Online Appendix for:
The Welfare Impact of Second Best Uniform-Pigouvian Taxation: Evidence from Transportation

Christopher R. Knittel and Ryan Sandler

A Proofs of Propositions

Proposition 1. *The second-best-optimal uniform per-gallon gasoline tax, τ^*, is* (from Diamond (1973))*:*

$$\tau^* = \frac{-\sum_i \sum_{h \neq i} \frac{\partial U^h}{\partial \alpha_i} \alpha_i'}{\sum_i \alpha_i'}. \tag{1}$$

where α_i' is the derivative of consumer i's demand for gasoline with respect to the price of gasoline.

Proof. Consumers have quasi-linear utility functions, given as:

$$\max_{\alpha_h} \quad U^h(\alpha_1, \alpha_2, ..., \alpha_h, ..., \alpha_n) + \mu_h, \tag{2}$$

$$s.t. \quad (p_g + \tau)\alpha_h + \mu_h = m_h, \tag{3}$$

where p_g is the price, τ the tax per gallon, α_h the consumption of the polluting good by consumer h, μ_h consumption of a numeraire, and m_h consumer h's income. Assuming an interior solution, we have:

$$\frac{\partial U^h}{\partial \alpha_h} = (p + \tau). \tag{4}$$

This yields demand curves, which we represent by α_h^*, given by:

$$\alpha_h^* = \alpha_h \left(p_g + \tau\right). \tag{5}$$

The SBO gasoline tax maximizes social welfare, or the sum of utilities:

$$W(\tau) = \sum_h U^h[\alpha_1^*, ..., \alpha_h^*, ...\alpha_n^*] - p_g \sum_h \alpha_h^* + \sum_h m_h. \tag{6}$$

The first-order condition for the SBO gasoline tax is given as:

$$W'(\tau) = \sum_i \sum_h \frac{\partial U^h}{\partial \alpha_i} \alpha_i' - p_g \sum_h \alpha_h' = 0. \tag{7}$$

Rewriting this and plugging in the result from the consumers' problem, $\frac{\partial U^h}{\partial \alpha_h} - p_g = \tau$, we have:

$$W'(\tau) = \sum_i \sum_{h \neq i} \frac{\partial U^h}{\partial \alpha_i} \alpha_i' + \tau \sum_i \alpha_i' = 0. \tag{8}$$

Solving for the second-best tax yields:

$$\tau^* = \frac{-\sum_i \sum_{h \neq i} \frac{\partial U^h}{\partial \alpha_i} \alpha_i'}{\sum_i \alpha_i'}. \tag{9}$$

□

Proposition 2. *Suppose drivers are homogenous in their demand for gasoline, but vehicles' per-gallon emissions differ. In particular, let β denote the derivative of the demand for gasoline with respect to the price of gasoline.*

If the distribution of the per-gallon externality, E, is log normal, with probability density function:

$$\varphi(E_i) = \frac{1}{E_i \sqrt{2\sigma_E^2}} \exp\left(\frac{-(E_i - \mu_E)^2}{2\sigma_E^2} \right), \tag{10}$$

the DWL absent any market intervention will be given as:

$$D = \frac{1}{2\beta} e^{2\mu_E + 2\sigma_E^2}.$$

Proof. Given these assumptions, the deadweight loss absent any market intervention will be given as:

$$\begin{aligned} D &= \int_0^\infty \frac{(E_i)^2}{2\beta} \varphi(E_i) dE_i \\ &= \frac{1}{2\beta} \mathbb{E}[E_i^2] \\ &= \frac{1}{2\beta} e^{2\mu_E + 2\sigma_E^2}. \end{aligned} \tag{11}$$

□

Proposition 3. *Under the assumptions in Proposition 1, the ratio of the remaining DWL with the deadweight loss after the tax is:*

$$R = \frac{D - \frac{e^{2\mu_E + \sigma_E^2}}{2\beta}}{D} = 1 - \frac{e^{2\mu_E + \sigma_E^2}}{e^{2\mu_E + 2\sigma_E^2}} = 1 - e^{-\sigma_E^2}. \tag{12}$$

Proof. The level of the externality is given as:

$$\overline{E} = \tau = e^{\mu_E + \sigma_E^2/2}. \tag{13}$$

The deadweight loss associated with all vehicles is given as:

$$\begin{aligned}
D(\tau) &= \int_0^\infty \frac{(\tau - E_i)^2}{2\beta} \varphi(E_i) dE_i \\
&= \frac{1}{2\beta} \mathbb{E}[\tau^2 - 2\tau E_i + E_i^2] \\
&= \frac{1}{2\beta} (\tau^2 - 2\tau \mathbb{E}[E_i] + \mathbb{E}[E_i^2]) \\
&= \frac{1}{2\beta} (\tau^2 - 2\tau e^{\mu_E + \frac{\sigma_E^2}{2}} + e^{2\mu_E + 2\sigma_E^2}) \\
&= \frac{1}{2\beta} (\tau^2 - 2\tau e^{\mu_E + \frac{\sigma_E^2}{2}}) + D \\
&= D - \frac{e^{2\mu_E + \sigma_E^2}}{2\beta}.
\end{aligned} \tag{14}$$

The ratio of remaining DWL with the deadweight loss absent the tax is therefore:

$$R = \frac{D - \frac{e^{2\mu_E + \sigma_E^2}}{2\beta}}{D} = 1 - \frac{e^{2\mu_E + \sigma_E^2}}{e^{2\mu_E + 2\sigma_E^2}} = 1 - e^{-\sigma_E^2}. \tag{15}$$

□

Proposition 4. *When $B_i = \frac{1}{\beta_i}$ and E_i are distributed lognormal with dependence parameter ρ, the optimal tax, represented by τ^*, is:*

$$\tau^* = e^{\mu_E + \frac{\sigma_E^2}{2} + \rho \sigma_E \sigma_B}$$

Proof. The slope of the demand curve with respect to the cost of driving, defined as $B_i = \frac{1}{\beta_i}$, where β_i is the VMT elasticity for the vehicle owned by consumer i is distributed lognormal with parameters μ_B and σ_B^2. ρ is the dependence parameter of the bivariate lognormal distribution (the correlation coefficient of $\ln E$ and $\ln B$). The optimal tax is:

$$\begin{aligned}
\tau^* &= \frac{\sum E_i \beta_i}{\sum \beta_i} \\
&= \frac{\frac{1}{N} \sum E_i \beta_i}{\frac{1}{N} \sum \beta_i} \\
&= \frac{\mathbb{E}[E_i \beta_i]}{\mathbb{E}[\frac{1}{B_i}]} \\
&= \frac{e^{\mu_E + \frac{\sigma_E^2}{2} - \mu_B + \frac{\sigma_B^2}{2}} e^{\rho \sigma_E \sigma_B}}{e^{-\mu_B + \frac{\sigma_B^2}{2}}} \\
&= e^{\mu_E + \frac{\sigma_E^2}{2} + \rho \sigma_E \sigma_B}.
\end{aligned} \tag{16}$$

□

Proposition 5. *When $B_i = \frac{1}{\beta_i}$ and E_i are distributed lognormal with dependence parameter ρ, the ratios of the remaining deadweight loss after the SBO gasoline tax to the original deadweight loss will be:*

$$R(\tau^*) = 1 - e^{-\sigma_E^2}, \tag{17}$$

and, the ratios of the remaining deadweight loss after the naive uniform tax to the original deadweight loss will be:

$$R(\tau_{naive}) = 1 - e^{-\sigma_E^2}(2e^{-\rho\sigma_E\sigma_B} - e^{-2\rho\sigma_E\sigma_B}). \tag{18}$$

Proof. The deadweight loss with no gasoline tax is:

$$\begin{aligned}
\mathcal{D} &= \int_0^\infty \left(\int_0^\infty \frac{(E_i)^2 B_i}{2} \varphi(E_i) dE_i \right) \varphi_B(B_i) dB_i \\
&= \frac{1}{2}\mathbb{E}[E_i^2 B_i] \\
&= \frac{1}{2} e^{2\mu_E + 2\sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + 2\rho\sigma_E\sigma_B}.
\end{aligned} \tag{19}$$

The deadweight loss with the optimal uniform tax is:

$$\begin{aligned}
\mathcal{D}(\tau^*) &= \int_0^\infty \left(\int_0^\infty \frac{(\tau - E_i)^2 B_i}{2} \varphi(E_i) dE_i \right) \varphi_B(B_i) dB_i \\
&= \frac{1}{2}\mathbb{E}[\tau^2 B_i - 2\tau E_i B_i + E_i^2 B_i] \\
&= \frac{1}{2}(\tau^2 \mathbb{E}[B_i] - 2\tau\mathbb{E}[E_i B_i] + \mathbb{E}[E_i^2 B_i]) \\
&= \frac{1}{2}(\tau^2 e^{\mu_B + \frac{\sigma_B^2}{2}} - 2\tau e^{\mu_E + \frac{\sigma_E^2}{2} + \mu_B + \frac{\sigma_B^2}{2} + \rho\sigma_E\sigma_B} + e^{2\mu_E + 2\sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + 2\rho\sigma_E\sigma_B}) \\
&= \frac{1}{2} e^{2\mu_E + \sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + 2\rho\sigma_E\sigma_B} - e^{2\mu_E + \sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + 2\rho\sigma_E\sigma_B} + \mathcal{D} \\
&= \mathcal{D} - \frac{1}{2} e^{2\mu_E + \sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + 2\rho\sigma_E\sigma_B},
\end{aligned} \tag{20}$$

while the deadweight loss with the naive tax, equal to the average externality level is:

$$\mathcal{D}(\tau_{naive}) = \mathcal{D} - \frac{1}{2}(2e^{2\mu_E + \sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + \rho\sigma_E\sigma_B} - e^{2\mu_E + \sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2}}). \tag{21}$$

Then the ratios of the remaining deadweight loss after a tax to the original deadweight loss will be:

$$\begin{aligned}
R(\tau^*) &= 1 - \frac{e^{2\mu_E + \sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + 2\rho\sigma_E\sigma_B}}{e^{2\mu_E + 2\sigma_E^2 + \mu_B + \frac{\sigma_B^2}{2} + 2\rho\sigma_E\sigma_B}} \\
&= 1 - e^{-\sigma_E^2},
\end{aligned} \tag{22}$$

$$
\begin{aligned}
R(\tau_{naive}) &= 1 - \frac{2e^{2\mu_E+\sigma_E^2+\mu_B+\frac{\sigma_B^2}{2}\rho\sigma_E\sigma_B} - e^{2\mu+\sigma_E^2+\mu_B+\frac{\sigma_B^2}{2}}}{e^{2\mu_E+2\sigma_E^2+\mu_B+\frac{\sigma_B^2}{2}+2\rho\sigma_E\sigma_B}} \\
&= 1 - e^{-\sigma_E^2}(2e^{-\rho\sigma_E\sigma_B} - e^{-2\rho\sigma_E\sigma_B}). \qquad (23)
\end{aligned}
$$

□

B Steps to Clean Smog Check Data

California implemented its first inspection and maintenance program (the Smog Check Program) in 1984 in response to the 1977 Clean Air Act Amendments. The 1990 Clean Air Act Amendments required states to implement an enhanced inspection and maintenance program in areas with serious to extreme non-attainment of ozone limits. Several of California's urban areas fell into this category, and in 1994, California's legislature passed a redesigned inspection program was passed by California's legislature after reaching a compromise with the EPA. The program was updated in 1997 to address consumer complaints, and fully implemented by 1998. Among other improvements, California's new program introduced a system of centralized "Test-Only" stations and an electronic transmission system for inspection reports.[1] Today, more than a million smog checks take place each month.

Since 1998, the state has been divided into three inspection regimes (recently expanded to four), the boundaries of which roughly correspond to the jurisdiction of the regional Air Quality Management Districts. "Enhanced" regions, designated because they fail to meet state or federal standards for CO and ozone, fall under the most restrictive regime. All of the state's major urban centers are in Enhanced areas, including the greater Los Angeles, San Francisco, and San Diego metropolitan areas. Vehicles registered to an address in an Enhanced area must pass a biennial smog check in order to be registered, and they must take the more rigorous Acceleration Simulation Mode (ASM) test. The ASM test involves the use of a dynamometer, and allows for measurement of NO_x emissions. In addition, a randomly selected two percent sample of all vehicles in these areas is directed to have their smog checks at Test-Only stations, which are not allowed to make repairs.[2] Vehicles that match a "High Emitter Profile" are also directed to Test-Only stations, as are vehicles that are flagged as "gross polluters" (those that fail an inspection with twice the legal limit of one or more pollutant in emissions). More recently some "Partial-Enhanced" areas that require a biennial ASM test have been added, but no vehicles are directed to Test-Only stations.

Areas with poor air quality not exceeding legal limits fall under the Basic regime. Cars in a Basic area must have biennial smog checks as part of registration, but they are allowed to take the simpler Two Speed Idle (TSI) test and are not directed to Test-Only stations. The least restrictive regime, consisting of rural mountain and desert counties in the east and north, is known as the Change of Ownership area. As the name suggests, inspections in these areas are only required upon change of ownership; no biennial smog check is required.

[1] For more detailed background see http://www.arb.ca.gov/msprog/smogcheck/july00/if.pdf.

[2] Other vehicles can be taken to Test-Only stations as well if the owner chooses, although they must get repairs elsewhere if they fail.

Our data from the Smog Check Program essentially comprise the universe of test records from January 1, 1996 to December 31, 2010. We were able to obtain test records only going back to 1996 because this was the year when the Smog Check Program introduced its electronic transmission system. Because the system seems to have been phased in during the first half of 1996, and major program changes took effect in 1998 we limit our sample to test records from January 1998 on. For our analyses, we use a 10 percent sample of VINs, selecting by the second to last digit of the VIN. We exclude tests that have no odometer reading, with a test result of "Tampered" or "Aborted" and vehicles that have more than 36 tests in the span of the data. Vehicles often have multiple smog check records in a year, whether due to changes of ownership or failed tests, but we argue that more than 36 in what is at most a 12 year-span indicates some problem with the data.[3]

A few adjustments must be made to accurately estimate VMT and emissions per mile.

First, we adjust odometer readings for "roll overs" and typos. Many of the vehicles in our analysis were manufactured with 5-digit odometers–that is, five places for whole numbers plus a decimal. As such, any time one of these vehicles crosses over 100,000 miles, the odometer "rolls over" back to 0. To complicate matters further, sometimes either the vehicle owner or smog check technician notices this problem and records the appropriate number in the 100,000s place, and sometimes they do not. To address this problem, we employ an algorithm that increases the hundred thousands place in the odometer reading whenever a rollover seems to have occurred. The hundred thousands are incremented if the previous test record shows higher mileage, or if the next test record is shows more than 100,000 additional miles on the odometer (indicating that the odometer had already rolled over, but the next check took this into account). The algorithm also attempts to correct for typos and entry errors. An odometer reading is flagged if it does not fit with surrounding readings for the same vehicle–either it is less than the previous reading or greater than the next–and cannot be explained by a rollover. The algorithm then tests whether fixing one of several common typos will make the flagged readings fit (e.g., moving the decimal over one place). If no correction will fit, the reading is replaced with the average of the surrounding readings. Finally, if after all our corrections any vehicle has an odometer reading above 800,000 or has implied VMT per day greater than 200 or less than zero, we exclude the vehicle from our analysis. All of our VMT analyses use this adjusted mileage.

Emissions results from smog checks are given in either parts per million (for HC and NO_x) or percent (O_2, CO, and CO_2). Without knowing the volume of air involved, there is no straightforward way to convert this to total emissions. Fortunately, as part of an independent evaluation of the Smog Check Program conducted in 2002-2003, Sierra Research Inc. and Eastern Research Group estimated a set of conversion equations to convert the proportional measurements of the ASM test to emissions in grams per mile traveled. These equations are reported in Morrow and Runkle (2005), NO_x, and CO, and estimate grams per mile for each pollutant as a non-linear function of all three pollutants, model year, and vehicle weight.

[3]For instance, there is one vehicle in particular, a 1986 Volvo station wagon, which has records for more than 600 smog checks between January 1996 and March 1998. The vehicle likely belonged to a smog check technician who used it to test the electronic transmission system.

The equations for vehicles of up to model year 1990 are

$$FTP\_HC = 1.2648 \cdot exp(-4.67052 \quad +0.46382 \cdot HC^* + 0.09452 \cdot CO^* + 0.03577 \cdot NO^* +0.57829 \cdot \ln(weight) - 0.06326 \cdot MY^* +0.20932 \cdot TRUCK)$$

$$\begin{aligned} FTP\_CO = 1.2281 \cdot exp(-2.65939 \quad & +0.08030 \cdot HC^* + 0.32408 \cdot CO^* + 0.03324 \cdot CO^{*2} \\ & +0.05589 \cdot NO^* + 0.61969 \cdot \ln(weight) - 0.05339 \cdot MY^* \\ & +0.31869 \cdot TRUCK) \end{aligned}$$

$$\begin{aligned} FTP\_NOX = 1.0810 \cdot exp(-5.73623 \quad & +0.06145 \cdot HC^* - 0.02089 \cdot CO^{*2} + 0.44703 \cdot NO^* \\ & +0.04710 \cdot NO^{*2} + 0.72928 \cdot \ln(weight) - 0.02559 \cdot MY^* \\ & -0.00109 \cdot MY^{*2} + 0.10580 \cdot TRUCK) \end{aligned}$$

Where

$$\begin{aligned} HC^* &= \ln((Mode1_{HC} \cdot Mode2_{HC})^{.5}) - 3.72989 \\ CO^* &= \ln((Mode1_{CO} \cdot Mode2_{CO})^{.5}) + 2.07246 \\ NO^* &= \ln((Mode1_{NO} \cdot Mode2_{NO})^{.5}) - 5.83534 \\ MY^* &= modelyear - 1982.71 \\ weight &= \text{Vehicle weight in pounds} \\ TRUCK &= 0 \text{ if a passenger car, } 1 \text{ otherwise} \end{aligned}$$

And for model years after 1990 they are:

$$\begin{aligned} FTP\_HC = 1.1754 \cdot exp(-6.32723 \quad & +0.24549 \cdot HC^* + 0.09376 \cdot HC^{*2} + 0.06653 \cdot NO^* \\ & +0.01206 \cdot NO^{*2} + 0.56581 \cdot \ln(weight) - 0.10438 \cdot MY^* \\ & -0.00564 \cdot MY^{*2} + 0.24477 \cdot TRUCK) \end{aligned}$$

$$\begin{aligned} FTP\_CO = 1.2055 \cdot exp(-0.90704 \quad & +0.04418 \cdot HC^{*2} + 0.17796 \cdot CO^* + 0.08789 \cdot NO^* \\ & +0.01483 \cdot NO^{*2} - 0.12753 \cdot MY^* - 0.00681 \cdot MY^{*2} \\ & +0.37580 \cdot TRUCK) \end{aligned}$$

$$\begin{aligned} FTP\_NOX = 1.1056 \cdot exp(-6.51660 \quad & +0.25586 \cdot NO^* + 0.04326 \cdot NO^{*2} + 0.65599 \cdot \ln(weight) \\ & -0.09092 \cdot MY^* - 0.00998 \cdot MY^{*2} + 0.24958 \cdot TRUCK) \end{aligned}$$

Where:

$$\begin{aligned}
HC^* &= \ln((Mode1_{HC} \cdot Mode2_{HC})^{.5}) - 2.32393 \\
CO^* &= \ln((Mode1_{CO} \cdot Mode2_{CO})^{.5}) + 3.45963 \\
NO^* &= \ln((Mode1_{NO} \cdot Mode2_{NO})^{.5}) - 3.71310 \\
MY^* &= modelyear - 1993.69 \\
weight &= \text{Vehicle weight in pounds} \\
TRUCK &= 0 \text{ if a passenger car, } 1 \text{ otherwise}
\end{aligned}$$

C Steps to Clean DMV Data

We deal with two issues associated with DMV data. The main issue is that DMV entries for the same addresses will often have slightly different formats. For example, "12 East Hickory Street" may show up as "12 East Hickory St," "12 E. Hickory St.", etc. To homogenize the entries, we input each of the DMV entries into mapquest.com and then replace the entry with the address that mapquest.com gives.

Second, the apartment number is often missing in DMV data. Missing apartment numbers has the effect of yielding a large number of vehicles in the same "location." We omit observations that have over seven vehicles in a given address or more than three last names of registered owners.

D Robustness Checks

In this appendix, we report the results of several robustness checks to our main results on the intensive margin. Table A.2 reports elasticities by quartile for all five categories of externality.

Our base specification controls for the fixed effect of a each NO_x quartile on miles traveled. One might be concerned, however, that variation in dollars per mile (DPM) might be correlated with other characteristics such as age, odometer, and demographics, and that the DPM-quartile interactions may be picking up this correlation, rather than true heterogeneity. To test for this, in Table A.5 we present results with vehicle fixed effects and interactions between NO_x quartiles and various control variables. Adding these interaction terms actually makes the heterogeneity in the effect of DPM more pronounced. The final column in Table A.5 includes month-by-year fixed effects, therefore allowing for a completely flexible time trend. Our point estimates suggest that the degree of heterogeneity increases when we include these fixed effects, although we lose statistical power due to the limited remaining variation on DPM.

Table A.6 repeats the same exercise, but uses levels rather than logs of DPM as the variable of interest. The results are qualitatively similar, with substantial heterogeneity in every specification. However, with a log-linear specification we do not observe the cleanest vehicles having a positive coefficient.

We also investigate the functional forms of these relationships in a semi-parametric way. For each externality, we define vehicles by their percentile of that externality. We then estimate Equation (8) with separate elasticities for vehicles falling in the zero to first percentile, first to second, etc. Appendix Figure A.3 plots a LOWESS smoothed line through these 100 separate elasticity estimates. For the three criteria pollutants, we find that the relationship is quite linear with the elasticity being positive for the cleanest 10 percent of vehicles. The dirtiest vehicles have elasticities that are roughly 0.4. For fuel economy, the relationship is fairly linear from the 60th percentile onwards, but begins steeply and flattens out from the 20th percentile to the 40th. The elasticity of the lowest fuel economy vehicles is nearly 0.6. To put these numbers into context across the different years, the average fuel economy of the 20th percentile is 18.7, while the average for the 40th percentile is 21.75. The variation in elasticities across weight is not monotonic. The relationship begins by increasing until roughly the 20th percentile, and then falls more or less linearly thereafter. The elasticity of the heaviest vehicles is roughly 0.3.

Note that the roughly linear relationship between criteria pollutant emissions and the elasticity is not due to "over smoothing." Appendix Figure A.4 plots the LOWESS smoothed lines for HCs under different bandwidths. The top left figure simply reports the 100 elasticities. There is some evidence that the relationship is not monotonic early on, but from the 5th percentile on, the relationship appears monotonic. Doing this exercise for the other criteria pollutants yields similar results.

E Details of the Gasoline Tax Policy Simulation

For the intensive margin, we estimate a regression as in column 6 of Tables **??** and A.2, except that we interact ln(DPM) with quartile of fuel economy, vehicle weight, and emissions of HC, NO_x, and CO, and dummies for vehicle age bins, again using bins of 4-9, 10-15, and 16-29 years, and control for the direct effects of quartiles of HC, NO_x, and CO emissions. We use quartiles calculated by year and age bin. The coefficients are difficult to interpret on their own, and too numerous to list. However, most are statistically different from zero, and the exceptions are due to small point estimates, not large standard errors.

As in Section G, we compress our dataset to have at most one observation per vehicle per year. Each vehicle is then assigned an elasticity based on its quartiles and age bin. Vehicle i's VMT in the counterfactual with an additional \$1 tax on gasoline is calculated by:

$$VMT^i_{counterfactual} = VMT^i_{BAU} \cdot \left(\frac{P_i + 1}{P_i} \cdot \beta_i \right),$$

where VMT^i_{BAU} is vehicle i's actual average VMT per day between its current and previous smog check, P_i is the average gasoline price over that time, and β_i is the elasticity for the fuel economy/weight/HC/NO/CO/age cell to which i belongs.

For the extensive margin, we estimate a Cox regression on the hazard of scrappage for vehicles 10 years and older, stratifying by VIN prefix and interacting DPM with all five type of quartiles and age bins 10-15 and 16-29. Similar to the intensive margin, we assign each vehicle a hazard coefficient based on its quartile-age cell. Cox coefficients can be transformed into hazard ratios, but to simulate the affect of an increase in gasoline prices on the composition of the vehicle fleet, we must convert these into changes in total hazard.

To do this, we first calculate the actual empirical hazard rate for prefix k in year t as:

$$OrigHazard_{kt} = \frac{D_{kt}}{R_{kt}},$$

where D_{kt} is the number of vehicles in group k, that are scrapped in year t, and R_{kt} is the number of vehicle at risk (that is, which have not previously been scrapped or censored). We then use the coefficients from our Cox regression to calculate the counterfactual hazard faced by vehicles of prefix k in quartile-age group q during year t as:[4]

$$NewHazard_{qkt} = OrigHazard_{kt} \cdot \exp\left\{ \frac{1}{MPG_k} \cdot \gamma_q \right\},$$

where MPG_k is the average fuel economy of vehicle of prefix k and γ_q is the Cox coefficient associated with quartile group q. We then use the change in hazard to construct a weight H_{qkt} indicating the probability that a vehicle of prefix k in quartile group q in year t would be in the fleet if a \$1 gasoline tax were imposed. Weights greater than 1 are possible, which should be interpreted as a $H_{qkt} - 1$ probability that another vehicle of the same type would be on the road, but which was scrapped under "Business as Usual." Because the hazard is the probability of scrappage in year t, conditional on survival to year t, this weight must be calculated interactively, taking into account the weight the previous year. Specifically, we have:

$$H_{qkt} = \prod_{j=1998}^{t} (1 - (NewHazard_{qkj} - OrigHazard_{kt})).$$

We also assign each vehicle in each year a population weight. This is done both to scale our estimates up to the size of the full California fleet of personal vehicles, and to account for the ways in which the age composition of the smog check data differs from that of the fleet. We construct these weights using the vehicle population estimates contained in CARB's EMFAC07 software, which are given by year, vehicle age, and truck status. Our population weight is the number of vehicles of a given age and truck status in a each year given by EMFAC07, divided by the number of such vehicle appearing in our sample. For instance, if EMFAC07 gave the number of 10-year-old trucks in 2005 as 500, while our data contained 50, each 10-year-old truck in our data would have a population weight of 10. Denote the population weight by P_{tac}, where t is year, a is age, and c is truck status.

There is an additional extensive margin that we have not estimated in this paper: new car purchases. To ensure that the total vehicle population is accurate, we apply an *ad hoc* correction based on Busse, Knittel and Zettelmeyer (forthcoming), who find that a \$1 increase in gasoline prices would decrease new car sales by 650,000 per year. Because California's vehicle fleet makes up about 13 percent of the national total, we decrease the population

[4] Note that age group is determined by model-year and year.

of model years 1998 and later by 84,500 when constructing the population weight for the counterfactual. We apply 40 percent of the decrease to trucks, and 60 percent to passenger cars. Denote the "new car effect" n_c.

We estimate the total annual emissions by passenger vehicle in California of NO_x, HC, CO, and CO_2 as actually occurred, and under a counterfactual where a $1 gasoline tax was imposed in 1998. Let i denote a vehicle, a vehicle age, c truck status. Then the annual emissions of pollutant p in year t under "business as usual" are:

$$Emission^{pt}_{BAU} = \sum_i P_{tac} \cdot VMT^i_{BAU} \cdot r_i(p) \cdot 365,$$

and under the counterfactual they are:

$$Emission^{pt}_{counterfactual} = \sum_i (P_{tac} - 1(\text{model year} >= 1998) \cdot n_c) \cdot H_{qkt} \cdot VMT^i_{counterfactual} \cdot r_i(p) \cdot 365,$$

where $r_i(p)$ is the emissions rate per mile of pollutant p for vehicle i. For NO_x, HC, and CO, this is the last smog check reading in grams per mile, while for CO_2 this is the vehicle's gallons per mile multiplied by 19.2 pounds per gallon.

F California versus the Rest of the United States

Given that our empirical setting is California, it is natural to ask whether our results are representative of the country as a whole. At the broadest level, the local-pollution benefits from carbon pricing are a function of the per-capita number of miles driven, the emission characteristics of the fleet of vehicles, and the marginal damages of the emissions. We present evidence that the benefits may, in fact, be larger outside of California. The reason for this is that while the marginal damages are indeed larger in California, the vehicle stock in California is much cleaner than the rest of the country because California has traditionally led the rest of the U.S. in terms of vehicle-emission standards.

The results in Muller and Mendelsohn (2009) provide a convenient way to test whether California differs in terms of marginal damages. Table A.9 presents points on the distribution of marginal damages for NO_x, HCs, and the sum of the two, weighted by each county's annual VMT.[5] Figure A.5 plots the kernel density estimates of the distributions. We present the sum of because counties are typically either "NO_x constrained" or "VOC (HC) constrained," and the sum is perhaps more informative. As expected, the marginal damages are higher in California for HCs, but lower for NO_x, as California counties tend to be VOC-constrained. The sum of the two marginal damages is 78 percent higher in California. Higher points in the distribution show an even larger disparity.

Larger marginal damages are offset, however, by the cleaner vehicle stock within California—a result of California's stricter emission standards. To illustrate this, we collected county-level average per-mile emission rates for NO_x, HCs, and CO from the EPA Motor Vehicle

[5] All of the points on the distribution and densities discussed in this section weight each county by its total VMT.

Emission Simulator (MOVES). MOVES reports total emissions from transportation and annual mileage for each county. Table A.9 also presents points on the per-mile emissions, and Figure A.6 plots the distributions.[6] Mean county-level NO_x, HCs, and CO are 67, 36, and 31 percent lower in California, respectively. Other points in the distributions exhibit similar patterns.

Finally, we calculate the county-level average per-mile externality for each pollutant, as well as the sum of the three. Table A.9 and Figure A.7 illustrates these. As expected, the HC damages are higher, but the average county-level per-mile externality from the sum of the three pollutants is 30 percent lower in California than the rest of the country; the 25th percentile, median, and 75th percentile are 35, 30, and 9 percent lower, respectively. These calculations suggest that, provided the average VMT elasticities are not significantly smaller outside of California and/or the heterogeneity across vehicle types is not significantly different (in the reverse way), our estimates are likely to apply to the rest of the country.

G Scrappage Decisions

Our next set of empirical models examines how vehicle owners' decisions to scrap their vehicles due to gasoline prices. Again we will also examine how this effect varies over emissions profiles.

We determine whether a vehicle has been scrapped using the data from CARFAX Inc. We begin by assuming that a vehicle has been scrapped if more than a year has passed between the last record reported to CARFAX and the date when CARFAX produced our data extract (October 1, 2010). However, we treat a vehicle as being censored if the last record reported to CARFAX was not in California, or if more than a year and a half passed between the last smog check in our data and that last record. As well, to avoid treating late registrations as scrappage, we treat all vehicles with smog checks after 2008 as censored. Finally, to be sure we are dealing with scrapping decisions and not accidents or other events, we only examine vehicles that are at least 10 years old.

Some modifications to our data are necessary. To focus on the long-term response to gasoline prices, our model is specified in discrete time, denominated in years. Where vehicles have more than one smog check per calendar year, we use the last smog check in that year. Also, because it is generally unlikely that a vehicle is scrapped at the same time as its last smog check, we create an additional observation for scrapped vehicles either one year after the last smog check, or six months after the last CARFAX record, whichever is later. For these created observations, odometer is imputed based on the average VMT between the last two smog checks, and all other variables take their values from the vehicle's last smog check. An exception is if a vehicle fails the last smog check in our data. In this case, we assume the vehicle was scrapped by the end of that year.

[6] We note that the emissions reported in MOVES exceed the averages in our data. This may reflect the fact that smog checks are not required for vehicles with model years before 1975, and these vehicles likely have very high emissions because this pre-dates many of the emission standards within the U.S.

Because many scrapping decisions will not take place until after our data ends, a hazard model is needed to deal with right censoring. Let T_{jivg} be the year in which vehicle i, of vehicle type j, vintage v, and geography g, is scrapped. Assuming proportional hazards, our basic model is:

$$\Pr[t < T_{ijvg} < t + 1 | T > t] = h^0_{jv}(t) \cdot \exp\{\beta x DPM_{igt} + \gamma D_{fail_{it}} + \psi G_{igt} + \alpha X_{it}\},$$

where DPM_{igt} is defined as before; $D_{fail_{it}}$ is a dummy equal to one if the vehicle failed a smog check any time during year t; G is a vector of demographic variables, determined by the location of the smog check; X is a vector of vehicle characteristics, including a dummy for truck and a sixth-order polynomial in odometer; and $h^0_{ijv}(t)$ is the baseline hazard rate, which varies by time but not the other covariates. In some specifications, we will allow each vehicle type and vintage to have its own baseline hazard rate.

We estimate this model using semi-parametric Cox proportional hazards regressions, leaving the baseline hazard unspecified. We report exponentiated coefficients, which may be interpreted as hazard ratios. For instance, a 1 unit increase in DPM will multiply the hazard rate by $\exp\{\beta\}$, or increase it by $(\exp\{\beta\} - 1)$ percent. In practice, we scale the coefficients on DPM for a 5-cent change, corresponding to a \$1.00 increase in gasoline prices for a vehicle with fuel economy of 20 miles per gallon.

Tables A.3 and A.4 show the results of our hazard analysis. Models 1 and 2 of Table A.3 assign all vehicles to the same baseline hazard function. Model 1 allows the effect of gasoline prices to vary by whether or not a vehicle failed a smog check. Model 2 also allows the effect of gasoline prices to vary by quartiles of NO_x.[7] Models 3 and 4 are similar, but stratify the baseline hazard function, allowing each VIN prefix to have its own baseline hazard function. Model 5 allows the effect of gasoline prices to vary both by externality quartile and age group, separating vehicles 10 to 15 years old from vehicles 16 years and older.

Models 1 and 2 indicate that increases in gasoline prices actually decrease scrapping on average, with the cleanest vehicles seeing the largest decreases. The effect is diminished once unobserved heterogeneity among vehicle types is controlled for, but is still statistically significant. However, the true heterogeneity in the effect of gasoline prices on hazard seems to be over age groups. Model 5 shows that when the cost of driving a mile increases by five cents, the hazard of scrappage decreases by about 23 percent for vehicles between 10 and 15 years old, while it increases by around 3 percent for vehicles age 16 and older, with little variation across NO_x quartiles within age groups. These results suggests that when gasoline prices rise, very old cars are scrapped, increasing demand for moderately old cars and thus reducing the chance that they are scrapped.

Table A.4 presents the quartile by age by DPM interactions for each of the 5 externality dimensions. Hydrocarbons and CO have the identical pattern to NO_x, with no heterogeneity within age-group. With fuel economy and vehicle weight, there is within-age heterogeneity, although the form is counter-intuitive. The heaviest and least fuel-efficient vehicles are relatively less likely thank the lightest and most fuel-efficient vehicles to be scrapped when gasoline prices increase. That is, while all 10- to 15-year-old vehicles are less likely to be

[7]Quartiles in these models are calculated by year among only vehicles 10 years and older.

scrapped, the decrease in hazard rate is larger for heavy, gas-guzzling vehicles. For vehicles 16 years and older, the heaviest quartile is less likely to be scrapped when gasoline prices increase, even though the lightest (and middle quartiles) are more likely. As the model stratifies by VIN prefix, this cannot be simply that more durable vehicles have lower fuel economy.

In summary, increases in the cost of driving a mile over the long term increase the chance that old vehicles are scrapped, while middle-aged vehicles are scrapped less, perhaps because of increased demand. Although vehicle age is highly correlated with emissions of criteria pollutants, there is little variation in the response to gasoline prices across emissions rates within age groups.

H Income Distribution Adjustment

In section VI, we assign income brackets to individual consumers using the method of Borenstein (2012). Here we briefly describe the details of that procedure; for more details see Borenstein (2012).

Borenstein shows that household consumption levels of some commodity (gasoline in our case) within income brackets can be bounded between the case where the ranking of household incomes is sorted by consumption levels (usage-ranking), and the case where the ranking of household incomes is random with respect to consumption levels (random ranking). If one can calculate the average consumption by income bracket, one can calculate a weighting between usage-ranking and random-ranking that correctly assigns households to income brackets based on their consumption. Borenstein proposes calculating these averages from a separate dataset that contains individual level income and consumption, if one can be found. We utilize the 2009 NHTS for this purpose.

Formally, let $\bar{g}_b$ denote the average gasoline consumption for consumers living in California in income group b in the 2009 NHTS. The N vehicles registered in each census block group (CBG) in California and appearing in the Smog Check data are to be assigned an integer rank from 1 to N, intended to correspond to the income ranking of the household those vehicles belong to. If s_b^c denotes the number of households falling into income bracket b in CBG c in the 2000 Census, and h_b denotes the number of vehicles per household in income bracket b, then, for instance, vehicles ranked from 1 to $fracNs_b * h_b$ will fall into bracket 1. The ranking for vehicle i will be $v_i(w) = (1 - w) \cdot r_{rr} + w \cdot r_{ur}$, where r_{rr} is drawn randomly from a uniform distribution over $[1, N]$, producing random-ranking, while r_{ur} sorts vehicles by gasoline consumption, producing usage ranking. Any choice of w will produce a joint ranking within CBGs, leading to a statewide average within-bracket gasoline consumption level of $\tilde{g}_b$ within the Smog Check data.[8]

The income brackets given at the CBG level in the 2000 census can be pooled into groups roughly approximating deciles of the total income distribution in California. The NHTS gives income in brackets as well, which can be grouped into 8 groups corresponding to the first 7

[8]Ideally this calculation would use a CBG-specific w, however the NHTS does not provide geographic data at that level. Borenstein (2012) has the same limitation.

"deciles" in the Census data, plus the top 3 deciles topcoded into one income bracket. We calculate w based on the 8 groupings in the NHTS data, but when using that w to assign vehicles to income brackets, we use the ranking implied by w to distribute vehicles across the top 3 deciles. We choose w to minimize the following goodness-of-fit measure:

$$G = \sum_{b=1}^{8} s_b(\tilde{g}_b - \bar{g}_b)$$

That is, we choose w such that when vehicles in the Smog Check Data are ranked into income brackets, the average gasoline usage in each income bracket matches the average gasoline usage for that income bracket in the 2009 NHTS.

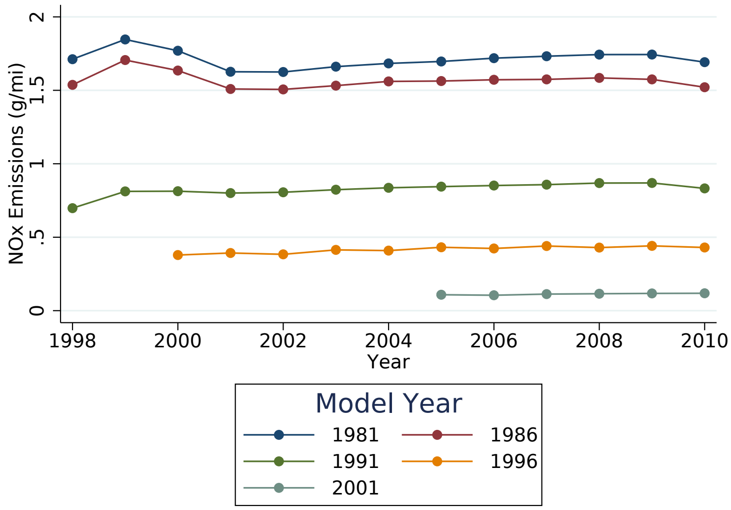
References

Borenstein, Severin. 2012. "The Redistributional Impact of Nonlinear Electricity Pricing." *American Economic Journal: Economic Policy*, 4(3): 56–90.

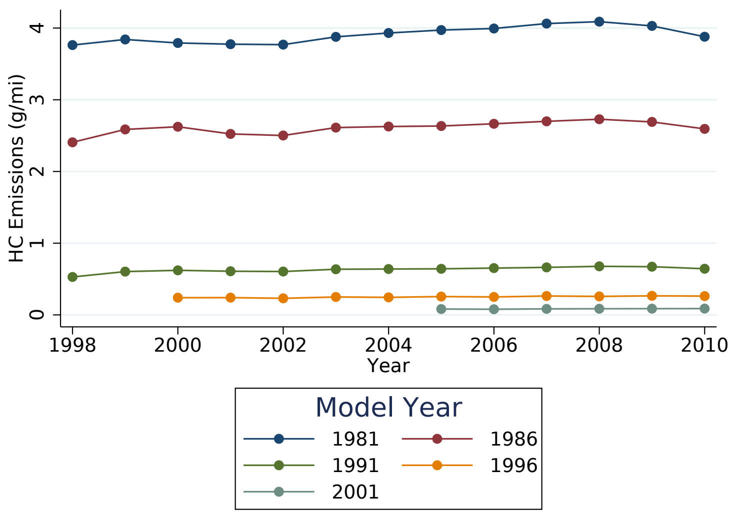
Busse, Meghan, Christopher R. Knittel, and Florian Zettelmeyer. forthcoming. "Are Consumers Myopic? Evidence from New and Used Car Purchases." *The American Economic Review*.

Morrow, Silvia, and Kathy Runkle. 2005. "April 2004 Evaluation of the California Enhanced Vehicle Inspection and Maintenance (Smog Check) Program." Air Resources Board Report to the Legislature.

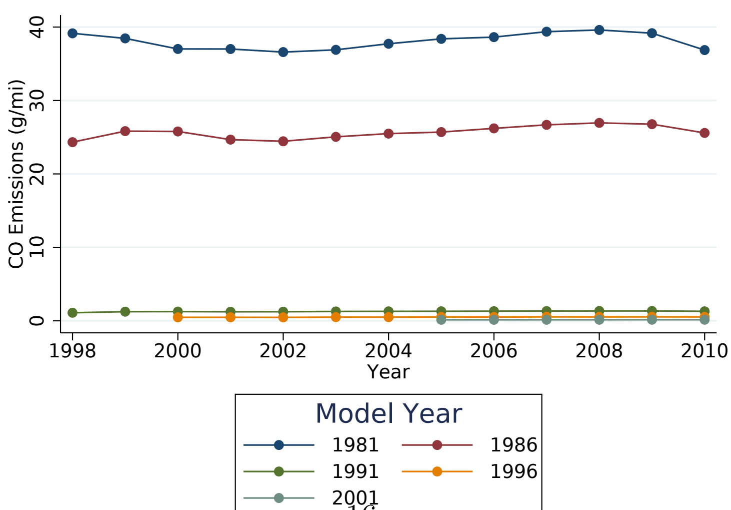
Muller, Nicholas Z., and Robert Mendelsohn. 2009. "Efficient Pollution Regulation: Getting the Prices Right." *American Economic Review*, 99(5): 1714–39.



(a) NOx



(b) HCs



(c) CO

Figure A.1: Emissions rates over time by model year

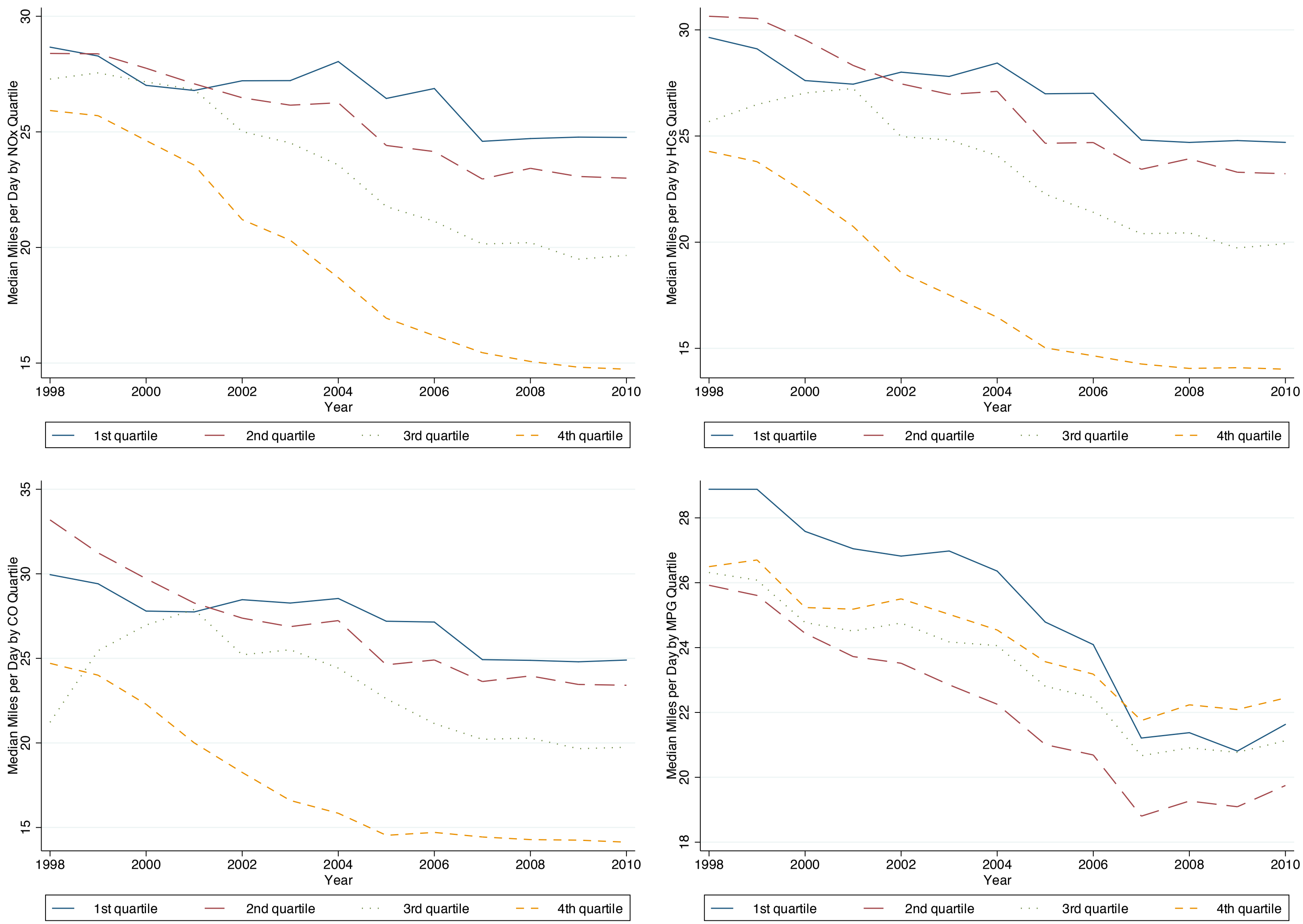


Figure A.2: Change in VMT over sample by pollutant quartile

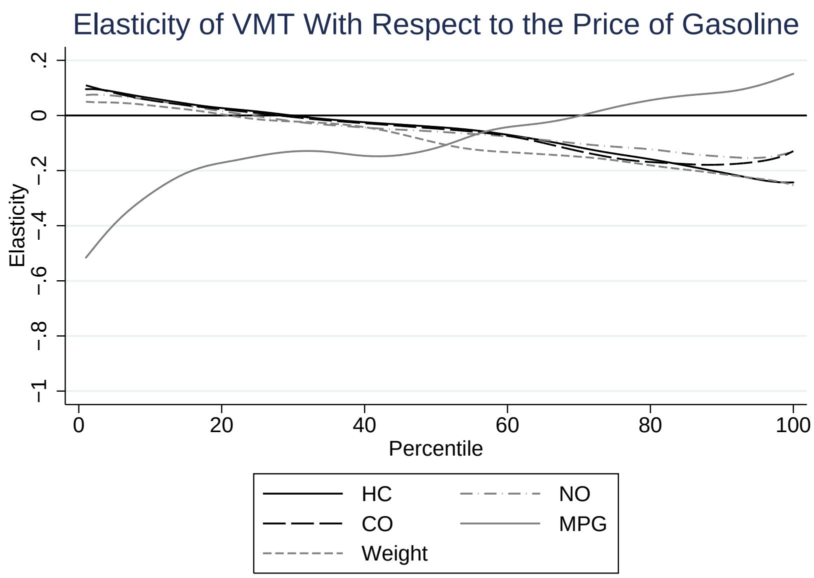


Figure A.3: Non-parametric relationships between elasticity and externality

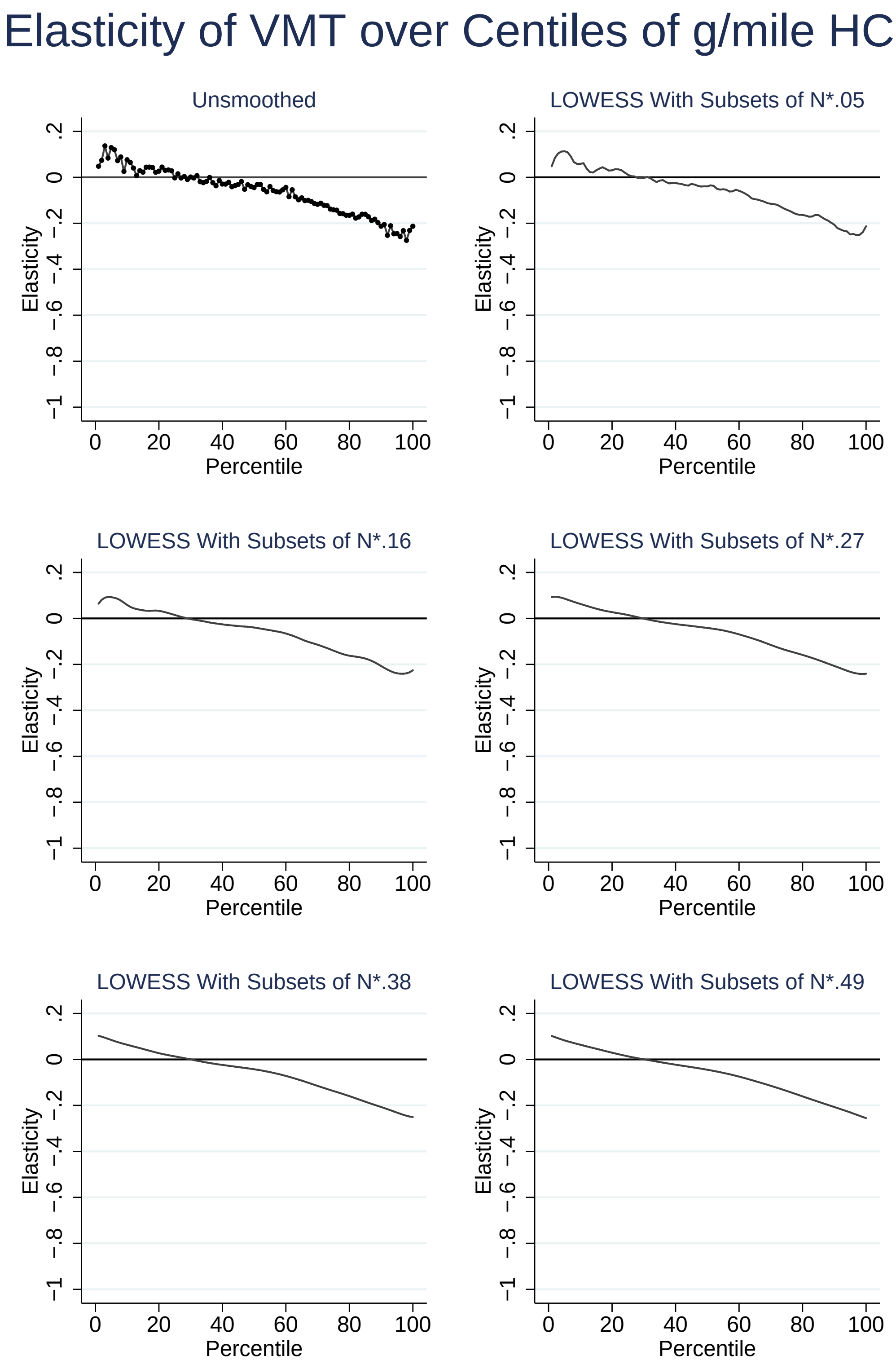


Figure A.4: The effect of bandwidth on the non-parametric function

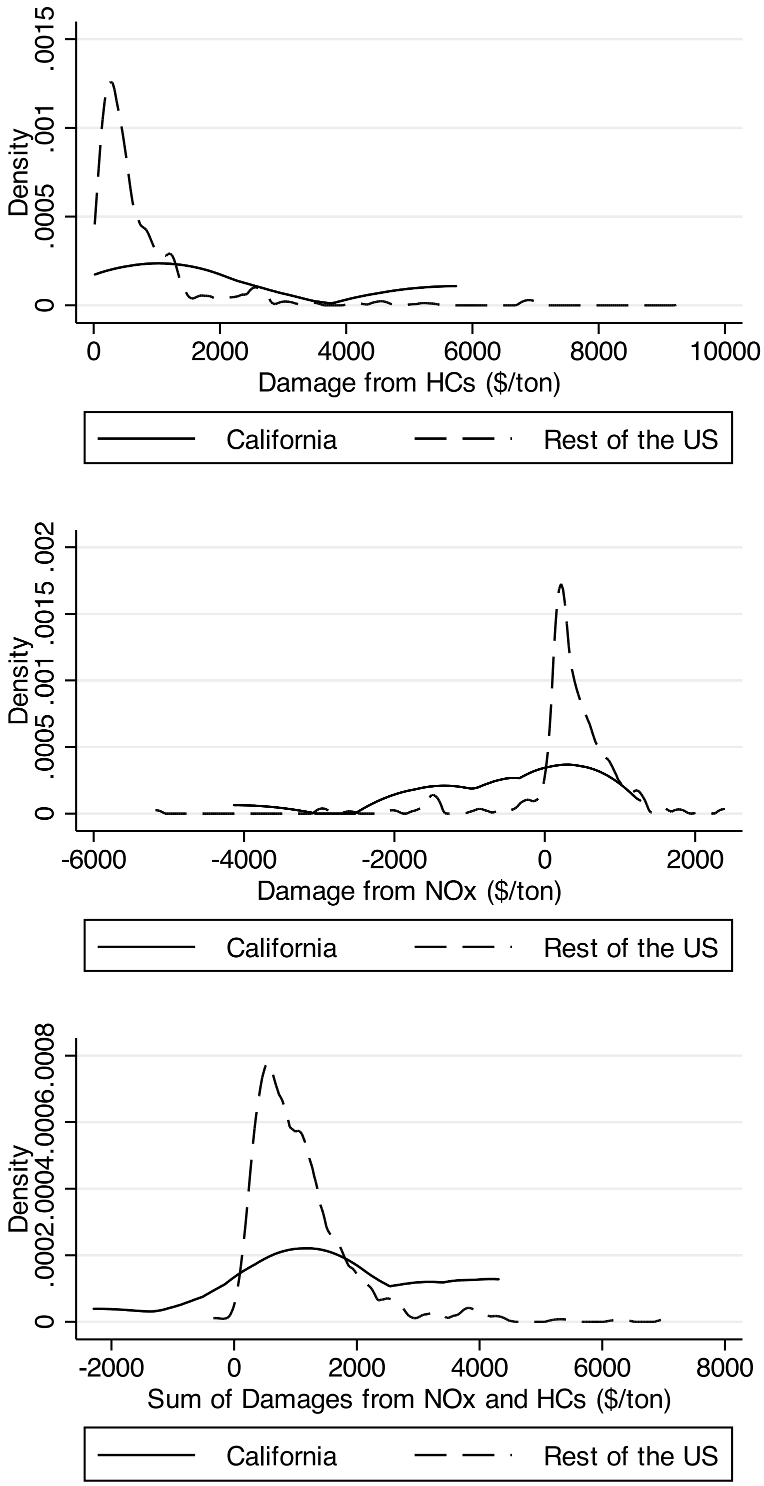


Figure A.5: Distributions of marginal damages from Muller and Mendelsohn (2009) for California and the rest of the U.S.

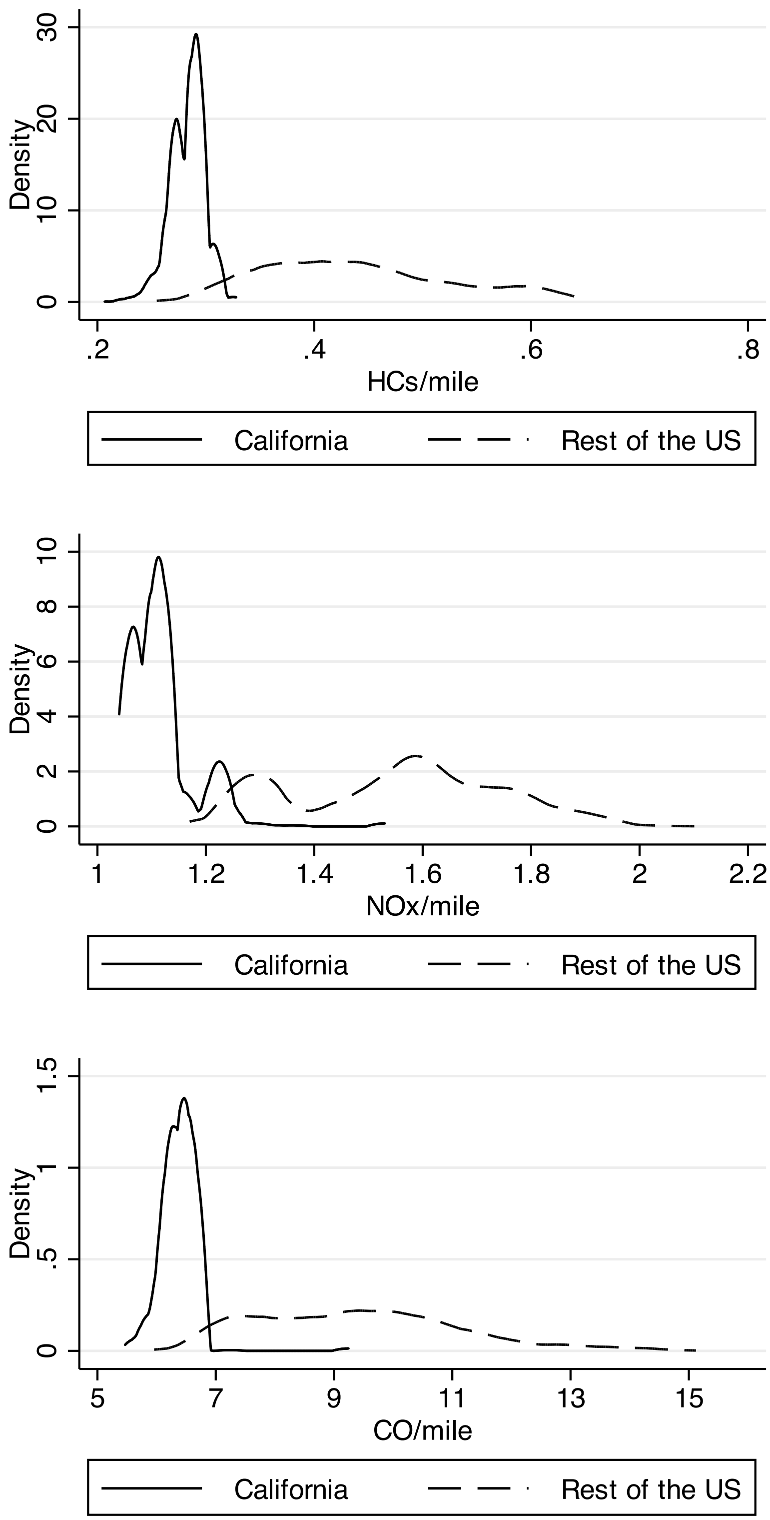


Figure A.6: Distributions of per-mile emissions for California and the rest of the U.S.

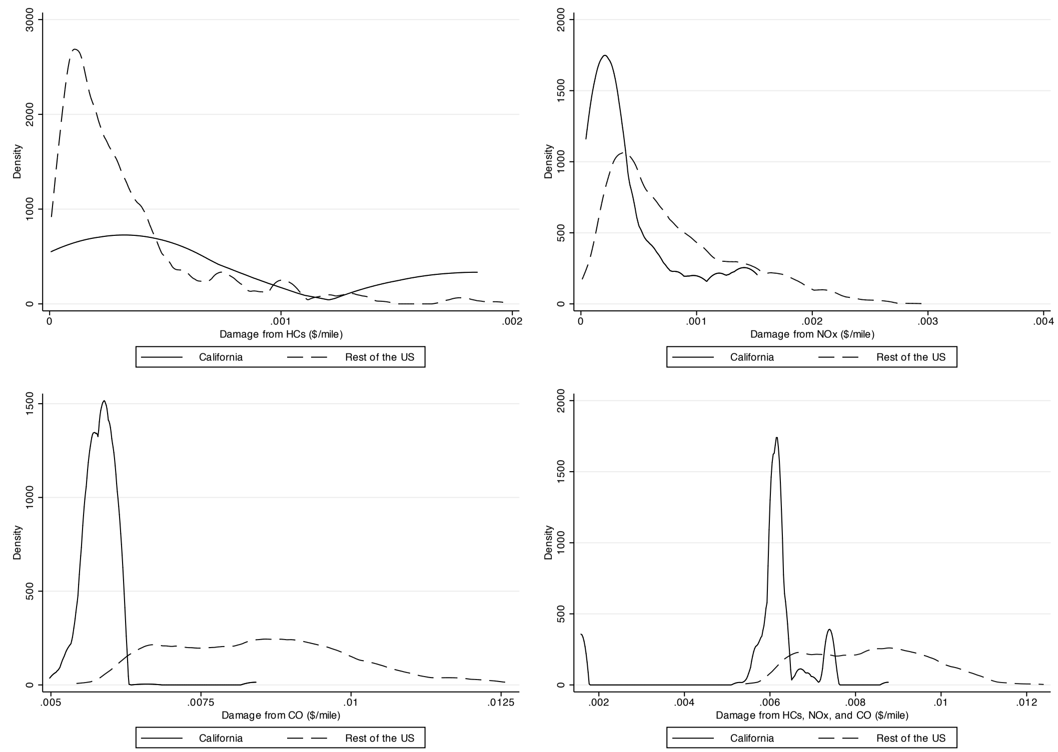


Figure A.7: Distributions of per-mile damages for California and the rest of the US

Table A.1: Average Pollutant Rates Per Mile Traveled by Year

	Nitrogen Oxides			Hydrocarbons			Carbon Monoxide			Gasoline	
Year	Mean	SD	Mean CV	Mean	SD	Mean CV	Mean	SD	Mean CV	Mean	SD
1998	1.161	1.051	0.536	1.662	1.866	0.504	15.400	24.739	0.516	0.043	0.010
1999	1.187	0.983	0.455	1.665	1.864	0.464	15.227	24.952	0.485	0.043	0.010
2000	1.094	0.915	0.441	1.535	1.826	0.460	13.539	23.849	0.476	0.044	0.010
2001	0.982	0.857	0.427	1.354	1.769	0.464	11.689	23.123	0.461	0.044	0.010
2002	0.876	0.816	0.418	1.145	1.679	0.446	9.694	21.381	0.430	0.044	0.010
2003	0.791	0.780	0.401	0.997	1.563	0.432	7.940	19.503	0.395	0.045	0.010
2004	0.715	0.742	0.380	0.855	1.469	0.421	6.561	17.581	0.363	0.045	0.010
2005	0.735	0.713	0.393	0.852	1.444	0.455	6.375	17.519	0.379	0.045	0.010
2006	0.638	0.667	0.382	0.718	1.351	0.430	5.157	15.887	0.350	0.045	0.010
2007	0.572	0.634	0.377	0.628	1.261	0.431	4.308	14.509	0.334	0.045	0.010
2008	0.512	0.602	0.373	0.545	1.185	0.400	3.556	13.064	0.317	0.046	0.010
2009	0.478	0.590	0.379	0.496	1.148	0.412	3.120	12.147	0.316	0.046	0.011
2010	0.462	0.566	0.402	0.460	1.002	0.427	2.741	10.901	0.323	0.046	0.010
N	10432374			10432374			10666348			13397795	

Note: Mean CV is the average VIN Prefix-level coefficient of variation (SD/Mean). Gasoline is measured in gallons per mile, while the remaining pollutant rates are measured in grams per mile.

Table A.2: Vehicle Miles Travelled, Dollars Per Mile, and Externality Quartiles

Quartile	Nitrogen Oxides	Hydrocarbons	Carbon Monoxide	Fuel Economy	Vehicle Weight
1	0.0425	0.0486	0.0466	-0.169	-0.111
2	-0.0540	-0.0550	-0.0527	-0.159	-0.114
3	-0.152	-0.149	-0.149	-0.104	-0.145
4	-0.280	-0.305	-0.307	-0.0986	-0.167

Coefficients are elasticities calculated by regressing the log of average daily VMT between Smog Checks on the log of the gas price in dollars per mile, interacted with quartiles of the pollutants indicated. Quartiles are based on rankings of within the calendar year in which the Smog Check occurs. All regressions control for direct effects of the quartiles, a quadratic time trend, a sixth-order polynomial in lagged odometer, demographics of the zip code where the Smog Check occurs, calendar-year fixed effects, vehicle age fixed effects, and vehicle fixed effects.

Table A.3: Hazard of Scrappage: Cox Proportional Hazard Model

	Model 1	Model 2	Model 3	Model 4	Model 5
Dollars per Mile	0.920*		0.965*		
	(0.039)		(0.018)		
DPM * Failed Smog Check	1.105**	1.074**	1.063**	1.043*	
	(0.029)	(0.026)	(0.021)	(0.020)	
Failed Last Smog Check	7.347**	7.800**	7.639**	8.155**	
	(0.242)	(0.246)	(0.161)	(0.173)	
DPM * NO Quartile 1		0.801**		0.893**	
		(0.044)		(0.035)	
DPM * NO Quartile 2		0.862**		0.923**	
		(0.038)		(0.026)	
DPM * NO Quartile 3		0.883**		0.956*	
		(0.034)		(0.017)	
DPM * NO Quartile 4		0.929*		0.983	
		(0.033)		(0.010)	
Vehicle Ages 10-15					
DPM * NO Quartile 1					1.285**
					(0.023)
DPM * NO Quartile 2					1.287**
					(0.018)
DPM * NO Quartile 3					1.291**
					(0.014)
DPM * NO Quartile 4					1.254**
					(0.012)
Failed Smog Check					8.732**
					(0.380)
DPM * Failed Smog Check					0.910**
					(0.014)
Vehicle Ages 16+					
DPM * NO Quartile 1					0.751**
					(0.016)
DPM * NO Quartile 2					0.745**
					(0.014)
DPM * NO Quartile 3					0.745**
					(0.011)
DPM * NO Quartile 4					0.737**
					(0.008)
Failed Smog Check					7.765**
					(0.286)
DPM * Failed Smog Check					1.185**
					(0.026)
Station ZIP Code Characteristics	Yes	Yes	Yes	Yes	Yes
Quadratic Time Trend in Days	Yes	Yes	Yes	Yes	Yes
Vehicle Characteristics	Yes	Yes	Yes	Yes	Yes
Quartiles of NO	No	Yes	No	Yes	Yes
Stratified on Vin Prefix	No	No	Yes	Yes	Yes
Observations	31567473	26720283	31567473	26720283	26720283

Note: Coefficients on dollars per mile scaled for a 5-cent change

Table A.4: Effect of a 5-cent/mile Increase in Driving Cost on the Hazard of Scrappage

	Nitrogen Oxides	Hydrocarbons	Carbon Monoxide	Fuel Economy	Vehicle Weight
Vehicle Ages 10-15					
Quartile 1	1.285	1.297	1.301	1.108	1.516
Quartile 2	1.287	1.288	1.302	1.237	1.349
Quartile 3	1.291	1.286	1.272	1.479	1.265
Quartile 4	1.254	1.236	1.240	1.707	1.174
Vehicle Ages 16+					
Quartile 1	0.751	0.754	0.749	0.674	0.866
Quartile 2	0.745	0.753	0.763	0.735	0.773
Quartile 3	0.745	0.757	0.745	0.797	0.715
Quartile 4	0.737	0.751	0.751	0.822	0.629

Statistics are exponentiated coeficients of a Cox proportional hazards model. Interpret as hazard ratios.

Table A.5: Robustness Check—Intensive Margin Interacting NOx Quartiles With Other Controls

	(1)	(2)	(3)	(4)	(5)	(6)
ln(DPM) * NO Q1	0.0406	0.0381	0.0678*	0.0605	0.0590	0.0666
	(0.0231)	(0.0250)	(0.0339)	(0.0335)	(0.0333)	(0.121)
ln(DPM) * NO Q2	-0.0617*	-0.0581*	-0.0453	-0.0478	-0.0484	-0.0410
	(0.0261)	(0.0269)	(0.0309)	(0.0310)	(0.0308)	(0.121)
ln(DPM) * NO Q3	-0.158***	-0.155***	-0.166***	-0.165***	-0.165***	-0.157
	(0.0271)	(0.0272)	(0.0282)	(0.0291)	(0.0294)	(0.120)
ln(DPM) * NO Q4	-0.288***	-0.298***	-0.355***	-0.353***	-0.351***	-0.344**
	(0.0300)	(0.0302)	(0.0325)	(0.0332)	(0.0331)	(0.120)
NO Q2	0.378	0.327	-2.622	-3.925*	-3.954*	-4.916**
	(0.800)	(0.735)	(1.622)	(1.693)	(1.673)	(1.732)
NO Q3	-1.246	-1.447	-5.233***	-6.846***	-6.793***	-7.987***
	(1.012)	(0.899)	(1.447)	(1.524)	(1.508)	(1.566)
NO Q4	-2.297*	-2.951**	-9.696***	-11.39***	-11.26***	-12.60***
	(1.116)	(1.084)	(2.257)	(2.253)	(2.271)	(2.301)
Quartile-Time Trend Interactions	Yes	Yes	Yes	Yes	Yes	Yes
Vintage-Quartile Interactions	No	Yes	Yes	Yes	Yes	Yes
Quartile-Year Interactions	No	No	Yes	Yes	Yes	Yes
Quartile-Lagged Odometer Interactions	No	No	No	Yes	Yes	Yes
Quartile-Demographics Interactions	No	No	No	No	Yes	Yes
Calendar Month Fixed-Effects	No	No	No	No	No	Yes
N	2979289	2979289	2979289	2979289	2979289	2979289

Note: All regressions include vehicle fixed-effects, year fixed effects, vintage/truck fixed effects, a quadratic time trend, a sixth order polynomial in the odometer reading at previous Smog Check, and ZIP code level demographic characteristics.

Table A.6: Robustness Check—Intensive Margin Interacting NOx Quartiles With Other Controls

	(1)	(2)	(3)	(4)	(5)	(6)
DPM * NO Q1	-2.676***	-2.807***	-2.294***	-2.412***	-2.421***	-5.089***
	(0.359)	(0.350)	(0.301)	(0.347)	(0.345)	(0.696)
DPM * NO Q2	-3.337***	-3.358***	-3.075***	-3.128***	-3.129***	-5.339***
	(0.359)	(0.357)	(0.334)	(0.355)	(0.354)	(0.631)
DPM * NO Q3	-3.925***	-3.941***	-3.858***	-3.881***	-3.875***	-5.728***
	(0.389)	(0.391)	(0.397)	(0.395)	(0.394)	(0.631)
DPM * NO Q4	-4.642***	-4.720***	-4.970***	-4.974***	-4.957***	-6.482***
	(0.425)	(0.433)	(0.444)	(0.440)	(0.442)	(0.653)
NO Q2	0.958	0.821	-5.997***	-7.404***	-7.433***	-4.917**
	(0.674)	(0.613)	(1.330)	(1.391)	(1.384)	(1.567)
NO Q3	0.242	-0.00702	-8.999***	-10.74***	-10.69***	-6.708***
	(0.889)	(0.798)	(1.563)	(1.619)	(1.605)	(1.527)
NO Q4	0.615	0.124	-12.63***	-14.43***	-14.34***	-9.222***
	(1.015)	(0.999)	(2.173)	(2.181)	(2.205)	(2.227)
Quartile-Time Trend Interactions	Yes	Yes	Yes	Yes	Yes	Yes
Vintage-Quartile Interactions	No	Yes	Yes	Yes	Yes	Yes
Quartile-Year Interactions	No	No	Yes	Yes	Yes	Yes
Quartile-Lagged Odometer Interactions	No	No	No	Yes	Yes	Yes
Quartile-Demographics Interactions	No	No	No	No	Yes	Yes
Calendar Month Fixed-Effects	No	No	No	No	No	Yes
N	2979289	2979289	2979289	2979289	2979289	2979289

Note: All regressions include vehicle fixed-effects, year fixed effects, vintage/truck fixed effects, a quadratic time trend, a sixth order polynomial in the odometer reading at previous Smog Check, and ZIP code level demographic characteristics.

Table A.7: Within and Between Standard Deviations of Vehicle Emissions Rates

	Total	Within Zip Code	Between Zip Code
Grams/mile HC emissions	1.177	0.996	0.250
Grams/mile CO emissions	12.912	10.788	2.220
Grams/mile NOx emissions	0.638	0.535	0.182

Table A.8: Ratio of Remaining Deadweight Loss With Tax to Deadweight Loss with No Tax: Calibration

	σ^2	σ_B^2	ρ	$R(\tau_{naive})$	$R(\tau^*)$
1998	1.407	1.465	0.322	0.789	0.755
1999	1.408	1.471	0.299	0.785	0.755
2000	1.438	1.486	0.308	0.794	0.763
2001	1.457	1.496	0.311	0.799	0.767
2002	1.492	1.506	0.283	0.802	0.775
2003	1.517	1.535	0.283	0.807	0.781
2004	1.525	1.531	0.265	0.806	0.782
2005	1.474	1.539	0.265	0.796	0.771
2006	1.482	1.539	0.251	0.795	0.773
2007	1.487	1.547	0.247	0.796	0.774
2008	1.498	1.533	0.252	0.799	0.777
Average	1.471	1.513	0.281	0.797	0.770

Table A.9: Percentage Difference Between California and the rest of the US

	25th Percentile	Median	75th Percentile	Mean
NOx g/mi	-0.230	-0.291	-0.338	-0.282
NOx Damage/ton (MM)	-0.439	-0.525	-0.558	-0.685
NOx Damage/mi	-0.595	-0.657	-0.712	-0.761
HC g/mi	-0.262	-0.321	-0.410	-0.354
HC Damage/ton	1.475	2.558	5.318	1.821
HC Damage/mi	0.602	1.134	3.358	1.035
CO g/mi	-0.226	-0.321	-0.366	-0.320
CO Damage/mi	-0.226	-0.321	-0.366	-0.320
NOx + HC Damage/ton (MM)	0.0191	0.994	2.337	0.787
NOx + HC + CO Damage/mi	-0.353	-0.299	-0.0883	-0.295

Notes: The table reports the coefficient on the California dummy divided by the constant. All differences are statistically significant at the 0.001 level, except for NOx g/mi and HC Damage/mi at the 25th percentile (significant at the 0.05 level), and NOx Damage/mi.