling local pollutants since actual emissions (i.e., $e_a x$) cannot be accurately observed by the regulator (Molina and Molina, 2002; Knittel and Sandler, 2018). Consequently, regulators must rely on alternative instruments.

In addition to different forms of driving restrictions, we consider other instruments that have either been used or received some attention by policy makers: namely, scrappage subsidies, annual registration/circulation fees (i.e., motor taxes), and gasoline taxes. With the exception of driving restrictions, the way these alternative instruments enter into our model is relatively simple. A scrappage subsidy, a payment for the retirement of old vehicles, enters by increasing the scrappage value in the polluted/restricted area from its baseline value of v to $v + \sigma$. Since the price differential σ creates incentives for drivers in the non-restricted area to scrap their vehicles in the restricted area, the regulator may try to prevent this arbitrage by requiring the scrapping vehicle to have a number of years of registration history in the restricted area, as currently done in California for example. A gasoline tax also enters the model in a straightforward manner, by increasing the variable cost of using a car in the restricted area from the baseline value of ψ_a to $\psi_a + g/mpg_a$, where g is the gasoline tax (in dollars per gallon) and mpg_a is a car's fuel economy measured in miles per gallon.[41] Unlike scrappage subsidies and gasoline taxes, annual registration/circulation fees can be made vintage specific.[42] They enter into the model by increasing the rental price of all polluting vehicles in the restricted area from p_a to $p_a + r_a$.

The way a driving restriction enters into the model is more involved depending on its design, which must specify the extent of the restriction and the car vintages affected. The extent of the restriction is captured by the parameter $R_a \leq 1$, which says that an a-year-old car can only be used in the polluted area a fraction of the time, say, 4 of 5 weekdays each week. Since drivers can move some trips from one weekday to another, R_a should not be read as 4/5 in this example, but possibly more. In any case, a fraction $1 - R_a$ of the trips that a driver would have otherwise made are rationed by the restriction, some more valuable than others. To model this inefficient rationing, we adopt the conservative assumption —less favorable to the driving restriction option— that the driving restriction destroys an equal fraction of car

[41] We omit from the model the possibility that households may drive outside the restricted area only to fill their tanks with cheaper gasoline.

[42] In theory, scrappage subsidies can also be made vintage specific; in particular, smaller for cars older than the youngest car being scrapped in equilibrium. This would help reduce the government's total bill, specially at the beginning, without affecting scrappage decisions.

trips of different values during the relevant period.[43] Some of these car trips may be replaced by using public transport.

Since we have not allowed yet for any temporal distinction as to when (i.e., hour of the day, day of the week, and month of the year) pollution is emitted, the latter assumption implies that a driving restriction reduces the number of car trips a driver would otherwise make uniformly over the week, or year, for that matter. Formally, a driver θ who rents an a-year-old car that faces an effective restriction R_a will now drive:

$$x(\theta, a, R_a) = R_a \left(\frac{\theta s_a}{\psi_a}\right)^{\alpha} \tag{19}$$

miles. As this travel reduction falls indistinctively over trips of different values, her utility reduces to:

$$u(\theta, a, R_a) = R_a \kappa_a (\theta s_a)^{\alpha} - p_a \tag{20}$$

per period (recall that the utility from using public transport, whether alone or in combination with private transport, is normalized to zero).[44]

By examining how these instruments work, we expect them to have different impacts on rental and usage decisions, and hence, on welfare. In general, restoring efficiency requires acting upon both decision margins, which makes the planner's problem of instrument design and choice far from trivial. We characterize this problem next, but in a simplified version of the general setting of sections 3.1 and 3.3, leaving the general setting for the estimation and simulations that follow.

3.5. *Planner's problem in a two-period world*

Closed-form solutions to a planner's problem are always useful to highlight the main forces at play behind those solutions and how they change to changes in parameter values. Unfortunately, no such solutions are available for the general setting developed so far, mainly because market transitions from one steady-state to another lack a simple recursive form one can build upon. Without losing much content, while keeping the main forces at play, we simplify the planner's problem by taking her to a situation where $h_{nr} = 0$ and cars last only two periods, i.e., $a \in \{0, 1\}$. In the second period, when $a = 1$, car dealers have the option to either scrap their cars for a value of v (or $v + \sigma$ if a scrappage subsidy is in place) or rent them to households, after which they are discarded for no value (which is equivalent to assuming $\gamma_1 = 0$).

A two-period setting makes the analysis tractable for two reasons. The first is that only

[43] This is the equivalent to using a proportional rationing rule (Tirole, 1988, p. 213).

[44] Note that plugging (19) in (12) does not lead to (20). Doing so would be equivalent to assuming that the restriction destroys the least valuable trips, replicating the work of a driving fee and departing from how a proportional rationing rule works.

rental and usage decisions in the polluted area are affected by policy. Rental prices announced by dealers remain largely unchanged to policy interventions; the only exception is rental prices in the polluted/restricted area after the planner introduces a scrappage subsidy (which is only available to cars with full registration history in that area). In all other cases, prices remain at their no-intervention levels. Assuming $\delta = \gamma_0 = 1$ (in this section only) these prices can be obtained from (10) and (11): $p_0^n = c - v$ and $p_1^n = v$, where superscript "n" denotes no intervention.

The second reason is that any market adjustment to any policy intervention takes place immediately.[45] These two reasons reduce the planner's social-welfare function to:

$$\begin{aligned} W = & -c[F_r(\bar{\theta}) - F_r(\theta_0)] + v[F_r(\bar{\theta}) - 2F_r(\theta_0) + F_r(\theta_1)] \\ & + \int_{\theta_0}^{\bar{\theta}} \left\{ \frac{\alpha}{\alpha - 1} \theta s_0 x_0(\theta)^{(\alpha-1)/\alpha} - (\psi_0 + h_r e_0) x_0(\theta) \right\} f_r(\theta) d\theta \\ & + \int_{\theta_1}^{\theta_0} \left\{ \frac{\alpha}{\alpha - 1} \theta s_1 x_1(\theta)^{(\alpha-1)/\alpha} - (\psi_1 + h_r e_1) x_1(\theta) \right\} f_r(\theta) d\theta \end{aligned} \tag{21}$$

where $F_r(\bar{\theta}) - F_r(\theta_0)$ and $F_r(\bar{\theta}) - 2F_r(\theta_0) + F_r(\theta_1)$ are, respectively, the number of cars brought to the polluted area in each period at cost c and the number of cars scrapped in that same area in each period for a value v, and the second and third lines capture the social gain from driving new and old cars in the polluted area, respectively.[46]

The planner's problem can then be seen as a two-stage problem: first, to maximize (21) for each of the available policy interventions in the polluted area (optimal design), and then, to choose the intervention that delivers the highest value of W (optimal choice). In this design-and-choice problem, we start by looking at the no-intervention and the first-best benchmarks. According to (15) and (18), the no-intervention benchmark is characterized by the mileage schedules $x_a^n(\theta) = [\theta s_a / \psi_a]^\alpha$, for $a \in \{0, 1\}$, and the rental cutoffs:

$$\theta_0^n = \left[\frac{c - 2v}{\kappa_0^n s_0^\alpha - \kappa_1 s_1^\alpha} \right]^{1/\alpha} \text{ and } \theta_1^n = \left[\frac{v}{\kappa_1^n s_1^\alpha} \right]^{1/\alpha} \tag{22}$$

where $\kappa_a^n = (\alpha - 1)^{-1} \psi_a^{1-\alpha}$. New cars are rented by households with $\theta \geq \theta_0^n$ while old cars are rented by households with $\theta \in [\theta_1^n, \theta_0^n)$. Households with $\theta \in [0, \theta_1^n)$ ride the subway. To avoid corner solutions, we assume throughout that (17) holds strictly, that is, $F_r(\theta_0) - F_r(\theta_1) < F_r(\bar{\theta}) - F_r(\theta_0)$ for any policy intervention (including no intervention). Plugging cutoffs (22) and schedules (13) into (21) yields $W = W^n$.

[45] This immediate adjustment is also in Adda and Cooper (2000).

[46] This two-period setting also serves to illustrate that the presence of a non-polluted area is not essential for our results. This is worth keeping in mind if, for example, transaction costs in the second-hand market are believed to be sufficiently high to eliminate trade across regions. In any case, and as mentioned in footnote 31, the exercise in the online Appendix (Section A.2, Table A.3) suggests that trade across regions is fairly active.

On the other hand, the first-best solution to the planner's problem in the polluted area is characterized by the mileage schedules $x_a^*(\theta) = [\theta s_a/(\psi_a + h_r e_a)]^\alpha$, for $a \in \{0,1\}$, and cutoffs

$$\theta_0^* = \left[\frac{c-2v}{\kappa_0^* s_0^\alpha - \kappa_1^* s_1^\alpha}\right]^{1/\alpha} \text{ and } \theta_1^* = \left[\frac{v}{\kappa_1^* s_1^\alpha}\right]^{1/\alpha} \tag{23}$$

where $\kappa_a^* = (\alpha-1)^{-1}(\psi_a + h_r e_a)^{1-\alpha} < \kappa_a^n$. While θ_0^* identifies the driver that leaves the social planner indifferent as to whether this driver opts for a new or old unit, θ_1^* identifies the driver whose private benefit from driving an old car is exactly equal to the external (pollution) cost this driving imposes on others. The welfare gain from implementing the first-best is $W^* - W^n$; gain that can be achieved in the hypothetical scenario that the planner observes emission rates and miles driven (or the product of the two).

Proposition 1. *In a hypothetical scenario where e_a and $x_a(\theta)$ are observed by the planner, with $a \in \{0,1\}$, the planner can implement the first-best with a Pigouvian tax (in dollars per mile) equal to $h_r e_a$.*

Proof. See online Appendix, Section C.■

Although Pigouvian taxation only affects rental decisions indirectly through its impact on usage, it is capable of implementing the first-best since the only market failure is the pollution externality. While it is evident that is in the planner's first-best interest to have all households in the polluted area driving less (i.e., $x_a^*(\theta) < x_a^n(\theta)$ for $a \in \{0,1\}$) and to move some of them to public transport (i.e, $\theta_1^* > \theta_1^n$), it is less evident that may be also in her first-best interest to increase the fraction of households riding new cars in the polluted area (i.e., $\theta_0^* < \theta_0^n$).

Evaluating how the different policy interventions fall between these two benchmarks determines their relative performance. One way or another, all instruments have the capability of affecting cutoff levels θ_0 and θ_1, which can be written more generally as

$$\theta_0^i = \left[\frac{p_0^i - p_1^i}{\kappa_0^i s_0^\alpha - \kappa_1^i s_1^\alpha}\right]^{1/\alpha} \text{ and } \theta_1^i = \left[\frac{p_1^i}{\kappa_1^i s_1^\alpha}\right]^{1/\alpha} \tag{24}$$

where $i = \{s, f, g, vr\}$ identifies the intrument, namely, scrappage subsidies (s), registration fees (f), gasoline taxes (g), and vintage-specific driving restrictions. Some instruments will affect cutoffs (24) directly through changes in rental prices (p_a^i) while others indirectly through changes in the cost of driving (κ_a^i). In fact, scrappage subsidies and registration fees work exclusively through changes in rental prices; the former by increasing the opportunity cost of renting an old vehicle ($p_0^s = c - v - \sigma$ and $p_1^s = v + \sigma$) and the latter by increasing the rental cost perceived by households ($p_0^f = p_0^n + r_0 = c - v + r_0$ and $p_1^f = p_1^n + r_1 = v + r_1$). Gasoline taxes and driving restrictions are different in that they work through changes in the cost of driving; the former with a price instrument ($\kappa_a^g = (\alpha-1)^{-1}(\psi_a + g/mpg_a)^{1-\alpha} < \kappa_a^n$) and the latter with rationing ($\kappa_a^{vr} = R_a \kappa_a^n$).[47]

Since instruments' relative performance ultimately depends on their ability to internalize vehicle emissions, it helps to consider two polar cases of emission rates e_0 and e_1.

Proposition 2. *If $e_0/e_1 = mpg_1/mpg_0$, then a gasoline tax $g = h_r e_a mpg_a$ (in dollars per gallon) implements the first best, i.e., $W^g = W^*$, where W^g denotes welfare under an optimally-designed gasoline tax. On the contrary, if $e_0 = 0$ and $e_1 \geq \underline{e}_1$, where $\underline{e}_1$ solves $\theta_0^*(\underline{e}_1) = \theta_1^*(\underline{e}_1)$, then both registration fees and vintage-specific restrictions implement the first best, i.e., $W^f = W^{vr} = W^*$, where W^f and W^{vr} denote welfare under optimally-designed registration fees and vintage-specific restrictions, respectively.*

Proof. The proof for the first case (i.e., $e_0/e_1 = mpg_1/mpg_0$) is immediate from Proposition 1 and from the fact that neither of the remaining instruments can implement $x_a^*(\theta)$, either because they cannot alter the intensive margin (sccrappge subsidies and registration fees) or because they can but inefficiently (vintage-specific restrictions). For the second case (i.e., $e_0 = 0$ and $e_1 \geq \underline{e}_1$), notice that it is first-best to ban old cars entirely ($\theta_0^* = \theta_1^*$) and let new cars run freely (because $e_0 = 0$). This can be achieved either with vintage-specific restrictions $R_0 = 1$ and $R_1 \in [0, \bar{R}_1)$, where $\bar{R}_1$ solves $\theta_1^{vr}(\bar{R}_1) = \theta_0^*$, or with registration fees $r_0 = 0$ and $r_1 \in (\underline{r}_1, +\infty)$, where $\underline{r}_1$ solves $\theta_1^f(\underline{r}_1) = \theta_0^*$. A gasoline tax fails to implement the first-best in this case because it reduces travel of new cars below first-best levels for any $g > 0$ and because some old cars may still remain in circulation. Finally, a scrappage subsidy also fails to implement the first best because one instrument (σ) cannot handle two problems (θ_0^* and θ_1^*): any σ that solves $\theta_1^s(\sigma) = \theta_1^*$ does not necessarily solve $\theta_0^s(\sigma) = \theta_0^*$.■

A key observation that follows from Proposition 2 is that the relative performance of available interventions is not invariant to parameter values; it depends on whether correcting the intensive margin (first case) is more important than correcting the extensive margin (second case). By not differentiating between new and old cars, a gasoline tax falls more heavily on cars that are run more often, ultimately reducing their rental demand. As the first case illustrates, this is not a problem when a gallon pumped in a new car's tank is equally harmful than in an old car's (i.e., when $e_0 mpg_0 \approx e_1 mpg_1$). However, as new cars become increasingly cleaner relative to old cars (i.e., $e_0/e_1 \to 0$), the relative advantage of a gasoline tax is diminished in favor of interventions better at handling the extensive margin. In the first case the planner may have no reason to increase the fraction of households riding new cars in the polluted area,[48] but in the second case it is first-best do so (i.e., $\theta_0^* < \theta_0^n$). This explains why a gasoline tax perfoms so badly in the second case (because $\theta_0^g > \theta_0^n$). In fact, if $\underline{e}_1 \leq e_1 \leq \psi_1/\alpha h_r$, the optimal tax is $g = 0$ despite it is first-best to get rid of old cars entirely. In this case,

[47] In the general setting these instruments also work through changes in rental prices.

[48] It may actually want to decrease such fraction (i.e., $\theta_0^* > \theta_0^n$) if fuel economy differences are not enough to compensate for quality and other cost differences, i.e., if $mpg_0/mpg_1 < (s_0/\psi_0)^\alpha/(s_1/\psi_1)^\alpha$.

the (marginal) private cost of imposing a fuel tax on old cars (i.e., $1/\psi_1^{\alpha}$) is greater than its (marginal) environmental benefit (i.e., $h_r e_1 \alpha/\psi_1^{\alpha+1}$).

Though reducing pollution calls for correcting both margins, attention to the extensive margin becomes particularly important when $\theta_0^* > \theta_0^n$, i.e, when it is in the planner's first-best interest to increase the fraction of households riding new cars. The exact condition for this to happen is

$$\frac{1-(1+h_r e_1/\psi_1)^{1-\alpha}}{1-(1+h_r e_0/\psi_0)^{1-\alpha}}\left(\frac{s_1}{s_0}\right)^{\alpha}\left(\frac{\psi_0}{\psi_1}\right)^{\alpha-1} > 1 \tag{25}$$

which comes from comparing (22) and (23). Inequality (25) is likely to hold as new cars become cleaner and quality does not deteriorate much with age, so that households stick longer (than socially optimal) to their old cars before switching to a new car. Our estimation of the different parameters in (25) indicates that it holds for any plausible range of parameter values (see section 4.2), so we will adopt it as a working assumption for the rest of the section.

Another important observation in Proposition 2 is that the driving restriction in the second case is vintage specific: new cars face no restriction ($R_0 = 1$) while old cars face a complete ban ($R_1 = 0$).[49] It is worth emphasizing that this "all-or-nothing" structure, where some cars face a complete ban while others face none, is not specific to the two-period setting. In a more general setting, with a very large number of vintages, we can identify $\underline{a}$ as the oldest car circulating in the polluted area in which the private benefit derived by household $\theta_{\underline{a}}^{vr}$ of doing so is at least as large as its external cost:

$$w(\theta_{\underline{a}}^{vr}, \underline{a}) \equiv \kappa_{\underline{a}}(\theta_{\underline{a}}^{vr} s_{\underline{a}})^{\alpha} - p_{\underline{a}} - h_r e_{\underline{a}}(\theta_{\underline{a}}^{vr} s_{\underline{a}}/\psi_{\underline{a}})^{\alpha} \geq 0$$

Consequently, any older car should be completely ban from circulation (i.e., $R_a = 0$ for all $a > \underline{a}$) in that $w(\theta < \theta_{\underline{a}}^{vr}, a > \underline{a}) < 0$.

In our two-period setting, with only two vintages, we cannot rely on such fine vintage partition to precisely identify the first old car to face a complete ban. We can, however, identify the lowest-type househould θ_1^{vr} that should rent an old car from a social standpoint:

$$\kappa_1(\theta_1^{vr} s_1)^{\alpha} - v - h_r e_1(\theta_1^{vr} s_1/\psi_1)^{\alpha} = 0 \tag{26}$$

Since $\theta_1^{vr} > \theta_1^n$, to make sure that households $\theta < \theta_1^{vr}$ do not end up renting a car, the planner must ration the number of old cars in the polluted area to $F_r(\theta_0^{vr}) - F_r(\theta_1^{vr}) \geq 0$ by placing a full restriction on some of them.[50] We denote this restriction by $R_{1-} = 0$, which is set separately from any restriction the planner may establish upon the remaining old cars going

[49]Setting $R_1 < \bar{R}_1$ is enough to discourage any driver from renting old cars, where $\bar{R}_1$ is defined in the proof of Proposition 2.

[50]To simplify the analysis we assume the fraction $F_r(\theta_0^{vr}) - F_r(\theta_1^{vr})$ of old cars to be allocated efficiently, that is, to be rented to those that value them most. In the general setting, with a large number of vintages, there is no need for this assumption because rental prices adjust accordingly.

to households $\theta \in [\theta_1^{vr}, \theta_0^{vr})$. We denote this latter by $R_{1+} \in [0,1]$. The restriction design in Proposition 2, for instance, corresponds to the situation in which $R_{1-} = R_{1+} = 0$, but this is rather an exception.

Proposition 3. *An optimally-designed driving restriction is vintage specific: It is defined by the triplet* $\{R_{1-} = 0, R_{1+} \in [0,1], R_0 = 1\}$, *where* R_{1-} *is the restriction faced by any household* $\theta < \theta_1^{vr}$ *renting an old car, with* θ_1^{vr} *given by (26);* R_{1+} *is the restriction faced by any household* $\theta \in [\theta_1^{vr}, \theta_0^{vr})$ *renting an old car; and* R_0 *is the restriction set upon any new car.*

Proof. See online Appendix, Section C.■

The reason new cars run freely ($R_0 = 1$) is because doing otherwise can only reduce the demand for these vehicles and destroy socially value trips (i.e., $\kappa_0(\theta_0^{vr} s_0)^\alpha - h_r e_0(\theta_0^{vr} s_0/\psi_0)^\alpha - p_0 > R_0[\kappa_0(\theta_0^{vr} s_0)^\alpha - h_r e_0(\theta_0^{vr} s_0/\psi_0)^\alpha] - p_0$ for any $R_0 < 1$). For this same reason, it may also be optimal to set $R_{1+} = 1$: not to destroy socially valuable trips made by $\theta \in [\theta_1^{vr}, \theta_0^{vr})$. In this two period setting, there is a caveat however. Since letting $R_0 = R_{1+} = 1$ leaves the demand for new vehicles in the polluted area unchanged (i.e., $\theta_0^{vr} = \theta_0^n$), at times —when θ_1^{vr} is not much smaller than θ_0^{vr}— it may be optimal to increase the demand for these vehicles a bit by setting $R_{1+} < 1$ at the cost of destroying some socially valuable trips of households $\theta \in [\theta_1^{vr}, \theta_0^{vr})$.[51]

Unlike a uniform driving restriction, which acts primarily on the intensive margin by imposing the same restriction $R < 1$ upon all cars regardless of their emission rates, a vintage-specific restriction acts exclusively on the extensive margin. The reason R_{1+} may be less than 1 in Proposition 3 is not to reduce pollution from some old cars, quite the opposite, to induce more drivers to adopt new models. Another way of contrasting the work of a vintage-specific restriction to that of a uniform restriction is by understanding what happens as e_0 goes to zero (which is when a vintage-specific restriction may implement the first-best if e_1 is high enough; see Proposition 2). In this case, a uniform restriction can only destroy welfare, so it is best to set $R = 1$, i.e., to have no restriction whatsoever (see online Appendix, Section C.3, for a formal proof). Setting $R < 1$ would not only destroy valuable trips in new cars, but, for that same reason it would also discourage drivers from renting these cars, resulting in $\theta_0^{ur} > \theta_0^n > \theta_0^*$, where "*ur*" denotes a uniform restriction.

Having established that a driving restriction ought be designed to act exclusively on the extensive margin, we conclude this discussion by comparing the work of a vintage-specific restriction to that of instruments that, by construction, also act exclusively on the extensive margin.

Proposition 4. $W^f \geq \max\{W^s, W^{vr}\}$, *where* W^f *and* W^s *are the social welfare levels under optimally designed registration fees* r_0 *and* r_1 *and a scrappage subsidy* σ, *respectively.*

[51] In the more general setting, there is less need to set $R_a < 1$ for some $a < \underline{a}$ because the price of middle-age cars is automatically increased as old cars are removed from the market (this price mechanism is ruled out in the two-period setting). Our simulations confirm to be optimal to set $R_a = 1$ for all $a < \underline{a}$.

Proof. Suppose r_0 and r_1 are chosen optimally leading to cutoffs $\theta_0^f(r_0, r_1)$ and $\theta_1^f(r_1)$; in particular, r_1 must be chosen so that $\kappa_1(\theta_1^f s_1)^\alpha - v - h_r e_1(\theta_1^f s_1/\psi_1)^\alpha = 0$. This latter condition is also satisfied with a vintage-specific restriction, by setting $R_{1-} = 0$ so that $\theta_1^{vr} = \theta_1^f$ (see Proposition 3), and with a scrappage subsidy, by setting σ so that $\theta_1^s(\sigma) = \theta_1^f$. But $\theta_0^{vr}(R_0 = R_{1+} = 1)$ and $\theta_0^s(\sigma)$ do not necessarily coincide with θ_0^f, unless some socially costly distorsions are introduced. The subsidy may need to go above σ in order to reduce θ_0^s at the cost of decreasing the number of old cars in the market (i.e., increasing θ_1^s), and the restriction R_1 may need to drop below 1 in order to reduce θ_0^{vr} at the cost of destroying some valuable old-car rides.■

Unlike restrictions and scrappage subsidies, registration fees involve a complete array of prices to influence extensive-margin decisions without affecting usage decisions. Therefore, for coming closer to the registration-fee design, both scrappage subsidies and vintage-specific restrictions may need to include costly adjustments, but of different nature, which make their ranking ambiguous: the former may reduce the number of old cars in the market and the latter may reduce their use.

While this two-period setting has served to illustrate key factors at work in policy design, particularly the role of e_0 and e_1, it is also useful for what follows (estimation and simulations) to understand the implications of relaxing some of its underlying assumptions. One assumption concerns the mute role played by the non-polluted area so far. This is relaxed in the simulations where the non-polluted area is shown to absorb an important fraction of old cars displaced from the polluted city by the policy, except for the scrappage subsidy. By permanently removing from the market cars with otherwise value in non-polluted areas, scrappage subsidies are shown, for instance, to do strictly worse than vintage restrictions.[52]

Another assumption is that vehicle emissions are equally harmful no matter *when* a car is used (e.g., peak vs. off-peak hours, weekdays vs. weekends, summer vs. winter months). In later simulations, we relax this assumption and show that by imposing restrictions only when emissions are harmful, a vintage restriction improves not only relative to the scrappage subsidy and gasoline tax, which by construction are incapable of handling such temporal variation, but also relative to registration fees unless these fees are offered in a menu format with choices of fees and car-use limitations.

Finally, some assumptions about emission rates e_a deserve discussion. So far, we have assumed that emission rates are explained entirely by vintage and perfectly observed by the regulator. Neither assumption perfectly matches reality. In the estimation section, we show that while vintage explains much of the variation in emission rates in our database, other factors such as model (e.g., Toyota Corolla) also explain a good part of it. This new source of heterogeneity could be captured by introducing some variation within each vintage of the

[52]Another reason subsidies do worse is dealers' incentive to arbitrage any difference in scrappage value created by the subsidies.

form $e_a^j = e_a + \nu_a^j$, where j identifies a particular car and $E[\nu_a^j] = 0$. As for implications for our results, there are considerations depending on whether this new source of variation is available for policy design and whether it is correlated with other variables in the model, particularly driving, $x_a(\theta)$.

For instance, if ν_a^j is not correlated with $x_a(\theta)$, as our data suggest (see online Appendix, Section D.4), and is unavailable to the regulator,[53] all our results (i.e., designs and rankings) go through but for a notational change: e_a should be interpreted now as the average emission rate. As ν_a^j becomes available to the regulator, all the "extensive-margin" instruments necessarily improve relative to the gasoline tax since now there is greater emission-rate variation to be exploited by these instruments; and rather than separating cars by vintage, they are now to be separated by emission rates, say, by some combination of vintage and model.[54]

4. Estimation

The two-period model has served to illustrate the role played by some parameters in the planner's problem. We now use different data sources and methodologies to obtain numerical values for the parameters that enter into the general model of sections 3.1 and 3.3. We organize the estimation in two parts. In the first part we obtain values for parameters related to car characteristics, households's preferences, and policy response, so as to capture the market equilibrium of 2006 (when applies, parameter values are presented in 2006 U.S. dollars). In the second part we do likewise for the pollution-related parameters with one difference: as we want these parameters to capture pollution damages from Santiago's current fleet, where the fraction of pre-1993 models is virtually zero, we use the most recent smog-check data and damage estimates. Thus, the policy simulations that follow would be informative in any effort dealing with Santiago's current pollution problem.

The following simplifying assumptions apply to both parts. First, we cluster car vintages in six age/quality groups, $a = 0, 1, ..., 5$, centered around the 1992-93 discontinuity: 2001-04, 1997-2000, 1993-96, 89-92, 85-88, and 81-84.[55] This grouping essentially assumes that people trade their cars every four years. Second, we cluster the more than 300 municipalities in 60

[53] Either because it is not observed or because it can only be obtained from smog checks, which are subject to manipulation, as reported by Oliva (2015). Another concern, distinct from the one raised by Oliva (2015), is that manufacturers can game the regulation by artificially improving the official measures of emission rates (Reynaert, 2017).

[54] Congestion is also absent, or optimally internalized by road pricing, in our model. If this were not the case, the gasoline tax should improve relative to any of the "extensive-margin" instruments because it can be used to further adjust the intensive margin. The rankings of the "extensive-margin" instruments (see Proposition 2) remain invariant, however, since ψ_a can always be interpreted more generally as the total cost per mile of using a car including the external cost of congestion, whether that is optimally internalized. This same reasoning explains why the optimal driving-restriction design remains as in Proposition 3: setting $R_{1+} < 1$ may only be used for "extensive-margin" reasons at the cost of destroying some socially valuable trips (although these trips are less valuable now because of congestion).

[55] Model years 1980 and earlier, which in any case are very few, are grouped with 1981 models.

markets, $i = 1, ..., 60$, each corresponding to an electoral district. Electoral districts group municipalities by geographic area, sharing similar characteristics. And third, we normalize the country's population to the unity, so our relevant car-holding observation q_a^i becomes the fraction of cars age-group a in district i relative to the district's number of households.

4.1. *Car characteristics, households' preferences, and policy response*

Here we discuss how we obtain values for the following parameters: (i) scrappage value v, (ii) survival rates γ_a, (iii) price of a new car c, (iv) cost per mile ψ_a, (v) car initial quality s_0, (vi) quality deterioration rate ς, (vii) household's decreasing return to driving α, (viii) consumers' marginal valuation for quality $F_i(\theta)$, and (ix) policy response R_a^i, which is assumed to be the same for all pre-1993 models registered in Santiago, that is, $R_a^i = R < 1$ for all i's in Santiago and $a \geq 3$, and $R_a^i = 1$ otherwise.

Values for (i), (ii) and (iii) can be obtained directly from observed data. Based on informal conversations with car dealers, we fix $v = \$600$, the lowest trade-in value that the car dealers we interviewed recall having seen in recent years (we do not see prices this low in our sample of newspaper ads). Next, and by exploiting the fact that the import of second-hand cars is forbidden in the country since 1985, γ_a is estimated directly from changes in the stock of cars observed in the circulation-permit datasets from 2006 through 2012.[56] Similarly, using the car-price database described in Section 2.2, we fix $c = \$16,000$, the (weighted average) price of a new car ($a = 0$).

The cost per mile, ψ_a, is assumed to have two components, one that is invariant to location and vintage, $\bar{\psi}$, and a second one that is observable and depends on a car's fuel economy:

$$\psi_a = \bar{\psi} + P_g/mpg_a \tag{27}$$

where P_g is the price of gasoline and mpg_a is vintage-group a's fuel economy measured in miles per gallon. Based on MTT (2016), we let $P_g = \$3$, $mpg_0 = 23.66$, and mpg_a be declining with age at an anual rate of 2 percent.[57] In turn, $\bar{\psi}$ is recovered from (13) and (15) as follows:

$$\psi_0 = \bar{\psi} + \frac{P_g}{mpg_0} = \frac{1}{x(\theta_0)} \frac{(p_0 - p_1)(\alpha - 1)}{1 - \varsigma^\alpha} \tag{28}$$

where $x(\theta_0)$ is the average travel of the last driver to rent one of the youngest cars (i.e., $a = 0$). The value we adopt for $x(\theta_0)$ is 50,489 miles, which, according to the 2015 and 2016 smog-check databases, is the average travel during the first four years of a car's life.

[56] By comparing stock changes across two consecutive years of data, we obtain six data points with survival rates for each car age. Imposing $\gamma_a \leq 1$ and $\gamma_{a+1} \leq \gamma_a$, an OLS fit to these data points delivers average survival rates for cars with ages ranging from 0 to 36 years, which we average at our vintage-group level leading to the survival numbers γ_a presented in the online Appendix (Section A.2, Table A.4).

[57] A similar rate is found in recent reports by the U.S. EPA (www.epa.gov/fueleconomy).

As seen from (28), the fixed component $\bar{\psi}$ requires rental prices p_0 and p_1 and values of α and ς. All rental prices p_a used in the estimation, which are supposed to reflect 2006 (relative) equilibrium prices, are also obtained from the car-price database described in Section 2.2.[58] On the other hand, values of α and ς are jointly estimated with s_0, $F_i(\theta)$ and R, as we describe next.

Taking the model to the data requires two additional assumptions. First, we let $F_i(\theta)$ to vary by district i in a semi-parametric way. In particular, we let $F_i(\theta) \equiv F(\theta|z_i)$ be a cubic in θ with each coefficient in the cubic (b^1, b^2, and b^3) varying across districts according to the linear function $b_i^j = \chi^j + z_i'\zeta^j$, where $j = 1, 2, 3$ denotes the coefficient in the cubic, χ^j is a constant and z_i is a vector that includes the following district characteristics: income per capita ($INCOME_i$), distance to Santiago ($DISTANCE_i$), and level of urbanization ($URBANIZATION_i$). Thus, the distribution $F(\theta|z_i)$ is characterized by three χ^j parameters and nine ζ^j parameters.

Secondly, we extend the utility function in (20) to include a demand shifter, so that the utility of driver θ in district i who rents an a-year-old car with an effective restriction R_a^i becomes:

$$u(\theta, i, a, R_a^i) = R_a^i \kappa_a (\theta s_a)^\alpha - p_a + \xi_a^i \tag{30}$$

where $\kappa_a = [(\alpha - 1)\psi_a^{\alpha-1}]^{-1}$. ξ_a^i is assumed to capture district-wide shocks that are orthogonal to observable district characteristics included in z_i and whether the district is subject to the driving restriction ($DR_i = 1$) or not ($DR_i = 0$).

According to our model we should observe cutoff levels θ_a^i that mark a driver's indifference between driving an a-year-old car and an $a+1$-year-old car. These cutoffs can be obtained from (30) and the equilibrium condition (15), and are given by:

$$\theta_a^i = \left(\frac{p_{a+1} - p_a + \xi_{a+1}^i - \xi_a^i}{R_{a+1}^i \kappa_{a+1} s_{a+1}^\alpha - R_a^i \kappa_a s_a^\alpha} \right)^{1/\alpha} \tag{31}$$

and from which we obtain car-holding predictions for all a and i, $\hat{q}_a^i = F(\theta_{a-1}^i|z_i) - F(\theta_a^i|z_i)$. Note that since $s_a = \varsigma^a s_0$ enters multiplicatively in (31), we cannot separately obtain estimates for s_0 and the equilibrium cutoffs θ_a's. Hence, we normalize $s_0 = 10$.

We are left with 15 parameters to be jointly estimated: α, ς, R, and the twelve (χ^j, ζ^j)

[58] More precisely, we exploit the no-arbitrage condition given by:

$$p_a = P_{a,t} - \delta P_{a+1,t+1} \tag{29}$$

where $P_{a,t}$ is the price of an a-year-old car in year t and δ is the discount factor, which we set at 0.9 per year. If P_{imat} is the price offer in newspaper ad i published in year t for model m that is a years old (see Section 2.2), we run an OLS regression of $\ln(P_{imat})$ on a constant and year, model, and age fixed effects to predict $\hat{P}_{mat}$. With these predictions and (29), we obtain (weighted average) rental prices for each of the six vintage groups identified above.

parameters that characterize $F(\theta|z_i)$. For their estimation we consider a GMM estimator based on aggregate data, much in the spirit of Berry and Pakes (2007). The estimator is based on the moment conditions:

$$E[X_i \Delta\xi_a^i] = 0 \tag{32}$$

for all $a = 0, ..., 5$ and where X_i is a 1×5 dimensional vector that includes a constant, the three districts characteristics included in z_i, and whether the district is affected by the driving restriction or not (note that this gives us a total of 30 (5×6) moment conditions to estimate 15 parameters). The only difference of (32) with the typical moment conditions inspired by Berry and Pakes (2007) is that we use $\Delta\xi_a \equiv \xi_{a+1} - \xi_a$, instead of ξ_a, which makes (32) more credible by eliminating any district specific effects that are constant across vintages and that may be correlated with X.

By letting Ψ be the vector of the 15 parameters, the estimation proceeds in two steps. First, for a given value of Ψ, the difference of unobserved demand terms ($\Delta\xi_a^i(\Psi)$) are computed so as to satisfy the condition $\hat{q}_a^i \equiv F(\theta_{a-1}^i|z_i) - F(\theta_a^i|z_i) = q_a^i$ for all a and i, and using (31) to relate θ_a^i with $\Delta\xi_a^i(\Psi)$. And second, using these $\Delta\xi_a^i(\Psi)$, the GMM objective function is computed as:

$$Q(\Psi) = \left(\sum_{i=1}^{60} X_i \Delta\xi_0^i(\Psi), ..., \sum_{i=1}^{60} X_i \Delta\xi_5^i(\Psi) \right) W \left(\sum_{i=1}^{60} X_i \Delta\xi_0^i(\Psi), ..., \sum_{i=1}^{60} X_i \Delta\xi_5^i(\Psi) \right)',$$

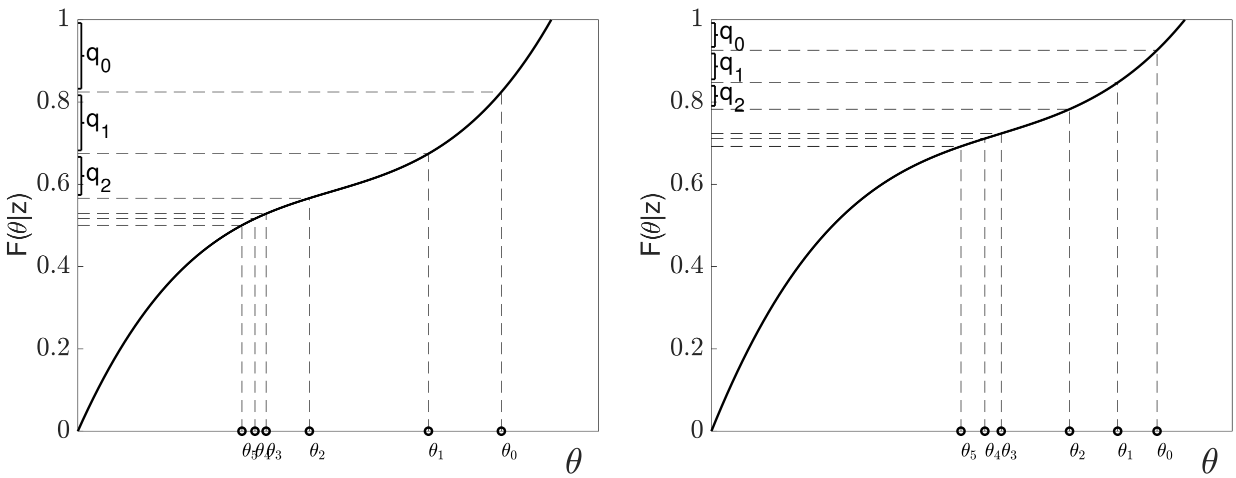
where $W = \hat{\Omega}^{-1}$ is a symmetric positive definite matrix, and $\hat{\Omega}$ is an estimator of the asymptotic variance of the sample moment conditions.[59]

Before commenting on the resulting parameter values, let us use them to generate Figure 4, which helps convey some intuition about where parameter identification is coming from. The figure depicts cutoff levels θ_a^i, the estimated function $F(\theta|z_i)$, and the quantities q_a^i for districts 22 and 32. District 22 is home to Santiago's city center and has a relatively high income per capita. District 32, on the other hand, corresponds to a city in the non-restricted zone of much lower income. On the horizontal axis, we mark with small circles the values of θ_a^i, which are constructed with rental prices p_a, the estimated values of α, ς, and R, and the residuals $\Delta\xi_a^i$. Note that the distance between θ_1 and θ_2 is relatively large for district 22 and small for district 32 whereas the distance between θ_2 and θ_3 is relatively small for district 22 and large for district 32. Such differences are mostly captured by R, which enters in equation (31) for district 22, but not for district 32. This highlights the identification of R, coming from the moment condition that assumes orthogonality of $\Delta\xi_a^i$ to whether the district is located in the restricted zone or not (and from the fact that this is not included in z_i).

The figure also shows $F(\theta|z_i)$, which corresponds to the thick black line crossing each box

[59] We estimate the model using iterated GMM and allow for intra-district correlation of the shocks $\Delta\xi_a$ when calculating $\hat{\Omega}$.

and constructed with the estimated values of χ^j and ζ^j. Because district 22 has a higher income per capita, we find that $F(\theta|z_{22}) < F(\theta|z_{32})$, meaning that drivers in district 22 tend to be of higher θ. This is explained by a larger share of new cars that we observe in richer districts. Without accounting for such differences in $F(\theta|z_i)$, we would find a positive correlation between $\Delta\xi_0^i$ and $INCOME_i$, which would then violate our moment conditions. Parameters α and ς enter the moment conditions through κ_a in equation (31) and, together with $F(\theta|z_i)$, need to be such that $\hat{q}_a^i = F(\theta_a^i|z_i) - F(\theta_{a+1}^i|z_i)$ exactly matches the observed holding q_a^i for all i and a.



(a) District 22 (b) District 32

Figure 4: Distribution of θ and number of cars in districts 22 and 32

Notes: The figure presents θ's cumulative distribution function for districts 22 and 32. The former is home to Santiago's city center and the latter corresponds to a lower-income (non-restricted) town south of Santiago. On the horizontal axis we mark the cutoff levels θ_a^i implied by the model, and on the vertical axis we plot observed quantities q_a^i obtained from the data.

Note that the estimated parameter values used to generate Figure 4, which are summarized in Table 3, are comparable to numbers from other sources/studies. For instance, the value of $\bar{\psi}$ we obtain from plugging α and ς into (28) implies that $\psi_0 = 0.413$, which is not that different from what would result from adding gasoline costs (0.127 \$/mile) to other monetary costs (e.g., parking, maintenance) and non-monetary costs (i.e., travel time costs).[60] Similarly, the value of α leads to a concave utility function $u(\theta, a, x)$, where $(\alpha - 1)/\alpha = 0.52$, close to the logarithmic utility in Gavazza et al. (2014); and the value for the decay rate ς is almost identical to the value used in that same study.[61] It is true that the policy response we find, $R = 0.968$, is

[60] According to studies by the Victoria Transport Policy Institute (www.vtpi.org), gasoline accounts for roughly 50% of monetary costs, and travel time costs (at peak hours) are usually valued at 50% of the median wage. A median wage of 8 \$/hr and an average speed of 22 miles/hr would lead to travel time costs of approximately 0.182 \$/mile. This would add up to a total cost of 0.436, quite close to our estimate of 0.413.

[61] Their annual decay rate is 0.976 while ours is $0.896^{1/4} = 0.973$.

less than the response in the absence of inter-temporal substitution ($R = 0.918$).[62] But notice that pre-1993 models were run much less often in 2006 than when the policy was enacted 13 years earlier, leaving room for inter-temporal substitution. As an additional validity check of our estimation, we run an out-of-sample exercise that contrasts model predictions for 2012 with the empirical estimation obtained from regressing (1) on the 2012 circulation-permit data. As shown in the online Appendix (section D.1), the model captures reasonably well policy effects on fleet composition both around the 1992-93 discontinuity and before.[63]

Table 3: Estimated parameters

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
α	2.063	$\bar{\psi}$	0.286	ς	0.896	R	0.968
	(0.0182)		(0.1348)		(0.0011)		(0.0012)

Notes: The parameters of the model are estimated using iterated GMM. Standard errors allow for auto-correlated unobserved demand shocks within districts. The standard error of $\bar{\psi}$ is estimated using the delta method.

4.2. *Pollution parameters*

Before moving to the policy simulations, we need an estimate of the external cost per mile of an age-group $a = 0, ..., 5$ car in area $k \in \{r, nr\}$. We denote this external cost by ϵ_a^k, which is the product of h_k, the marginal harm, and e_a, the car's emission rate. As explained earlier, we want pollution parameters to capture external damages from Santiago's current fleet, when the fraction of pre-1993 models is virtually zero, so we combine the predictions of our model with external costs in the literature and the most recent smog-check data.[64]

Our point of departure in the estimator of ϵ_a^k, which we denote by $\hat{\epsilon}_a^k$, is to make sure that the total pollution harm predicted by our model is as close as possible to actual harm. We do this by matching the actual harm generated by cars in each vintage-group in area k to the harm predicted by our model, that is:

$$\epsilon_a^k x_a q_a^k = \int_{\theta_a^k}^{\theta_{a-1}^k} \hat{\epsilon}_a^k x_a(\theta) dF_k(\theta) \tag{33}$$

for all $a = 0, ..., 5$ and where q_a^k is the total number of age-group-a cars in area k according to the circulation-permit data, x_a are miles traveled by those cars during the period according the

[62]See footnote 28.

[63]It fails, however, to capture the larger fraction of newer cars in Santiago relative to the rest of the country, partly because of the substantial shift to private transport due to the poorly implemented public transport reform in Santiago in February 2007, known as Transantiago (see Gallego et al., 2013).

[64]We only use smog-check data of post-1992 models, all equipped with a catalytic converters. In the online Appendix (Section E.1), we use our estimated model to provide a more comprehensive evaluation of the 1992 policy to complement the evaluation carried out in section 2.3.

smog-check data, θ_a^k corresponds to the equilibrium cutoff level of a driver in area k indifferent to driving an age-group-a car and an age-group-$a+1$ car according to the model (recall that $\hat{q}_a^k = F_k(\theta_{a-1}^k) - F_k(\theta_a^k)$ is the number of age-group-a cars in area k according to the model), and $x_a(\theta)$ is the amount of travel during the period by type θ according to the model. As $\hat{q}_a^k$ and $x_a(\theta)$ are not exactly equal to q_a^k and x_a (note that $\hat{q}_a^k$ is a prediction at the area, not district, level, so it does not need to match q_a^k), equation (33) ensures that the contribution of each vintage-group to total damages is properly accounted for in the simulations that follow.

We still need to compute ϵ_a^k in (33) from available data. Since cars emit multiple pollutants, we follow Knittel and Sandler (2018) and Fullerton and West (2010) and consider the contribution of all local pollutants with smog-check readings:

$$\epsilon_a^k = h_k^{HC} \upsilon^{HC} \varrho_a^{HC} + h_k^{NO_x} \upsilon^{NO_x} \varrho_a^{NO_x} + h_k^{CO} \upsilon^{CO} \varrho_a^{CO} \tag{34}$$

where h_k^p is the marginal damage (in dollars per gram) of pollutant $p \in \{\text{HC}, \text{NO}_x, \text{CO}\}$ in area k, ϱ_a^p is p's average smog-check (ASM-2525) reading of an age-group-a car (in parts per million for HC and NO_x and in percent of the exhaust for CO), and υ^p is a conversion factor (the product of υ^p and ϱ_a^p gives e_a^p, the car's emissions rate of pollutant p in grams per mile).

We do not have separate values of h_k^p and υ^p, but we can obtain their product, $h_k^p \upsilon^p$, from two different sources. From Rizzi and De La Maza (2017), we can calculate total harm in area k in a given period as follows:

$$H_k = d_k \sum_a x_a q_a^k \tag{35}$$

where d_k is the average external cost per mile in area k reported in their study. And from Knittel and Sandler (2018), we can obtain the relative contribution of the different pollutants to total damages:

$$H_k^p = \sum_a h_k^p \upsilon^p \varrho_a^p x_a q_a^k = \omega^p H_k \tag{36}$$

where ω^p is pollutant p's relative contribution to external costs ($\sum_p \omega^p = 1$).[65] Combining (35) and (36) we obtain $h_k^p \upsilon^p$, that plugged into (34) and then (33) yields $\hat{\epsilon}_a^k$.

There is one remaining caveat. Rizzi and De La Maza (2017) only report a value of d_k for Santiago, ¢4 per mile, so we only have estimations of ϵ_a^r for Santiago. A value for $\hat{\epsilon}_a^{nr}$ is obtained as the product of d_{nr}/d_r and $\hat{\epsilon}_a^r$, where d_{nr}/d_r comes from Parry and Strand (2012). Values of $\hat{\epsilon}_a^r$ and $\hat{\epsilon}_a^{nr}$ are summarized in the online Appendix (Section A.2, Table A.4). Values of $\hat{\epsilon}_a^r$, along with those in Table 3, can also be used to establish that expression (25) holds for different possible "new" and "old" car combinations, suggesting that extensive-margin instruments such as vintage-specific restrictions are expected to perform comparably well in the simulations that follow.[66] But before doing so, there are some concerns about ϵ_a^k that

[65] We use the average relative contributions that are reported at the bottom of their Table A.11: 82%, 15% and 3% for HC, CO and NO_x, respectively.

[66] For example, if new cars correspond to cars in age-group 1 (i.e., 5-8 years old) and old cars correspond to

deserve attention.

One concern is the extent to which today's emission rates can serve as proxy for rates of future models. As explained in the online Appendix (Section D.3), emission rates increase with vehicle age as a result of newer cars entering the market under more stringent standards and also pollution-control technologies wearing out. Our model does not include technological progress, but we posit that the results should not change if the relative difference between emission rates of new and old models is preserved over time and society's valuation of clean air increases with time (or income) as technology improves accordingly, so that ϵ_a^k remains more or less unchanged. If anything, a more drastic technological advance, such as electric vehicles, can only widen that relative difference, strengthening the case for vintage-specific restrictions (see Propositions 2 and 3).

Another concern, already raised in section 3.5, is the extent to which emission rates do in fact depend so fundamentally on age as we have assumed so far. As discussed in the online Appendix (Section D.2), age explains a significant portion of the variation in emission rates (see also Knittel and Sandler, 2018), but not all the variation. Other car characteristics such as size and make also play a role. Our model could easily be extended to account for these other characteristics by relabeling vintage/age into a bundle of observable car characteristics that can be ordered vertically in the preference space if emission rates correlate negatively with car quality. If for some reason, however, the planner cannot use these characteristics in policy design, our emission rate function would be noisier than assumed here. This would have no implications for optimal policy design, as emissions enter linearly in the social damage function, but our policies then would be moved further from the Pigouvian (first-best) solution.[67]

5. Policy simulations

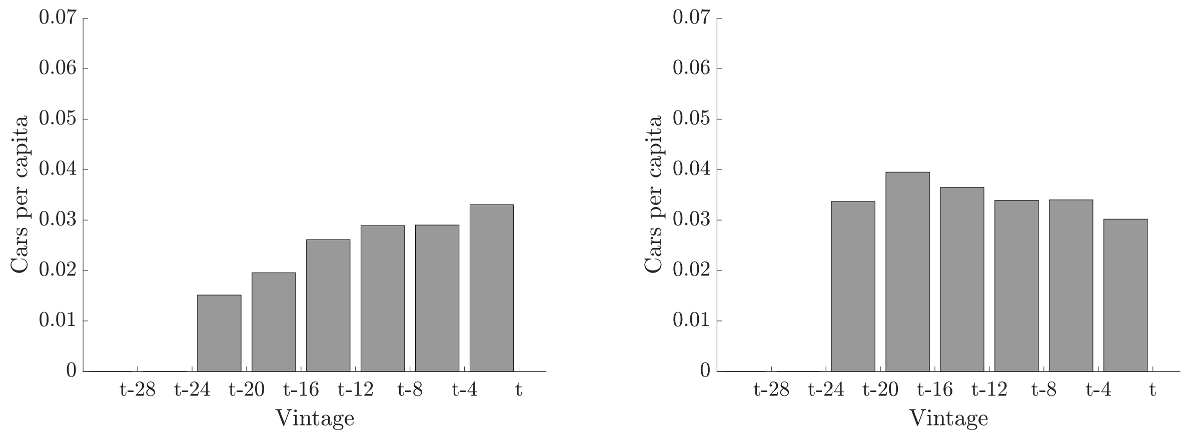
Guided by insights gained from the two-period model, we are now ready to use the general model and the (estimated) parameter values to corroborate and expand on those insights with some numerical simulations. Although the numbers that emerge from the simulations are specific to Santiago's current pollution problem, their qualitative implications should apply more broadly. In all the simulations that follow, parameter values are kept constant over time, including the speed at which a car's emissions rate deteriorates with age, and population and household characteristics.

We start by constructing the no-intervention scenario. Figure 5 shows that, in the absence of any government intervention, the city of Santiago (the restricted or polluted area) already

cars in age-group 3 (i.e., 13-16 years old), the left-hand-side of (25) is $10.9 > 1$.

[67] A third concern might be whether households in higher-income municipalities own not only newer cars but also larger and potentially more polluting ones. If car quality, as perceived by households, is found to be negatively correlated with emission rate, our conclusions would not hold. The online Appendix (Section D.5) provides evidence showing that is not the case: cars in richer municipalities are newer and cleaner than in poorer municipalities.

exhibits a newer fleet than that in the rest of the country (the non-restricted or non-polluted area). While Santiago's smaller population (36% of the country's total) explains its smaller overall fleet, its higher income per capita explains why it nevertheless has 10% more of the newest models (0 to 4 years: vintage-group 0 in our simulations) than the rest of the country. Santiago also has fewer older models.[68] Notice that in both locations, we find cars running until scrapped, somewhere between 20 and 24 years old. This will change as the government intervenes in the market.



(a) Santiago

(b) Rest of the country

Figure 5: Steady-state fleet composition under no intervention

Notes: The figure shows the steady-state profile of car fleets in Santiago and the rest of the country under the no-intervention scenario. Since total population has been normalized to unity, each bar indicates the number of cars per capita for a particular vintage group.

The next set of exercises estimates welfare gains made possible by moving away from the no-intervention scenario. We are particularly interested in the welfare gap between the first-best and the outcome of alternative policy interventions. The regulator's problem is to maximize social welfare's present value, $W = \sum_{t=0}^{\infty} \delta^t w_t$, subject to available instruments, where w_t is social surplus in period t and δ is the discount factor, common to all agents in the economy (including the regulator) and equal to $\delta = 0.9$ per year. Since there is perfect competition in the car market, w_t can be written as:

$$w_t = -cq_{0,t} + vq_{T(t),t} + \sum_{k=p,np} \int_{\theta^k_{T(t),t}}^{\bar{\theta}^k} [u^k_t(\theta) - \epsilon^k_t(\theta) x^k_t(\theta)] dF_k(\theta)$$

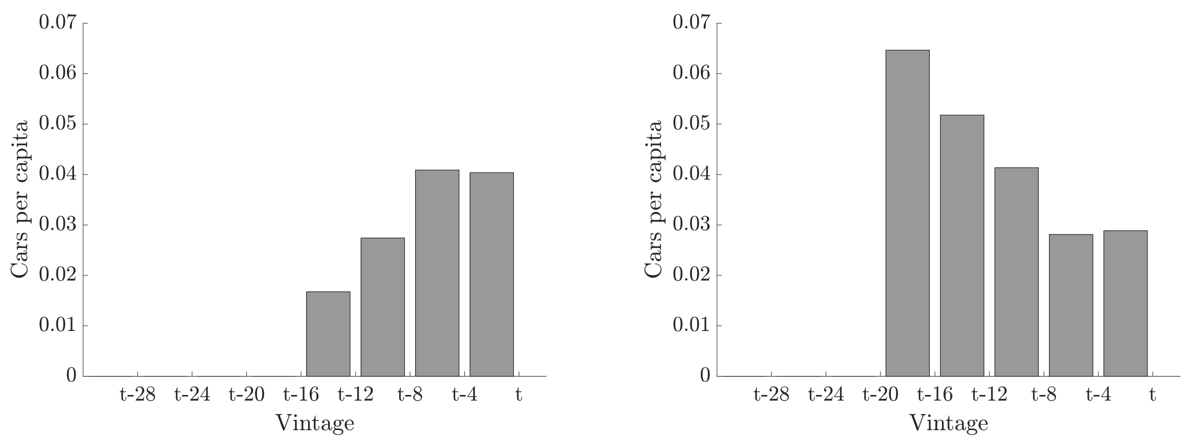
where $q_{0,t}$ is the total number of new models that are (rationally) expected to enter the market in period t; $q_{T(t),t}$ is the total number of cars expected to exit the market in period t (exiting at age $T(t)$); $\theta^k_{T(t),t}$ is the last household to rent a car in area k during period t (see (18));

[68] If, while comparing fleets across regions, one were to eliminate any size effect, leaving only income effects, one would need to multiply the height of each bar in Figure 5(a) by $1.78 = 0.64/0.36$.

and $u_t^k(\theta)$, $\epsilon_t^k(\theta)$, and $x_t^k(\theta)$ are, respectively, a household θ's utility, external cost per mile, and miles traveled in region k during period t. These expressions vary with time because households adjust rental and travel decisions over time in response to changes in rental price. Furthermore, the specific forms of $x_t^k(\theta)$ and $u_t^k(\theta)$ vary with the policy scenario; compare, for instance, equations (13) and (14) with (19) and (20).

5.1. *The Pigouvian benchmark*

As observed in the steady-state outcome of Figure 6, the effect on fleet composition of levying a Pigouvian tax equal to ϵ_a^k per mile driven is dramatic. Over the long run, households in Santiago have no incentive to hold cars older than 16 years—an 8-year reduction compared to the no-intervention case. While sales of new cars in Santiago increase by 25%, fewer households drive cars there; those that do, however, drive cleaner cars. This major adjustment also has large impacts outside Santiago: The scrappage age of a car in the rest of the country is reduced by 4 years. This may seem surprising at first because there is no direct intervention in the non-restricted area, but everything works through the second-hand market, as it does in all existing restriction programs. Instead of scrapping cars, Santiago is now exporting a large fraction of 16-year-old cars to the rest of the country. This increase in supply reduces the rental price of all 20-year and older models on the market to the point that scrapping them much sooner becomes optimal.



(a) Santiago (b) Rest of the country

Figure 6: Steady-state fleet composition under the first best

Notes: The figure shows the steady-state profile of car fleets in Santiago and the rest of the country under Pigouvian taxation. Since total population has been normalized to unity, each bar indicates the number of cars per capita for a particular vintage group.

This adjustment has profound welfare implications. Estimating them is far from trivial because the transition from one steady state to the other is not only very long, so it cannot be omitted from any welfare estimation, but it is also non-monotonic (this applies to any policy intervention). As illustration, in the online Appendix (Section A.1, Figure A.3), we show the (non-monotonic) dynamics of new car sales (q_0) for both Santiago and the rest of the country.

In present-value terms, the welfare gain of moving from no intervention to the first-best amounts to \$433.1 per household: a 9.7% gain from the no-intervention baseline of \$4467.4 (see first two rows of Table 4). The drop in pollution costs (\$634.8) is more than three times larger than the loss in transport surplus (\$200.7). At the country level, this net welfare gain adds a total of \$1.8 billion—comparable to the gain from Germany's LEZs (Wolff, 2014), for example. In any case, we do not want to push these welfare numbers too much. Other than being a rough approximation of the potential gains from curbing vehicle emissions, these numbers serve our purpose as a benchmark for evaluating the relative performance of real-world policies like driving restrictions, registration fees, scrappage subsidies, and gasoline taxes.

Table 4: Welfare under various policy simulations

#	Counterfactual	Transport surplus (in 2006 dollars)	Pollution cost (in 2006 dollars)	Welfare (in 2006 dollars)	Welfare gain/loss (relative to first-best)
1.	No intervention	5579.6	-1112.2	4467.4	0%
2.	First best	5378.9	-477.4	4900.5	100%
3.	Driving restriction with no exemptions ($R = 0.9\ \forall\tau$)	5192.4	-1122.3	4070.1	-92%
4.	Driving restriction with some exemptions ($R = 0.9$, $a \geq 3$)	5542.7	-1005.7	4537.0	16%
5.	Complete-ban vintage driving restriction ($R = 0$, $a \geq 4$)	5491.1	-800.9	4690.2	51%
6.	Scrappage subsidy (\$2420), full arbitrage	5276.2	-648.1	4628.1	37%
7.	Scrappage subsidy (\$3240), no arbitrage	5465.2	-786.4	4678.8	49%
8.	Circulation fees	5379.1	-571.0	4808.1	79%
9.	Gasoline tax (\$1.06 per gallon)	5520.6	-985.9	4534.7	16%
10.	Vintage restriction & gas tax ($R = 0$, $a \geq 4$, ¢80 per gallon)	5456.3	-726.6	4729.7	61%

Notes: The table shows present-value welfare calculations under a number of different policy counterfactuals. All calculations are in per capita terms, in 2006 U.S. dollars. The first column presents household surplus from renting and driving cars, ignoring pollution costs (recall that surplus from using public transport is normalized to zero). As expected, transport surplus is maximized in the absence of any intervention. The second column presents pollution costs. The third column corresponds to welfare calculations as the sum of transport surplus and pollution costs. The fourth column presents welfare gains/losses as a fraction of the welfare gain under the first best (i.e., Pigouvian taxation).

At least two elements separate real-world policies from first-best instruments. The first is that, for either political or technical reasons, the instruments involved are never first-best. The second is that instrument's use is restricted to geographic areas that have been declared "in non-attainment" with existing air-quality standards. Consequently, the regulator cannot introduce policies in attainment areas only to contain any eventual pollution leakage from regulations imposed elsewhere. We adopt this geographic limitation in the simulations that follow, focusing on policies that are exclusive to Santiago.

Furthermore, given how long it takes to move from one steady state to another, it is natural to think that the optimal policy, whether quantity- or price-based, may vary over time. Given the dynamics of the first-best outcome, it appears the regulator would like to start with a tougher policy to gradually relax as the steady state is approached. Following current practice, however, we focus on time-invariant policies in what follows.[69]

5.2. *Driving restrictions*

Designing a driving restriction requires that both the intensity of the restriction (R_a) and the vintages affected be defined. Before presenting the optimal (vintage-specific) design, it is instructive to go over less than optimal designs, as they help to clarify why these policies can sometimes inflict so much harm —if poorly designed— but can nevertheless be improved by introducing vintage differentiation. Take, for instance, the HNC design, as implemented in Mexico City in 1989, and, following Gallego et al. (2013), assume a uniform restriction intensity of $R = 0.9$ upon all cars. Even neglecting the second-car effect, such a uniform design leads to a welfare *loss* that amounts to 92% of welfare gain from implementing the first-best, as shown in the third row of Table 4.

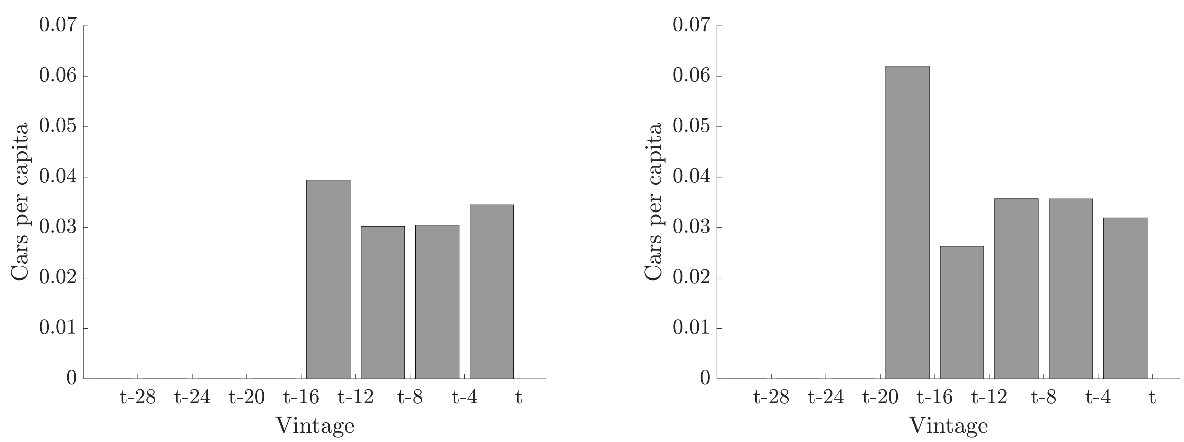
An HNC-1989 design not only fails to remove old cars from the road (actually, it extends their lives by reducing their rental prices); it also reduces sales of new cars in Santiago (a figure with the steady-state fleet composition can be found in the online Appendix, Section E.3, Figure E.2). Since new cars are driven by households that value quality the most, a uniform reduction in quality is felt more heavily in these new cars, by destroying a good fraction of trips that are highly valuable (see column one of Table 4). As a result, demand for them falls, and with that, their rental prices and sales. As the demand for cars shifts toward older models, the life of the existing stock is extended, so pollution may end up higher than in the no-intervention baseline (see column two of Table 4).

One way for the regulator to reverse the unfortunate outcome of a uniform restriction is to follow the reforms introduced in Santiago and Mexico City: exempting some cleaner cars from the restriction. In fact, if we impose $R = 0.9$ only on cars that are 12 or more years of age, this vintage-specific design results in a welfare gain, although that is only 16% of the gain under the first-best, as indicated in the fourth row of Table 4. Extending the exemption to cleaner cars solves one part of the problem: it boosts new car sales in Santiago (a figure with the steady-state fleet composition of this particular vintage-specific design can be found in the online Appendix, Section E.3, Figure E.3). The second part of the problem, however—the removal of high-emitting vehicles from the road—requires a tougher restriction on these cars. As shown in column 2 of the table, this vintage-specific design yields only a mild reduction in

[69]Besides, computing time-varying (optimal) policies can be quite demanding based on the non-monotonic dynamics described above.

pollution costs.

As suggested by Proposition 3, and following the "complete-ban" schemes seen in Paris and Germany's LEZs, it turns out that within the class of time-invariant restriction designs the one that delivers the highest welfare takes a complete-ban form: it imposes a total circulation ban on vehicles that are 16 or more years of age and extends a full exemption on newer vehicles. As indicated in row 5 of Table 4, the resulting welfare gain is significant. Figure 7 can help explain this result. Such vintage restriction design works at both ends of the fleet spectrum, prompting the removal of old cars and boosting sales of new ones. Because scrapping 16–20-year-old cars in areas where local pollution is not a problem is socially inefficient, a driving restriction works its way through the second-hand market to reallocate these cars from pollution-affected areas to pollution-free areas. But there is more: The export of these older cars to the rest of the country does not result in a sharp increase of high-emitting vehicles in this region; quite the opposite. Similar to what is behind the first-best profile of Figure 6(b), the export of 16–24-year-old models to the rest of the country puts downward pressure on the price of very old cars (20–24 years old), ultimately inducing car dealers to retire them from the market. This market dynamics may help to explain why Wolff (2014) fails to find pollution leakage from LEZs to non-affected areas.



(a) Santiago

(b) Rest of the country

Figure 7: Steady-state fleet composition under a complete-ban vintage driving restriction

Notes: The figure shows the steady-state profile of car fleets in Santiago and the rest of the country under a complete-ban vintage-specific driving restriction. Since the total population in the country has been normalized to unity, each bar indicates the number of cars per capita for a particular vintage group.

Before we move on to compare vintage-specific restrictions to alternative instruments, it is worth highlighting a few observations. First, there is a significant welfare improvement that results from moving from the vintage-specific design of row 4 of Table 4 to the vintage-design of row 5. This is important because some existing driving restrictions, while vintage-specific, impose only mild restrictions on old models, as the design in row 4 does. The second observation is that, by imposing a complete ban on old vehicles, this complete-ban design closes any possibility for a second-car effect to occur. The third observation deals with distributional implications: By placing a total ban on old cars, which are mostly owned by lower-income drivers, such vintage design may raise distributional concerns. Almost all drivers and public-transport users in Santiago are better off under the vintage design of row 5 vis-à-vis the no-intervention scenario. A relatively small group of households driving cars soon to be retired, however, are strictly worse off (we present these results in the online Appendix, Section A.1, Figure A.4). The gain in air quality, which is valued equally by all households in the economy, is not enough to compensate these drivers for the loss that implies moving to either public transport or newer but more expensive cars. In the absence of transfers, the government can still prevent this outcome, at the cost of some efficiency loss, by slightly relaxing the complete ban on old vehicles.

A fourth observation regards the role played by the shape of the emissions-age relationship in our results. As established in Proposition 2, the faster emission rates increase with age, the more to gain from a policy aimed at the extensive margin. Indeed, if we marginally increase the slope of the emissions-age relationship,[70] welfare under the vintage restriction goes up slightly, to 52% of the first-best gain.

Our last observation concerns some limitations of our welfare estimations. So far we have assumed that emission rates are entirely explained by vehicle age, and that age and quality, as perceived by households, are highly correlated. Since we know that emission rates are also explained by other factors, a main concern for our results would be whether these and other factors can eliminate the convexity in the emissions-age/quality relationship enough to invalidate our results. One way would be for these omitted factors to be positively correlated with travel, particularly in newer models, so that these cars would be dirtier than they appear, on average. Another way would be for these omitted factors to give rise to a negative correlation between quality and emission rates so that households in higher-income municipalities own not only newer cars but also larger and potentially more polluting ones. Fortunately, neither possibility is supported by the data (see the online Appendix, Sections D.4 and D.5).

If these omitted factors are available to the regulator for policy design (for example, by establishing restrictions that depend not only on age but also on other observables such as horse power), this can only add convexity to the emissions-age relationship, where "age" must now be interpreted more generally as a collection of observables that separate cars by their emission rates. If for some reason, however, these omitted factors are not available for policy design, our emission rate function e_a would be noisier than it is here. This would have no implications for (optimal) policy design, as emissions enter linearly in the social damage

[70] We do this by placing a bit more weight to NO_x and less to HC in the estimation of ϵ_a^k: $\omega^{HC} = 0.7$, $\omega^{NOx} = 0.2$, and $\omega^{CO} = 0.1$.

function, but it would move our policies further from the Pigouvian benchmark, which in any case is out of the regulator's reach.

5.3. *Vintage restrictions vs alternative instruments*

We now use the model to study how the optimal vintage-specific design performs relative to alternative instruments—namely, scrappage subsidies, annual registration/circulation fees, and gasoline taxes. According to our simulations, the optimal scrappage subsidy varies from \$2420 to \$3240, depending on how much the authority can prevent car dealers with cars outside the restricted area from arbitraging the price gap in scrap values created by the subsidy. This can be done, although at the cost of introducing some friction in the car market, by requiring any given vehicle to have a number of years of registration history in the restricted area. The two simulations reported in rows 6 and 7 of Table 4 correspond, respectively, to the extreme cases of requiring either no registration history or a full history. In any case, the numbers in the table indicate that preventing arbitrage does not make much difference for welfare (although it surely does for the government budget).

More importantly for the purpose of our study, scrappage subsidies provide no efficiency advantage over a vintage-specific restriction. One reason is that vintage restrictions allow the reallocation of cars from areas where they are of no social value to areas where they still are. Another is that implementation constraints should favor the use of vintage-specific restrictions not only because subsidies are subject to arbitrage problems, but also because restrictions are much cheaper to implement for the government under any reasonable estimate of the shadow cost of public funds. Ultimately, this fiscal cost may explain why these subsidies are used only rarely, and when they are, for a very short time.

This fiscal cost could be avoided if, instead of paying for the removal of these old cars, the government could increase their annual registration/circulation fees to reflect their (expected) external pollution costs during the year. Moreover, if the government levies pollution-based circulation fees not only on the oldest, most polluting cars but on all cars, including new ones, the welfare gains exceed those achieved with a vintage restriction (see row 8 in Table 4), provided these fees are set at their welfare-maximizing levels, which roughly equal the expected Pigouvian bills. Interestingly, and like any other instrument, these circulation fees reduce pollution costs close to their first-best level. Introducing these circulation fees is, in any case, a major policy challenge for any authority, as they imply a complete reversal of existing circulation-fee profiles, under which older cars pay much less than newer cars. This contrasts sharply with existing driving restrictions, under which older cars are already subject to much tougher restrictions than newer cars.

Another policy alternative is for the authority to increase gasoline taxes. As shown in row 9 of Table 4, by making no distinction between high- and low-emitting vehicles, a gasoline tax is a poor instrument for handling local pollution, even if set at the welfare-maximizing level of \$1.06 per gallon (which raises the price of gasoline to \$4.06 per gallon). Gasoline taxes impose a heavier burden on newer vehicles because those typically are run more intensely than older vehicles, doing little to move the fleet composition toward cleaner cars (see the online Appendix, Section E.3, Figure E.7, for the steady-state fleet composition).

The best way to understand why gasoline taxes do so poorly in our context is to start by running the following short-run exercise: to compare the performance of a gasoline tax to Pigouvian taxes when rental/ownership decisions are fixed, that is, when only intensive-margin decisions matter. It turns out that a gasoline tax set at the optimal level of \$2 per gallon delivers 50% of the welfare gain under Pigouvian taxes (and 40% if the tax is set at \$1.06 per gallon). Interestingly, Knittel and Sandler (2018) carry out the exact same short-run exercise finding that their (county-level) gasoline tax delivers 35% of the Pigouvian gain (the reason our tax delivers more is probably because our pollution damages are higher).

This exercise illustrates that gasoline taxes are not that bad when it comes to correcting the intensive margin (in fact they do better than a vintage restriction in this case, i.e., provided the fleet is fixed). But in contexts where changes in fleet composition matter a great deal for social welfare, like ours (see Figures 5 and 6), failing to incorporate policy effects on extensive-margin decisions can lead to wrong conclusions, so much that instrument relative performance can reverse.[71] Understanding the role of both intensive and extensive margins raises yet another question: how much is to be gained from combining policies that aim at different margins? Row 10 of Table 4 has the answer: combining a vintage restriction with a gasoline tax delivers 61% of the first-best gain, a 20% increase above the gain from the vintage restriction alone. As far as the political environment permits it, a gasoline tax should always be part of the policy design.

5.4. *Temporal variation in pollution harm*

While the optimal vintage-specific restriction compares reasonably well to alternative instruments in the above results, an additional element strengthen its case. So far, we have assumed that vehicle emissions of local pollutants are equally harmful regardless of the hour, day, or month they are emitted. In reality, however, emissions' impact differs at different points in time, which has been recognized somewhat in existing restriction programs. São Paulo, for instance, places restrictions only during peak hours, from 7 to 10 am and 5 to 8 pm; Paris restricts pre-1997 models only during weekdays; and Santiago extends restrictions only from March through September.

[71] Note also that because extensive-margin adjustments can take a long time before reaching the steady state, the importance of transitory states when evaluating policy should not be understimated. In fact, the contribution of the gasoline tax in row 9 of Table 4 to welfare is 21% of the first-best gain during the transition phase (first 12 years after policy implementation) and only 5% at the steady state. The contribution of the complete-ban restriction of row 5, on the other hand, is 53% during the transition phase and 49% at the steady state.

Extending our model to address this situation is relatively straightforward. Suppose the harm caused by a unit of pollution in the polluted area is h_r during a fraction $\lambda \in (0, 1]$ of the time, and 0 otherwise. In such a setting, driving restrictions appear particularly flexible, going into force only when pollution is a problem. As for alternative instruments, it is evident that, by construction, scrappage subsidies cannot cope with this temporal variation, as scrapped cars are removed permanently from the market. Gasoline taxes face a similar problem because drivers will arbitrate any price difference created by these taxes if they are adjusted daily and/or weekly. In contrast, circulation/registration fees can potentially cope with temporal variation, as driving restrictions do: The authority must offer each year a menu of registration fees that vary by vintage, giving drivers the option to either pay a fee for unlimited car use or pay no fee for car use only during the $1 - \lambda$ hours during which pollution is no problem. However, if for some reason the option of offering these circulation menus is not available, the advantage of circulation fees over vintage restrictions reduces as λ drops and, according to the exercise in the online Appendix (Section E.2), completely disappears when $\lambda = 0.25$.

6. Conclusions

Evidence from many cities around the world experiencing local air pollution problems suggests that driving restrictions are becoming increasingly popular tools to control vehicle pollution. Previous literature (e.g., Eskeland and Feyzioglu, 1997; Davis, 2008; Gallego et al., 2013), as well as this paper, show that these policies perform particularly poorly when designed to affect a driver's *intensive* margin (i.e., amount of travel) with restrictions that treat all cars equally, regardless of how much they pollute. In this paper, we have instead focused on the potential of these policies to affect a driver's *extensive* margin (i.e., type of car driven). By introducing vintage-specific restrictions, or more precisely, restrictions that differentiate cars by their pollution rates, this paper shows that these policies are effective in moving the fleet composition toward lower-emitting vehicles. The vintage-specific design should take the "complete-ban" structure already utilized in Paris and Germany's low emission zones: a (moving) vintage threshold separates cars between complete restriction and full exemption.

As these vintage-specific restrictions prove to be effective and practical for fighting local air pollution, it would be interesting to apply our model toward exploring their potential to accelerate the introduction of electric vehicles at a much lower cost to government than the subsidies currently being offered in the developed world. An important design issue is to make sure the price of the pollution-free (and fully exempt) option is not much higher than the price of existing, polluting alternatives—not enough that drivers might opt to buy a second, polluting car instead of the pollution-free option, to bypass the restriction.[72] The

[72] One important reason why the Santiago-1992 restriction worked reasonably well was precisely because the clean option at that time, which was to switch to a car with a catalytic converter sooner than otherwise, was affordable to many households.

optimal design might be combination of subsidies on pollution-free vehicles and restrictions on polluting ones that changes over time as the fraction of electric vehicles in the market evolves.

References

Adamou, A., S. Clerides, and T. Zachariadis (2014). Welfare implications of automobile feebates: A simulation analysis. *The Economic Journal 124*(578), F420–F443.

Adda, J. and R. Cooper (2000). Balladurette and juppette: A discrete analysis of scrapping subsidies. *Journal of political Economy 108*(4), 778–806.

Angrist, J. D. and M. Rokkanen (2015). Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association 110*(512), 1331–1344.

Bento, A. M., L. H. Goulder, M. R. Jacobsen, and R. H. Von Haefen (2009). Distributional and efficiency impacts of increased US gasoline taxes. *American Economic Review 99*(3), 667–99.

Berry, S. and A. Pakes (2007). The pure characteristics demand model. *International Economic Review 48*(4), 1193–1225.

Currie, J. and M. Neidell (2005). Air pollution and infant health: What can we learn from California's recent experience? *The Quarterly Journal of Economics 120*(3), 1003–1030.

Davis, L. W. (2008). The effect of driving restrictions on air quality in Mexico City. *Journal of Political Economy 116*(1), 38–81.

d'Haultfoeuille, X., P. Givord, and X. Boutin (2014). The environmental effect of green taxation: the case of the French Bonus/Malus. *The Economic Journal 124*(578), F444–F480.

Drummond, P. and P. Ekins (2016). Tackling air pollution from diesel cars through tax: options for the UK. Technical report, UCL.

EMBARQ (2017). Informe de la evaluación externa al sistema de transporte público remunerado de pasajeros de la provincia de santiago y de las comunas de san bernardo y puente alto. Technical report, prepared by the Centro de Transporte Sustentable de México A.C. and the World Resources Institute, México City, February.

Eskeland, G. S. and T. Feyzioglu (1997). Rationing can backfire: The "day without a car" in Mexico City. *The World Bank Economic Review 11*(3), 383–408.

Fullerton, D. and S. E. West (2010). Tax and subsidy combinations for the control of car pollution. *The BE Journal of Economic Analysis & Policy 10*(1).

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.

Gavazza, A., A. Lizzeri, and N. Roketskiy (2014). A quantitative analysis of the used-car market. *American Economic Review 104*(11), 3668–3700.

Hahn, R. W. (1995). An economic analysis of scrappage. *The RAND Journal of Economics 26*(2), 222–242.

Hoekstra, M., S. L. Puller, and J. West (2017). Cash for Corollas: When stimulus reduces spending. *American Economic Journal: Applied Economics 9*(3), 1–35.

Kahn, M. E. (1996). New evidence on trends in vehicle emissions. *The RAND Journal of Economics 27*(1), 183–196.

Knittel, C. R. and R. Sandler (2018). The welfare impact of indirect Pigouvian taxation: Evidence from transportation. *American Economic Journal: Economic Policy*, forthcoming.

Liu, Y., Z. Yan, S. Liu, Y. Wu, Q. Gan, and C. Dong (2017). The effect of the driving restriction policy on public health in Beijing. *Natural Hazards 85*(2), 751–762.

Lu, S. (2006). Vehicle survivability and travel mileage schedules. Technical report, US DOT, Springfield, VA.

Mian, A. and A. Sufi (2012). The effects of fiscal stimulus: Evidence from the 2009 cash for clunkers program. *The Quarterly Journal of Economics 127*(3), 1107–1142.

Molina, L. and M. Molina (Eds.) (2002). *Air quality in the Mexico megacity: An integrated assessment.* Kluwer Academic Publishers, Dordrecht.

MTT (2016). Información de transporte urbano en ciudades de Chile: Período 2010-2015. Technical report, prepared by the MTT (Ministerio de Transporte y Telecomunicaciones) Office of Development and Planning. Santiago.

Oliva, P. (2015). Environmental regulations and corruption: Automobile emissions in Mexico City. *Journal of Political Economy 123*(3), 686–724.

Onursal, B. and S. Gautam (1997). Vehicular air pollution: Experiences from seven Latin American urban centers. *World Bank Technical Paper No. 373, Washington, DC*.

Parry, I. and J. Strand (2012). International fuel tax assessment: An application to Chile. *Environment and Development Economics 17*(2), 127–144.

Posada, F., D. V. Wagner, G. Basnal, and R. Fernandez (2015). Survey of best practices in reducing emissions through vehicle replacement programs. *ICCT White Paper. ICCT, Washington, DC*.

Reynaert, M. (2017). Abatement strategies and the cost of environmental regulation: Emission standards on the European car market. *Working paper*.

Rizzi, L. I. and C. De La Maza (2017). The external costs of private versus public road transport in the metropolitan area of santiago, chile. *Transportation Research Part A: Policy and Practice 98*, 123–140.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

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.

Wolff, H. (2014). Keep your clunker in the suburb: Low-emission zones and adoption of green vehicles. *The Economic Journal 124*(578), F481–F512.

Zhang, W., C.-Y. C. L. 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.