

 Open access • Journal Article • DOI:10.1093/RESTUD/RDZ031

Vintage-Specific Driving Restrictions — [Source link](#)

[Nano Barahona](#), [Francisco Gallego](#), [Juan-Pablo Montero](#)

Institutions: [Stanford University](#), [Pontifical Catholic University of Chile](#)

Published on: 01 Jul 2020 - [The Review of Economic Studies](#) (Oxford Academic)

Related papers:

- [The Effect of Driving Restrictions on Air Quality in Mexico City](#)
- [Adopting a Cleaner Technology: The Effect of Driving Restrictions on Fleet Turnover](#)
- [The effect of transport policies on car use: Evidence from Latin American cities](#)
- [A practical approach for curbing congestion and air pollution: Driving restrictions with toll and vintage exemptions](#)
- [Evaluating the benefits and effectiveness of public policy](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/vintage-specific-driving-restrictions-4rq5tv7zxi>

VINTAGE-SPECIFIC DRIVING RESTRICTIONS

Nano Barahona

Francisco Gallego

Juan-Pablo Montero

LATIN AMERICAN AND THE CARIBBEAN ECONOMIC ASSOCIATION

May 2018

The views expressed herein are those of the authors and do not necessarily reflect the views of the Latin American and the Caribbean Economic Association. Research published in this series may include views on policy, but LACEA takes no institutional policy positions.

LACEA working papers are circulated for discussion and comment purposes. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

© 2018 by Nano Barahona, Francisco Gallego and Juan-Pablo Montero. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.



LACEA

+57 1 3259777
lacea@fedesarrollo.org.co

www.lacea.org
Bogotá, Colombia

Vintage-specific driving restrictions

Nano Barahona
Economics Department, Stanford University
hbaraho@stanford.edu

Francisco Gallego
Economics Department, Pontificia Universidad Catolica de Chile (PUC-Chile)
fgallego@uc.cl

Juan-Pablo Montero
Economics Department, Pontificia Universidad Catolica de Chile (PUC-Chile)
jmontero@uc.cl

ABSTRACT

Local air pollution has led authorities in many cities around the world to impose limits on car use, increasingly by means of driving restrictions or license-plate bans. With some exceptions, these restrictions tend to be poorly designed creating incentives for drivers to buy additional, more polluting cars. We study vintage-specific restrictions that place heavy limits on older, polluting vehicles and none on newer, cleaner ones. A novel model of the car market and evidence from Santiago's 1992 program, the earliest attempt to use vintage-specific restrictions, are used to show that these restrictions can be welfare enhancing by accelerating fleet turnover toward cleaner cars. These policies compare well to alternative instruments such as scrappage subsidies and pollution-based registration fees.

JEL Classification: R41, Q53, Q58

Keywords: Driving restrictions, Local Pollution, Car Turnover

ACKNOWLEDGEMENTS AND FINANCIAL DISCLOSURE

We have benefited from the comments of Lucas Davis, Liran Einav, Juan Escudero, Claudio Ferraz, Matthew Gentzkow, Marco Gonzalez-Navarro, Larry Goulder, Enrique Ide, Gaston Illanes, Kelsey Jack, Shaun McRae, Sebastian Otero, Mar Reguant, James Sallee, Brandon Shaufele, Paulo Somaini 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. Gallego acknowledges financial support from CONICYT/Programa de Investigacion Asociativa (Project SOC 1102) and Montero likewise from Fondecyt (Grant No. 1130998) and the ISCI Institute (Basal FBO-16). We also thank Jose D. Salas and Felipe Sepulveda for research assistance.

Vintage-specific driving restrictions

Nano Barahona Francisco Gallego Juan-Pablo Montero*

May 9, 2018

Abstract

Local air pollution has led authorities in many cities around the world to impose limits on car use, increasingly by means of driving restrictions or license-plate bans. With some exceptions, these restrictions tend to be poorly designed creating incentives for drivers to buy additional, more polluting cars. We study vintage-specific restrictions that place heavy limits on older, polluting vehicles and none on newer, cleaner ones. A novel model of the car market and evidence from Santiago's 1992 program, the earliest attempt to use vintage-specific restrictions, are used to show that these restrictions can be welfare enhancing by accelerating fleet turnover toward cleaner cars. These policies compare 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 drivers to give up their cars in favor of public transport, authorities increasingly rely on

*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 have benefited from the comments of Lucas Davis, Liran Einav, Juan Escudero, Claudio Ferraz, Matthew Gentzkow, Marco González-Navarro, Larry Goulder, Enrique Ide, Gastón Illanes, Kelsey Jack, Shaun McRae, Sebastián Otero, Mar Reguant, James Sallee, Brandon Shaufele, Paulo Somaini 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. Gallego also acknowledges financial support from CONICYT/Programa de Investigación Asociativa (Project SOC 1102) and Montero likewise from Fondecyt (Grant No. 1130998) and the ISCI Institute (Basal FBO-16). We also thank José D Salas and Felipe Sepúlveda for research assistance.

¹Cars are major contributors of carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NO_x). These local pollutants, unlike global pollutants such as carbon dioxide (CO₂), 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.

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-Chile (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 et al., 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 October 2017 with a reform to the program that now extends the one-day-a-week restriction to 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

²Authorities 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 could potentially alter a city’s fleet composition.

³Zhang 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.

⁴LEZ 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 ’92-93 vintage discontinuity. 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.⁵ 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 enough about the welfare implications of these policies and says nothing at all about the socially optimal vintage-specific design and how that compares to optimal designs of 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. 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.⁶ 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-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-

⁵Similar results are in [Wolff \(2014\)](#) for the LEZ programs in Germany.

⁶Using 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).

to-pay driver decides to scrap the unit. We are the first to model these long equilibrium transitions and their implications for policy evaluation.⁷

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 is that vintage-specific restrictions can be an effective tool to reduce local air pollution by helping to accelerate fleet turnover toward lower-emitting vehicles. The optimal vintage-specific design follows closely the complete-ban structure seen in Paris and Germany’s LEZs: an optimal (moving) vintage threshold separates cars between complete restriction and (almost) full exemption. The reason for this complete-ban structure is that a driving restriction’s relevant margin of action is ownership (i.e., extensive margin), never use (i.e., intensive margin). Some middle-aged cars may face some restriction, but only to hasten their owners’ switch to newer cars. By working exclusively on the extensive margin, a well-designed driving restriction can be particularly effective when emission rates vary widely with (observable and enforceable) car characteristics, most notably vintage.^{8,9}

On the contrary, a driving restriction that aims exclusively 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

⁷Gavazza et al. (2014) omit any market dynamics by focusing on the steady-state equilibrium. Adda and Cooper (2000) consider dynamics but 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.

⁸As discussed in the estimation section below, emission-rate variation is also explained by other observable car attributes. Including these observables in the restriction design would only reinforce our results, as they only add emission-rate heterogeneity to the analysis. So, despite vintage is our focus in much of the analysis, nothing fundamental prevents a broader interpretation, say, any collection of observables that separate cars by their emission rates.

⁹We 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.

cars from the road; it also reduces their prices, extending their lives and dampening sales of new cars. In fact, our base simulation shows that a uniform one-day-a-week restriction leads to a welfare *loss* of 146% of the welfare gain from implementing the first-best. In contrast, the optimal (vintage-specific) restriction, which establishes a complete ban on cars at least 16 years old, leads to a welfare *gain* of 58% 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).¹⁰ Much has been written on the use of scrappage subsidies, also known as cash-for-clunker programs, to help accelerate the retirement of old vehicles and stimulate the purchase of new ones (e.g., [Hahn, 1995](#); [Adda and Cooper, 2000](#); [Mian and Sufi, 2012](#); [Hoekstra et al., 2017](#)). Our model shows that the optimal subsidy does not present any efficiency advantage over the optimal vintage-specific restriction. The reason is simple: both instruments are designed to work exclusively on the extensive margin. If anything, implementation constraints appear to favor the use of vintage-specific restrictions. First, for a scrappage subsidy to reach its full potential, the regulator must prevent old cars from outside the restricted 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 restricted area. More important, even 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. This is explained in part by the high fiscal cost incurred by the government, but also because in most cases these programs were conceived as a temporary stimulus to boost the local auto industry, not as a permanent environmental policy.

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 hours 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

¹⁰We also consider a gasoline tax despite it is recognized to be a poor instrument for handling local pollutants, as recently stressed by [Knittel and Sandler \(2018\)](#). Our (base) simulation numbers are similar to theirs in that the optimal uniform gasoline tax, used in isolation, delivers 32% of the welfare gain from implementing the first-best. This is not a call to disregard this instrument but rather to use it in combination with instruments that act on the extensive margin such as optimal vintage-specific restrictions.

¹¹So far, these fees/subsidies cover only CO₂ 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.

effectively since they involve a complete set of prices to adjust this margin at each vintage level, which optimal restrictions and subsidies may fail to do. 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). But even in the absence of these latter, the results of this paper show that well-designed, vintage-specific restrictions are a good (sometimes superior) 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.

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,¹² 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.

¹²A 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 323 of the 346 municipalities in the country,¹³ 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)

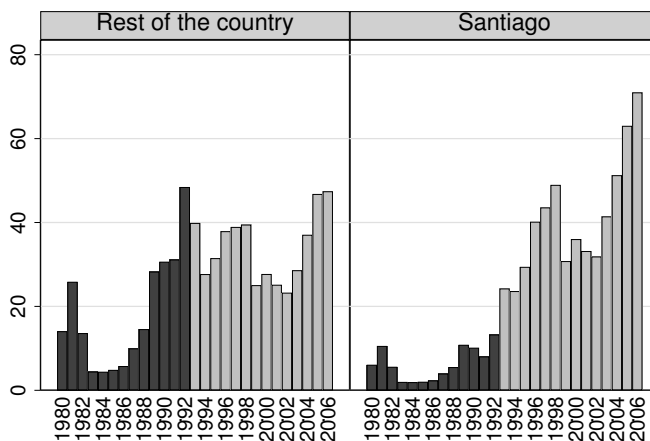


Figure 1: Fleets in Santiago vs the rest of country in 2006

Notes: Each bar represents number of cars (in thousands) of each vintage.

Pre-1993 vintages are highlighted as darker bars.

¹³The municipalities missing information are in remote areas with low population density. More details about the database are in the online Appendix (Section B.1).

One way to test whether there was a causal effect of the policy on q_τ^i , the number of cars of vintage- τ in municipality i in a given year, is to focus on vintages around the 92-93 discontinuity and estimate the following equation:

$$\log(y_{92,93}^i/(1 - y_{92,93}^i)) = \beta_{92,93}DR_i + x_i'\gamma + \varepsilon_i \quad (1)$$

where $y_{92,93}^i = q_{92}^i/(q_{92}^i + q_{93}^i)$, 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), and x_i is a vector with the municipality's characteristics such as income per capita, population, distance to Santiago, income dispersion and level of urbanization.¹⁴ Columns 1 and 2 of Table 1 present the *marginal effects* of estimating equation (1) for year 2006.¹⁵ The effects we find are statistically as well as economically significant. If, in a given municipality not affected by the restriction, we observe one 93 model for each 92 model (i.e., $y_{92,93}^i = 0.5$), in a similar municipality in Santiago, that ratio would be 3.05 ($= [0.5 - 0.253]^{-1} - 1$), according to the point estimate in column 2.

This exercise assumes that, controlling for all the variables included in x , β captures the causal effect of the driving restriction on $y_{92,93}^i$. A simple check of this assumption is to run placebo regressions of the ratio for pairs of vintages other than the 92-93 dyad. Under this assumption, we should find a zero effect for other dyads: for them, the policy should not create incentives for jumps in car ownership in Santiago relative to the rest of the country. Results in columns 3 and 4 of Table 1 confirm this to be the case for the $y_{91,92}^i$ and $y_{93,94}^i$ ratios, but it actually extends to other dyads (see the online Appendix, Section A.2, Table A.1).

While the previous identification strategy focuses on estimating policy effects on the discontinuity between the 1992 and 1993 vintages in treated (i.e., Santiago) versus non-treated (i.e., the rest of the country) municipalities, we also estimate a more general model by which we study the effect of the driving restriction on other vintages. This is important because, based on what we know from other restriction programs, it could well be that the exodus of 1992 models was completely undone if a sizable number of drivers by-passed the restriction not with the purchase of a post-1992 model but with the purchase of a second and possibly much older pre-1993 model. Thus, we estimate the following regression:

$$\log(q_\tau^i) = \beta_\tau DR_i + \alpha_\tau \log(INCOME_i) + \gamma_\tau \log(POP_i) + z_i'\zeta + \delta_\tau + \varepsilon_\tau^i \quad (2)$$

where $INCOME_i$ is municipality i 's income per capita, POP_i is the municipality's total population, z_i is a vector that includes the remaining controls used in (1), and δ_τ is a vintage

¹⁴This equation comes from a logistic model where consumers choose between cars vintage $\tau = 1992$ and $\tau = 1993$. In the online Appendix (Section A, Table A.1) we present estimates coming from a linear probability model with $y_{92,93}^i$ as the dependent variable. Results for the marginal effects are very similar to the ones in Table 1.

¹⁵We use 2006 because that is the year closest to policy implementation, but similar results are obtained for more recent years.

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

	92-93	92-93	91-92	93-94
DR_i	-0.284*** (0.013)	-0.253*** (0.037)	0.00241 (0.015)	0.0145 (0.015)
Controls	No	Yes	Yes	Yes
Observations	268	268	268	268
R^2	0.480	0.527	0.090	0.320

Notes: OLS regressions with one observation per municipality. Municipalities with less than 300 cars were dropped from the sample. The dependent variable corresponds to $\log(y_{\tau,\tau+1}^i/(1 - y_{\tau,\tau+1}^i))$ with $y_{\tau,\tau+1}^i = q_{\tau}^i/(q_{\tau}^i + q_{\tau+1}^i)$, where q_{τ}^i is the total number of cars of vintage τ found in municipality i in 2006. The first two columns correspond to the case of $\tau = 1992$, while columns 3 and 4 correspond to $\tau = 1991$ and $\tau = 1993$, respectively. We report marginal treatment effects on $y_{\tau,\tau+1}^i$. Standard errors are calculated via block bootstrap at the province level (53 provinces in total). Municipality controls include: Income per capita (in linear and quadratic form), population, coefficient of variation of income per capita, urbanization ratio, distance to Santiago (in linear and quadratic forms), and dummies for municipalities in northern and far away regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

fixed effect. Assuming that the driving-restriction dummy (DR_i) captures only effects of the driving restriction on the car stock of each vintage,¹⁶ this approach allows us to also estimate the effect of the policy further from the 92-93 discontinuity.

Unlike α_{τ} and γ_{τ} , we expect the evolution of β_{τ} to exhibit a discontinuous jump around the 92 and 93 vintages. Figure 2 presents regression results for year 2006 with estimates for both α_{τ} and β_{τ} (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, the 92-93 discontinuity is clear in panel (b). The point estimate for vintage 92 is -1.008 (statistically significant at the 1% level). This implies that for each 1992 model circulating in a given municipality in Santiago, 2.74 such models will be in a similar municipality not subject to restriction. Conversely, the point estimate for vintage 93 of 0.239 (statistically significant at the 10% level) indicates that for each 1993 model circulating in a given municipality in Santiago, only 0.79 such models will be in a similar municipality not subject to restriction.¹⁷

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.,

¹⁶This is similar in spirit to 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.

¹⁷In the online Appendix (Section B.2) we also exploit the 1992-93 threshold in a regression discontinuity design. Note that the difference $\beta_{93} - \beta_{92} = 1.247$ is very close to the estimate of the same effect found using an RDD approach, which is 1.222.

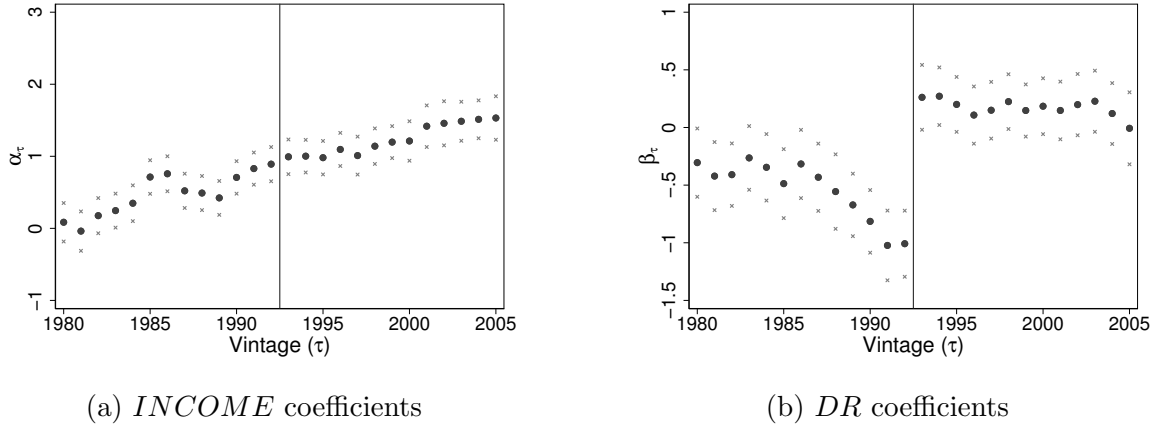


Figure 2: Vintage effects of driving restrictions and income

Notes: This figure presents estimated vintage effects after estimating equation (2) 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.

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.¹⁸

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 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,

¹⁸In 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.

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 for each of the seven models mentioned above:

$$\log(P_{irt}^m) = \beta^m DR_\tau + g(\tau) + \delta_a^m + \delta_t^m + \varepsilon_{irt}^m \tag{3}$$

where P_{irt}^m 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 ,¹⁹ δ_a^m and δ_t^m are age of the car (where age $a = t - \tau$) and date of the ad fixed effects, respectively, and ε_{irt} 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 $g(\tau)$.

We also pool observations for all the models and include model fixed effects in equation (3) to estimate the average effect of the policy on prices. Table 2 presents results for the pooled data (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 pay more for

¹⁹We use three different specifications for $g(\tau)$: (i) using 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), (ii) using a linear function of vintage that takes a different slope for before and after 1993, and (iii) using interactions of age dummies with linear trends in time to allow for different depreciation rates.

cars exempted from the restriction.^{20,21}

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

	(1)	(2)	(3)	(4)
DR_τ	0.048*** (0.006)	0.065*** (0.004)	0.065*** (0.004)	0.064*** (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 19.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 driving restriction 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.²² 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).²³

²⁰Notice that the cost of \$265 for replacing a catalytic reported in Onursal and Gautam (1997) cannot explain the effects we report in Table 2 because this is such a small share of the total price of a new car. Moreover, this cost difference should also be captured by our control for the price differences of new cars reported in column 2. Besides, if differences were explained by this 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. In addition, 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.

²¹We 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 exploits the fact that our database contains a few ads for some pre-1993 Honda Accords with indication that they were already equipped with a catalytic converter.

²²Emission 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 inspection test reports test location and 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.

²³Figure 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 entering with cleaner technologies (although much attenuated after 2003) and (ii) pollution-control technologies wearing out overtime.

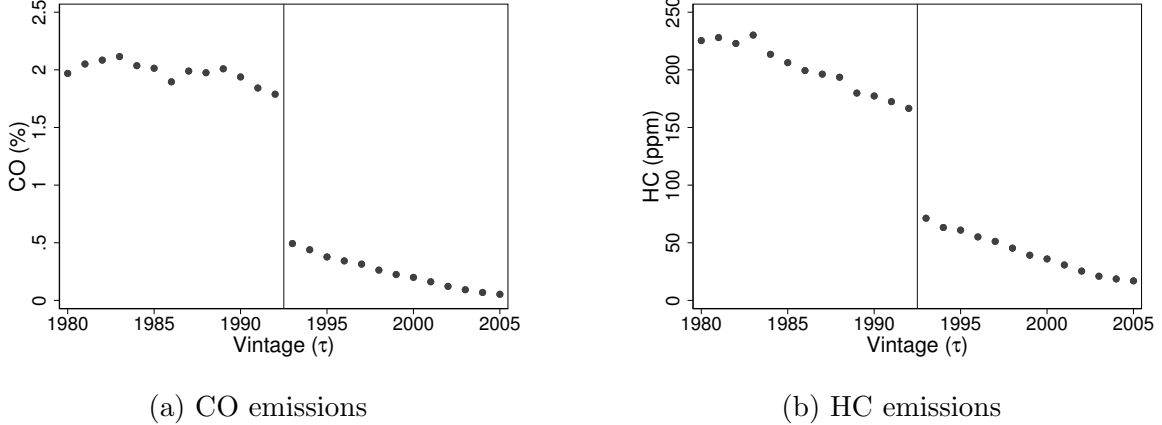


Figure 3: CO and HC emissions

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.

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 local pollutants in Figure 3 for both Santiago (the restricted area) and the rest of the country (the non-restricted area).

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:

$$\Delta E_\tau^k = (q_\tau^k x_\tau^k - q_\tau^{k'} x_\tau^{k'}) e_\tau \quad (4)$$

where q_τ^k is the total number of cars of vintage τ that are actually in area k , $q_\tau^{k'}$ 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 , $x_\tau^{k'}$ 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:

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

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. While q_τ^k is obtained directly from the circulation-permit data described in Section 2.1, $q_\tau^{k'}$ is estimated using results from equation (2). 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.²⁴ Thus, taking as given the overall fleet in 2006, we can obtain estimates of $q_\tau^{k'}$ from (2) as follows

$$q_\tau^{r'} = q_\tau^r - \frac{q_\tau^{nr} q_\tau^r (1 - \exp(-\beta_\tau))}{q_\tau^{nr} + q_\tau^r \exp(-\beta_\tau)} \quad (6)$$

and $q_\tau^{nr'} = q_\tau^{nr} - (q_\tau^{r'} - q_\tau^r)$, where β_τ are the coefficients presented in Figure 4(b).²⁵

Emission rates e_τ for CO and HC correspond to the values plotted in Figure 3.²⁶ Miles traveled, x_τ^k and $x_\tau^{k'}$, 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 $x_\tau^{nr'}$ and x_τ^{nr} for all vintages, and to equal $x_\tau^{r'}$ and x_τ^r for $\tau \geq 1993$. For $\tau \leq 1992$, we let $x_\tau^{r'}$ and x_τ^r differ anywhere between 0 and 8.2%.²⁸

Based on this information for the different components in ΔE_τ^k , we estimate that CO emissions in Santiago dropped anywhere between 21 and 28% in 2006 because of the 1992 policy (the reduction in HC emissions is anywhere between 14 and 21%). 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 16 and 22% for CO and between 11 and 17% 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.

²⁴We 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.

²⁵Controlling for $INCOME_i$, POP_i , z_i , and $\varepsilon_{i\tau}$, we have that $q_\tau^{nr'} = q_\tau^{r'}$. This and $q_\tau^{nr} + q_\tau^r = q_\tau^{nr'} + q_\tau^{r'}$ yield (6).

²⁶Emission rates for 2006 are not available, but this is less of a problem if they are thought to differ from the 2008 rates 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_τ^k 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.

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

²⁸Given 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, i.e., the intensity in the absence of intertemporal substitution.

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? What is the socially optimal driving restriction design? How does this latter 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 applied to Santiago and the rest of the country 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. Notation

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.²⁹ A large number of car dealers buy new cars from car producers and rent them, together with second-hand cars, to drivers.³⁰ The (annual) rental price for a car of age $a = \{0, 1, 2, \dots\}$ 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 $\gamma_a \geq \gamma_{a+1}$ (to simplify the exposition we assume throughout this section, but not in the estimation and simulations, that $\gamma_a = \gamma$ for all a). 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, if any, get scrapped for a value v .³¹ The remaining

²⁹We 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.

³⁰Note 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.

³¹This scrappage value can be interpreted, for example, as the value a dealer gets for selling remaining parts

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 γ), we have that:

$$v = p_{T(t),t} + \gamma\delta v \quad (7)$$

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:

$$c = p_{0,t} + \sum_{i=1}^{\Gamma(t)} (\gamma\delta)^i p_{i,t+i} + (\gamma\delta)^{\Gamma(t)+1} v \quad (8)$$

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). Note that both $\Gamma(t)$ and $T(t)$ depend on the existing stock of vehicles, so only in steady state, $\Gamma(t) = T(t)$.

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. We expect the willingness to pay to be positively correlated with income. A driver of type θ who benefits from driving a car of quality s_a for x miles has to pay a variable cost ψx and a rental price p_a every period. Every period she obtains a utility of (to save on notation henceforth, in many places we will omit the subscript “ t ” unless it is strictly necessary):

$$u(\theta, a, x) = \frac{\alpha}{\alpha - 1} \theta s_a x^{\frac{\alpha-1}{\alpha}} - \psi x - p_a \quad (9)$$

where θ is distributed according to the cumulative distribution function $F(\theta)$ over the interval $[0, \bar{\theta}]$, $s_a > 0$ is the quality of a car that is a years old, x corresponds to miles traveled during the period, $\alpha > 1$ is a parameter that captures decreasing returns on car use (as seen from equation (10) below, it also corresponds to the demand elasticity of driving), ψ is the per-mile cost of using the car (parking, gasoline, maintenance, insurance, inspections, etc.),³² and p_a is the rental price. Car quality falls with age according to $s_{a+1} = \varsigma s_a$ with $\varsigma \in (0, 1)$, either because older cars are more likely to break down or because they lack the latest technological advances.³³

The problem of a type- θ driver is to decide what car to rent and how much to use it so as to maximize equation (9). If she happens to rent an a -year-old car, utility maximization

or exporting a car to another country (we assume v to be insensitive to $T(t)$).

³²If congestion is a problem, ψ may also include (socially optimal) congestion charges, which we do not model explicitly.

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

leads to:

$$x(\theta) = \left(\frac{\theta s_a}{\psi} \right)^\alpha \quad (10)$$

miles of driving per period. Anticipating this, her problem reduces to renting the car a that maximizes:

$$u(\theta, a) = \kappa (\theta s_a)^\alpha - p_a \quad (11)$$

where $\kappa = [(\alpha - 1)\psi^{\alpha-1}]^{-1}$. Our formulation captures with a single parameter two empirical regularities that we observe in our data (and in [Lu, 2006](#) for U.S. data): that households which value quality more tend to drive newer cars and that newer cars are, on average, run more often.³⁴

Cars emit local pollutants (e.g., CO, HC) at a rate of e_a per mile, where $e_a \leq e_{a+1}$. How harmful this local pollution is to society depends on where the car is used.³⁵ We assume that drivers live and use their cars in two distinct areas: “polluted” and “non-polluted” areas.³⁶ The harm per mile generated by a car of vintage a is $h_p e_a$ in the “polluted” area (e.g., Santiago) and $h_{np} e_a$ in the “non-polluted” area (e.g., rest of the country), with $h_p \gg h_{np} \approx 0$. 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 some passages of the paper we refer to the polluted area as the “restricted” area. Drivers’ valuation θ in area $k \in \{p, np\}$ is distributed according to the cumulative distribution $F_k(\theta)$ over the interval $[0, \bar{\theta}]$, where $\mu F_p(\theta) + (1 - \mu) F_{np}(\theta) = F(\theta)$ for all $\theta \in [0, \bar{\theta}]$ and μ is the fraction of households living in the polluted area. Finally, a driver θ who does not rent a car obtains an outside utility from riding pollution-free public transport, which is assumed constant across households and normalized to zero.

3.2. *Equilibrium benchmark: No intervention*

We first characterize the market equilibrium in the absence of any pollution-control policy. At the beginning of any given period, say, year t , there will be a stock of used cars $\mathbf{S}_t = (q_{1,t}, q_{2,t}, \dots)$. As a function of that stock, the market equilibrium for the year 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 $\{\theta_{0,t}, \theta_{1,t}, \dots\}$ precisely determines the prices at which certain drivers rent certain cars. Denote by $\theta_{a,t}$ the driver who, at time t , is indifferent to

³⁴We abstract from the possibility of households renting multiple cars mainly because the optimal (vintage-specific) driving restriction eliminates this possibility by construction, as we shall see.

³⁵It also depends on when the car is used (e.g., weekends, peak hours). We cover this temporal distinction in section 5.4.

³⁶The 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.

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:

$$\kappa (\theta_a s_a)^\alpha - p_a = \kappa (\theta_a s_{a+1})^\alpha - p_{a+1} \quad (12)$$

for all $a = 0, 1, \dots, T - 1$, 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 (12) is that a higher valuation consumer obtains strictly more surplus than a lower valuation consumer.

In equilibrium, the series of cutoff levels $\{\theta_0, \theta_1, \dots\}$ must be consistent with the 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

$$q_0 = 1 - F(\theta_0) \text{ and } q_a = F(\theta_a) - F(\theta_{a+1}) \quad (13)$$

for all age- a vehicles that are rented in equilibrium.

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

$$F(\theta_{T-1}) - F(\theta_T) \leq \gamma q_{T-1} \quad (14)$$

where γq_{T-1} is the number of age T vehicles that survived from the last period.³⁷

One last condition must hold in equilibrium: The lowest-valuation household to rent a car today, θ_T , obtains its outside utility:

$$\kappa (\theta_T s_T)^\alpha - p_T = 0 \quad (15)$$

If (15) 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 (7) and (8), conditions (12) through (15) determine the unique equilibrium for any given stock of used cars \mathbf{S}_t , that is, rental prices of new and used cars and sales of new cars. Unlike some previous work, we are interested not only in the steady-state equilibrium, but also in the equilibrium during the transition phase after a policy shock. Transitions can be particularly long in car markets, so they cannot be neglected in policy evaluation and design.

³⁷Note 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 (7) 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.

3.3. Available policy instruments

Clearly, in the absence of any policy intervention the market equilibrium above leads to socially inefficient levels of pollution. As prescribed by Pigou 90 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 et al., 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 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 area from its baseline value of v to $v + \Delta$. Since the price differential Δ creates incentives for drivers in the non-polluted area to scrap their vehicles in the polluted 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 polluted area. A gasoline tax also enters the model in a straightforward manner, by increasing the variable cost of using a car in the polluted area from the baseline value of ψ to $\psi + g$.³⁸ Unlike scrappage subsidies and gasoline taxes, annual registration fees can be made vintage specific. They enter into the model by increasing the rental price of all polluting vehicles in the polluted area according to their emissions rate e_a and expected mileage; say, 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, such as 4 of 5 weekdays each week. Since drivers can move some trips from one weekday to another, R should not be read as 4/5 in this example, but more. It is less obvious whether the trips that can be moved are the most valuable to the driver. We adopt the conservative assumption —less favorable to the driving restriction option— that the driving restriction destroys an equal fraction of car trips of different values during the day of the restriction. Some of these car trips may be replaced by using public transport.

Since we have not made, at least not yet, 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 owns an

³⁸We omit from the model the possibility that households may drive outside the restricted area only to fill their tanks with cheaper gasoline.

a -year-old car that faces an effective restriction R_a will now drive:

$$x(\theta, a, R_a) = R_a \left(\frac{\theta s_a}{\psi} \right)^\alpha \quad (16)$$

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

$$u(\theta, a, R_a) = R_a \kappa (\theta s_a)^\alpha - p_a \quad (17)$$

per period (recall that the utility from using public transport, whether alone or in combination with private transport, is normalized to zero).³⁹

By examining how these instruments work, we expect them to have different impacts on rental and usage decisions, and hence, on welfare. In fact, some instruments act exclusively on the extensive margin (e.g., scrappage subsidies, circulation/registration fees) while others (e.g., driving restrictions) potentially also affect the intensive margin when setting $R_a \in (0, 1)$. In general, restoring efficiency requires acting upon both 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 3.2, leaving the general setting for the estimation and simulations that follow.

3.4. Planner's problem in a two-period world

In order to provide intuition and grasp the relevant trade-offs in the planner's problem, here we take the planner to a situation where cars last only two periods. In the second period, car dealers have the option to either scrap their cars for a value of v (or $v + \Delta$ if a scrappage-subsidy policy is in place, and provided the car has a full registration history in the polluted area) or rent them to households for p_1 , after which they are discarded for no value.

Without losing much content, this setting makes the conceptual analysis tractable for two reasons. First, it leaves rental prices unchanged to any policy intervention except in the polluted area after the introduction of a scrappage subsidy. This implies that only rental and usage decisions in the polluted area are affected by policy. Following (7) and (8), we have that, in the absence of intervention, $p_0^n = c - v$ and $p_1^n = v$ (note that, to simplify notation, we also assume that $\delta = \gamma = 1$, in this section only), which, according to (12) and (15), leads to the following rental cutoffs:

$$\theta_0^n = \frac{1}{s_0} \left[\frac{c - 2v}{\kappa(1 - \varsigma^\alpha)} \right]^{1/\alpha} \quad \text{and} \quad \theta_1^n = \frac{1}{\varsigma s_0} \left[\frac{v}{\kappa} \right]^{1/\alpha} \quad (18)$$

where superscript “ n ” denotes no intervention and $\varsigma = s_1/s_0$. New cars are rented by house-

³⁹Note that (17) is not obtained by plugging (16) in (9). Doing so would be equivalent to assuming that the restriction destroys the least valuable trips, replicating the work of a driving fee set at the proper level.

holds with $\theta \geq \theta_0^n$ while old cars are entered by households with $\theta \in [\theta_1^n, \theta_0^n)$. To avoid corner solutions, we assume throughout that (14) holds strictly, that is, $F_p(\theta_0) - F_p(\theta_1) < F_p(\bar{\theta}) - F_p(\theta_0)$ for any policy intervention (including no intervention).

The second reason is that any market adjustment to policy intervention takes place immediately. This reduces the planner's problem to maximize the following social-welfare function:

$$\begin{aligned} W = & -c[F_p(\bar{\theta}) - F_p(\theta_0)] + v[F_p(\bar{\theta}) - 2F_p(\theta_0) + F_p(\theta_1)] \\ & + \int_{\theta_0}^{\bar{\theta}} \left\{ \frac{\alpha}{\alpha - 1} \theta s_0 x_0(\theta)^{\frac{\alpha-1}{\alpha}} - (\psi + h_p e_0) x_0(\theta) \right\} f_p(\theta) d\theta \\ & + \int_{\theta_1}^{\theta_0} \left\{ \frac{\alpha}{\alpha - 1} \theta s_1 x_1(\theta)^{\frac{\alpha-1}{\alpha}} - (\psi + h_p e_1) x_1(\theta) \right\} f_p(\theta) d\theta \end{aligned} \quad (19)$$

where $F_p(\bar{\theta}) - F_p(\theta_0)$ and $F_p(\bar{\theta}) - 2F_p(\theta_0) + F_p(\theta_1)$ are, respectively, the number of cars that enter the market in each period at cost c and the number of cars scrapped in each period for a value v , and the second and third lines capture the social gain from driving new and old cars, respectively. In the absence of intervention, $W = W^n > 0$.⁴⁰

The first-best solution to the planner's problem is characterized by the rental cutoffs

$$\theta_0^* = \frac{1}{s_0} \left[\frac{c - 2v}{\kappa_0 - \kappa_1 \varsigma^\alpha} \right]^{1/\alpha} \quad \text{and} \quad \theta_1^* = \frac{1}{\varsigma s_0} \left[\frac{v}{\kappa_1} \right]^{1/\alpha} \quad (20)$$

where $\kappa_a = (\alpha - 1)^{-1}(\psi + h_p e_a)^{1-\alpha}$ for $a \in \{0, 1\}$, and the mileage schedule $x_a^*(\theta) = [\theta s_a / (\psi + h_p e_a)]^\alpha$. The welfare gain from implementing the first-best is $W^* - W^n$.

Proposition 1. *In a hypothetical scenario where $e_a x_a(\theta)$ is observed, a planner can implement the first-best with Pigouvian taxes equal to $h_p e_a x_a(\theta)$ for $a \in \{0, 1\}$.*

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. Even if the first-best is beyond the planner's reach, it serves as a benchmark to guide the design and choice of alternative instruments.

Proposition 2. *It is socially optimal to increase the number of new vehicles in the polluted area, i.e., $\theta_0^* < \theta_0^n$, if*

$$\Lambda(e_0, e_1, \alpha, \varsigma) \equiv \frac{1 - (1 + h_p e_1 / \psi)^{1-\alpha}}{1 - (1 + h_p e_0 / \psi)^{1-\alpha}} \varsigma^\alpha > 1 \quad (21)$$

⁴⁰This 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 30, the exercise in the online Appendix (Section A.2, Table A.3) suggests that trade across regions is fairly active.

Proof. From (18) and (20), it is immediate that $\theta_0^* < \theta_0^n$ implies (21). \square

$\Lambda(e_0, e_1, \alpha, \varsigma)$ captures the impact that different parameters have on the social need to accelerate fleet turnover toward cleaner vehicles in the polluted area. $\Lambda(\cdot)$ is likely to be greater than 1 when (i) emission rates grow rapidly with age (i.e., $e_0 \ll e_1$), (ii) quality drops slowly with age (i.e., ς not too far from unity), and (iii) the demand for driving is not too elastic (i.e., α not too large).⁴¹ While it is intuitive that $\Lambda(\cdot)$ goes up as new cars become relatively cleaner (note that (21) holds easily as $e_0 \rightarrow 0$ and never as $e_0 \rightarrow e_1$), it is less intuitive that it does so as ς increases and/or as α drops. In these two latter cases, new cars become less attractive relative to old cars either because the quality gap shrinks or because new and old cars are used less often, so any quality difference becomes less important.

The following set of results shows that policies acting primarily on the extensive margin (i.e., rental decisions) can be particularly attractive when (21) holds, which is what we find in our estimation (see section 4.2).

Proposition 3. *Suppose (21) holds. Then, the optimal driving restriction is vintage specific: $R_0 = 1$, $R_1 \leq 1$ and $R_1[\kappa(\theta_1^{vr} s_1)^\alpha - h_p e_1(\theta_1^{vr} s_1/\psi)^\alpha] \geq v$, where “vr” denotes a vintage-specific driving restriction.*

Proof. See online Appendix, Section C. \square

In this two-period world a driving restriction may not only impose restrictions on new and old models (R_0 and R_1 , respectively) but may also establish the number of old cars that can circulate in the market. Condition $R_1[\kappa(\theta_1^{vr} s_1)^\alpha - h_p e_1(\theta_1^{vr} s_1/\psi)^\alpha] \geq v$ says that the last car to circulate (and the last household θ_1^{vr} to rent a car) must generate at least a social-welfare gain equal to its outside option, v . Here, in the absence of any heterogeneity across old cars, this condition is satisfied by setting θ_1^{vr} directly or indirectly through R_1 , but in the more general model, with a very large number of vintages, this condition is satisfied with equality by imposing a total ban on any vintage that, given its prospective driver, reports a social-welfare gain lower than v in equilibrium.

Unlike a uniform driving restriction, which acts primarily on the intensive margin by imposing the same restriction on all cars, the optimal (vintage-specific) restriction acts exclusively on the extensive margin. The reason R_1 may be less than 1 is only to induce more drivers to adopt new models at the cost of destroying some socially valuable trips by drivers of old cars (recall that in the optimal design $\kappa(\theta_1^{vr} s_1)^\alpha - h_p e_1(\theta_1^{vr} s_1/\psi)^\alpha \geq v$). In principle, the loss of valuable trips in older vehicles may also be part of the optimal design in the more general model, but again, only for affecting rental decisions.

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 and (21) holds easily. This is

⁴¹Note that $\lim_{\alpha \downarrow 1} \Lambda(\cdot) = \varsigma \ln(1 + e_1 h/\psi) / \ln(1 + e_0 h/\psi)$ and $\lim_{\alpha \uparrow \infty} \Lambda(\cdot) = \varsigma^\alpha = 0$.

when an optimal vintage-specific restriction becomes particularly valuable, as established in Proposition 3. In contrast, a uniform restriction can only destroy welfare in this case, 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 $R < 1$ not only destroys valuable trips in new cars, but, for that same reason it also discourages drivers from renting these cars, resulting in $\theta_0^{ur} > \theta_0^n$, where “ur” denotes a uniform restriction, which goes against the recommendation in Proposition 2.

As the optimal vintage restriction works exclusively through the extensive margin, we can now compare it to instruments that work through the same margin since this only requires us to look at cutoffs θ_0 and θ_1 .

Proposition 4. $W^* > W^f \geq \max\{W^s, W^{vr}\}$, where W^s and W^f are the social welfare levels under an optimal scrappage subsidy Δ and optimal registration fees r_0 and r_1 , respectively.

Proof. First, $W^* > W^f$ because any registration fee fails to correct for car use. Second, to demonstrate that $W^f \geq \max\{W^s, W^{vr}\}$ notice that r_0 and r_1 can be freely used to set $\theta_0^f(r_1, r_0) = [(c - 2v + r_0 - r_1)/\kappa(1 - \zeta^\alpha)]^{1/\alpha} / s_0$ and $\theta_1^f(r_1) = [(v + r_1)/\kappa]^{1/\alpha} / s_1$ at their optimal levels; in particular, r_1 is set such that $v = [\kappa(\theta_1^f s_1)^\alpha - h_p e_1 (\theta_1^f s_1 / \psi)^\alpha]$ holds. This latter condition is also satisfied in the optimal vintage restriction from setting $\theta_1^{vr} = \theta_1^f$ (see Proposition 3) and is satisfied in the scrappage-subsidy design when setting $\Delta = \Delta'$ so that $\theta_1^s = [(v + \Delta')/\kappa]^{1/\alpha} / s_1 = \theta_1^f$. Notice also from $\theta_0^{vr}(R_0, R_1) = [(c - 2v)/\kappa(R_0 - \zeta^\alpha R_1)]^{1/\alpha} / s_0$ and $\theta_0^s(\Delta) = [(c - 2v - 2\Delta)/\kappa(1 - \zeta^\alpha)]^{1/\alpha} / s_0$ that $\theta_0^{vr}(R_0 = R_1 = 1)$ and $\theta_0^s(\Delta = \Delta')$ do not necessarily coincide with θ_0^f , unless some socially costly distortions are introduced. The subsidy Δ may need to go above Δ' in order to reduce θ_0^s at the cost of reducing 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. \square

The reason we cannot precisely rank the scrappage subsidy and the vintage restriction is because their designs include distortions of different natures: the former reduces the number of old cars in the market and the latter reduces their use.⁴² Comparing an optimal vintage restriction to an optimal gasoline tax is also difficult because gasoline taxes act on extensive as well as intensive margins, both of which are important for implementing the first best. The next proposition illustrates with two cases that the relative advantage of one instrument over the other is, in the end, an empirical matter.

Proposition 5. *If $e_0 = e_1 > 0$, then a gasoline tax $g = h_p e_a$ implements the first best and $W^g = W^* > W^f \geq \max\{W^s, W^{vr}\}$. 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 either registration fees or a vintage-specific restriction could be used to implement the first best and $W^g < W^* = W^f = W^{vr} \geq W^s$.*

⁴²In the simulations we consider two additional reasons that improve the vintage-restriction’s performance over the subsidy’s: the absence of arbitrage and the possibility of exporting old vehicles to the non-polluted area.

Proof. The proof for the first case (i.e., $e_0 = e_1 > 0$) is immediate from Propositions 1 and 4. For the second case (i.e., $e_0 = 0$ and $e_1 \geq \underline{e}_1$), notice that it is first-best optimal to ban old cars entirely and let new cars run freely, so $\kappa(\theta_0^* s_0) = c - v$. This can be achieved by setting either $R_0 = 1$ and $R_1 \leq \underline{R}_1$ or $r_0 = 0$ and $r_1 \geq \underline{r}_1$, where \underline{R}_1 solves $\theta_1^{vr}(R_1) = [v/\kappa R_1]^{1/\alpha} / s_1 = \theta_0^*$ and \underline{r}_1 solves $\theta_1^f(r_1) = \theta_0^*$. Also, $W^* \geq W^s$ because the subsidy Δ that solves $\theta_1^s(\Delta) = \theta_1^*$ does not necessarily deliver $\theta_0^s(\Delta) = \theta_0^*$, except in the particular case that $c = (1 + \zeta^\alpha)(c - v)$, which clearly does not hold when $\zeta \approx 1$ since $c - 2v > 0$ by assumption. \square

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. This is not a problem when $e_0 \approx e_1$ since each mile driven is equally harmful. But as $e_0/e_1 \rightarrow 0$, so that (21) holds, the relative advantage of a gasoline tax is greatly diminished. An important reason for this is the decline in the number of new cars in the market, i.e., $\theta_0^g > \theta_0^n$, as can be seen by looking at (18) and

$$\theta_0^g = \frac{1}{s_0} \left[\frac{c - 2v}{\kappa^g (1 - \zeta^\alpha)} \right]^{1/\alpha}$$

where $\kappa^g = (\alpha - 1)^{-1}(\psi + g)^{1-\alpha}$. Furthermore, if $e_0 = 0$ and $\underline{e}_1 \leq e_1 \leq \psi/\alpha h_p$, the optimal tax is $g = 0$ despite it is first-best to get rid of all old cars. In this case, the (marginal) private cost of imposing a fuel tax on old-car drivers (i.e., $1/\psi^\alpha$) is greater than its (marginal) environmental benefit (i.e., $h_p e_1 \alpha / \psi^{\alpha+1}$). As demand for driving becomes more elastic (higher α), the environmental benefit becomes relatively more important than the private cost since the last trip demanded is equally damaging from an environmental perspective but increasingly less beneficial from a private perspective.

While this two-period model has served to illustrate key factors at work in policy design, particularly, e_0/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 changes 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, the optimal scrappage subsidy is shown, for instance, to do strictly worse than the optimal vintage restriction.

Another assumption is that all cars, regardless of vintage, share the same fuel efficiency, so their cost ψ is independent of the emission rate e_a , giving rise to a negative correlation between annual mileage x_a and e_a . Since what matters for policy is the relative cost of operation across vehicles of different vintage, the way we control in the simulations for the fact that, on average, new models may be more fuel efficient than older models is by permitting gasoline taxes to enter somewhat differentiated across vintages, i.e., $g_a < g_{a+1}$. Because this fuel-efficiency correction is barely enough to correct for (large) differences in emission rates, simulations show that the gasoline tax does not improve welfare much relative to the other instruments.

It is also assumed that vehicles 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 since their variation is a main force behind our results. 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. In terms of our model, this new source of variation could be captured by introducing within each vintage some variation of the form $e_a^j = e_a + \epsilon_a^j$, where j identifies a particular car and $E[\epsilon_a^j] = 0$. 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 ϵ_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,⁴³ 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 ϵ_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 (imagine the case in which on average there is no variation at the vintage level, i.e., $E[e_1^j] = E[e_0^j]$). And rather than separating cars by vintage, they are now to be separated by emission rates, say, by some combination of vintage and model.⁴⁴

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

⁴³Either 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](#)).

⁴⁴Congestion 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 4) remain invariant, however, since ψ 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).

values for the different parameters that enter into the more general model of sections 3.1 and 3.2. Parameters related to individual preferences are obtained from matching the model predictions to actual fleet compositions observed in Santiago (the restricted or polluted area) and the rest of the country (the non-restricted or non-polluted area) following the 1992 reform. Parameters related to costs and scrappage values are taken directly from the data. Finally, to estimate emission rates, we exploit the smog-check dataset described in Section 2.3.

4.1. Households' characteristics, car parameters, and policy response

Four important inputs are key to understanding consumer preferences in our model: (i) the extent to which households exhibit decreasing returns on driving ($\alpha > 1$), (ii) cost per kilometer driven (ψ), (iii) perceived car quality (s_0, ς), and (iv) the cdf $F(\theta)$ of consumers' marginal valuation for quality $\theta \geq 0$. Neither input is directly observable, similar to the actual response to the policy, which is assumed to be the same over all pre-1993 models, that is, $R_a = R < 1$ for all vintage $\tau \leq 1992$. These four inputs along the policy response are obtained by matching the model's predictions to car allocations observed in the circulation permits database for year 2006. Throughout our estimation we follow [Berry and Pakes \(2007\)](#) framework to estimate vertical differentiation models.

We adapt our model to accommodate the existing data in several ways. First, while the 92-93 discontinuity introduces a clear partition in car quality for those two adjacent vintages, drivers more generally tend to regard cars of slightly different vintages to be of similar quality. We address this quality overlap in a simple way by clustering car vintages in six vintage/quality groups centered around the 92-93 discontinuity: 1981-84, 85-88, 89-92, 93-96, 97-2000, and 2001-04.⁴⁵ This grouping essentially assumes that people trade their cars every four years.

Second, we reduce the problem dimensionality from more than 300 municipalities to 60 electoral districts. Electoral districts group municipalities that are located in the same geographic areas and therefore share similar characteristics, most importantly, income. Since the country's population is normalized to 1, our relevant car-holding variable q_a^i becomes the fraction of cars of age-group $a = 0, \dots, 5$ in district $i = 1, \dots, 60$ relative to the district's number of households.

Third, we add a component to the utility function for factors not observed by the econometrician that affect demand. Without this, it would be impossible to find parameter values that make implications of the model consistent with the data ([Berry and Pakes, 2007](#)). We let the utility of a driver θ in district i who rents an a -year-old car that faces an effective restriction of R_a^i be:

$$u(\theta, i, a, R_a^i) = R_a^i \kappa (\theta s_a)^\alpha - p_a + \varepsilon_a^i \quad (22)$$

where $\kappa = [(\alpha - 1)\psi^{\alpha-1}]^{-1}$. The only difference from (17) is the extra term ε_a^i that shifts

⁴⁵Model years 1980 and earlier, which in any case are very few, are grouped with 1981 models.

the demand for cars of age a in district i . We assume ε_a^i to be orthogonal to observable district characteristics such as income, urbanization rate, distance to Santiago, and whether the district is subject to the driving restriction or not.

The rental price of an a -year-old car, p_a , is obtained from our car prices database described in Section 2.2, where the (annual) rental price of a given car is assumed to be the difference between its purchasing price today and next year (all our numbers are in 2006 U.S. dollars).⁴⁶ Following Gavazza et al. (2014), cost ψ is assumed to be invariant to location and age. From (10) and (12), ψ can be expressed as:

$$\psi = \frac{1}{x(\theta_0)} \frac{(p_0 - p_1)(\alpha - 1)}{1 - \varsigma^\alpha}$$

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 46,827 miles, which, according to the smog-check database, is the average travel during the first four years of a car's life in that region.

We also need an estimate of $F(\cdot)$. Instead of imposing a restrictive functional form, we use our data and district characteristics to estimate the function in a semi-parametric way. We let $F(\cdot|z_i)$ be a cubic in θ with each coefficient in the cubic (b^1 , b^2 , and b^3) varying across districts $i = 1, \dots, 60$ according to the linear function $b_i^j = \zeta_0^j + z_i' \zeta^j$, where $j = 1, 2, 3$ denotes the coefficient in the cubic, ζ_0^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 $F(\cdot|z_i)$ is characterized by 12 ζ^j parameters (four for each of the three coefficients in the cubic) that need to be estimated along with α , s_0 , ς , and R_a .

According to our model, we should observe cutoff levels θ_a^i that mark drivers' indifference between driving an a -year-old car and an $a + 1$ -year-old car. These cutoffs can be obtained from (17) and the equilibrium condition (12), and are given by:

$$\theta_a^i = \left(\frac{p_{a+1} - p_a + \varepsilon_{a+1}^i - \varepsilon_a^i}{R_{a+1}^i \kappa s_{a+1}^\alpha - R_a^i \kappa s_a^\alpha} \right)^{\frac{1}{\alpha}} \quad (24)$$

where $R_a^i = 1$ for all age- a cars registered in a "non-restricted" district (i.e., outside Santiago), and $R_a^i = R < 1$ for $a > 3$ and $R_a^i = 1$ for $a \leq 3$ for cars registered in a "restricted" district (i.e., in Santiago). Note that since $s_a = \varsigma^a s_0$ enters multiplicatively in (24), we cannot separately

⁴⁶More precisely, we exploit the no-arbitrage condition given by:

$$p_{at} = P_{at} - \delta P_{a+1,t} \quad (23)$$

where P_a is the price of an a -year-old car 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, we run an OLS regression of $\ln(P_{imat})$ on a constant and year, model, and age fixed effects to predict \hat{P}_{imat} . With these predictions and (23), we obtain (weighted average) rental prices for each of the six vintage groups identified above.

obtain estimates for s_0 and the equilibrium cutoffs θ_a 's. Hence, we normalize $s_0 = 10$.⁴⁷

The estimation procedure is simple. For each set of parameters α , ς , and R_a , residuals $\Delta\varepsilon_a^i = \varepsilon_{a+1}^i - \varepsilon_a^i$ and prices p_a we can find cutoff levels θ_a^i for each district. Moreover, values of ζ^j will produce distributions $F(\cdot|z_i)$ that need to match observed data on quantities such that $F(\theta_a^i|z_i) - F(\theta_{a+1}^i|z_i) = q_a^i$. Since we assume that ε_a^i is orthogonal to observable district characteristics, using two-step GMM, we look for values of α , ς , R_a , ζ^j , and $\Delta\varepsilon_a^i$ that minimize the following five moments for each age-group a :⁴⁸ $\sum_{i=1}^{60} \Delta\varepsilon_a^i = 0$, $\sum_{i=1}^{60} INCOME_i \times \Delta\varepsilon_a^i = 0$, $\sum_{i=1}^{60} DISTANCE_i \times \Delta\varepsilon_a^i = 0$, $\sum_{i=1}^{60} URBANIZATION_i \times \Delta\varepsilon_a^i = 0$, and $\sum_{i=1}^{60} DR_i \times \Delta\varepsilon_a^i = 0$.

It is important to point out the rationale behind the identification assumptions. The orthogonality assumption between ε_a^i and observable district characteristics z_i comes from the fact that higher demand for some models in some districts, due to higher income, for example, should be captured by differences in $F(\cdot|z_i)$. Any differences not explained by $F(\cdot|z_i)$, and therefore not related to z_i , are attributed to ε_a^i . Note also that the moment conditions include orthogonality to whether the district is located in the restricted area ($DR_i = 1$) or not ($DR_i = 0$), which is excluded from z_i . This moment helps to identify R_a . Including DR_i in z_i would allow $F(\cdot|z_i)$ to absorb the differences between restricted and non-restricted cities around the policy threshold that should be captured by R_a .

Figure 4 conveys some intuition for our estimation method by showing the cutoff levels θ_a^i , the estimated function $F(\cdot|z_i)$, and the data 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 is located out of the restricted area and corresponds to a city of lower income per capita. On the horizontal axis, we mark with small circles the values of θ_a^i , which are constructed with actual prices, the optimal values of α , ς , and R_a , and the residuals $\Delta\varepsilon_a^i$. Note that the distance between θ_1 and θ_2 is relatively large for district 22 and small for district 32. On the other hand, 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 (24) for district 22, but not for district 32. The thick black line that crosses each box corresponds to $F(\theta|z_i)$, which is constructed with the optimal values of ζ^j . Because district 22 has a larger 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 . Finally, the vertical axis shows the values for q_a^i observed in the data, such that our equilibrium condition $F(\theta_a^i|z_i) - F(\theta_{a+1}^i|z_i) = q_a^i$ holds.

The resulting estimated parameters are presented in panel (a) of Table 3. Note that the actual policy intensity in 2006, $R = 0.967$, is weaker than one would expect in the absence of intertemporal substitution ($R = 0.918$).⁴⁹ This is not surprising as pre-1993 models

⁴⁷Note that $\kappa = [(\alpha - 1)\psi^{\alpha-1}]^{-1}$, which allows us to separately identify α , ς and the distribution of θ .

⁴⁸This gives us a total of 30 moments to estimate 15 parameters.

⁴⁹See footnote 28 in section 2.3.

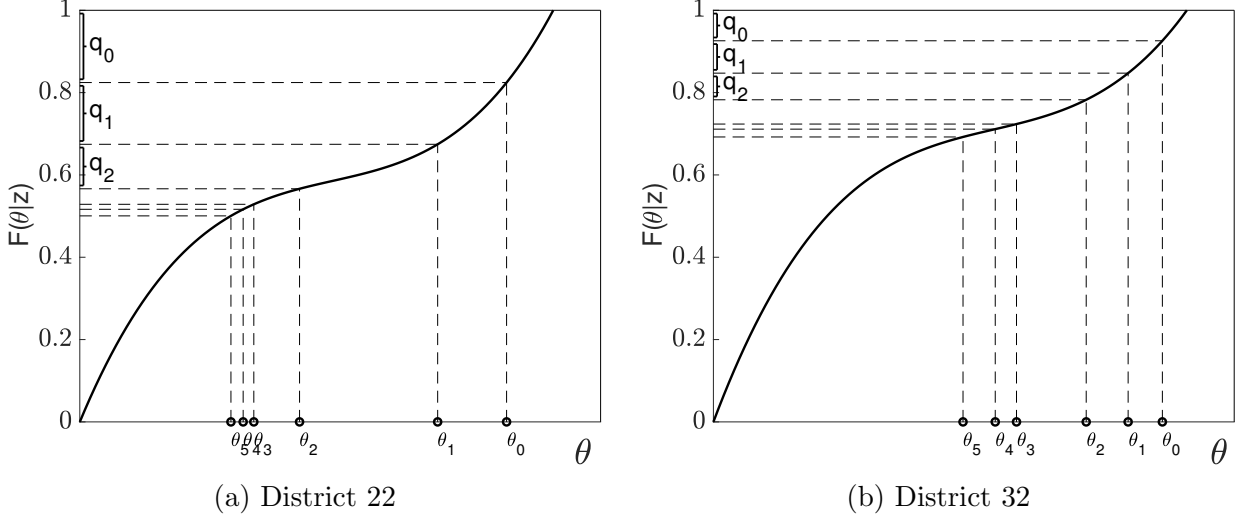


Figure 4: Distribution of θ and number of cars in districts 22 and 32

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.

were run much less often in 2006 than when the policy was enacted 13 years earlier, leaving room for intertemporal substitution. Other values in the table are similar to those used in previous studies. The value we obtain for α leads to a concave utility function $u(\theta, a, x)$, where $(\alpha - 1)/\alpha = 0.502$, which is not that different from the logarithmic utility in [Gavazza et al. \(2014\)](#). The value for the decay rate ζ is also very close to the value they use.⁵⁰ Our estimated cost ψ , however, is twice their number, partly explained by our higher gas prices.⁵¹

The remaining parameters are the scrappage value of cars v , new car price c , and survival rates γ_a . Based on informal conversations with car dealers, we set $v = \$600$, the lowest trade-in value some of them recall having seen in recent years (we do not see prices this low in our sample of newspaper ads, however). We use our car prices database to obtain the (weighted average) price of a new car (i.e., $a = 0$), which we set at $c = \$16,000$. Since importing used cars is forbidden, we estimate survival rates, γ_a , directly from stock changes observed in the circulation-permit data from 2006 through 2012. 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

⁵⁰Their annual decay rate is 0.976 while ours is $0.8764^{1/4} = 0.9681$.

⁵¹We check the validity of our estimation by running an out-of-sample exercise that contrasts model predictions for 2012 with the empirical estimation obtained from regressing (2) on the 2012 circulation-permit data. As explained in the online Appendix (section [D.1](#)), the model captures reasonably well the policy effects on fleet composition both around the 92-93 discontinuity and before. It fails, however, to capture the larger fraction of newer cars in Santiago relative to the rest of the country. One reason for this latter is 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](#)).

Table 3: Parameters

Parameter	Value	Parameter	Value	Parameter	Value
Panel (a): Households' characteristics and car parameters					
α	2.007 (0.0034)	ψ	0.3161 (0.0320)	ς	0.8784 (0.0006)
s_0	10	c	16000	v	600
Panel (b): Pollution parameters					
ω	0.8479	h_r	0.0096	h_{nr}	0.0011

Notes: Estimated parameters of the model using data sources including circulation permits and smog checks. See the text for more details. Standard errors allow for autocorrelated unobserved demand shocks within districts. The standard error of ψ is estimated with the delta method.

for cars with ages ranging from 0 to 36 years. Averaging these numbers at our vintage-group level leads to the survival numbers γ_a in the online Appendix (Section A.2, Table A.4).

4.2. Pollution parameters

The model's pollution-related parameters are the external cost of local pollution in area $k \in \{r, nr\}$, h_k , and the function e_a that relates age a to emission rate e . We combine two information sources to estimate both parameters simultaneously. The first source is Parry and Strand (2012), which contains specific estimates of vehicle external costs associated with emissions of local pollutants for various cities in the country. Their damage estimate (in 2006 U.S. dollars) for an average vehicle in Santiago is ¢2 per mile; that for an average vehicle in the rest of the country is ¢0.23 per mile.⁵²

The second information source is the smog-checks dataset described in Section 2.3. Figure 3 from that section plots average smog-check readings of CO and HC as a function of age for year 2008. One concern that might stem from using the 2008 data to obtain e_a is the “presence” of a technology —absence of converter— that is no longer in the current fleet but can be found in a handful of very old units. As we are interested in the design of policies for handling existing pollution problems, our estimation of e_a is therefore based instead on CO readings from the most recent data, that is, the 2016 sample, covering a range of 24 years (equivalent to 6 age intervals in our model), with no pre-1993 models in it.^{53,54}

Some possible concerns with our estimation of e_a are worth indicating. One is the extent to

⁵²Differences in marginal damage rates are driven by higher emission levels in Santiago, which are mostly explained by higher population density and topographical conditions favorable to pollution formation.

⁵³We nevertheless use the 2008 sample in the online Appendix (Section E.1) to provide a more comprehensive evaluation of the 1992 policy to complement the evaluation carried out in section 2.3.

⁵⁴Note that we only use CO readings under an engine speed of 2500 rpm for our estimations of e_a . Since what matters is relative emission rates, results barely change if we use other pollutants (e.g., HC) and/or engine speeds (e.g, idle speed).

which existing 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, h_r , increases with time (or income) as technology improves. 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 3 and 5).

Another concern 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, 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.⁵⁵

The estimation procedure to obtain h_k and e_a is as follows. First, we use existing data to calculate the average external damage that an a -year-old car generates in a given time period. Second, we use the model to calculate the average external damage of an a -year-old car as a function of the model parameters. Finally, we find the parameters that minimize the distance between the data- and model-driven damage figures.

Let ϱ_a be the average emissions rate of an a -year-old car that we observe in the smog-check readings of 2016, x_a the average miles traveled by an a -year-old car according to the 2015 and 2016 odometer readings contained in the smog-check dataset (see section 2.3), q_a^k the number of a -year-old cars in area k in 2006 as reported in the circulation-permit dataset, and d_k the average external cost per mile traveled in area k as reported in Parry and Strand (2012). Using that total damage in area k is $d_k \sum_a x_a q_a^k = h_k \sum_a \varrho_a x_a q_a^k$, we can express the average damage imposed on area k by an a -year-old car during a given period as:

$$D_a^k = h_k \varrho_a x_a = \frac{d_k \sum_a x_a q_a^k}{\sum_a \varrho_a x_a q_a^k} \varrho_a x_a$$

⁵⁵A 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.

We can do an analogous calculation using our model:

$$D_a^{k'} = \int_{\theta_a^k}^{\theta_{a-1}^k} h_k e_a x_a(\theta) dF_k(\theta)$$

where θ_a^k corresponds to the equilibrium cutoff level of a driver in area k indifferent to driving a car of age a or a car of age $a + 1$, and $x_a(\theta)$ is the amount of travel according to the model. Assuming that emissions increase with age at an exponential rate, i.e., $e_a = \exp(\omega a)$, we search for values of h_r , h_{nr} , and ω that minimize the loss function $L = \sum_a [q_a^r (D_a^r - D_a^{r'})^2 + q_a^{nr} (D_a^{nr} - D_a^{nr'})^2]$.⁵⁶ Results for Santiago and the rest of the country are presented in panel (b) of Table 3.

Before turning to the simulations, we can use these latter results and those in panel (a) of Table 3 to quickly check whether expression (21) in Proposition 2 holds or not. This is important for policy design (Proposition 3) and choice (Proposition 5). For example, if in that expression we let “new” cars correspond to cars in age-group 0 (i.e., 4 years old or less) and “old” cars correspond to cars in age-group 3 (i.e., 12-16 years old), we obtain $\Lambda(\cdot) = 4.33$. If, instead, we let new cars be those in age-group 1 (4-8 years old) and old cars be those in age-group 2 (8-12 years old), we obtain $\Lambda(\cdot) = 1.65$. Any value of $\Lambda(\cdot)$ within this range suggests that “extensive-margin” instruments such as vintage-specific restrictions can be expected to perform comparably well in the simulations that follow.

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 exhibits a fleet newer 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 18% more of the newest models (0 to 4 years: vintage-group 0 in our simulations) than the rest of the country.⁵⁷

⁵⁶We present the fit of the model to the data in the online Appendix (Section A.1, Figure A.5(a)). Note that a similar fit is obtained if, in the loss function, q_a^k is taken to be the number of cars predicted by the model.

⁵⁷Income enters into the model through $F(\theta|z_i)$.

Santiago also has fewer older models.⁵⁸ Notice that in both locations, we find cars running until scrapped, somewhere between 24 and 28 years old.⁵⁹ This will change as the government intervenes in the market.

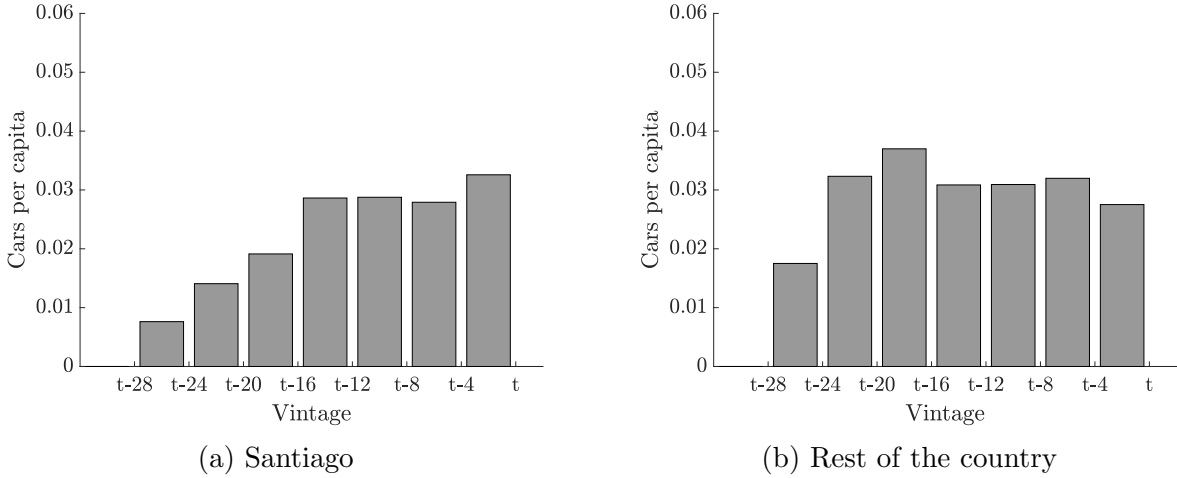


Figure 5: Steady-state fleet composition under no intervention

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 outcome 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:⁶⁰

$$w_t = -cq_{0,t} + vq_{T(t),t} + \sum_{k=p,np} \int_{\theta_{T(t),t}^k}^{\bar{\theta}^k} [u_t^k(\theta) - h_k e_t^k(\theta) x_t^k(\theta)] dF_k(\theta)$$

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 $T(t)$); $\theta_{T(t),t}^k$ is the last household to rent a car in area k during period t (see (15)); and $u_t^k(\theta)$, $e_t^k(\theta)$, and $x_t^k(\theta)$ are, respectively, a household θ ’s utility, emissions per mile,

⁵⁸If, 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 $1.78 = 0.64/0.36$.

⁵⁹Not surprisingly, this age of scrappage differs from that in the estimation (i.e., 20-24 years old). The equilibrium in the estimation is subject to intervention and is not a steady state.

⁶⁰Note that policy intervention affects the value of the existing stock of used vehicles by altering future rental prices. Any (unanticipated) change this may cause in dealers’ revenues is not in our estimation.

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 (10) and (11) with (16) and (17).

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 h_k upon cars circulating in region k is dramatic. Over the long run, households in Santiago have no incentive to hold cars older than 20 years—an 8-year reduction compared to the no-intervention case. While sales of new cars in Santiago increase by 37%, 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.

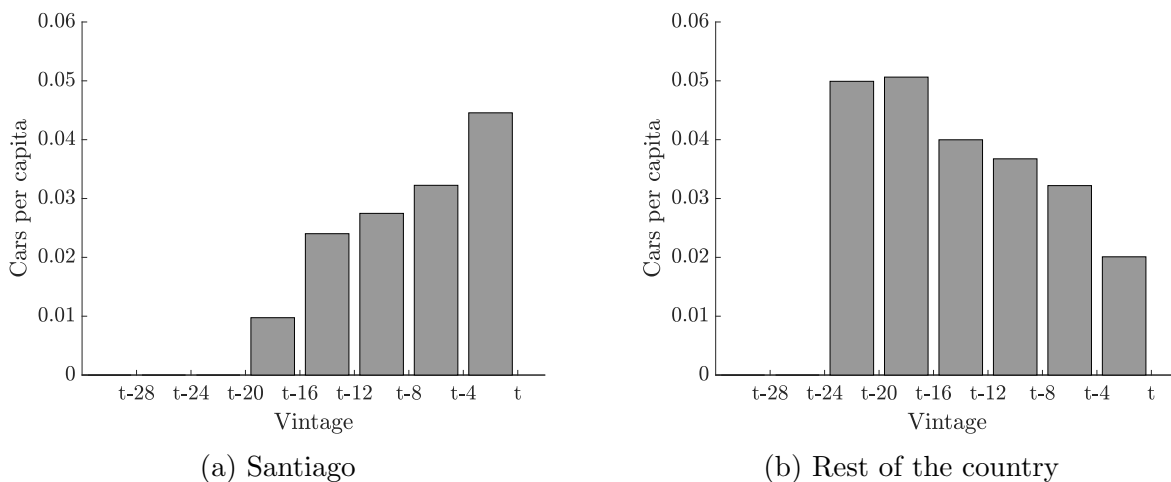


Figure 6: Steady-state fleet composition under the first best

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).⁶¹ This non-monotonicity introduces computational challenges for determining the equilibrium dynamics, particularly when we need to find the optimal policy intervention.

In present-value terms, the welfare gain of moving from no intervention to the first-best amounts to \$265.4 per household: a 5.8% gain from the no-intervention baseline of \$4581.7 (see first two rows of Table 4). The drop in pollution costs (\$365.6) is almost four times larger than the loss in transport surplus (\$100.2). At the country level, this net welfare gain adds a total of \$1.1 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	5286.9	-705.2	4581.7	0%
2.	First best	5186.7	-339.6	4847.1	100%
3.	Driving restriction with no exemptions ($R = 0.9 \forall \tau$)	4905.5	-711.8	4193.7	-146%
4.	Driving restriction with some exemptions ($R = 0.9, a > 3$)	5247.4	-656.3	4591.0	4%
5.	Optimal driving restriction ($R = 0, a > 4$)	5252.4	-515.8	4736.5	58%
6.	Scrappage subsidy (\$1820), full arbitrage	5202.7	-506.8	4695.8	43%
7.	Scrappage subsidy (\$2332, \$350), no arbitrage	5198.1	-481.9	4716.1	51%
8.	Circulation fees	5139.0	-385.2	4753.8	65%
9.	Gasoline tax (¢19.5 per gallon)	5276.8	-644.6	4632.2	19%
10.	Gasoline tax with efficiency correction (¢22 per gallon)	5272.8	-607.4	4665.4	32%

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

⁶¹We illustrate the dynamics of new car sales (q_0) for both Santiago and the rest of the country in the online Appendix (Section A.1, Figure A.3).

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. For simplicity, however, we focus on time-invariant policies in what follows.⁶²

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 is 146% higher than the 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 4% 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

⁶²The non-monotonic dynamics described above and illustrated in the online Appendix (Section A.1, Figure A.3) make the computation of time-varying (optimal) policies quite demanding. We nevertheless attempted some departures from and around the time-invariant (optimal) design. The additional welfare gains achieved do not qualitatively change any results that follow.

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 pollution costs.

As suggested by Proposition 3, and following the “complete-ban” schemes seen in Paris and Germany’s LEZs, it turns out that the optimal vintage-specific design imposes a total circulation ban on vehicles that are 16 or more years of age and extends a full exemption on newer vehicles.⁶³ As indicated in row 5 of Table 4, the resulting welfare gain is significant. Figure 7 can help explain this result. An optimal driving restriction 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 (24–28 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.

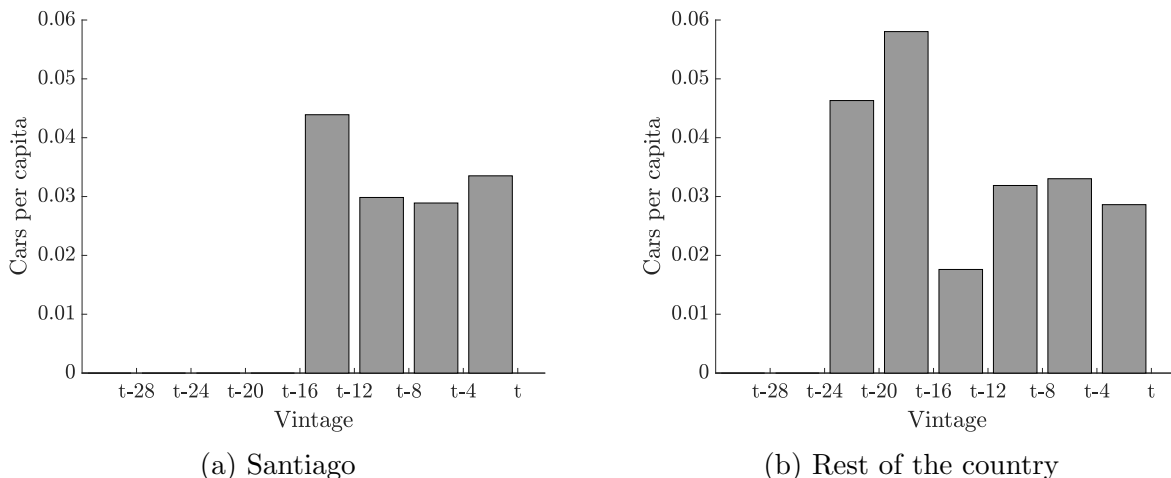


Figure 7: Steady-state fleet composition under the optimal driving restriction

Notes: The figure shows the steady-state profile of car fleets in Santiago and the rest of the country under the optimal (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.

⁶³Following Proposition 3, we explored welfare changes to introduce mild restrictions on 16 year-old and younger models that could encourage wider adoption of newer models. Finding virtually no benefit, we chose to stay with the simple design separating cars into just two categories. In addition, this separation eliminates altogether any possibility of buying an older and partially-restricted car to bypass a partial restriction on another car.

Before we move on to compare the optimal vintage-specific design to optimal alternative instruments, it is worth highlighting a few observations. First is the significant welfare improvement that results from moving from the vintage-specific design of row 4 of Table 4 to the optimal 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, the optimal 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, the optimal design may raise distributional concerns. Almost all drivers and public-transport users in Santiago are better off under the optimal driving restriction 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 key role played by the shape of the emissions-age relationship in our results. As established in Proposition 5, the faster emission rates increase with age, the more to gain from a policy aimed at the extensive margin. In fact, if we add convexity to the emissions-age relationship by increasing ω from its current value of 0.85 to 1, the optimal driving restriction performs significantly better, delivering 79% of the first-best gain. On the other hand, if we suppress some convexity by letting ω drop to 0.7, the optimal restriction performs worse, but still delivers 48% of the first-best gain.⁶⁴

This comparative-static exercise takes to our last observation, that of 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

⁶⁴To facilitate the comparison with the numbers in Table 4, the values of h_r and h_{nr} in these comparative-static exercises were adjusted so welfare under no-intervention would remain unchanged at its \$4582 level. In the first exercise, for example, h_r and h_{nr} changed to 0.0058 and 0.0007, respectively, so welfare under the first-best increased from \$4847 to \$4862, and welfare under the optimal driving restriction increased from \$4736 to \$4804.

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 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 \$1820 to \$2332, 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 optimal (vintage-specific) driving restrictions. The reason is simple: both instruments seek to affect fleet turnover by aiming to remove the most polluting cars; one with a prohibition, the other with a reward.⁶⁵ If anything, it appears that implementation constraints should favor the use of vintage-specific restrictions over scrappage subsidies, not only because of the arbitrage issue just described, but also because restrictions are much cheaper to implement for the government under any reasonable estimate of the shadow cost of public funds (see, for example, [Laffont, 2005](#)). 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)

⁶⁵As the restricted area becomes much larger than the non-restricted area so the “subsidy arbitrage” becomes less of a problem, welfare under the optimal scrappage subsidy approaches that under the optimal driving restriction.

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 the optimal driving restriction (see row 8 in Table 4), provided these fees are set at their optimal levels, which roughly equal the expected Pigouvian bills. Interestingly, and like any other instrument, optimal 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 bad proxy for handling local pollution, even if optimally set at ¢19.5 per gallon (equivalent to a cost increase of ¢1.02 per mile). In fact, 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).⁶⁶ Since one may argue that newer cars are not only cleaner but also more fuel efficient than older cars, in row 10 of Table 4 we attempt a correction in this direction with a consequent increase in welfare.⁶⁷

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 strengthens 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.

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_p 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

⁶⁶A similar result, but based on U.S. data, is in Knittel and Sandler (2018).

⁶⁷Since our model abstracts from technical progress, the way we introduce this fuel-efficiency consideration is by artificially allowing the fuel economy of a car to deteriorate over time, enough to reflect the fuel economy differences observed in the 2016 smog-check data.

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 the (optimal) vintage-specific restriction reduces as λ drops and, according to the exercise in the online Appendix (Section E.2), completely disappears when $\lambda = 0.43$.

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 sort of policies can effectively help to accelerate fleet turnover towards lower-emitting vehicles. The optimal vintage-specific design takes the “complete-ban” structure already utilized in Paris and Germany’s low emission zones: an optimal vintage threshold separates cars between complete restriction and (nearly) 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.⁶⁸ The 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.

⁶⁸One 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.

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).
- Drummond, P. and P. Ekins (2016). Tackling air pollution from diesel cars through tax: options for the UK. Technical report, UCL.
- 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.
- 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*, 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*, 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.
- Laffont, J.-J. (2005). *Regulation and Development*. Cambridge University Press.
- 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. T., M. J. Molina, et al. (2002). *Air quality in the Mexico megacity: An integrated assessment*. SciELO Chile.
- 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*, Volume 373. World Bank Publications.
- 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*.
- 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).
- 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.