INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

469
2016

Adopting a Cleaner Technology: The Effect of Driving Restrictions on Fleet Turnover

Hernán Barahona, Francisco Gallego y Juan-Pablo Montero

www.economia.puc.cl • ISSN (edición impresa) 0716-7334 • ISSN (edición electrónica) 0717-7593

Adopting a cleaner technology: The effect of driving restrictions on fleet turnover

Hernán Barahona, Francisco Gallego and Juan-Pablo Montero*

February 26, 2016

Abstract

Driving restrictions –limits on car use based on the last digit of a car's license plate– are increasingly popular forms of pollution and congestion control, notwithstanding the literature has shown they typically result in more pollution by moving the fleet composition toward higher-emitting vehicles. We study a design feature present in some restriction programs but much overlooked in the literature: that cleaner cars be exempted from the restriction. Based on evidence from Santiago-Chile's 1992 program, we find this exemption feature to have a large impact on fleet composition toward cleaner vehicles. We also develop and calibrate for Santiago a vertical-differentiation model of the car market to show that driving restrictions that make optimal use of these exemptions can be way more effective in the fight against local air pollution than alternative instruments such as scrappage subsidies and gasoline taxes.

1 Introduction

Air pollution and congestion remain serious problems in many cities around the world because of the steady increase in car use. In an effort to contain such trend and persuade drivers to give up their cars in favor of public transport, authorities have tried different

*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, Juan Escudero, Marco González-Navarro, Larry Goulder, Gastón Illanes, Kelsey Jack, Shaun McRae, Mar Reguant, James Sallee, Brandon Shaufele and audiences at Berkeley, Chicago, EAERE-2015, Ecole Polytechnique, Harvard, LSE, Michigan, PUC-Chile, Toronto, Toulouse School of Economics, Stanford, Universidad de Montevideo and Universidade de Vigo. Gallego also thanks financial support from CONICYT/Programa de Investigación Asociativa (Project SOC 1102) and Montero from Fondecyt (Grant No. 1130998) and the ISCI Institute (Milenio P05-004F and Basal FBO-16).

policies, in particular, driving restrictions or limits on car use on the basis of some combination of the last digit of a vehicle's license plate and colored stickers displayed on its windshield. Driving restrictions have been adopted in various cities that with no exception, but for some adjustments, continue in operation as of today, including Athens (first introduced in 1982), Santiago-Chile (1986), Mexico-City (1989), São Paulo (1996), Manila (1996), Bogotá (1998), Medellín (2005), San José-Costa Rica (2005), Beijing (2008), Tianjin (2008), Berlin, Frankfurt, Munich and several other German cities (2008),[1] Quito (2010), Hangzhou (2011) and Chengdu (2012). The population of the metropolitan areas surrounding these cities add up to way above 100 million people, which suggests that the study of these regulations is relevant from a quantitative point of view.[2]

An obvious political economy question is why these driving restrictions are increasingly popular among environmental and transport authorities as opposed to alternative policies (on this latter see Fullerton and Gan 2005). We do not attempt to answer this question here. But given the popularity of these restrictions and the ongoing experimentation with them,[3] our goal is to try to understand how they actually work and see if there is room for improvement. We are not the first in this quest. There is already a series of papers looking at the "Hoy-No-Circula" or HNC program in Mexico-City. As implemented in 1989, some believe that HNC had a good start (e.g., Onursal and Gautam 1997; Gallego-Montero-Salas, GMS, 2013a), but most agree that over the longer term it lead to an increase in the number of vehicles on the road and in pollution levels (e.g., Eskeland and Feyzioglu, 1997; Onursal and Gautam 1997; Molina and Molina 2002; Davis 2008; GMS 2013a). In fact, Davis (2008) documents an increase of 20 percent in the car fleet due to HNC (the increase in GMS (2013b) is somewhat smaller but still important and quite fast, within a year).[4]

[1]The restriction programs in Germany are best known as low-emission zones or LEZs, which differ from the other listed programs in that they impose a total circulation ban, as opposed to a partial ban, to higher-emitting vehicles in a prestablished area of the city (see Wolff 2014). As we argue below, this distinction is only semantic, because in the design of an optimal restriction program the extent of the restriction should be treated as an endogneous variable.

[2]Unlike all these programs that run on a permanent basis, authorities in other big cities such as Brussels, London, Milano and Paris had also turned to driving restrictions to fight critical episodes of air pollution, but on a temporary basis.

[3]As documented, for example, in Davis (2015) for Mexico-City, Lin et al. (2014) for Bogota and Beijing, and this paper for Santiago, the implementation of these policies have shown a significant degree of experimentation.

[4]Lin et al. (2014) also failed to find air quality improvements from restrictions elsewhere, namely, Bogotá, São Paulo and Tianjin. They did find some in the restriction program implemented in Beijing during the Olympic games, and that was later decided to be extended indefinitely. This initial gain in air quality is confirmed by the recent work of Liu et al. (2015), but who in addition show that it rapidly disappeared over a period of a year, entirely consistent with the unfortunate pattern found by GMS (2013a) for HNC. Interestingly, the approach used by Liu et al (2015) is quite different from GMS

The take-away message from the existing literature is that while driving restrictions can lead to some pollution reduction in the very short run, over the long run they create perverse incentives for drivers to buy a second and older vehicle increasing the size of the car fleet and also moving the fleet composition toward higher-emitting vehicles. However, a crucial design aspect much overlooked in the literature —and that is present in later reforms in some restriction programs, including Mexico-City and Santiago— is that in some programs cleaner cars are exempted from the restriction. Santiago, in particular, reformed its restriction program in 1992 to allow all cars equipped with a catalytic converter, a technical device that transform toxic pollutants into less toxic gases and that became mandatory in all post-1992 models, be entitled to a green sticker indicating their exemption from the restriction.[5] Since this design feature provides drivers with a cleaner alternative to "by-pass" the restriction, the objective of this paper is to use evidence from Santiago's 1992 reform to study its effect on fleet turnover and the extent to which it can alleviate or eliminate altogether the previously documented "second-car" effect.[6]

Since its focus on fleet turnover, the paper also relates to the literature on the use of fiscal instruments to affect the fleet composition toward cleaner cars; whether scrappage subsidies to 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, and Hoekstra et al 2015) or the combination of taxes and subsidies, the so-called feebates, to favor the purchase of new cars that are more fuel efficient (e.g., d'Haultfoeuilley et al. 2014, Adamou et al. 2014, and Rivers and Sahufele 2015). One problem with these fiscal policies is that they have rarely been tried in developing and emerging economies (Posada et al. 2015).

But even when scrappage subsidies have been used, whether in the US or Europe, they have tended to be short-lived, lasting only a few months. This is partly explained 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 hit-hard by economic downturns, not as a long-term environmental policy. Similarly, feebates, even if designed to be revenue neutral for the government, do little to accelerate the scrap

(2013a); for one, they handle confounding factors by cleverly exploiting the fact that Chinese drivers tend to avoid license plates ending in digit four.

[5]In addition to operate on unleaded gasoline, a converter at the time was expected to reduce carbon monoxide (CO) and hidrocarbon (HC) emissions by 68 and 52 percent, respectively (Onursal and Gautam 1997, p. 176).

[6]Wolff (2014) also looks at the effect of driving restrictions on fleet composition, but for the low-emission-zone (LEZ) programs in Germany. We separate from this work in two important aspects. First, LEZ programs do not suffer from a second-car effect by construction, i.e., because they impose a total circulation ban on older models. And second, we go beyond empirical estimations and ask how existing restriction designs (including LEZs, HNC-1989 and Santiago-1992) compare to the optimal restriction design and to alternative instruments such as scrappage subsidies.

of the more polluting and less fuel efficient vehicles, only indirectly through the natural turnover of the fleet. This may take a very long time in most emerging economies where drivers tend to keep their clunkers much longer.[7]

Driving restrictions are also closely related to emission inspection tests, or so-called smog checks, that vehicles must regularly undergo in order to renew its registration and permission to circulate. In California, for example, all vehicles are required to be smog checked every 2 years, except those that are 6 years old or less.[8] Vehicles that do not pass these tests, which are year-model specific and vary across geographic areas, are completely banned from circulation in the area of the test but not necessarily in other geographic areas. As we will see below, this all-or-nothing feature of smog checks make them, in theory, superior to the best driving restriction policy one could possibly design because it can be tailored at the individual-car level and to be city-specific (or district-specific within a city).

Despite these smog checks are also popular in cities implementing driving restrictions, there are good reasons why they cannot come to replace restrictions as a policy tool to accelerate fleet turnover. One reason is enforcement. It is well known that smog checks are prone to corruption (e.g., Oliva 2015), at least much more so than driving restrictions, which have been shown to be well enforced (see Davis 2008 for Mexico-City, our results below for Santiago, and Liu et al. 2015 for Beijing). This is not surprising: to alter a smog check you would need to bribe a technician, and only once a year, while to get away with a driving-restriction violation you would need to bribe one or more policemen, and on a weekly basis. The second reason is political feasibility. Driving restrictions can be designed, as usually are, to impose a partial ban on circulation, as oppose to a total ban, on older cars, which by and large are owned by lower-income families.

Since everything indicates that driving restrictions will continue playing a major role in the fight against (local) air pollution in many cities around the world, we believe that the Santiago-1992 reform constitutes a great opportunity to advance our understanding of how these restrictions actually work and could be improved, in particular, regarding the role that exemptions to cleaner cars may play. We divide our analysis into two but complementary parts. In the first part we present a positive analysis that seeks to explain how the 1992 program actually worked in terms of its effect on fleet composition. Using information on annual vehicle circulation fees, we compare fleet age profiles of the 30 municipalities affected by the driving restriction —the ones located in Santiago's Metropolitan Area— against the remaining 293 municipalities in the rest of the country,

[7] According to Davis and Kahn (2008), for example, the average vehicle age in the Mexican fleet is 14 years while in the US is less than 9.

[8] The Japanese *shaken* system is another good example, perhaps the toughest of all, which explains why many drivers get rid off their vehicles within 5 years.

not affected by the restriction. We find for the 2006 sample, the first year for which we have detailed information on annual car registrations, that for every vehicle vintage-model 1992 circulating in any given municipality in Santiago (so subject to the restriction), there are 2.7 of such models in a similar municipality not subject to the restriction, that is, in a municipality outside Santiago of similar income, population/density, access to public transport and level of urbanization.

Evidence that Santiago's fleet composition had moved toward cleaner vehicles because of the 1992 restriction is only reinforced when looking at evidence coming from other data sources. Data from nationally representative household surveys taken in 1998 and 2006, discard any second-car effect the restriction may have had on households' purchasing decisions. Conditional on owning at least one car, and controlling for household and municipality characteristics, results imply that living in Santiago does not make it more likely to own two cars. It could still be argued that these results may not be explained by the Santiago-1992 program but by elements specific to Santiago that we do not account for, including that the restriction may have not been well enforced in the first place. While all the anecdotal information we have points otherwise, the formal evidence we have to refute this argument comes from price data of the used-car market. Based on a large sample of price quotes collected from newspaper ads for different years and models, our results indicate that drivers were willing to pay a premium to avoid the restriction, premium that was increasing in the value of the car, from 4 to 18 percent.

The second part of the analysis takes a normative direction. Using data and the previous results from the Santiago-1992 experiment, we develop and calibrate a vertical-differentiation model of the car market subject to a pollution externality. Our model shares important elements of some existing models, in particular, the dynamics of the one-household model of Adda and Cooper (2000) and the vertical-differentiation aspect of the steady-state model of Gavazza et al (2014). Both aspects are crucial in our analysis. Like in any secondary market, vertical product differentiation is what explains trade among consumers that differ in their willingness to pay for quality. In a car market, high willingness-to-pay drivers upgrade to a new car when they decide to sell their used units to medium willingness-to-pay drivers, which in turn, sell their used units to lower willingness-to-pay drivers and so on. This trading process over the life-time of a unit ends when a low willingness-to-pay driver decides to scrap it.

Our model not only separates from existing models by paying attention to both dynamics and vertical differentiation but also by its focus on pollution control. Following what motivates existing driving restrictions, we limit attention to local air pollution, which adds an extra element of product differentiation since the restriction is not placed upon all cars traded in the used-car market, which clears at the national level, but only upon those operated in the geographic area subject to the restriction. This introduces an interesting differentiation between the two policy instruments we study in detail with the calibrated model: driving restrictions and scrappage subsidies.[9] Unlike driving restrictions, scrappage subsidies cannot discriminate across cars operated in different locations; any effort to do that is immediately undone since cars freely move across locations. Since local pollution problems are location-specific, this discrimination is desirable because the social value of a car varies with location.

Results from our model show that well designed driving restrictions not only perform far better than scrappage schemes (and more so if we consider the shadow cost of public funds associated to subsidies) but also can come remarkably close to implementing the first-best. The reason is that the first best requires not much to persuade households to drive less but to drive cleaner cars. And a driving restriction can achieve this reasonably well by extending full exemptions on relatively clean cars and nearly complete restrictions on the use of dirtier models. This optimal restriction design is not entirely new to policy makers. At least qualitatively, it is already present in the low-emission-zone (LEZ) programs implemented in Germany, which impose a total ban on vehicles that fail to pass some emission-technology criteria (e.g., Euro standard). Since such all-or-nothing design may face implementation problems —politicians tend to favor restrictions that affect all cars equally—, we also use the model to see how much is lost as we move away from the optimal all-or-nothing design to designs that make no technological distinction, like the HNC program as implemented in 1989. Even if we abstract from the second-car effect already documented by Davis (2008) and GMS (2013a), our model shows that a HNC-1989 design leads to significant welfare destruction. The reason is that uniform driving restrictions place too strict a restriction on clean cars without creating incentives to scrap high-emitting ones.

The rest of the paper is organized as follows. In Section 2 we present Santiago's 1992 driving restriction and document our empirical results. The vertical-differentiation model of the car market is developed in Section 3. After presenting the no-intervention equilibrium and the first-best outcome (Pigouvian taxation), we explain how driving restrictions enter into the model and elaborate on their potential effects on fleet composition and how these effects are consistent with some of the empirical evidence presented in Section 2. Using data from the Santiago-1992 experiment, in Section 4 we calibrate the model and obtain parameter values about car and household characteristics and policy intensity. With these and other parameter values, in Section 5 we present the policy simulation

[9] Despite the political economy of gasoline taxes severely restricts their use, we also consider them in our simulations. Not surprisingly, the optimal tax is significantly outperformed by both the optimal scrappage subsidy and the optimal driving restriction. The reason is that gasoline taxes not only are quite an imperfect instrument to deal with local pollution (Fullerton and Gan 2005), as opposed to global pollution where they can be quite effective, but also appear to have a limited effect on fleet turnover (Bento et al. 2009).

exercises for different driving restrictions, and for scrapping subsidies and gasoline taxes. Distributional implications are also discussed. We conclude in Section 6 with ideas on how to extend our model and a discussion of how our positive and normative results make a strong case for the extension of driving restrictions from local to global air pollution problems; if well designed, they can help accelerate the transition towards low- or free-carbon-emission vehicles at a lower cost for the government.

2 Santiago's 1992 driving restriction

Santiago's Metropolitan Area, Chile's capital and home to 40 percent of the country's 17 million people, shares the strong correlation between growth and car ownership of any emerging economy, but in addition, it has one of the worst air pollution problems of any urban center in Latin America. Efforts to control vehicle emissions date back at least to the mid 80's, first in 1985 with the total prohibition to 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 percent of the vehicle fleet only in those days in which air pollution was expected to reach critical levels. Over time these restriction episodes were called upon ever more often, and by 1990 the restriction program applied virtually every weekday from 6.30 am to 8.30 pm during the months of April thru September. In 1992, the government established by executive order that starting in 1993 any new vehicle must be equipped with a catalytic converter if it were to circulate in Santiago. And to help accelerate the turnover toward these cleaner cars, in September 1992 the government also decided to reform the existing driving restriction program to exempt all post-1992 cars from the restriction.

In addition to the geographic dimension of the policy (Santiago versus the rest of the country), our empirical approach exploits the intertemporal discontinuity in car age created by the 1992 reform to study its effect on fleet turnover. Only households in Santiago should present a large difference in valuation between 1992 and 1993 models. One advantage of using the 1992 reform to study changes in fleet composition due to a policy shock is that the car market in the country is well integrated, so ownership decisions can only be explained by policy exposure and households characteristics, including location, and not by variation of car prices in different markets. The other advantage of using the 1992 intervention is that it has remained practically unchanged ever since, so we should be able to identify its effect even years after implementation.

2.1 Effect on fleet composition

Our main database to study changes in fleet composition comes from vehicle registration at the municipality level. The registration is implemented through the payment of circulation fees every March in the car owner's home municipality. We collected data for 323 municipalities, 93 percent of the country's total, for years 2006-2012.[10] Each data collection contains information on the number of cars of each vintage by municipality. With few exceptions (e.g., cars traded during the year), these data capture the age-profile of the fleet of cars that households own and drive across municipalities in any given year of our sample.

Figure 1 shows the fleet composition at the national level for selected years: 2006, 2009 and 2012. Bars in dark grey correspond to pre-1992 models (i.e., 1992 and older), which are the ones subject to the 1992 restriction, and in light grey to post-1992 models, which most, but not all, are equipped with catalytic converters. Since the 1992 reform only required catalytic converters on new models going to Santiago (and neighboring regions V and VI), it took three years for the rest of the country to have all new models be imported with a catalytic converter. According to Onursal and Gautam (1997, p. 177), only 79 percent of all new models registered in 1993 came with a catalytic converter, 87.6 percent in 1994 and 94.8 percent in 1995.[11]

Besides the rapid growth in the number of vehicles, there are a couple of facts that arise from Figure 1 that are relevant for our analysis. Because of the ban on the import of used cars, years of economic downturns (e.g. 1983-1985, 1999-2002, 2008) show a permanent reduction in the number of cars that were added to the existing stock during those years. This carry-over effect is important in our model as this requires agents to anticipate the long-lasting effect that today's sales will have on future equilibrium prices. The second is that one cannot help but notice that non-catalytic cars are now a relatively small fraction of the national fleet, which calls for an adjustment to the existing policy. Our numerical exercises shed light on the type of adjustments needed and their corresponding gains.

*** INSERT FIGURE 1 HERE OR BELOW ***

To get some idea of the effect of the 1992 restriction on fleet composition we can

[10] We miss just a few municipalities with low population and located in remote rural areas. The administrative organization of the country includes 16 regions (numbered I through XV in addition to the Santiago's Metropolitan Region), which, in turn, are divided into municipalities that overall add to 346.

[11] Since our registration samples only identify car vintages, we correct our data by assigning the fraction of post-1992 models without a converter to regions other than Santiago, V and VI following the proportions observed in the actual data.

compare Santiago's fleet with that of the rest of the country, as done in Figure 2 for the 2006 sample. The numbers in Figure 2 tell us that the fleet in Santiago is indeed cleaner (i.e., larger fraction of post-1992 cars), but it is not obvious how much of this is due to the 1992 policy and how much to characteristics specific to Santiago that can be affecting car-purchasing decisions (Santiago is home to the wealthiest households in the country). While most jumps in the figure in the number of cars by vintage are positively correlated between Santiago and the rest of the country (which are mostly related to macro shocks), the jump for the 1992 vintage (and vintages around it) is negatively correlated between Santiago and the rest of the country, suggesting that there may be something special related to that vintage that is different in Santiago than in the rest of the country.

In a first attempt to test for that, the left panel of Figure 3 controls for income by plotting each municipality's income per capita against the fraction of pre-1992 models in the municipality's fleet. Black dots correspond to municipalities affected by the 1992 policy, which on average show a much larger fraction of post-1992 models.

*** INSERT FIGURE 2 HERE OR BELOW ***

*** INSERT FIGURE 3 HERE OR BELOW ***

Some readers may still argue that the higher fleet-turnover shown in the figure for the city of Santiago could be mostly explained by the fast turnover observed in the very rich municipalities of the city. If the used-car market is subject to some frictions, so that the market in Santiago is not fully integrated with markets in the rest of the country, a fast turnover in Santiago's highest-income municipalities could have very well accelerated the turnover in neighboring middle- and low-income municipalities as well. If this is true, these same drivers should see a 1993 model almost as old as a 1992 model. In the right panel of figure we test for this possibility for our 2006 sample by plotting the ratio $q_i^{92}/(q_i^{92}+q_i^{93})$ as a function of income, where q_i^{τ} is the total number of vintage-τ cars in municipality i. The numbers in the figure confirm that there is a clear discontinuity between pre- and post-1992 models in municipalities subject to the 1992 policy, discontinuity that can be only attributed to the driving restriction. In fact, when we extend the same exercise to other adjacent vintages, we find no difference between municipalities that are subject to the restriction and those that are not (see Figure A1 in the Appendix).

Since income is not the only variable affecting purchasing decisions, Table 1 presents the results of estimating the ratio $y_i^{\tau,\tau+1} \equiv q_i^{\tau}/(q_i^{\tau}+q_i^{\tau+1})$ on a number of variables

$$y_i^{\tau,\tau+1} = \beta DR_i + x_i'\gamma + \varepsilon_i \tag{1}$$

where DR_i is a dummy that takes the value of 1 if municipality i is affected by the driving restriction (i.e., if it is located in Santiago) and x_i is a vector with municipality's characteristics such as income per capita, population, distance to Santiago, income dispersion, level of urbanization, and two dummies indicating whether the municipality is located north of Santiago and in any of the country's furthest north and south regions (I, XI, XII and XV), as purchases of new cars in these extreme regions are entitled to tax breaks. As can be seen in the table, the coefficient of DR is negative and significant only for the 92-93 ratio (in column 3),[12] and economically important as well. In fact, if in a given municipality not affected by the restriction we observe one 93 model for each 92 model (i.e., $y_i^{92,93} = 0.5$), in a similar municipality in Santiago that ratio would be 2.15 ($= [0.5 - 0.183]^{-1} - 1$).

*** INSERT TABLE 1 HERE OR BELOW ***

Another interesting result in Table 1 is that the effect of distance from Santiago is only statistically significant for the 92-93 ratio (column 3). Specifically, the estimated coefficients imply an inverted-U relationship between the 92-93 ratio and distance that reaches its peak at 764 kms from Santiago. This relationship is the result of two effects that influence the export of 1992 models from Santiago to the rest of the country as a result of the 1992 policy. The first is a neighborhood effect. Drivers living closer to Santiago are more likely to travel to Santiago for work or personal reasons (e.g., medical attention), which makes them less likely to own a pre-1992 car that may face a restriction just when they need to travel to Santiago. And the second is a transaction cost effect. Arbitrating used-car prices is costly because it requires distant parties to negotiate and cars to be moved around. We expect this transaction-cost effect to go up with distance. Overall, these results speak of a well integrated second-hand market, which is what we assume in the construction of our model and policy exercises.

So far we have focused on the 92-93 discontinuity to show that the 1992 policy did have a distinctive effect on fleet composition, but it is clearly insufficient to evaluate its overall effect. Based on what we know from other restriction programs, it could well be that the exodus of 1992 models was completely undone if a good fraction of drivers were by-passing the restriction not with the purchase of a post-1992 model but with the purchase of a second and possibly much older pre-1992 model. To test for this possibility we run the following regression:

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

[12] We do not report results for all combinations of $\tau, \tau+1$ to save space in the Table. Results are very similar for the unaffected ratios as it is evident from the figures in Appendix A1.

where q is the number of cars of vintage τ in municipality i, $INCOME$ is the municipality's income per capita, DR is a dummy that indicates whether the municipality is subject to the restriction or not, POP is the municipalities's total population, z is a vector that includes the remaining controls included in x in (1), and δ is a vintage fixed effect.

Specification (2) captures how the different coefficients vary across vintages, in particular those of $INCOME$, DR, and POP. The results do not vary much if we run (2) on any of the 2006-2012 samples or as a panel and allowing the different coefficients to vary with time. Thus, in Figure 4, we present the estimates for the 2006 sample, which is the closest sample to the time of policy implementation. The figure depicts coefficients of $INCOME$ and DR. Consistent with the notion that income is a main factor behind purchasing decisions, the correlation of $INCOME$ with vintage shows that newer models are indeed concentrated on richer municipalities. More importantly, the relationship is smooth with no discontinuity between the 92 and 93 vintages, providing additional evidence that income has nothing to do with the 92-93 discontinuity documented in Table 1.

*** INSERT FIGURE 4 HERE OR BELOW ***

In contrast, the 92-93 discontinuity is clearly present in panel (b) of Figure 4, which plots the evolution of the estimated DR coefficient by vintage. For example, the point estimate for vintage 92 (-1.008 and statistically significant at the 1 percent level) indicates that for each 1992 model circulating in a given municipality in Santiago, there will be 2.74 of such models in a similar municipality not subject to the restriction. Conversely, the point estimate for vintage 93 (0.239 and statistically significant at the 10 percent level) indicates that for each 1993 model circulating in a given municipality in Santiago, there will be only 0.79 of such models in a similar municipality not subject to the restriction (similar conclusions are obtained for the 91 and 94 coefficients, which are equal to -1.023 and 0.251, respectively).

Another interesting feature of Figure 4b is the evolution of the DR coefficient as we move away from the 92-93 discontinuity in either direction. In a market for products that are vertically differentiated and where consumers differ in their willingness to pay for quality (i.e., newer models), the null DR coefficients for the most recent post-92 models should not come as a surprise. Regardless of location, a driver's alternative to, say, a 2004 model is not a model that is ten years old or more but something closer to 2004. In other words, ownership decisions concerning models away from the discontinuity should be independent of location, and hence, of the restriction. A similar logic applies to the fact that DR coefficients reverts toward zero for the very old models, so it would be a mistake to interpret this reversal as an indication of the presence of a "second-car" effect (the fact that a fraction of drivers that own a pre-92 model do not by-pass the restriction by switching to a 93 or 94 model but rather by purchasing an additional car, e.g., a 83 model). As we will see shortly, the observed evolution of the DR coefficients is entirely consistent with the prediction of the model we develop in Section 3.

Since our results so far are based on data collected in 2006, 13 years after policy implementation, one may wonder how much of the policy effects documented in Figure 4 would stand, at least qualitatively, in years closer to implementation; in particular, the absence of a second-car effect. In the next two sections we work with two data sources closer to the implementation of the policy to show that to be the case. The first source is official national surveys, which contain information on households' characteristics including car ownership, and the second is a collection of used-car prices based on newspaper ads.

2.2 Is there a second-car effect?

The previous literature has emphasized the perverse incentives that driving restrictions might create for the purchase of a second and older vehicle (Davis, 2008; GMS 2013a and b). However, these results correspond to restrictions imposed upon all cars ruling out the option to bypass the restriction with the purchase of a cleaner car. In order to study the potential second-car effect in Santiago's 1992 reform, we use the Socioeconomic Characterization Surveys (CASEN) for years 1998 and 2006. Despite these are national surveys taken every two or three years to thousands of households, these are the only two years when the surveys included detailed questions on car ownership that can be used for our purpose here.

Figure 5 presents histograms of households owning zero, one or more than one car for each year. Whether in Santiago or in the rest of the country, the majority of households own no car and less than 5 percent own more than one. This latter number already indicates that this margin may not be of first order importance. It does suggest, though, that the fraction of households owning more than one car is twice as large in Santiago as in the rest of the country.

*** INSERT FIGURE 5 HERE OR BELOW ***

Since living in Santiago (and be affected by the driving restriction) is not the only variable affecting purchasing decisions, we run different regressions to capture how the number of cars owned by a household is affected by living in Santiago along with other household characteristics such as income, assets, age, gender and employment status of the head of the household, the composition of the household (in terms of number of members and also number of employed members), and the size of the municipality where household is located. The main econometric challenge has to do with the discrete nature of the data. Despite there may be households owning three or more cars, we are data-constraint to work with just three categories: households may either own zero, one, or more than one car. This corresponds to a right-censoring of the count data but empirically does not have much of an empirical implication as the share of household having more than two cars is really low.

We employ two types of models to estimate the probability of owning a second car, conditional on owning at least one (i.e., $\Pr[c > 1|c \geq 1]$, where c is the number of cars in the household). First, we work with "naive-models" in which we just run regressions of a dummy that takes a value of one if a household has more than one car conditional on having at least one. Results for the marginal effect of living in Santiago are presented in the first two rows of Table 2 for both OLS and probit models, respectively. Results indicate that living in Santiago does not change the probability of having more than one car, after controlling for relevant variables. This means that the differences seen in Figure 5 between Santiago and the rest of the country are mostly driven by households characteristics other than location, mainly income.

*** INSERT TABLE 2 HERE OR BELOW ***

The second model, a Poisson-logit hurdle model, takes a more structural approach. It is a count model that combines a logit process for the generation of the extensive margin (owning at least one car or not) and a Poisson process for the extensive margin (the actual number of cars) (see Wooldridge 2010 for more details). The third row of Table 2 presents the results of this hurdle model and again we find a statistically insignificant effect of living in Santiago on the probability of owning more than one car, conditional on having at least one.[13] In all, these results show no indication of a second-car effect; and if any, it fully evaporated shortly after policy implementation.[14]

[13] We implemented several additional empirical models and in all the cases we found zero effects for the dummy indicating whether a household is located in Santiago. We also estimated multinomial choice models in which households sort into the three ordered categories (i.e., owning 0, 1 or 2 or more cars) (e.g., Matas and Raymond 2008); Poisson count models; Negative Binomial models (e.g., Huang and Chao 2014); and models including right-censoring in the data (and found that right-censoring in two cars is not relevant for the estimations). Results available upon request.

[14] We have no form to rule out a second-car effect between 1993 and 1998. But even if so, its effect for policy evaluation should be minor given the long horizons involved.

2.3 The price of the restriction

If the driving restriction was actually well enforced and binding to drivers, then 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 (see Figure 2). This is important not only as a robustness check of our previous results, but also in itself, as it provides an estimate of the cost of the restriction to individuals. To analyze the effect on prices, we assembled a rich panel data of car prices on the basis of newspaper ads published in "El Mercurio" –Chile's main newspaper– during the period 1988-2000 with price offers for used and new cars of different models (e.g., Honda Accord, Toyota Corolla, Peugeot 205, etc.), sufficient to cover a wide price range.

Following the idea behind a RDD, we regress car prices (in log) on a dummy equal to one for cars equipped with a catalytic converter (i.e., post-1992) while controlling for age of the car and date of the price offer with fixed effects. Controlling with fixed effects is more flexible than with log-linear relationships, as we can calculate an average "catalytic effect" for cars offered in different years. The OLS results are reported in Table 3. The coefficient of Catalytic shows, after controlling for everything else, that drivers are willing to pay a significant premium for having a catalytic converter installed in their cars (and not be subject to the restriction). The fact that the premium tend to be higher in the more expensive models (16 percent in Honda Accord vs 5 percent in Fiat Uno) is consistent with a situation in which individuals that own more expensive cars have a greater opportunity cost of not driving every day and, therefore, are willing to pay more for cars exempted from the driving restrictions.[15]

*** INSERT TABLE 3 HERE OR BELOW ***

One of the main assumptions behind our identification strategy is that cars from "close" vintages are similar in all respects except that only some of them have installed a converter. One may argue that 1993 models could be more expensive than 1992 models not because of the driving restriction, but because of a discrete jump in quality or costs between these two vintages. We offer two exercises to rule out this possibility. The first is to simply compare the cost of replacing an existing converter for a new one to the premiums in Table 3. Onursal and Gautam (1997) report a cost of $265 for replacing a catalytic converter, equivalent to 1.8 percent of the price of a new Toyota Corolla

[15] Figure A2 in the Appendix diplays estimates for Toyota Corollas (a very popular model) as an example. Every point in the scatterplot represents an ad with the vintage of the car and its respective price (in log). In order to use comparable data we show prices from ads in October, November and December from 1991, 1995 and 1997.

in 1995.[16] Moreover, if differences were explained by a high 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.

The second exercise takes advantage of the fact that some pre-1992 Honda Accords were already equipped with catalytic converters, and therefore, exempted from the restriction. Our sample captures many instances in which this feature was explicitly reported in the ad along with the price quote, but not only for pre-1992 Accords but also for some post-1992 ones. Using price offers for Accords in ads published in October, November and December of 1995, we test for the effect of reporting a catalytic converter on the price offer by running four separate OLS regressions for vintages 1991 thru 1994. Since converters were required by law in all post-1992 models, reporting its existence in ads for post-1992 models should make no difference. This is precisely what we see in the last two columns of Table 4, where the coefficient of the dummy Catalytic is not statistically different from zero. This contrasts with the catalytic premiums observed in the first two columns of the table. Albeit somewhat larger, the 1991 and 1992 premiums are not statistically different from the 16 percent premium reported in Table 3 for the same model.

*** INSERT TABLE 4 HERE OR BELOW ***

3 A model of the car market

Two key messages emerge from the empirical evaluation of the Santiago-1992 reform. One is that driving restrictions tend to be well enforced. And the second is that exemptions to cleaner cars make these restrictions a potentially useful tool to fight air pollution by helping accelerate the fleet composition towards lower-emitting vehicles and without inducing the well-documented second-car effect of uniform restrictions (i.e., creating incentives to buy or keep an old and high-emitting vehicle to by-pass the restriction). This second message, however, raises a set of new questions that the empirical analysis cannot answer.

Looking at how the fraction of restricted vehicles shrinks over time (see Figures 1 and 2), an obvious question is for how long a post-1992 model should remain exempted from the restriction. It cannot be indefinitely because a vehicle's emission rate for any of the local pollutants (e.g., CO, particulates, NOx, HC) increases with age, even if the converter is properly maintained. Authorities in Mexico-City have addressed this

[16] As explained by Onursal and Gautam (1997), converters can only be installed in vehicles with spark-ignition engines, which explains why, at least in Santiago, we did not observe pre-1992 vehicles being retrofitted with converters.

problem by exempting new cars only for a limited number of years (no more than 8) while maintaining the one-day-a-week restriction over the rest of the fleet.[17] Is this an optimal restriction design? Would not be better to adopt a more drastic design as done in the LEZ programs in Germany, in which cars that do not comply with an emission-technology criteria (e.g. Euro standard) are completely ban from circulation while those that do enjoy full exemption? But if we adopt such all-or-nothing design, would not too much pollution be exported to neighboring regions that do not have restriction policies in place? More generally, how much do we lose/gain as we consider different restriction designs? What are the distribution implications of the different designs? Are there other instruments, such as scrappage subsidies and gasoline taxes, that can work better at accelerating the fleet turnover?

We address all these questions in the rest of the paper with a dynamic model of the car market that is then calibrated and applied to Santiago and the rest of the country based on data generated from the 1992 reform. We develop the model in this section and leave the calibration and application for the following two sections. It is a model of vertical product differentiation that shares some features of existing models (e.g., Gavazza et al 2014) but it separates from them in its emphasis on pollution control policies and their effect on fleet evolution. Car vintage is the only attribute that creates product differentiation. New models enter the market at perfectly competitive prices and used models are traded in a frictionless secondary market by drivers with different willingness-to-pay for quality (i.e., vintage). Our model also abstracts from multiple ownership and, hence, from the possibility of a second-car effect. This is not only because the empirical evidence from the Santiago-1992 program finds no support for this possibility but, more importantly, because the optimal restriction program eliminates this possibility by construction, as we shall see below.

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. There is a large number of car dealers that buy new cars from car producers and rent them together with second-hand cars to drivers.[18] The (annual) rental price for a car of age $\tau = \{0, 1, 2, ...\}$ at date t is denoted by $p_{\tau,t}$ ($\tau = 0$ corresponds to a

[17] They have also extended the restriction to Saturdays. See Davis (2015) for an evaluation of this latter reform.

[18] Notice that the renting assumption, which is also in Bento et al. (2009), is equivalent to assuming a frictionless secondary market that clears once within each period.

new car). Note that the rental price is not invariant to time as it depends on the stock of used cars which can vary in response to policy shocks, which are the only shocks we consider in our model. Cars exit the market at some exogenous rate due to crush, fatal malfunctioning, etc. This rate may vary with age, so the probability that at age τ car is still in the market next period is $\gamma_\tau \in (0,1)$, with $\gamma_\tau \geq \gamma_{\tau+1}$ (to simplify notation we will assume throughout the rest of this section, but not in the calibration and simulations, that $\gamma_\tau = \gamma$ for all τ). All surviving cars at time t are (endogenously) scrapped at age T_t for a value of v. This latter can be seen, for example, as the price a dealer gets for an old vehicle when exported to another country or, more importantly for the purposes of this paper, as the subsidy in a government's scrappage program.

There is a continuum of households/drivers of mass 1 that vary in their willingness to pay for quality but also in how much they drive. A driver that rents car τ obtains a per-period utility of (to save on notation in many places we will omit the subscript "t" unless is strictly necessary)

$$u(\theta,\tau,x) = \frac{\alpha}{\alpha-1}\theta s_\tau x^{\frac{\alpha-1}{\alpha}} - \psi x - p_\tau \tag{3}$$

where θ is the consumer's type, $s_\tau > 0$ is the quality of a vintage-τ car, x corresponds to miles traveled during the period, $\alpha > 1$ is a parameter that captures decreasing returns in car use, ψ is the per-mile cost of using the car (including parking, gasoline, maintenance, insurance, inspections, circulation fees, etc.),[19] and p_τ is the rental price. The quality of a car falls with age according to $s_{\tau+1} = \varsigma s_\tau$ with $\varsigma \in (0,1)$, either because older cars are more likely to break down or because they lack the latest technological advances.[20]

A consumer θ that rents a vintage-τ car anticipates that she will drive

$$x(\theta) = \left(\frac{\theta s_\tau}{\psi}\right)^\alpha \tag{4}$$

so, her utility from renting a vintage-τ car reduces to

$$u(\theta,\tau) = \kappa\left(\theta s_\tau\right)^\alpha - p_\tau \tag{5}$$

where $\kappa = [(\alpha-1)\psi^{\alpha-1}]^{-1}$.

Our formulation captures with a single parameter two empirical regularities: that households that value quality more tend to drive newer cars and that newer cars are, on average, run more often (e.g., Lu 2006). Households are distributed according to the cdf $F(\theta)$ over the interval $[0,\bar{\theta}]$. A driver θ that does not rent a car obtains an outside utility

[19]This may also include (socially optimal) congestion charges which we do not model explicitely.
[20]The linear quality decay rate is also in Gavazza et al (2014).

equal to u_0, which we interpret as the utility from using pollution-free public transport.

3.2 The market equilibrium

At the beginning of any period, say, year t, there will be some stock of used cars $\mathbf{S}_t = (q_{1t}, q_{2t}, ...)$. 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 with higher types rent newer cars. There will be a series of cutoff levels $\{\theta_{0t}, \theta_{1t}, ...\}$ that precisely determine the prices at which consumers are renting which cars. Denote by $\theta_{\tau t}$ the consumer that at time t is indifferent between renting a car of age τ at price $p_{\tau,t}$ and one of age $\tau+1$ at a lower price $p_{\tau+1,t}$, that is

$$\kappa\left(\theta_\tau s_\tau\right)^\alpha - p_\tau = \kappa\left(\theta_\tau s_{\tau+1}\right)^\alpha - p_{\tau+1} \tag{6}$$

for all $\tau = 0, 1, ..., T-1$, where T is the age of the oldest car that is rented. Consumers of type $\theta \geq \theta_\tau$ rent age-τ vehicles or newer while consumers of type $\theta < \theta_\tau$ rent pre-τ vehicles (or not at all for θ's sufficiently low). As in any vertical differentiation model, an obvious corollary from (6) is that a higher valuation consumer obtains strictly more surplus than a lower valuation consumer.

In equilibrium, the series of cutoff levels $\{\theta_0, \theta_1, ...\}$ must be consistent with the population of drivers and 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

$$q_0 = 1 - F(\theta_0) \text{ and } q_\tau = F(\theta_\tau) - F(\theta_{\tau+1}) \tag{7}$$

for all τ that are rented in equilibrium.

Since car dealers have always the option to scrap an old car and receive v, in equilibrium they must also be indifferent between renting an age T vehicle today (and scrap it tomorrow, if the vehicle still exits) and scrapping it today, i.e.,

$$p_T + \delta\gamma v = v \tag{8}$$

In general, only a fraction of vintage-T vehicles will be scrapped in equilibrium (while all pre-T vehicles will), so

$$F(\theta_{T-1}) - F(\theta_T) \leq \gamma q_{T-1} \tag{9}$$

where γq_{T-1} is the number of age T vehicles that survived from last period.[21]

In addition, in equilibrium (competitive) car dealers must break even, so the evolution

[21] Note that because quality drops discretely with age, it can happen that in equilibrium all $T-1$

of rental prices must satisfy

$$c = \sum_{i=0}^{\Gamma} (\gamma\delta)^i p_i + (\gamma\delta)^{\Gamma+1} v \tag{11}$$

where Γ is the age at which a car bought today, i.e., at date t, is expected to be retired (or rented for the last time). Note that both Γ and T depend on the existing stock $\mathbf{S}_t$ and in steady-state $\Gamma = T$.

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

$$\kappa \left(\theta_T s_T\right)^{\alpha} - p_T = u_0 \tag{12}$$

If (12) does not hold, a dealer would be strictly better off by renting a T vehicle at a price slightly above p_T instead of scrapping it.[22]

Conditions (6)–(12) 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 other papers, we are not only interested 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 despite they can be computationally demanding they cannot be neglected in policy evaluation and design.

3.3 The social optimum

We now characterize the social optimum. If cars pollute, then the market equilibrium described above is not socially optimal. Suppose that cars emit pollutants at a rate e per mile, which is increasing with age, that is, $e_{\tau+1} > e_\tau$. Denote by h the harm from pollution, so the cost to society of a vintage-τ car running x miles is $e_\tau x h$. If the social planner can monitor emissions, $e_\tau x$, he can restore the social optimum by levying a Pigouvian tax equal to h on each unit of pollution. This will affect decisions on car use

vintage are rented but all T vintage are scrapped, then the relevant scrapping condition is not (8) but

$$p_{T-1} + \delta\gamma v > v > p_T + \delta\gamma v \tag{10}$$

where p_T is the hypothetical price for the rental of a T vehicle.

[22] The same logic applies if we are in the corner (10) of the previous footnote: given the fixed supply of vintage $T-1$ vehicles, a dealer owning a $T-1$ vehicle could slightly raise its rental price above p_{T-1} and still find demand for it.

and ownership, i.e., it will affect (4) and (5), in the following way

$$x^{*}(\theta,\tau)=\left(\frac{\theta s_{\tau}}{\psi+e_{\tau}h}\right)^{\alpha} \tag{13}$$

and

$$u^{*}(\theta,\tau)=\kappa_{\tau}\left(\theta s_{\tau}\right)^{\alpha}-p_{\tau} \tag{14}$$

where $\kappa_{\tau}=[(\alpha-1)(\psi+e_{\tau}h)^{\alpha-1}]^{-1}$.

As we will illustrate later, in a market for a durable good, a policy intervention can affect the market in subtle ways. One may argue that if the government levies a tax on emissions, cars will become relatively more expensive than (pollution-free) public transport, reducing the demand for cars in the market, which in turn, should reduce the number of cars that enter the market each year. This intuition is only partially correct because a car is not a single product but a collection of different products providing different services over its life time. Newer (and cleaner) cars have become relatively cheaper than older cars so their demand has increased. The overall effect is that there will be more new cars coming to the market each year but each lasting fewer periods.

3.4 Real-world policy interventions

Since Pigouvian taxation of local air pollution is politically and technically unfeasible (e.g., Fullerton and Gan), policy makers tend to rely on imperfect instruments. Scrappage subsidies, gasoline taxes and driving restrictions in their different formats (including LEZs and smog checks) are good examples that we study here. The way scrappage subsidies and gasoline taxes enter into our model is relatively simple; by increasing v and ψ, respectively.

The way a driving restriction enters into the model is more involved because it depends on the specific design which must specify the extent of the restriction and the car vintages that are affected. The extent of the restriction is captured by the parameter $R_{\tau}<1$, which says that a vintage-τ car can only be used a fraction of the time, for example, 4 of the 5 weekdays of the week. We understand that drivers can move some trips from one weekday to another at a low cost, so R should not be read as $4/5$ in this example but probably more. It is less obvious whether the trips that can be moved around are the most valuable to the driver or not. We adopt the conservative assumption (i.e., that is less favorable to the driving restriction option) that the driving restriction destroys an equal fraction of trips of different values during the day of the restriction. Since our model does not distinguish by day of the week, this assumption implies that the driving restriction reduces the number of trips a driver would otherwise make uniformly over the week, or the year for that matter. Formally, a driver θ that owns a vintage-τ car that faces an effective restriction of $R_\tau < 1$ will now drive

$$x^r(\theta, \tau) = R_\tau \left(\frac{\theta s_\tau}{\psi} \right)^\alpha \tag{15}$$

miles and obtain a utility of

$$u^r(\theta, \tau) = R_\tau \kappa \left(\theta s_\tau \right)^\alpha - p_\tau \tag{16}$$

per period.

Expressions (15) and (16) raise two observations that are important for the analysis that follows. The first is that by comparing (15) and (13) it appears that the driving restriction could replicate the first-best amount of driving x^* by simply setting vintage-specific restriction levels, i.e., $R_\tau = [\psi/(\psi + e_\tau h)]^\alpha$. While this is true if households cannot adjust their renting decisions, it is not enough to restore the first-best because a policy affects not only the amount of driving but also the cars that households rent in equilibrium.

The second, which is closely related to the previous one, is that the optimal driving restriction takes an extreme form, that is, R_τ should be either 1 or 0. The reason is that since R_τ enters linearly in the social welfare estimate of type-θ household driving a vintage-τ car under restriction R_τ, $u^r(\theta, \tau) - he_\tau x^r(\theta, \tau)$, the optimal value of R_τ cannot be interior but the corner 1 or 0. One immediate implication of this corner solution is that an optimal driving restriction eliminates by construction the possibility of observing a second-car effect. A more practical implication is that this all-or-nothing structure is already in existing pollution-control policies, namely, LEZs and smog checks, which impose a complete ban on vehicles that do not pass a preestablished standard. This is the reason we view LEZs and smog checks as one of the many formats driving restrictions can take in practice. In our model the two formats are indistinguishable because new cars start equal and age homogeneously. In practice, however, cars are not only different when new but they age quite differently, as indicated, for example, by the large variation in their (natural) scrappage ages (see Jacobsen et al. 2015). The (uniform) technology-standards used in LEZs programs fail to account for this car heterogeneity, unlike smog checks that are car-specific. In theory, this makes smog checks, if set at the socially optimal levels, the best driving restriction design.

3.5 Application to Santiago's program

With the use of the model we can provide further support to some of the empirical evidence shown in Section 2, in particular, that in Figure 4b. If, following the Santiago-1992 reform, the driving restriction affects only a fraction of the drivers in a (well-integrated) car market, we can use (16) to obtain cutoff levels θ_τ for drivers in a municipality not affected by the restriction, θ_τ^{nr}, and compare them to those in a municipality affected by the restriction, θ_τ^r. Rearranging (16), we obtain

$$\theta_\tau^k = \left(\frac{p_{\tau+1} - p_\tau}{R_{\tau+1}^k \kappa s_{\tau+1}^\alpha - R_\tau^k \kappa s_\tau^\alpha} \right)^{\frac{1}{\alpha}} \tag{17}$$

where $k \in \{nr, r\}$, $R_\tau^{nr} = 1$ for all τ, $R_\tau^r = R < 1$ for all pre-92 models, and $R_\tau^r = 1$ for all post-92 models.

According to our model, a driving restriction like Santiago-1992 should only produce relative changes in car holdings for models just on either side of the 92-93 discontinuity, as depicted in Figure 4b. In fact, if we compare car holdings in two municipalities that are identical but for the restriction (i.e., $F^{nr}(\theta) = F^r(\theta)$, $\alpha^{nr} = \alpha^r$, $\varsigma^{nr} = \varsigma^r$, $\psi^{nr} = \psi^r$), we should see no difference in car holdings for newer models; strictly speaking for 94 and newer models since $\theta_\tau^{nr} = \theta_\tau^r$ for all those models. This is entirely consistent with Figure 4b, that shows that the DR coefficients in equation (2) for the most recent models are not statistically different from zero. The fact the DR coefficients for 94 and 95 vintages, and not just 93, are positive is not surprising because there will be always noise in car quality coming from individual preferences, cars aging differently, etc. This is the reason why in the calibration that follows we cluster cars in 4-vintage groups around the 92-93 discontinuity. Using the next section's calibration values for the different parameters, the model's prediction for Figure 4b is Figure 6a.

*** INSERT FIGURE 6 HERE OR BELOW ***

Following the same logic, we should also see no difference in car holdings between affected and non-affected municipalities for models older than 92. In this case $\theta_\tau^{nr} = (1/R)^{1/\alpha}\theta_\tau^r$ for both τ and $\tau+1$, so if F is linear in the relevant range, which is a good approximation given so many vintages, we should obtain $q_\tau^{nr} = F(\theta_\tau^{nr}) - F(\theta_{\tau+1}^{nr}) \approx q_\tau^r = F(\theta_\tau^r) - F(\theta_{\tau+1}^r)$.[23] This again explains why the DR coefficients for older models returns to zero as we move away from the 92-93 discontinuity. This provides additional evidence of the absence of a second-car effect in the Santiago program; otherwise, some

[23]The linearity in $F(\cdot)$ is not even necessary given our calibration results below: $(1/R)^{1/\alpha} = 1.02$.

of the DR coefficients of the older models should have been strictly positive. Given this latter result, together with the fact that optimal driving restrictions do not allow for a second-car effect, validates the use of the single-ownership assumption in our model since these are the restriction designs we will be considering in the calibration and policy exercises that follow.

4 Calibration

We use different data sources and methodologies to obtain numerical values for the different parameters that enter into the model. Some parameters are taken directly from the data while others are calibrated to match the distribution of cars across municipalities that we observe in the 2006 sample of car registrations. The order in which we discuss our choices follows roughly the order in which parameters appear in the presentation of the model.

4.1 Car-related parameters

While the 92-93 discontinuity introduces a clear partition in car quality for those two adjacent vintages, more generally drivers tend to regard cars of slightly different vintages of similar quality. We address this quality overlap in very 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.[24] This grouping is equivalent to assuming that people trade their cars not every year but every four years.[25]

Rental prices p_y are obtained from car prices according to the no-arbitrage condition

$$p_\tau = P_\tau - \delta P_{\tau+1} \tag{18}$$

where P_τ is the price of a vintage-τ car and δ is the discount factor that we set at 0.9 per year. Car prices are generated from the dataset described in Section 2.3. If P_{iyma} is the price offer in newspaper ad i published in year y for model m that is a years old, we run an OLS regression of $\ln(P_{iyma})$ on a constant and year, model and age fixed effects to predict $\hat{P}_{yma}$. With these predictions and (18), we obtain (weighted average) rental prices for each of the six vintage groups identified above. These predictions are also

[24] The few models year 1980 and older that are still around are grouped together with 1981 models.

[25] A more sensible approach for handling the quality overlap is perhaps to allow the quality of car i of vintage τ to vary stochastically such as $s_{i\tau} = \varsigma^\tau s_0 + \varepsilon_{i\tau}$, where $\varepsilon_{i\tau}$ is a random shock that correlates highly with earlier shocks and that has a variance that probably grows with age. Unfortunately, we do not have the information to do this.

used to obtain the (weighted average) price of a new car (i.e., $a < 1$), which we set at $c = \$16,000$ (all our numbers are in 2006 US dollars).

Since the import of used cars is forbidden, we estimate survival rates, γ_τ, directly from stock changes observed in the car registration samples from 2006 through 2012. By comparing stock changes across two consecutive samples, we obtain six data points with survival rates for each car age. Imposing $\gamma_\tau \leq 1$ and $\gamma_{\tau+1} \leq \gamma_\tau$, an OLS fit to these data points delivers average survival rates for cars with ages that go from 0 to 36 years old. Averaging out these numbers at our vintage-group level leads to the survival numbers γ_τ in Table B1 of the Appendix.

The four remaining parameters related to vehicles are the scrappage value v, the initial quality s_0, the quality decay rate ς, and the per-unit cost ψ. None of these parameters can be obtained directly from the data. Based on informal conversations with car dealers we set $v = \$700$, the lowest trade-in value some of them recall to have seen in recent years (we do not see prices this low in our sample of newspaper ads). On the other hand, since $s_\tau = \varsigma^\tau s_0$ enters multiplicatively in (3), we cannot separately obtain estimates for s_0 and the equilibrium cutoffs θ_τ's. Hence, we normalize $s_0 = 10$.

Similar to Gavazza et al (2014), the per-unit cost ψ is assumed to be invariant to location and vintage. From (4) and (6), 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 a car during the first (vintage) period ($\tau = 0$), which in our calibration corresponds to the first four years of the car's life. With values of ς and α, which are obtained from the calibration that we explain next, we estimate ψ to match the figure of $x(\theta_0) = 50,952$ miles that is in Lu (2006) for the average travel during the first four years of a car's life.[26]

4.2 Households' characteristics and policy response

Two important inputs for our analysis are the extent to which households exhibit decreasing returns to driving ($\alpha > 1$) and the distribution $F(\theta)$ for their marginal valuation for quality $\theta \geq 0$. Both inputs are not directly observable, neither is the actual response R_τ to the policy, which is assumed to be the same over all pre-92 models, that is, $R_\tau = R < 1$ for all $\tau \leq 1992$ (recall that by construction $R_\tau = 1$ for all $\tau \geq 1993$). These three inputs together with ς are calibrated to match the car-holding predictions of the model to the actual holdings in the 2006 sample for each location and vintage.

[26] We use Lu's (2006) numbers because we do not have a comparable study for the local fleet. According to domestic car dealers, however, numbers for the local fleet are very similar.

We start the calibration by reducing the dimensionality of the problem from more than 300 municipalities to 60 electoral districts. Electoral districts bunch municipalities located in the same geographic areas and, therefore, that share similar characteristics, most importantly, income. Since the country's population is normalized to 1, our relevant car-holding variable $q_{i\tau}$ becomes the fraction of cars of vintage-group $\tau = 0, ..., 5$ that is in district $i = 1, ..., 60$ relative to the district's number of households.

We need to define a functional form for $F_i(\cdot)$ and how it varies from district to district. We let F_i be a cubic in θ with each coefficient in the cubic (b^1, b^2 and b^3) varying across districts $i = 1, ..., 60$ according to the linear function $b_i^j = \beta_0^j + z_i' \beta^j$, where $j = 1, 2, 3$ denotes the coefficient in the cubic, β_0^j is a constant and z_i is a vector that includes the following district's characteristics: income per capita ($INCOME$), distance to Santiago ($DISTANCE$), and level of urbanization ($URBANIZATION$). Thus, the distribution F_i's are characterized by 12 β^j's parameters, four for each of the 3 coefficients in the cubic, that need to be calibrated along with α, ς and R.

Some of the calibrated F_i's are plotted in panel (b) of Figure 6. The lower F corresponds to the country's highest-income district (district 23), and the upper F corresponds to the country's lowest-income district (district 46). The plot also includes two intermediate distributions that will be extensively used in the policy simulations: the distribution for Santiago, which aggregates all 15 districts affected by the driving restriction, and the distribution for the rest of the country, which aggregates all 45 districts not affected by the restriction.

According to (17), our (deterministic) model predicts the exact same cutoff levels θ_τ for all districts within the driving restriction zone, θ_τ^r, and likewise for all districts outside the driving restriction zone, θ_τ^{nr}. In fact, the two rows of dots at the bottom of Figure 6b depicts the cutoff predictions for each of the six vintage groups for districts that are in Santiago (lower row) and in the rest of the country (upper row).

Obviously, if we plug the cutoffs and distribution F_i's that are in Figure 6b into (7) to obtain predictions $\hat{q}_{i\tau}$, these latter will not precisely match the shares $q_{i\tau}$ that are observed in the data. Thus, our calibration procedure looks for parameter values that minimize those mismatches. In particular, we introduce vintage-district shocks so that the cutoff in district $i = 1, ..., 60$ for vintage-group $\tau = 0, ..., 5$ is given by

$$\theta_{i\tau} = \theta_\tau^k + \varepsilon_{i\tau}$$

where θ_τ^k is given by (17) with $k \in \{nr, r\}$ and $\varepsilon_{i\tau}$ are error terms that make the model's prediction $\hat{q}_{i\tau} = F_i(\theta_{i\tau}) - F_i(\theta_{i,\tau+1})$ to exactly match the actual share $q_{i\tau}$.[27] Since there

[27] More precisely, the 60×6 $\varepsilon_{i\tau}$'s unknows are found by solving a system of 60×6 $\hat{q}_{i\tau} = q_{i\tau}$ equations.

is no reason for the error terms to be correlated with the district characteristics that define the F_i's distributions and whether the district is in a driving restriction zone or not (i.e., $DR = 1$ or 0), our calibration minimizes the following five moments for each vintage-group τ: $\sum_{i=1}^{60} \varepsilon_{i\tau} = 0$, $\sum_{i=1}^{60} INCOME_i \times \varepsilon_{i\tau} = 0$, $\sum_{i=1}^{60} DISTANCE_i \times \varepsilon_{i\tau} = 0$, $\sum_{i=1}^{60} URBANIZATION_i \times \varepsilon_{i\tau} = 0$, $\sum_{i=1}^{60} DR_i \times \varepsilon_{i\tau} = 0$.

We obtain a value of α of 2.08, which leads to a concave utility function $u(\theta, \tau, x)$ —note that $(\alpha - 1)/\alpha = 0.52$— that is not that different from the logarithmic utility used in Gavazza et al. (2014). We obtain a value of the decay rate ς of 0.891, which is also almost identical to the value they use.[28] Our calibrated unit-cost ψ is 0.2230, which is twice as large as their number, however, partly because of our higher gas prices but nevertheless not high enough to explain the difference. It is also interesting to notice that the actual policy intensity, $R = 0.967$, does not appear particularly large. We need to remember, however, that the policy was enacted 13 years earlier than the 2006 data used in the calibration. In fact, if we assume that the policy effect should fall linearly with age and practically disappear after 20 years of implementation,[29] a simple interpolation yields an immediate policy effect of $R = 0.906$, much closer to the short-run effect found by GMS (2013a) for HNC.

4.3 Pollution data

The pollution-related parameters in the model are the social harm from local pollution, h, and the emissions rate of a vintage-τ car, e_τ. We exploit different sources in the public domain to estimate both. The first source is Parry and Strand (2012), which contains specific estimates of vehicle emission damages for Chile using the usual methodology applied in the literature. Their damage estimate for an average vehicle in Santiago is ¢6 per mile while for an average vehicle in the rest of the country is ¢0.7 per mile. With this information and the travel figures in Lu (2006), we estimate an annual pollution damage per car of \$720 in Santiago and of \$84 in the rest of the country.

In order to disaggregate average damages at the vintage level, we use information from Molina and Molina (2002, pp. 236 and 255) for Mexico City on the contribution of cars of different vintages to local emissions (i.e., CO, NO_x, HC, SO_2 and particulates). The vintage-pollution relationship they obtain is summarized in the first two columns of Table 5, which, for example, shows that the newest 60 percent of the fleet contributes with only 15 percent of total emissions. The informal documentation that we have for Chile

[28] Their annual decay rate is 0.976 while ours is $\sqrt[4]{0.891} = 0.972$.

[29] As the policy ages, the discontinuity 92-93, which permits the estimation of R, hits older models, which tend to be run less often (Lu 2006) increasing the possibility to move trips around.

is consistent with these numbers,[30] which is not at all surprising given the similarities in fleet compositions (see column 3 in the table).

*** INSERT TABLE 5 HERE OR BELOW ***

According to our theoretical model, the damage contribution of a vintage-τ car is $xe_\tau h$, so the contribution of all vintage-τ cars in region k would be equal to

$$\int_{\theta_\tau^k}^{\theta_{\tau-1}^k} \left(\frac{\theta s_\tau}{\psi}\right)^\alpha e_\tau h_k dF_k(\theta)$$

where $k \in \{r, nr\}$ and $\tau = 0, ..., 5$.[31] Assuming that e_τ evolves according to the linear function $e_\tau = (1+\omega)e_{\tau-1} + \omega$, with $e_0 = 0$, the values of ω, h_r and h_{nr} that provide the best fit to the damage estimates of Parry and Strand (2012) and the vintage-pollution relationship of Molina and Molina (2002) are, respectively, 1.52, 0.0189 and 0.0016.

5 Policy exercises

With these parameter values we now perform different policy exercises to answer the questions posed in the introduction to Section 3. In all simulations that follow, parameters values are kept constant overtime including the speed at which a car's emissions rate deteriorates with age (one may argue that newer cars are equipped with technologies that may age differently). We start by constructing the no-intervention scenario. Figure 7 shows that in the absence of any government intervention, the city of Santiago (the restricted region) already exhibits a relatively newer fleet than in the rest of the country (the non-restricted region). While Santiago's smaller population (36 percent of the country's total) explains its smaller overall fleet, its higher income per capita explains why it nevertheless has 14 percent more of the newest models (0 to 4 years, or age-group 0 in our simulations) than the rest of the country. For any of the older models, Santiago has fewer of them.[32] Notice that in both locations we will find cars running up to when they are scrapped, somewhere between 24 and 28 years old.[33] This will change as the government intervene the market.

The next set of exercises estimate welfare gains of moving away from the no-intervention

[30] Luis Cifuentes of the Industrial Engineering Department of PUC-Chile, personal comunications, 2015.

[31] Note that $\theta_{-1}^k = \bar{\theta}^k$.

[32] If in comparing fleets across regions one were to eliminate any size effect and just leave income effects one would need to multiply the height of each of the bars in Figure 7a by $1.78 = 0.64/0.36$.

[33] It is not surprising that this scrappage age is different than the one in the calibration (20-24 years old). The equilibrium in the calibration is subject to an intervention and is not steady-state.

scenario. We are particularly interested in the welfare gap between the first-best outcome and the outcome from alternative policy interventions. Since there is perfect competition in the car market, welfare in any given period t is equal to (time index t has been omitted)[34]

$$-cq_0 + vq_T + \sum_{k=r,nr} \left(u_0 \int_0^{\theta_T^k} dF_k(\theta) + \int_{\theta_T^k}^{\bar{\theta}^k} [u^k(\theta) - h_k e^k(\theta) x^k(\theta)] dF_k(\theta) \right)$$

where q_τ is the total number of vintage-τ models in the market in period t, θ_T^k is the last household to rent a car in region k in period t (see (12)), and $u^k(\theta)$, $e^k(\theta)$ and $x^k(\theta)$ are, respectively, a household θ's utility, emissions per mile and miles traveled in region k and period t. Notice that the specific forms of $x^k(\theta)$ and $u^k(\theta)$ vary with the policy scenario; compare, for instance, eqs. (13) and (14) with (15) and (16). We assume that all agents (including the benevolent regulator) discount the future at $\delta = 0.9$ per year, as assumed in the calibration.

5.1 The first-best benchmark

As observed in the steady-state outcome of Figure 8, the effect on fleet composition of levying a Pigouvian tax equal to h_k upon cars circulating in region k is quite dramatic. Over the long-run, households in Santiago have no incentives to hold cars older than 16 years old; a 12-year reduction compared to the no-intervention case. While sales of new cars in Santiago increases by 36 percent, there are fewer households driving cars in Santiago, but those that do, drive cleaner cars. This dramatic adjustment has also large impacts outside Santiago. The scrappage age of a car in the rest of the country is reduced in 8 years. This may look surprising at first because there is no intervention in rest of the country, but everything works through the second-hand market. Instead of scrapping them, Santiago is now exporting a large fraction of 16-years-old cars to the rest of the country. This increase in supply reduces the rental price of all 20-year and older models in the market to the point that is optimal to scrap them much sooner.

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 long, so it cannot be omitted from any welfare estimation, but also non-monotonic (this applies to any policy intervention). Figure 9, for example, illustrates the dynamics of new cars sales (q_0) for

[34] Since a policy intervention affects the value of the existing stock of used vehicles, by altering future rental prices, there may be (unanticipated) changes in dealers' revenues that we are missing in our estimation.

both Santiago and the rest of the country. In the case of Santiago, an initial jump in sales of 53 percent (from 0.034 to 0.052) is followed by a sharp drop to the steady-state level of 0.046, but a few years later, there is a new jump to finally reach, although not exactly, the steady-state level 25 years after implementation (notice the similar non-monotonicity in Adda and Cooper 2000). This non-monotonicity introduces some computational challenges for the determination of the equilibrium dynamics, particularly, when we need to search for the optimal policy intervention.

In present-value terms, the welfare gain of moving from the no-intervention to the first-best amounts to \$323.8 per household, or a 5.9 percent gain from no-intervention baseline of \$5528.5 (see the first two rows of Table 6). At the country level, this adds to a total of \$1.3 billion; comparable, for example, to the gain of introducing LEZs in Germany (Wolff 2014). In any case, we do not want to push these welfare numbers too much. Other than to be a rough approximation of the potential gains from curbing vehicle emissions, our purpose here is to use them as benchmark for evaluating the relative performance of real-world policies like driving restrictions in their different formats, and scrappage subsidies and gasoline taxes.

*** INSERT TABLE 6 HERE OR BELOW ***

There are at least two elements that separate real-world policies from first-best implementation. The first is that for either political or technical reasons the instruments involved are never first-best. And the second is that their use is restricted to geographic areas that has been declared in non-attainment with existing air quality standards. Consequently, the regulator cannot introduce policies in other geographic areas to contain any eventual pollution leakage coming from the regulation imposed elsewhere. We also adopt this geographic limitation in the simulations that follow, so that policy measures 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 it is a scrappage subsidy, gasoline tax or a driving restriction, may vary overtime. Given the dynamics of the first-best outcome, it appears that regulator would like to start with a tougher policy to be gradually relaxed to its steady-state level. For simplicity, however, in what follows we focus on time-invariant policies.[35]

[35] The non-monotonic dynamics described above and illustrated in Figure 9 makes the computation of time-varying (optimal) policies quite demanding. We nevertheless attempted some departures from and around the time-invariant (optimal) design. The additional gains were not important to change any of the results that follow.

5.2 Scrappage subsidies and gasoline taxes

We start with these price policies only because they are easier to compute and analyze. A permanent scrappage subsidy enters in our model as an increase in v, the scrappage value, while a gasoline tax as an increase in ψ, the per-mile cost of using the car. As shown in rows 3 and 4 of Table 6, the two policies perform remarkably differently in terms of welfare. While the welfare gain from an optimal scrappage subsidy of $1575 is 73 percent of the first-best gain, the gain from the optimal gasoline tax (¢2.5 per mile) is only 35 percent (this number is probably an upper bound if we accept that some households living in Santiago may go outside the "tax area" to fill up their tanks).

The difference between the two policies does not come as a surprise after we understand their effect on fleet composition, which are shown in Figures 10 and 11, respectively. The scrappage subsidy results in a steady-state fleet with no car older than 20 years old, which is somewhat consistent with what we observe in the first-best outcome, since old cars are the ones that contribute the most to air pollution. An important reason the scrap subsidy does not come closer to the first-best is that it fails to discriminate across regions. Since the regulator cannot prevent used cars to freely move across regions, the subsidy ends up applying to all cars regardless of their location. This explains why, unlike Figure 8, the scrap age in Figure 10 is not only the same across regions but also higher. If the regulator were to replicate the first-best scrap age in Santiago with a subsidy, it would need to increase the subsidy to suboptimally high levels because of its effect on the rest of the country.

Another way to appreciate why subsidies work relatively well in fighting pollution is to understand why gasoline taxes work so poorly. Gasoline taxes impose a uniform restriction making no distinction between high and low emitting vehicles. Actually, they impose a heavier burden on newer vehicles because they are run more intensely (this could be somewhat offset if we allow for newer cars not only to be cleaner, in terms of local pollution, but also more fuel efficient). Because of this uniform treatment, a gasoline tax can do little to move the fleet composition towards cleaner cars (see Figure 11). Furthermore, one cannot rule cases in which the optimal gasoline tax is to have no tax at all, or perhaps, a negative one that can work as an indirect subsidy to the purchase of newer cars.

The reason then why scrap subsidies work this well, is because they are quite effective in removing old, high-emitting vehicles from the road. However, they suffer from a major shortcoming, which may explain why they are only rarely used, and when so, for very short periods of time: their fiscal cost. Since governments must pay these subsidies with additional (distortionary) taxation, any reasonable estimate of the shadow cost of public funds (see, e.g., Laffont 2005) would greatly reduce their net benefits. Fortunately, there is a policy instrument that is equally or more effective in removing old cars from the road —because it can perfectly discriminate across regions— and that does not require from any government subsidy: well designed driving restrictions. We turn to them now.[36]

5.3 Driving restrictions

The design of driving restrictions is more involved because it requires to decide about the intensity of the restriction (R) and the vintages to be affected. This flexibility can prove fatal, as the numbers in the fifth row of Table 6. If the regulator follows a HNC design, as implemented in 1989 in Mexico-City, and establishes a uniform restriction of $R = 0.9$ on all cars, the results are devastating.[37] Even neglecting the second-car effect, which is omitted in our model and simulations but present in comparable real-world applications (Davis 2008, GMS 2013a), such uniform design leads to a 9 percent welfare loss relative to the no-intervention baseline.

The reason for this loss is illustrated in the resulting (steady-state) fleet of Figure 12. A HNC-type restriction not only fails to remove old cars from the road (actually it extends their lives by reducing their rental prices) but also reduces sales of new cars in Santiago, the affected region. Since new cars are rented by households that value quality the most, a uniform reduction in quality is felt heavier on these new cars. As a result, the demand for them fall and with that their rental prices and sales. As the demand for cars gets shifted towards older models, the life of the existing stock gets extended as well. And because of this life extension, pollution may end up higher than in the no-intervention baseline, and without resorting to the second-car effect; unless, of course, the reduction in circulation prompted by the restriction is enough to offset the fleet ageing.

There is a relatively easy way for the regulator to reverse the perverse outcome of a uniform restriction: to follow the Santiago-1992 reform and allow some of the cleaner cars be exempt from the restriction. For example, if the $R = 0.9$ restriction falls only upon cars that are 12 or more years old, then this Santiago-type design results in welfare gains that are 24 percent of the first-best gains. As shown in Figure 13, the exemption on cleaner cars solves one part of the problem; it boosts sales of new cars in Santiago. The second part of the problem, the removal of high-emitting vehicles from the road, requires a tougher restriction upon these cars. Indeed, the optimal driving restriction is to impose a total ban on vehicles that are 16 and more years old and a full exemption on newer vehicles. As indicated in the sixth row 6 of Table 6, the welfare gain of doing so is

[36] The model also allows us to easily analyze other interventions such as a subsidy to the purchase of new cars (drop in price c) or a subsidy to public transport (increase in u_0). While the first is clearly suboptimal, we do not pursue the second here because it requires modelling public transport explicitly.

[37] The value of 0.9 is borrowed from GMS (2013a) who found an effective reduction in car-activity of 10 percent during the first month of HNC.

remarkably close to the first-best gain. Notice also that by imposing a complete ban on old vehicles, an optimal driving restriction closes any possibility for a second-car effect.

The reason why a well design driving restriction can come so close to the first-best pollution policy can be explained with the aid of Figure 14. Like the optimal scrappage subsidy, an optimal driving restriction works in both ends of the fleet spectrum. It directs the removal of old cars and boosts sales of new cars. But unlike scrap subsidies, it allows the regulator to "policy discriminate" across regions that share a common car market but face different (local) pollution realities. Because scrapping 16-20 years old cars in areas where local pollution is less of a problem is socially inefficient, a driving restriction works its way through the second-hand market to reallocate these cars 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 8b, the export of 16-24 years old models to the rest of the country puts a downward pressure on the rental price of the very old cars (24-28 years old), ultimately, inducing car dealers to retire them from the market. This market dynamics may help explain why Wolff (2014) fails to find pollution leakage to areas not covered by LEZ programs.

5.4 Properties of the optimal driving restriction

Our optimal driving restriction raises a couple of issues that deserve further discussion, from its connection to other real-world policies to its distributional implications. As indicated before, in our model and simulations the optimal restriction policy is not structurally different from LEZ programs and smog checks. An issue arises, however, when there is car heterogeneity that can only be captured during smog checks. If a driving restriction cannot rely on car-specific smog checks to grant exemptions, because of corruption concerns (see Oliva 2015), and therefore, must rely on easily observable technology standards, as done in Santiago-1992 or in existing LEZ programs, then the optimal restriction design has no longer an all-or-nothing structure but rather a partial-ban structure. In fact, if we introduce quality heterogeneity within each vintage in our model (except for the new models), the optimal restriction design would be an schedule of partial circulation bans, starting with none upon new cars and tightening with age until becoming complete upon cars that are older than the 16 years-old threshold in the optimal design of Figure 14b.

By placing a total ban on old cars, which are mostly owned by lower-income drivers, an all-or-nothing design may also rise serious distributional concerns. This is most reasonable, as seen in Figure 15a. Almost all drivers and public-transport users in Santiago are better off under the optimal driving restriction vis-a-vis the no-intervention scenario.[38] There is, however, a relatively small group of drivers of cars that are soon to be retired that are strictly worse off. 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 to be moved 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. This is another reason why complete bans may not be preferable from a practical point of view.

6 Conclusions

As the experience in many cities around the world shows, driving restrictions are becoming increasingly popular tools to control vehicle pollution and congestion. The empirical and numerical results of the paper suggest that these policies not only tend to be well enforced but also, provided they include exemptions to cleaner cars, can be quite effective in helping accelerate the fleet composition towards lower-emitting vehicles. The optimal restriction design, which extends full exemptions on relatively clean cars and complete restrictions on the use of dirty ones, comes remarkably close to implement the first-best. The reason is that implementing the first-best requires not much to persuade households to drive less but to drive cleaner cars. If properly designed, these restriction policies can be way more effective in the fight against local air pollution than alternative instruments such as scrappage subsidies and gasoline taxes.

Despite the optimal restriction design follows the same all-or-nothing structure of existing "restriction" policies such as LEZs and smog checks, practical (e.g., monitoring, within vintage heterogeneity, etc.) and distributional considerations may require to move away from complete to partial bans on older vehicles, as done today in Mexico-City and Santiago. The model developed here is not capable of handling these more complex restriction designs, in part, because this would open up the possibility of a second-car effect. We would need to extend the model to allow cars to age heterogeneously and to permit multiple ownership. In doing so, it would also be useful to explore how to introduce congestion and traffic considerations that so far have been neglected (e.g., that the marginal harm from weekdays' pollution is higher than weekends', that congestion is not fully taken care of with alternative instruments, etc.).

Motivate by all restriction programs around the world, our focus has been on local air pollution. Our results, however, make a strong case for the use of driving restrictions to fight global air pollution as well; if well designed, they can help accelerate the transition

[38]Households in the rest of the country are not much affected, as seen in Figure 15b.

towards low- or free-carbon-emission vehicles at a lower cost for the government. The only issue there is to make sure the price of the carbon-free (and exempt) option is not that much higher than the price of the existing (dirtier) alternatives, so much that drivers opt to buy a second and older car to by-pass the restriction instead of the clean option. One important reason the Santiago-1992 reform work so well was precisely because the clean option (which was to switch to a car with a catalytic converter sooner than otherwise) was affordable to many.

References

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

[2] Adda, J. and R. Cooper (2000), Balladurette and Juppette: A Discrete Analysis of Scrapping Subsidies, *Journal of Political Economy* 108, 778-806.

[3] Bento, A., L. Goulder, M. Jacobsen and R. von Haefen (2009), Distributional and Efficiency Impacts of Increased US Gasoline Taxes, *American Economic Review* 99, 667-99.

[4] d'Haultfœuille, X., P. Givord and X. Boutin (2014), The environmental effect of green taxation: The case of the French bonus/malus, *The Economic Journal* 124, F444-F480.

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

[6] Davis, L., and M. Kahn (2010), International trade in used vehicles: The environmental consequences of NAFTA, *American Economic Journal*: *Economic Policy* 2, 58-82.

[7] Davis, L. (2015), The effect of Saturday driving restriction on air quality in Mexico City, mimeo, UC Berkeley.

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

[9] Fullerton, D., and L. Gan (2005), Cost-effective policies to reduce vehicle emissions, *American Economic Review* 95, 300–304.

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

[11] Gallego, F., J.-P. Montero, and C. Salas (2013b), The effect of transport policies on car use: A bundling model with applications, *Energy Economics* 40, S85-S97.

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

[13] Hahn, R. (1995), An economic analysis of scrappage, *RAND Journal of Economics* 26, 222-242.

[14] Hoekstra, M., S. Puller and J. West (2015), Cash for Corollas: When stimulus reduces spending, NBER working paper.

[15] Huang, W., and M. Chao (2014), The Impacts of the Mass Rapid Transit System on Household Car Ownership in Taipei, *J. Sustain. Dev. Energy Water Environ. Syst.* 2, 191-207.

[16] Jacobsen, M., C. Knittel, J. Sallee, and A. van Benthem (2015), Sufficient Statistics for Imperfect Externality-Correcting Policies, mimeo, UC Berkeley.

[17] Laffont, J.-J. (2005), *Regulation and Development*, Cambridge University Press, UK.

[18] Lin, C.-Y., V. Umanskaya and W. Zhang (2014), On the Design of Driving Restrictions: Theory and Empirical Evidence, Working paper, UC Davis.

[19] Liu, Y., Z. Yan, S. Liu, Y. Wu, and Q. Gan (2015), The effect of the driving restriction policy on public health in Beijing, Working paper, Beijing Institute of Technology.

[20] Lu, S. (2006), Vehicle survivability and travel mileage schedules, NHTSA Technical Report, US DOT, Springfield, VA.

[21] Matas, A., and J.-L. Raymond (2008), Changes in the structure of car ownership in Spain, *Transportation Research Part A: Policy and Practice* 42, 187-202.

[22] 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, 1107-1142.

[23] Molina, L., and M. Molina (2002), Eds., *Air Quality in the Mexico Megacity: An Integrated Assessment*, Kluwer Academic Publishers, Dordrecht.

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

[25] Onursal, B. and S.P. Gautam (1997), Vehicular air pollution: Experiences from seven Latin American urban centers, World Bank Technical Paper, Washington, DC.

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

[27] Posada, F., D. Wagner, G. Bansal, and R. Fernandez (2015), Survey of best practices in reducing emissions through vehicle replacement programs, White paper, The International Council on Clean Transportation.

[28] Rivers, N., and B. Shaufele (2015), New vehicle feebates, mimeo, University of Western Ontario.

[29] Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data*, Second Edition, MIT Press, Cambridge, MA.

Figures and Tables

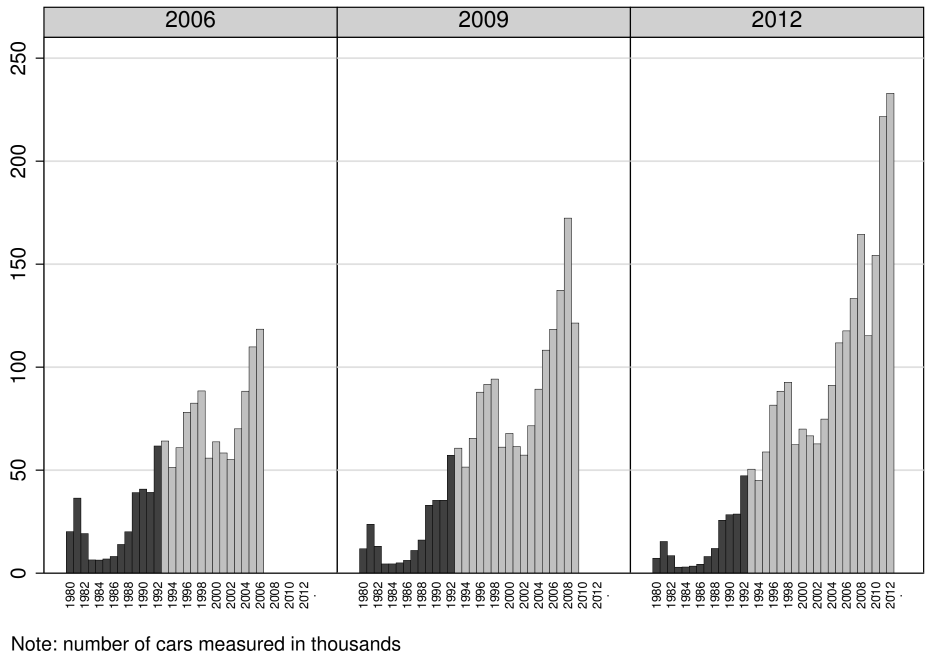


Figure 1: Number of cars by vintage for different years

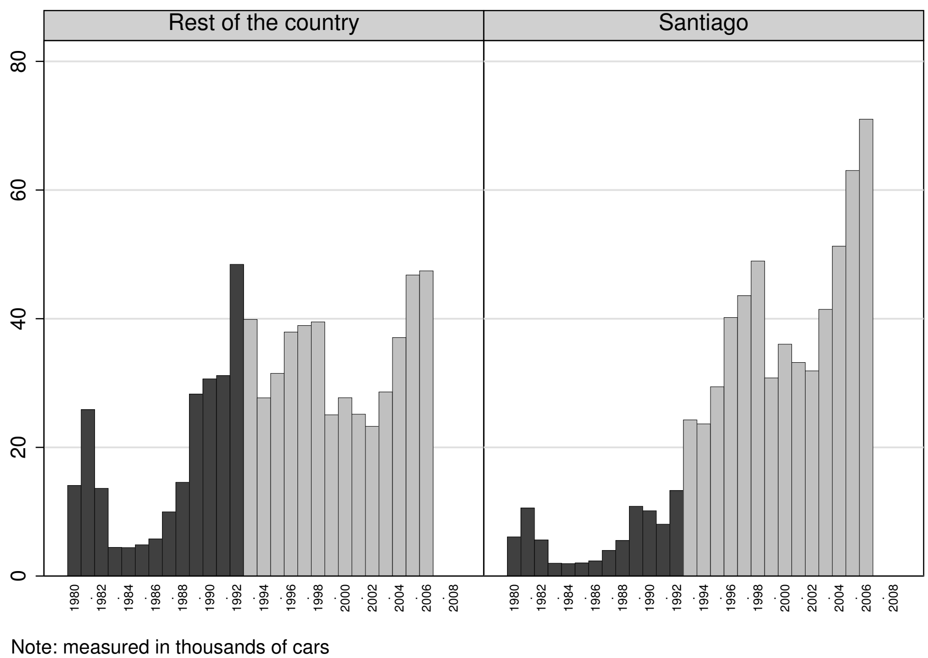


Figure 2: Santiago's fleet vs rest of country's in 2006

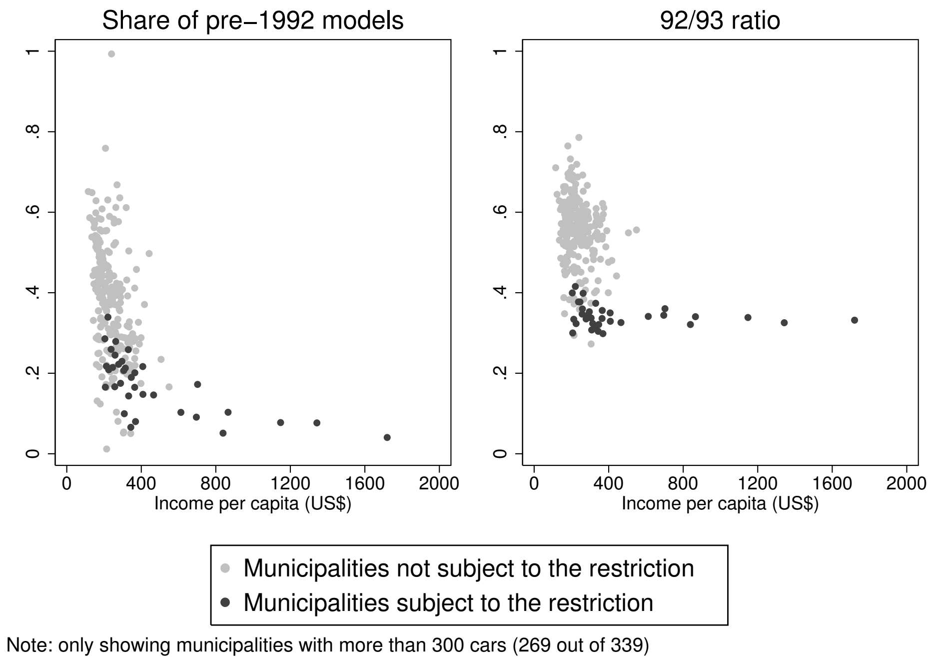
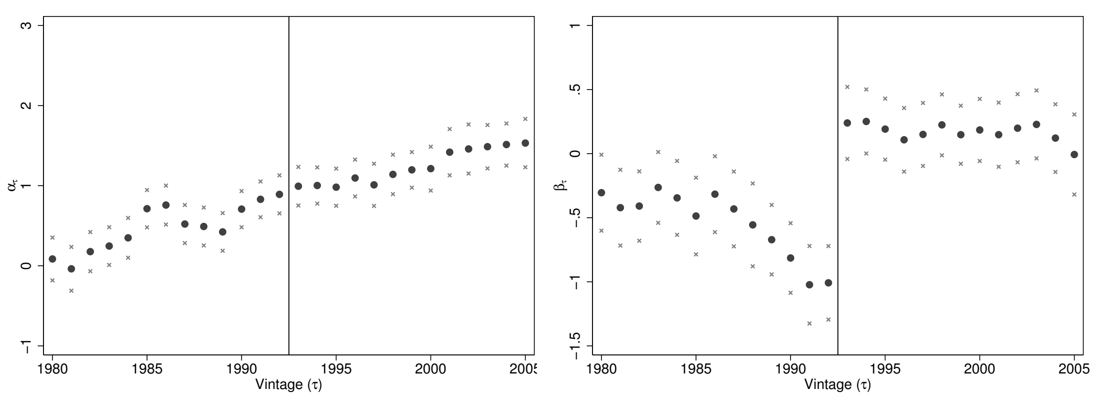


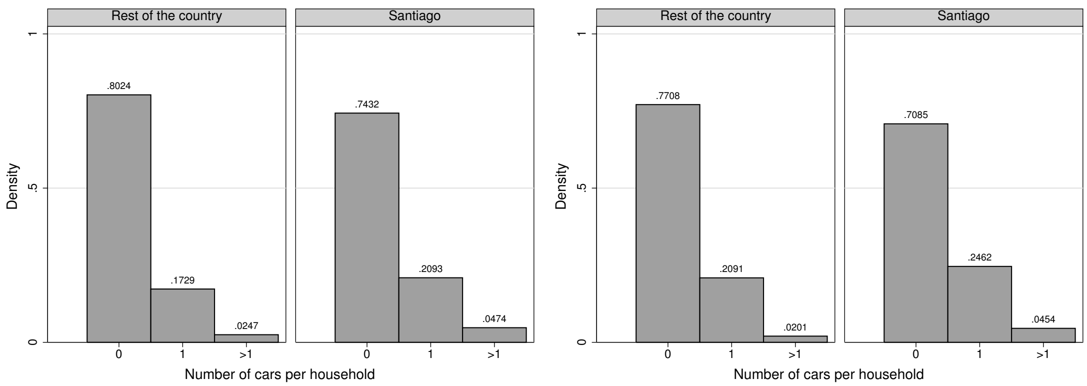
Figure 3: Municipalities' fleet composition as a function of income and restriction status



(a) $INCOME$ coefficients

(b) DR coefficients

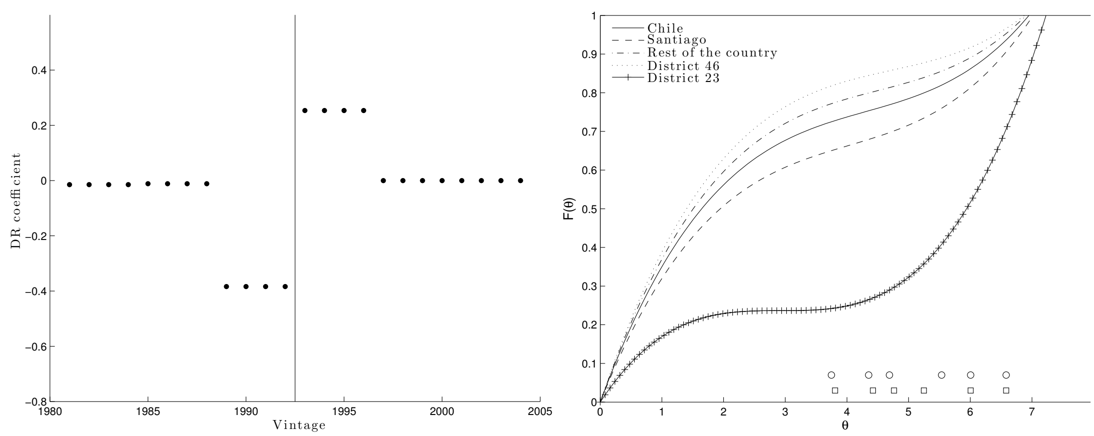
Figure 4: Income and driving-restriction effects by vintage



(a) 1998 sample

(b) 2006 sample

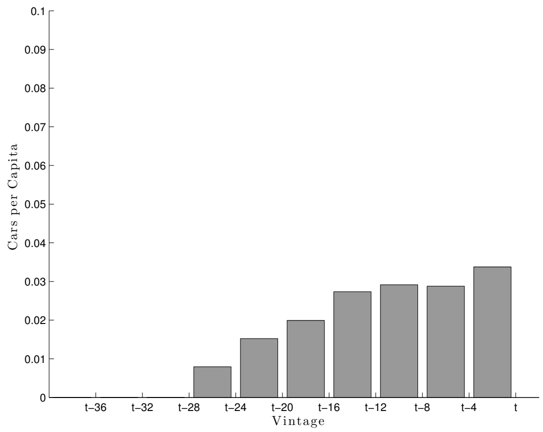
Figure 5: Histogram with numbers of cars per household



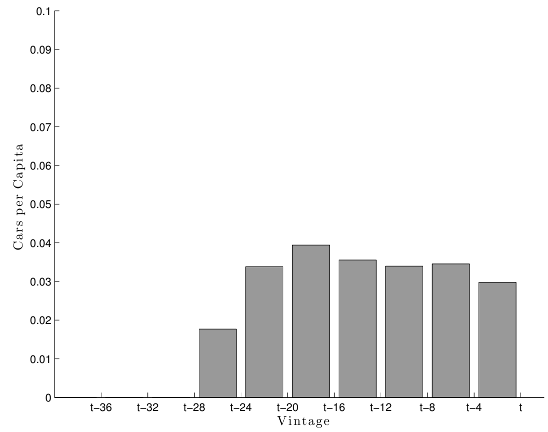
(a) Model prediction for DR coefficients

(b) Estimated CDF for different groups

Figure 6: Selected calibration results

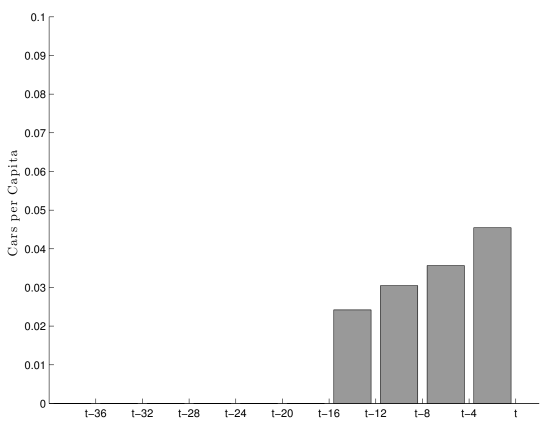


(a) Santiago

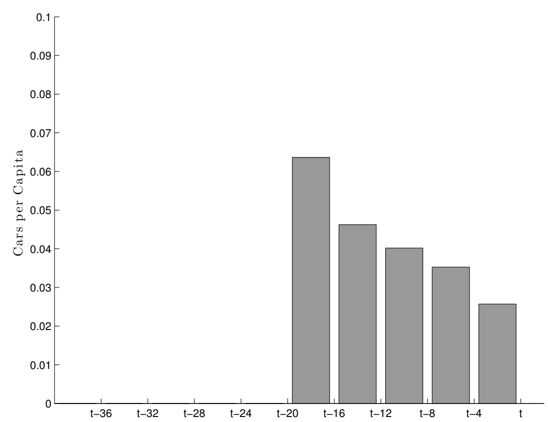


(b) Rest of the country

Figure 7: Steady-state fleet composition under no intervention



(a) Santiago



(b) Rest of the country

Figure 8: Steady-state fleet composition under the first best

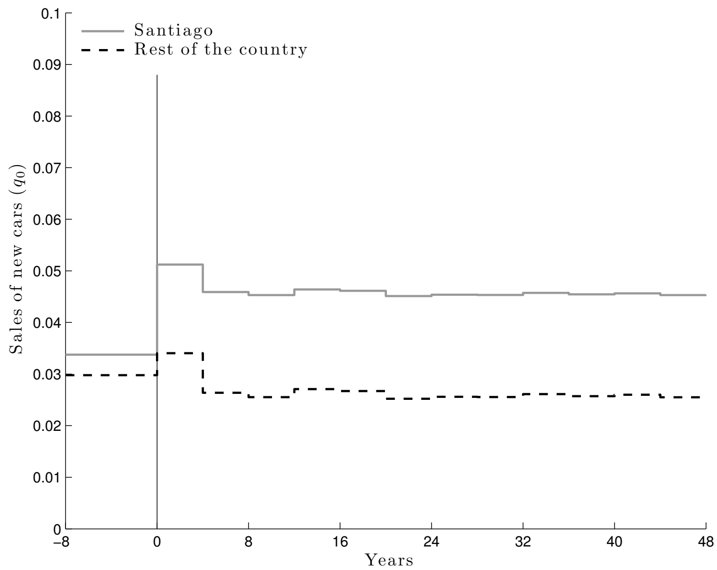
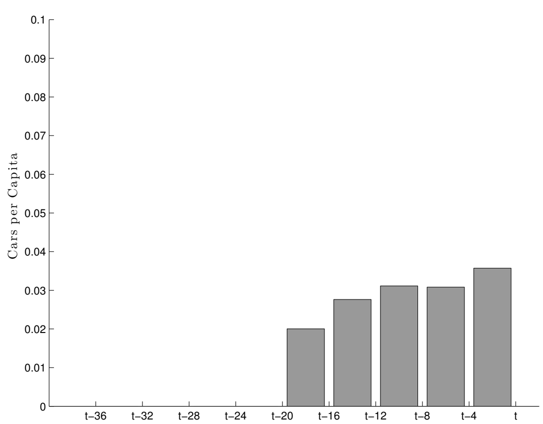
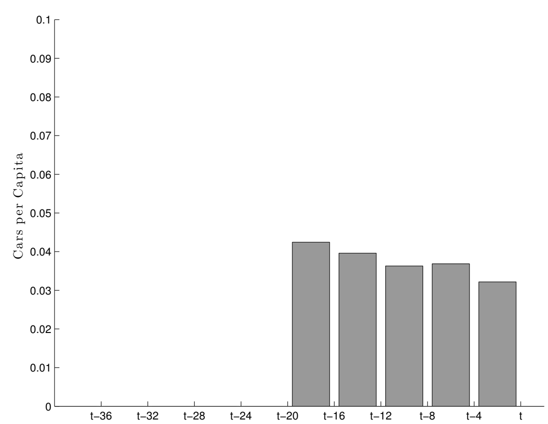


Figure 9: Transition phase: Sales of new cars

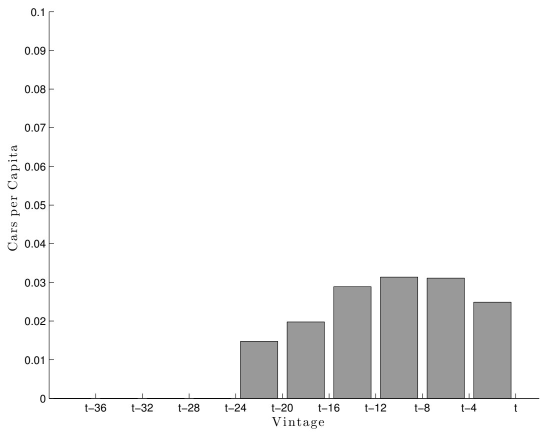


(a) Santiago

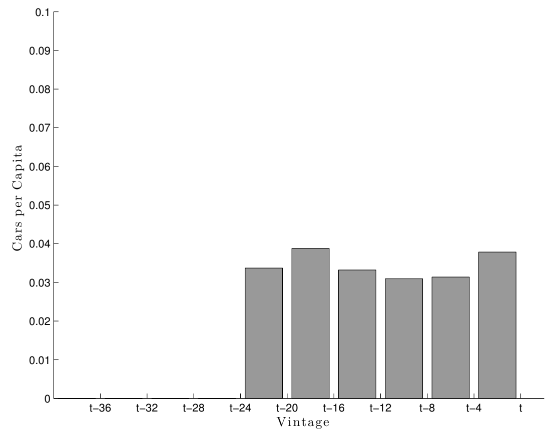


(b) Rest of the country

Figure 10: Steady-state fleet composition under the optimal scrappage subsidy ($1575)

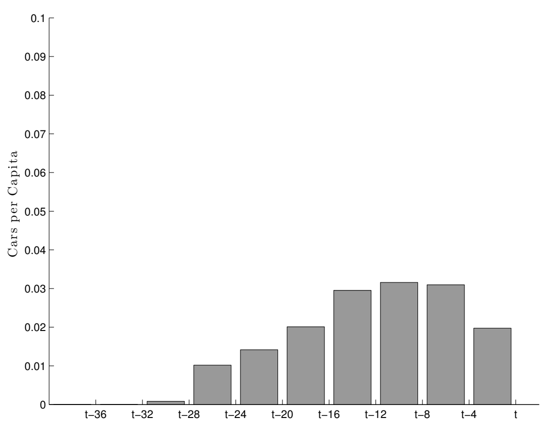


(a) Santiago

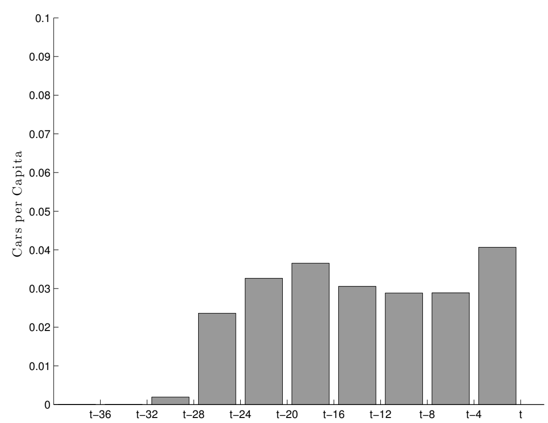


(b) Rest of the country

Figure 11: Steady-state fleet composition under an optimal gasoline tax in Santiago (¢2.5 per mile)

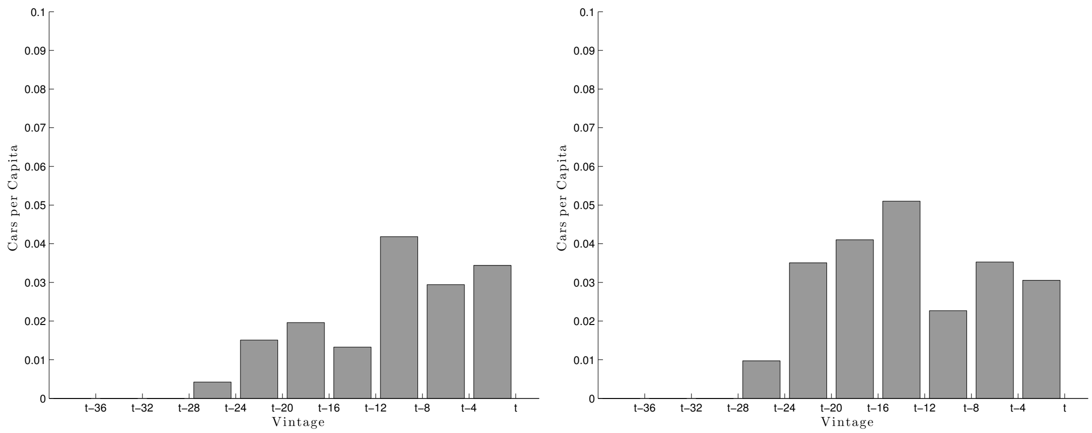


(a) Santiago



(b) Rest of the country

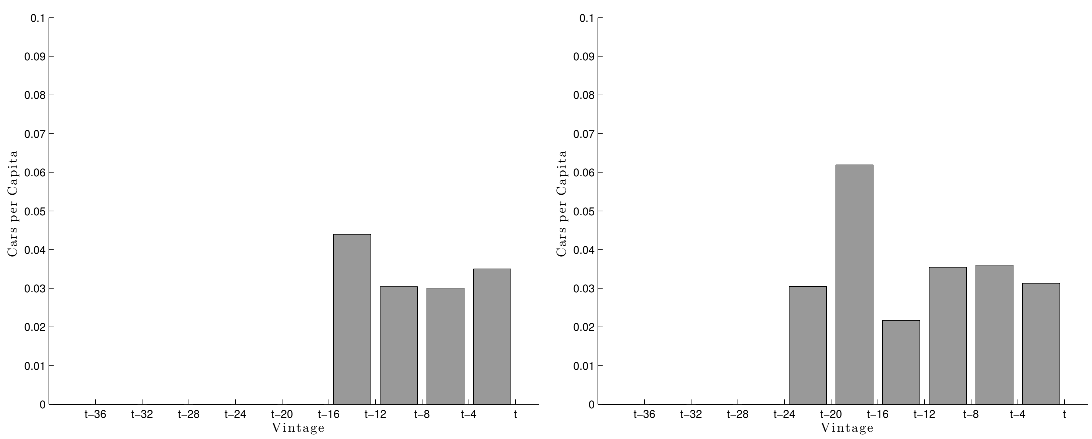
Figure 12: Steady-state fleet composition under a uniform driving restriction



(a) Santiago

(b) Rest of the country

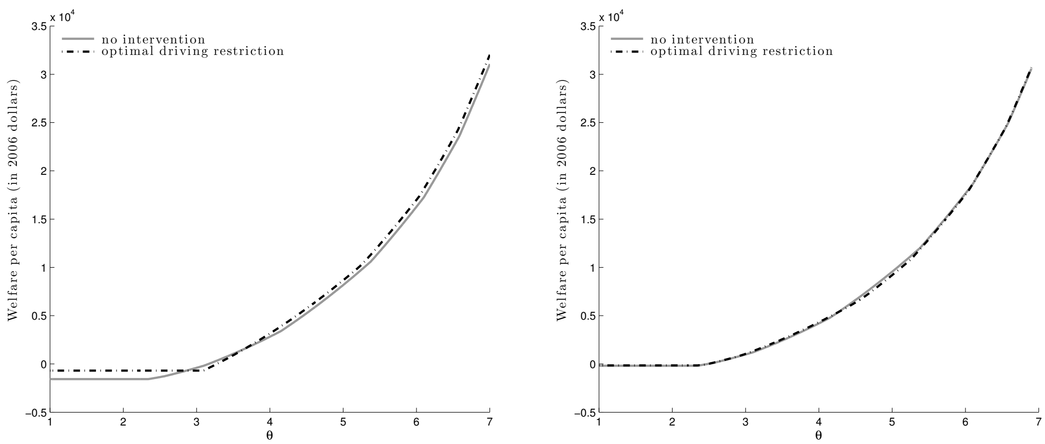
Figure 13: Steady-state fleet composition under a driving restriction that exempts cars 12 years old and younger



(a) Santiago

(b) Rest of the country

Figure 14: Steady-state fleet composition under the optimal driving restriction



(a) Santiago (b) Rest of the country

Figure 15: Distributional implications of the optimal driving restriction

Table 1: OLS results for different contiguous-year ratios

	(1) 88-89	(2) 91-92	(3) 92-93	(4) 93-94	(5) 95-96
Driving restriction	0.0176	0.00167	-0.183***	-0.0225	-0.00801
	(0.015)	(0.013)	(0.019)	(0.015)	(0.013)
Population	-0.00108	0.00228	0.00218	-0.0000718	0.000867
	(0.005)	(0.005)	(0.007)	(0.005)	(0.004)
Income per capita	-0.00112	-0.00527	-0.00694	-0.00659	-0.0104**
	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)
Distance to Santiago	-0.0578**	-0.0145	0.128***	0.0139	0.000246
	(0.026)	(0.024)	(0.034)	(0.027)	(0.023)
(Distance to Santiago)^2	0.0288	0.0197	-0.0838***	0.00385	0.00718
	(0.020)	(0.018)	(0.025)	(0.021)	(0.017)
Extreme regions	0.0835**	-0.0435	0.0265	0.127***	0.0852***
	(0.034)	(0.030)	(0.043)	(0.035)	(0.029)
Income dispersion	0.00262	-0.000744	-0.00132	-0.00599	0.00464
	(0.006)	(0.005)	(0.007)	(0.006)	(0.005)
North	0.0208*	0.0402***	-0.0232	0.0328***	-0.0229**
	(0.012)	(0.011)	(0.015)	(0.012)	(0.010)
Urbanization	-0.0478***	-0.0286*	-0.0123	-0.00460	0.0120
	(0.017)	(0.015)	(0.021)	(0.017)	(0.014)
Constant	0.370***	0.413***	0.552***	0.557***	0.443***
	(0.014)	(0.013)	(0.018)	(0.015)	(0.012)
Observations	268	268	268	268	267
R^2	0.152	0.086	0.523	0.320	0.178

Notes: OLS regressions with one observation per municipality. Municipalities with less than 300 cars were dropped from the sample. Standard errors in parentheses.

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

Table 2: Effect of living in Santiago on having more than one car

	1998 survey	2006 survey
OLS	0.0159	0.00999
	(0.015)	(0.014)
Probit	0.0103	0.00310
	(0.014)	(0.011)
Hurdle poisson-logit	0.062	0.0136
	(0.081)	(0.101)

Notes: We present only the marginal effects of living in Santiago but the models also include the following variables: household characteristics related to income, assets, age, gender and employment status of the head of the household, the composition of the household (in terms of number of members and also number of employed members), and the size of the county in which the household is located. OLS and probit estimations are on households with at least one car. The Hurdle poisson-logit model uses all the observations. Observations are weighted using expansion factors. Standard errors clustered at the county level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Effect of driving restriction on prices (1993-2000)

	Fiat Uno	Honda Accord	Honda Civic	Mazda 323	Peugeot 205	Peugeot 505	Toyota Corolla
Catalytic	0.0458***	0.162***	0.0633***	0.0459***	0.0378***	0.149***	0.180***
	(0.006)	(0.008)	(0.007)	(0.006)	(0.007)	(0.008)	(0.009)
Age f.e.	yes	yes	yes	yes	yes	yes	yes
Offer date f.e.	yes	yes	yes	yes	yes	yes	yes
Observations	4136	5980	5530	5796	3396	6788	5764
R^2	0.930	0.966	0.924	0.950	0.937	0.934	0.941

Notes: OLS regressions with age and date fixed effects. Standard errors clustered by offer date in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Effect of driving restriction on Honda Accord prices

	(1991)	(1992)	(1993)	(1994)
Catalytic	0.223***	0.189***	0.0206	-0.00487
	(0.054)	(0.035)	(0.035)	(0.010)
Constant	15.60***	15.68***	15.96***	16.40***
	(0.032)	(0.034)	(0.025)	(0.010)
Observations	47	53	58	49
R^2	0.245	0.309	0.006	0.001

Notes: OLS regressions without controls. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Fleet composition and pollution contribution

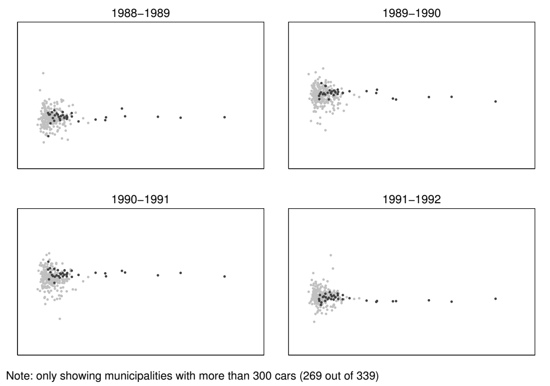
Car vintage	Fleet Percent Share (Mexico)	Emissions Contribution (Mexico)	Fleet Percent Share (Chile)
1993-2001	60%	15%	63.3%
1985-1992	28%	30%	24.1%
1980-1985	7%	25%	9.3%
1979 & older	5%	30%	3.3%

Source: information in columns 1 and 2 comes from Molina and Molina (2002).

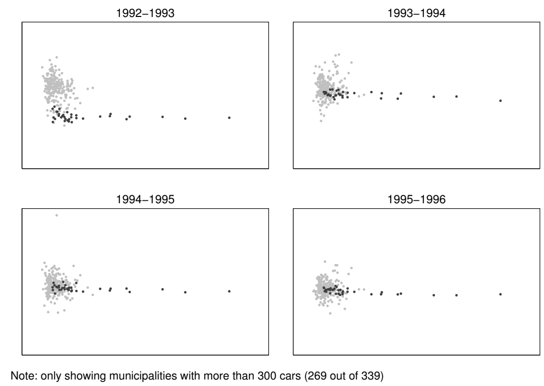
Table 6: Welfare calculations

Counterfactual	Welfare per capita (in 2006 dollars)	Welfare gain/loss (relative to first-best)
No intervention	5528.5	0
First best	5852.3	100
Optimal subsidy ($1575)	5765.1	73.06
Optimal gasoline tax (¢2.5 per mile)	5642.0	35.08
Driving restriction with no exemptions ($R = 0.9\ \forall\tau$)	5032.6	-153.13
Driving restriction with some exemptions ($R = 0.9$, $\tau > 3$)	5606.5	24.09
Optimal driving restriction ($R = 0$, $\tau > 4$)	5815.5	88.64

Appendix A: Figures



(a) Vintages 88 to 92



(b) Vintages 92 to 96

Figure A.1: Falsification excersice with different ratios

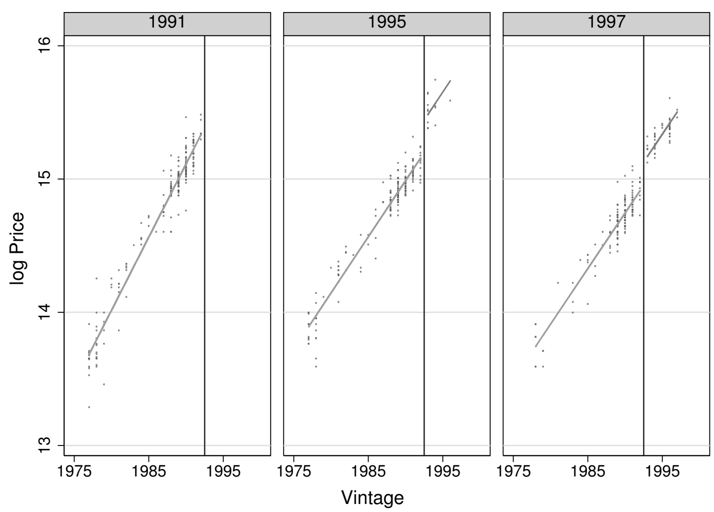


Figure A.2: Price of a second-hand Toyota Corolla

Appendix B: Tables

Table B.1: Survival rates

Age	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36
γ_τ	0.9966	0.9966	0.9966	0.9434	0.8267	0.7226	0.5828	0.5242	0.5242