

# Clunkers or Junkers?

## Adverse Selection in a Vehicle Retirement Program

Online Appendices

Ryan Sandler

February 2, 2012

## A Data

### A.1 Vehicle Buyback

I have obtained records of every vehicle scrapped by BAAQMD's Vehicle Buyback Program via a Public Records Act<sup>1</sup> request. These records contain the VIN and license plate of each vehicle, along with the date the vehicle was purchased by the dismantler, and the final odometer reading. The records also include make, model and year for the vehicle, although I do not use this information. Instead, I take the list of all 17-digit VINs in the data and feed them through a program which scrapes a website that decodes VINs.<sup>2</sup> This gives me the specific make, model, year, trim and body type of each vehicle, as well as information about the type of engine originally installed.

In addition to basic information on the vehicles, the records from BAAQMD contain the results of a survey given to vehicle owners at the time they sell their vehicle. Among other things, the former vehicle owners are asked how much they thought the car was worth, how long they think the car would have lasted if not scrapped, and what they would have done if the program were not available. Dill (2004) makes use of this information to calibrate her assumptions, but I do not. As it turns out, most of the information in the participant survey is inconsistent both with my counterfactual estimates and with the evidence in the Smog Check Data (i.e., owners say the vehicle was driven 5000 miles in

---

<sup>1</sup>The Public Records Act is California's equivalent to the Freedom of Information Act for state agencies

<sup>2</sup>I utilize a free "preview before buying" page generated by a popular website selling full vehicle histories.

the last 12 months, when Smog Check records show the vehicle had been driven only 100 miles in the last three years).

The BAAQMD data contains 17,759 VINs of post-1981 vehicles and 23447 pre-1981 vehicles. Of the 17-digit VINs, 835 are clearly invalid for some reason (e.g. invalid check digit) and an additional 1523 are technically valid but have no match in the Smog Check or DMV data. The pre-1981 vehicles have less coverage in the Smog Check data.<sup>3</sup> Of the pre-1981 VINs, 5263 do not have any Smog Check records, and of the remainder 2124 did not match any manufacturer's VIN patterns. An additional 2945 are missing geographic data. The resulting list contains 28,500 unique vehicles, of which 15546 have 17 digit VINs and are used in my analyses.

## A.2 Smog Check Data

My data from the Smog Check Program are essentially the universe of test records from January 1, 1996 to December 31, 2010. I was only able to obtain test records 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, I limit my data to test records from January 1997 on. I exclude tests which have no odometer reading, and vehicles which 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 I argue that more than 36 in what is at most a 12 year-span indicates some problem with the data.<sup>4</sup> Each test record contains odometer and emissions readings, the reason for the test, the date, time and location of the test, and information on the vehicle, including VIN, vehicle weight, and transmission type<sup>5</sup>

I use the information from these test records to construct the miles traveled and grams of pollutants emitted over time by each vehicle. To do this accurately, certain adjustments must be made first.

### A.2.1 Odometer Readings

Because all of the vehicles in my analysis were manufacturer prior to 1988, nearly all have 5-digit odometers—that is, five places for whole numbers plus a decimal. As such,

---

<sup>3</sup>As vehicles of 1975 model year and earlier have been exempt from Smog Checks since 1998, this is not entirely surprising

<sup>4</sup>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. I suspect the vehicle belonged to a Smog Check technician who used it to test the electronic transmission system.

<sup>5</sup>The data also contains make, model and model year, but again I instead use the VIN to obtain this information.

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. This presents complications for estimating vehicle miles traveled. The left panels of Figure A.3 illustrates this problem, showing the last known odometer readings of the vehicles retired by the Buyback and my control group. To address this problem, I 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 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 my corrections any vehicle has an odometer reading above 800,000 or has implied VMT per day greater than 200 or less than zero, I exclude the vehicle from my analysis. The right panels of Figure A.3 shows the same distributions after my algorithm has completed. All of my analyses use this adjusted mileage.

### A.2.2 Emissions readings

Having emissions test results from individual vehicles allows me to more accurately pin down total emissions per vehicle than would be possible using fleet average statistics. However, a few obstacles must be dealt with.

First, emissions results from both the ASM and TSI tests are given in either parts per million (for HC and NOx) or percent (O<sub>2</sub>, CO, and CO<sub>2</sub>). 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 CARB (2006) and are reproduced in Appendix B. The equations are for HCs, NOx and CO, and estimate grams per mile for each pollutant as a non-linear function of all three pollutants, model year and vehicle weight.

A second challenge is that I am missing NOx measurements for a large number of test records. This is because NOx can only be measured with the ASM test, and even then only when tested using the BAR97 test machine, which replaced the older BAR90 machine starting in 1998. Besides being interested in measuring NOx for itself, because the conversion equations for HC and CO take the NOx readings as an argument, these missing values become a problem. To resolve this, I impute NOx values where they are missing, using the average NOx reading by 10-digit VIN prefix, regression adjusted with a fifth-order polynomial in mileage.<sup>6</sup>, emissions readings of HC and CO, and dummies for overall test result, reason for test, year, month and month-of-year. Because the conversion equations are intended to take ASM test measurements as inputs, I estimate similar regressions for HC and CO, omitting HC and CO emissions as covariates and adding an indicator for the type of test. I adjust the HC and CO readings of TSI tests by the value of the test-type dummy. These corrections are reported in Table A.4

Finally, to account for the fact that total emissions are being made over time, while my emissions data captures single points in time, I calculate total emissions using the average of the current emissions reading and the one immediately previous. Specifically, total emissions between time  $t$  and time  $t - 1$  are given by:

$$totalemit_{t,t-1} = (miles_t - miles_{t-1}) \frac{gpm_t + gpm_{t-1}}{2}$$

Where  $miles_t$  and  $gpm_t$  are the adjusted odometer reading and grams/mile at time  $t$ , respectively. Implicitly, this assumes that instantaneous emissions per mile evolve linearly between Smog Checks. This should not cause problems in cases where a vehicle has multiple Smog Checks in a year, or even on the same day—because test records have time stamps, the first failed test would be used to calculate emissions over the longer period of time.

This assumes that vehicles which fail an inspection and are reinspected and eventually pass are truly repaired prior to the passed inspection. It may be instead that due to randomness in the inspection process or fraud on the part of the vehicle owner or technician<sup>7</sup> a failed vehicle passes a subsequent inspection with no repairs or only superficial repairs. In this case the “true” emissions level at that point in time should be the level observed at the first failed test. I have tried computing total emissions by assigning all Smog Checks within 90 days of each other the maximum emissions from those Smog Checks, and it makes very little difference for the final estimates.

---

<sup>6</sup>Mileage is adjusted for rollovers here.

<sup>7</sup>This behavior was documented by Hubbard (1998) using the Smog Check data.

### A.3 DMV Data

The Smog Check data gives me excellent information on vehicle usage but little on scrappage. While I know from my BAAQMD data when the Buyback vehicles were scrapped, I must also know when or if the control vehicles are scrapped. To address this I supplement the Smog Check data with information from California Department of Motor Vehicles.<sup>8</sup> For each vehicle Buyback and control vehicle, I have the latest date that registration was paid through on the vehicle. I also have the last date of any junk or salvage<sup>9</sup> certificate issued for the vehicle<sup>10</sup>, as well as the date the vehicle was registered for planned non-operation (PNO) or taken out of state, presuming DMV was notified of these actions. The DMV data for most of the 17-digit VINs used in this paper was accessed from DMV’s database in May of 2010 and the dates are accurate up to that time. Data for the pre-1981 VINs and additional post-1981 control vehicles used in appendix F were accessed in December 2010.

It is important to note, however, that scrapping a vehicle does not necessarily mean crushing it or otherwise putting it permanently out of operation. Receiving a junk certificate is not always the end of a vehicle’s life. Frequently vehicles are sold to dismantlers, but then are left sitting intact until they are later purchased and driven again. Indeed, this is source of the extreme version of the adverse selection story that many economists think of with respect to vehicle retirement programs. Some of the vehicles I mark as scrapped may in fact return to the road some time after my data ends.

Broadly, I use the last date I observe the vehicle—either its last Smog Check or latest date among its DMV records—to determine scrappage. I consider a vehicle scrapped if any of the following are true:

1. The final date is a junk, salvage, or PNO record from DMV.
2. The final date is the “paid thru” date or date of last correspondence, and it is before October 1, 2009.
3. The final date is the “paid thru” date or date of last correspondence, and there is also a junk certificate after the date of last Smog Check, but within a year of the final date.
4. The final date is the last Smog Check, and one of the following is also true:

---

<sup>8</sup>I am grateful to Jeffrey Williams and Rocky Carlyle for providing these data.

<sup>9</sup>A salvage certificate is issued when the insurance company insuring the vehicle is deems it too expensive to repair after a collision.

<sup>10</sup>“Last” is a key word here—a non-trivial number of vehicles have multiple junk certificates issued, generally to different license plates.

- (a) The date is before December 1, 2008.
- (b) The test was failed and the date is before November 1, 2010.

This tells me whether a vehicle is scrapped during my sample period, but not when. For cases 1 and 3 above, I assume the date of the junk, salvage or PNO certificate is also the scrap date. The vehicles owner may not get the certificate issued on the actual date the vehicle is scrapped, but it should be close enough for my purposes. When a car is considered scrapped due to cases 2 or 4, I must approximate. In case 2 the scrap date must be some time before the date registration was paid through, but after the last Smog Check. Since registration is required annually in California, in case 2 I assume that the scrap date is 6 months before the “paid thru” date, or halfway in between the last Smog Check and the “paid thru” date, whichever is later. If last date registration is paid through is more than 4 years after the last Smog Check, I drop the vehicle from the analysis. In case 4, if the test is passed, I only know that the car must have been scrapped some time between the date of the test and the date the next test is due. I split the difference and set the scrap date at one year after the last Smog Check, assuming this will be correct on average. If the test is failed, the final Smog Check date is the scrap date.

## **B Conversion Equations**

The following equations are used to convert emissions measurements of HC, NO<sub>x</sub> and CO from parts per million, as they are reported in Smog Check tests, to grams per mile, as is standard in the Federal Test Procedure (FTP). They are reproduced from CARB (2006), and were originally reported in BAR and CARB’s report, *2004 Evaluation of the California Enhanced Inspection and Maintenance (Smog Check) Program*.

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

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

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

Where:

$$HC^* = \ln((Mode1_{HC} \cdot Mode2_{HC})^{.5}) - 3.72989$$

$$CO^* = \ln((Mode1_{CO} \cdot Mode2_{CO})^{.5}) + 2.07246$$

$$NO^* = \ln((Mode1_{NO} \cdot Mode2_{NO})^{.5}) - 5.83534$$

$$MY^* = modelyear - 1982.71$$

$$weight = \text{Vehicle weight in pounds}$$

$$TRUCK = 0 \text{ if a passenger car, } 1 \text{ otherwise}$$

And where  $Mode1_x$  is the emissions reading from mode 1 of an ASM smog check for pollutant  $x$ , in parts per million for HC and NOx, and in percent for CO.

## C Carbon Dioxide Emissions

Most of the recent slate of national vehicle retirement programs were aimed in part at reducing emissions of carbon dioxide, a greenhouse gas responsible for climate change. While the Vehicle Buyback was not intended to reduce CO<sub>2</sub>, it is worth considering how the program may have affected emissions of this important pollutant. Although Smog

Checks measure CO<sub>2</sub> in percent, I have no conversion equation to convert this into grams per mile. Instead, I use EPA fuel economy and the stoichiometric properties of gasoline. One gallon of fuel contains 19.4 pounds of CO<sub>2</sub> burned. I value a ton of CO<sub>2</sub> at \$28, the figure estimated by the CBO for the Waxman-Markey cap and trade bill of 2009.

The average vehicle scrapped by the Vehicle Buyback program is actually more fuel efficient than the fleet average vehicle by about 2 miles per gallon.<sup>11</sup> The average Buyback Vehicle had a fuel economy of about 27 MPG and would have emitted 10300 pounds of CO<sub>2</sub> over it's remaining lifetime. The average replacement vehicle had a fuel economy of about 25 MPG, and would have emitted 10750 pounds of CO<sub>2</sub> over the remaining lifetime of the buyback vehicle. Based on the assumptions above, the average environmental benefit of retiring each vehicle would be lower by \$5 if CO<sub>2</sub> were taken into account. Of course, if the VMT of the buyback vehicles was actually replaced with a fleet average vehicle having lower fueleconomy, there would likely be a rebound effect. Driving would be relatively more expensive in the replacement vehicle, and as such the replacement VMT might be lower. However, even if the buyback vehicles were not replaced at all, an unlikely assumption, the environmental benefit from CO<sub>2</sub> reductions would amount to less than \$150 per vehicle.

## D Propensity Score

Here I discuss the details of my propensity score estimation. Section [D.1](#) presents basic procedure and variables used to calculate the propensity score. Section [D.2](#) shows how my counterfactual estimates change with different specifications of the propensity score, and [D.3](#) discusses the impact of the Abadie and Imbens (2007) bias correction on my counterfactual estimates.

### D.1 Base Specification

I calculate the propensity score by estimating a logit regression of the probability of participation on vehicle and owner characteristics as of the date the relevant buyback vehicle was scrapped. Vehicle characteristics are total odometer reading, average miles per day, and emissions in grams per mile. I also include a set of variables that might influence the behavior of the owners: the value of the scrapping subsidy being offered by the Buyback, the number of days until the vehicle's next registration is due, and the

---

<sup>11</sup>The fleet average is calculated by year based on the universe of the Smog Check data. I have tried using the total fleet average, the average of all vehicles post-1991, and the average of vehicles 10 years and younger, and the difference is approximately the same.



demographics of the ZIP code where the previous Smog Check took place. As there are only two prices offered during my sample (\$500 before October 2004, \$650 after), I scale the price to be a dummy variable. Vehicles in California are required to register at the same time every year (coincident with the first issuance of the license plate), regardless of changes of ownership, and thus the month and day of the last date of registration given in my DMV data tells me the next registration date for each vehicle. ZIP code demographics are taken from the 2000 Census. Although the ZIP code of the Smog Check station may not be the same as the ZIP code of the vehicle owner's residence, the demographics should be correlated, and in any event the demographics of the area where the vehicle owner chooses to take their car to be inspected is still information that may bear on their preferences.

I include only control vehicles from the eligible group, since owners in the ineligible group could never participate in the Buyback program. Many control vehicles match more than one Buyback vehicle, and when this occurs I include an observation for each matched purchase date, weighting by the inverse of the number of matches to prevent double-counting. Since very few control vehicles have a Smog Check on the exact date that their matched Buyback Vehicle was scrapped, I impute their characteristics by a linear interpolation of the values at the Smog Checks immediately before and after the purchase date. For vehicles sold to the Buyback, I use the emissions readings from their last Smog Check.

The full set of coefficients is too numerous to list, and so table A.5 shows the variables included in each specification and McFadden’s pseudo-R-squared as a measure of fit.<sup>12</sup> Adding station demographics adds a little explanatory power to the model, allowing the effect of vehicle characteristics to vary with the level of the scrappage subsidy adds more, and entering all vehicle characteristics as 3rd order polynomials and owner characteristics as quadratic adds more still<sup>13</sup> Adding model year and make (3-digit VIN prefix, to be precise) only improves the fit of the model a moderate amount. For reasons of computational tractability, rather than using the full 10-digit VIN prefix as a fixed effect, I use the first 8 digits of the 17-digit VINs (i.e., everything but model year and plant) for the model in column 6. The fact that adding prefix fixed-effects to the already flexible model in column 5 changes the model fit relatively little suggests that my logit model is explaining most of the explainable variation in program participation.

I use the model in column 6 of table A.5 to predict the probability of participation for all Buyback Vehicles and all control vehicles in both groups.

## D.2 Propensity Score Specification

I will now show the impact of the specification of the propensity score on my results. Table A.6 shows these results, with each column summarizing the counterfactual estimates based on the eligible group using the propensity score predicted by the corresponding model in table A.5, as compared to the baseline using model 6. Column 6 shows the results of the baseline, for comparison. The alternate specifications for estimating the propensity score do not change the predicted outcomes very much. Increasing the fit of the propensity score model decreases the estimated remaining VMT and emissions, confirming the intuition that the endogenous selection of vehicles into the program favors shorter lived vehicles.

## D.3 Abadie-Imbens Bias Correction

Abadie and Imbens (2006) show that matching estimators based on a simple average of “nearest neighbors” will be biased if there are not perfect matches. A solution to this is the bias correction suggested in Abadie and Imbens (2007). In essence, this involves adjusting the outcomes for each control vehicle to what they would be if they were an exact match

---

<sup>12</sup>A full table of coefficients is available from the author.

<sup>13</sup>I have tried higher order polynomials, but these proved statistically insignificant and did not improve the fit of the model.

for their matched buyback vehicle. Specifically, for a given outcome I estimate

$$\hat{\mu}(x) = E[Y(0)|X = x] = \hat{\beta}'x;$$

where  $Y(0)$  is the outcome for a vehicle if it does not participate in the Buyback Program, and  $X$  is a vector of covariates. Only vehicles which are among the four closest matches to their Buyback vehicle are used for this regression. Then, the bias-adjusted outcome for each control vehicle  $i$  matched to Buyback vehicle  $j$  is given by

$$\tilde{Y}_{0i} = Y_{0i} + \mu(x_j) - \mu(x_i).$$

That is, the actual outcomes for the control vehicle are adjusted by the difference between the predicted outcome for the Buyback vehicle's characteristics and the predicted outcome for its own characteristics. Then, the predicted counterfactual outcome for each Buyback vehicle  $j$  with set of matches  $\mathcal{J}_M(j)$  is:

$$Y_j(0) = \frac{1}{\#\mathcal{J}_M(j)} \sum_{i \in \mathcal{J}_M(j)} Y_{0i} + \mu(x_j) - \mu(x_i)$$

Abadie and Imbens (2007) recommend using a smaller and simpler group of covariates for the bias correction than is used to estimate the propensity score. I use all of the dummies for 8 digit VIN prefix, calendar year and model year, and specify the continuous variables linearly. Odometer readings and average miles per day seem to have an unusually large correlation with outcomes, and so I estimate the bias correction with and without those variables. Table A.7 shows the summary statistics of my estimates based on the eligible group, as they compare to my baseline estimates. Column 1 repeats my preferred specification from the first columns of table 2 and table 3. Column 2 shows the changes from applying the Abadie-Imbens bias correction using all of the control variables, and column 3 applies the bias correction without odometer reading and average miles per day. Column 4 includes all matches, instead of just the closest four, but still excludes control vehicles above the 90th percentile of distance from their Buyback vehicle. Column 5 uses the best 4 matches, but does not exclude outliers, and column 6 includes all matches by VIN prefix, ignoring the propensity score.

While choosing the best 4 matches and dropping outliers makes a moderate difference for the predicted outcomes, the bias correction changes the average predicted outcomes relatively little. The average predicted reduction in criteria pollutants decreases by about 2.1 kilograms. The most likely reason that the bias correction is small is my sample size.

A typical dataset used with matching estimators has treatment observations numbering in the hundreds and control observations in the low thousands at most. By contrast, depending on the control group I have over 13,000 treatment observations whose counterfactual estimates are computed using hundreds of thousands of potential matches. My sample is big enough that the bias is mostly averaged out. As well, the fact that I am matching on VIN prefix prior to estimating the propensity score, the control vehicles are necessarily closer matches for the Buyback vehicles than their time-varying characteristics might make it appear. As the bias correction produces unbelievable individual estimates without altering the mean very much, I feel confident skipping the bias correction in my main estimates. Further, since the bias correction adjusts the remaining miles and emissions downward, omitting the correction is consistent with my attempt to put an upper bound on the cost-effectiveness of the program.

## E Bootstrap Procedures

Here I will discuss alternate bootstrapping procedures for calculating the standard errors in tables 2, 3 and 4. For the standard errors presented in the tables, I sample from replacement from clusters of estimated outcomes, where clusters are defined by 10-digit VIN prefixes.

First I offer a bit of background on bootstrap-based inference with matching estimators. Generally a conventional bootstrap, sampling from the full set of data with buyback and control vehicles, would be invalid for a matching estimator such as the one I use (Abadie and Imbens 2008). The difficulty is that the distribution of the number of times each control vehicle serves as a match to a buyback vehicle in each resample does not match the distribution of match frequency in the real data: any time a buyback vehicle is sampled twice, the closest 4 control vehicles by propensity score automatically match it. However, de Luna, Johansson and de Luna (2010) show that this problem can be avoided by sampling from individual estimated outcomes post-match.

Normally, this would in turn create another problem: individual estimated outcomes which share the same controls will be dependent on each other, and failure to account for this dependence will lead to biased standard errors. de Luna et al. (2010) suggest a circular block bootstrap procedure to overcome the dependence of individual estimates. In the context of my analysis, however, it is more natural to perform cluster sampling. Because control vehicles for any given buyback are restricted to match within VIN prefix, sampling clusters of VIN prefixes will account for any dependence among individual estimates.

However, my estimated standard errors and in particular the inferences made in section

4 are robust to alternate specifications. Table A.8 shows the point estimates from the eligible control group, and various estimates of the standard errors. The first column of the table gives the point estimates from tables 2 and 3, and the second column gives the standard errors from those tables, calculated by a bootstrap sampling VIN prefix clusters.

The next column samples VIN prefixes, but adds a perturbation of the propensity score with each iteration, in the spirit of a wild bootstrap. For each control vehicle, there is a 50% change that its propensity score distance from its matched buyback vehicle is increase by 0.1, and a 50% chance that it is decreased by 0.1. If that Abadie-Imbens critique of the bootstrap for matching estimators also applied to sampling estimated outcomes, this might address the problem by ensuring that sampling an additional treatment vehicle does not deterministically cause a match to its closest controls. In fact, this makes very little difference in the estimated standard errors.

The next two columns change the level of the clusters I sample from. The fourth column of the table samples 3-digit VIN prefixes (i.e. makes) as clusters. This allows for more dependence among the individual estimates than sampling 10-digit prefixes, but may be less valid as well. I have only 92 unique VIN prefixes, and many have only one or two individual estimated outcomes in them. Indeed, the table shows that my standard errors inflate substantially when clustering by make, by a factor of 3 to 4. However, they are still small compared to the differences I observe between groups in section 4. The next column does not cluster at all, resampling from individual estimates independently. This shrinks the standard errors somewhat relative to clustering by 10-digit VIN. Finally, the last column presents naive standard errors of the means, with no bootstrapping. These standard errors are lower than my preferred bootstrap specification by about a third.

## F Hazard Analysis

To get the total remaining emissions and miles traveled for each control vehicle, I must know when the vehicle exits the car fleet. For about half the vehicles in my control groups, I observe some evidence of scrappage in the Smog Check or DMV data (see appendix A.3). The remaining vehicles are right-censored. As far as I can identify, these vehicles are still on the road as of the end of my data in October 2009, May 2010, or December 2010, depending on the source. To address this, I must estimate the underlying hazard function and use this to find the expected value of the scrappage date for each right-censored vehicle. Let  $T_i$  be the time when vehicle  $i$  is scrapped. Since I have vehicles of different vintages, time will be defined in terms of vehicles' age. Define the instantaneous hazard

function  $h(t, X_{it})$  as

$$h(t, X_{it}) = \lim_{\Delta t} \frac{\Pr[t \leq T_i < t + \Delta t | T_i, X_{it}]}{\Delta t},$$

the probability of scrapping occurring between time  $t$  and the next period as the interval length goes to zero, conditional on vehicle  $i$  surviving at least until time  $t$  and on a set of (potentially time-varying) covariates  $X_{it}$ .

One approach to estimating  $h(t, X_{it})$  is to assume proportional hazards, such that  $h(t, X_{it})$  can be decomposed as

$$h(t, X_{it}) = h_0(t) \exp(\alpha + X_{it}\beta).$$

Where  $h_0(t)$  is the baseline hazard and is the same for all individuals at any given time. In a proportional hazards model, the interpretation of  $\beta$  is that an increase in a covariate by one unit multiplies the hazard by  $\exp(\beta)$ . Alternately, the hazard can be estimated by modeling  $T$  in what is known as an accelerated failure time framework, such that

$$\ln T = X_{it}\beta + \sigma W,$$

where  $W$  is an appropriate (parametric) distribution. In this case the hazard is decomposed as

$$h(t, X_{it}) = h_0(t)(t \exp(\alpha + X_{it}\beta)) \exp(\alpha + X_{it}\beta).$$

In an accelerated failure time framework, an increase in  $X$  by one unit multiplies  $T$  by  $\exp(\beta)$ . Most distributions lend themselves to one framework or the other, although a Weibull model can be parameterized either way.

If the goal were to estimate the coefficient vector on  $X_{it}$ , the preferred method would be to apply a semi-parametric estimation method—either a Cox regression or a logit regression with time dummies are commonly used, either of which assumes proportional hazards and treats the baseline hazard  $h_0(t)$  non-parametrically. However, these semi-parametric methods only provide in-sample predictions of the hazard rate. My aim is not to estimate  $\beta$  but instead to make out-of-sample predictions regarding  $T$ , and so I must have an estimate of  $h(t, X_{it})$  for any point in time. To do this I must make a parametric assumption on the functional form of the hazard function.

While in principle analysis time could be denominated in days, I use quarters instead. The scrap dates I use for the non-censored vehicles are not truly accurate to the day, and in any event the time-horizon for scrapping decisions is surely longer than a day in most

cases. By construction, no vehicles in my sample of control vehicles are at risk of being scrapped prior to reaching 72 quarters of age, when 1981 vehicles become eligible for the Buyback Program in 1998. As such, for the purpose of estimating the hazard function, I define hazard time such that  $t = 0$  implies that age, in quarters, equals 72. To calculate vehicle age in quarters I use the DMV information on the date of last registration, as the month registration is due for a vehicle remains the same throughout the lifetime of the vehicle and corresponds to the date of first purchase. Thus, the quarter when age equals zero is the quarter when registration is due of the vehicle’s model year.<sup>14</sup> Where a vehicle has more than one Smog Check in a quarter, I take the maximum value of its covariates among those Smog Checks. The covariate vector  $X_{it}$  includes a polynomial of mileage (in 100,000s), a dummies for failing a Smog Check and for being flagged a gross polluter in addition to failing a Smog Check, and engine characteristics.

Since my control vehicles are further selected to be those which last at least until their matched buyback vehicle is scrapped, estimates of the hazard function using only these vehicles will be biased—there will seem to be too little scrappage in the earlier quarters. To avoid this problem, I include in my hazard analysis all vehicles which are VIN prefix matches to the Buyback vehicles, have half their Smog Checks in BAAQMD or neighboring counties and which are at least 18 years old, regardless of their last appearance in the Smog Check Data.

Rather than choose a parametric form arbitrarily, I estimate the baseline hazard parameters for a several hazard distributions, and compare these to a smoothed non-parametric hazard function calculated with a Cox regression. Each hazard function is generated by a regression model which also controls for a vector of covariates.

To determine the best choice of baseline hazard function, I estimate proportional hazard models using the Weibull and Gompertz and distributions, and accelerated failure time models assuming log-logistic, log-normal and generalized gamma distributions, as well as a Cox regression. Although they are not entirely comparable, coefficient estimates from the different models are very similar.<sup>15</sup> The baseline hazard functions are plotted in figure A.4 The parametric hazard functions track the non-parametric estimates fairly well until  $t \geq 22$ , except for the log-normal, which matches very poorly. After this point, the Weibull and generalized gamma distributions seem to have the closest fit, with the Weibull perhaps making a somewhat better fit until very late in analysis time, when  $t \geq 38$ . It is

---

<sup>14</sup>Since new model years are generally released in the fourth quarter of the previous year, I assume that vehicles which register in the fourth quarter were first sold the year before their model year.

<sup>15</sup>That is, the accelerated failure time coefficients are very similar in all four models which lend themselves to that interpretation, and the three models which can be interpreted as proportional hazards have similar hazard ratios.

not clear from visual inspection which of the two functions better captures the baseline hazard after this point. However, since I am attempting to place an upper bound on the cost-effectiveness of the Vehicle Buyback Program, I will assume a Weibull distribution for scrapping time, as it implies a lower hazard rate and longer remaining life for control vehicles with censored scrappage dates.

To estimate remaining life, I want to control for non-time varying unobservable characteristics. In the context of hazard analysis, this is typically done by stratifying the regression, allowing groups to have a separate baseline hazard function. Ideally I would stratify by the 10-digit VIN prefix, but this is computationally intractable. In a Weibull regression, stratifying means estimating a separate shape parameter  $\lambda_k$  and constant term  $\alpha_k$  for each strata. With more than 4000 10-digit prefixes in my sample, this is simply not feasible. As well, the cell sizes of the 10 digit prefixes are frequently quite small, making it difficult to estimate the distribution parameters.<sup>16</sup> Instead, I stratify the regression on the first 3 digits of the VIN (roughly corresponding to make), and control for engine size and model year as part of  $X_{it}$ .

Assuming a Weibull distribution for the baseline hazard, I estimate the hazard of vehicle  $i$ , of make  $k$  being scrapped between time  $t$  and  $t + \delta t$  as:

$$h_k(t, X_{it}) = \lambda_k t^{\lambda_k - 1} \cdot \exp(\alpha_k + X_{it}\beta)$$

Table A.9 shows the results of my hazard analysis. I present the exponentiated coefficients, which can be interpreted as hazard ratios. Increasing a covariate by 1 unit multiplies the hazard rate by the hazard ratio. Columns 1-3 are specified without stratifying on the make prefix, while columns 4 and 5 show results with stratification. The coefficients are quite similar across specifications. Increased mileage increases the probability of scrappage, although at a decreasing rate with higher mileage.

With the parameter vectors in hand, I estimate expected survival times for the right-censored vehicles as the sum of the vehicle's probability of survival over all future periods. Let  $S_k(t, X_{it}) = 1 - h_k(t, X_{it})$  be the probability that vehicle  $i$ , of make  $k$ , survives to period  $t + 1$  given survival to  $t$ , and let  $\bar{T}_i$  be the last period vehicle  $i$  is observed before being censored. The evolution of the hazard rate over time is determined by  $\lambda_k$  and  $X_{it}\beta$ . However, since increasing mileage impacts hazard rates, I assume that the vehicle continues to travel the same number of miles per day as at its latest Smog Check, and

---

<sup>16</sup>In practice I have found that the log-likelihood function fails to reach convergence in Stata when trying to estimate shape parameters for groups of fewer than 100 vehicles.



increase  $X_{it}\beta$  accordingly. This means that the survival probability is given by,

$$S_k(t, X_{it}) = 1 - \lambda_k t^{1-\lambda_k} \exp(\alpha_k + X_{it}\beta + \frac{90 \cdot mpd_i \bar{T}_i}{100,000} \cdot (t - \bar{T}_i) \cdot \beta_{mileage}).$$

Then ,

$$E(T_i) = \bar{T}_i + \sum_{j=\bar{T}_i}^{\infty} \prod_{l=\bar{T}_i}^j S_k(l, X_{il}).$$

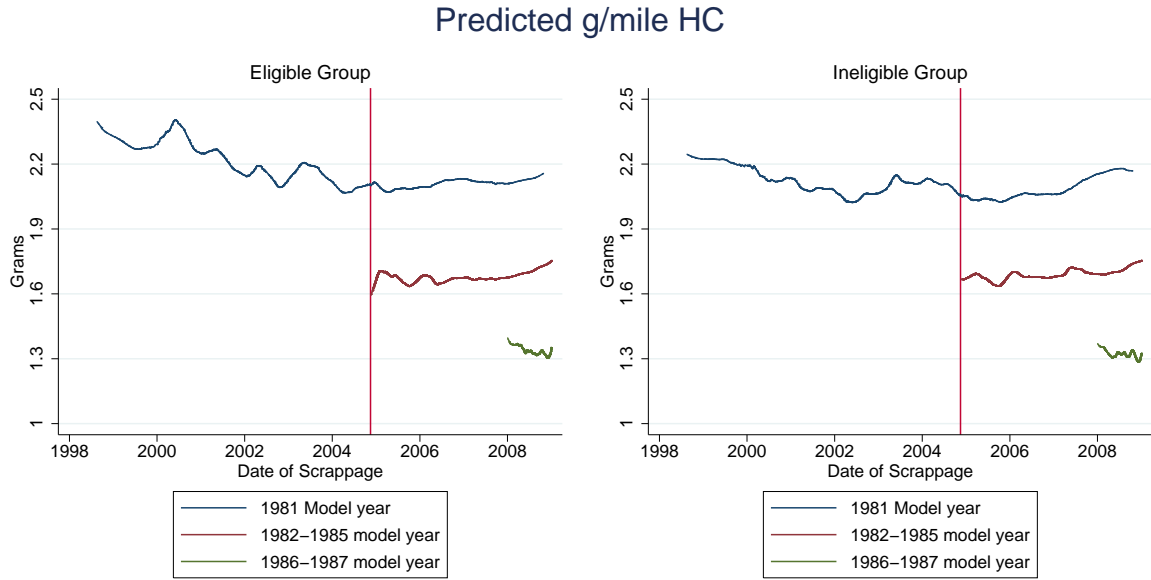
I calculate this sum iteratively for each vehicle until the expected scrappage time changes by less than .001. The bottom panel of Table A.9 shows summary statistics of the predicted remaining life calculated this way. Based on my preferred specification in column 5, the average censored vehicle would last an average of about 10.5 quarters (two and a half years), although the predictions range from essentially zero to more than 60 quarters (15 years).

## References

- Abadie, Alberto and Guido Imbens**, “On the Failure of the Bootstrap For Matching Estimators,” *Econometrica*, 2008, 76 (6), 1537–1557.
- and **Guido W. Imbens**, “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, January 2006, 74 (1), 235–267.
- and — , “Bias Corrected Matching Estimators for Average Treatment Effects,” August 2007. Working Paper.
- CARB, California Air Resources Board**, “The Carl Moyer Program Guidelines: 2006 Project Criteria for Light-Duty Vehicles,” 2006.
- de Luna, Xavier, Per Johansson, and Sara Sjstedt de Luna**, “Bootstrap Inference for K-nearest Neighbour Matching Estimators,” November 2010. IZA Discussion Paper No. 5461.
- Dill, Jennifer**, “Estimating emissions reductions from accelerated vehicle retirement programs,” *Transportation Research Part D*, 2004, 9, 87–106.
- Hubbard, Thomas**, “An Empirical Examination of Moral Hazard in the Vehicle Inspection Market,” *The RAND Journal of Economics*, 1998, 29 (2), 406–426.

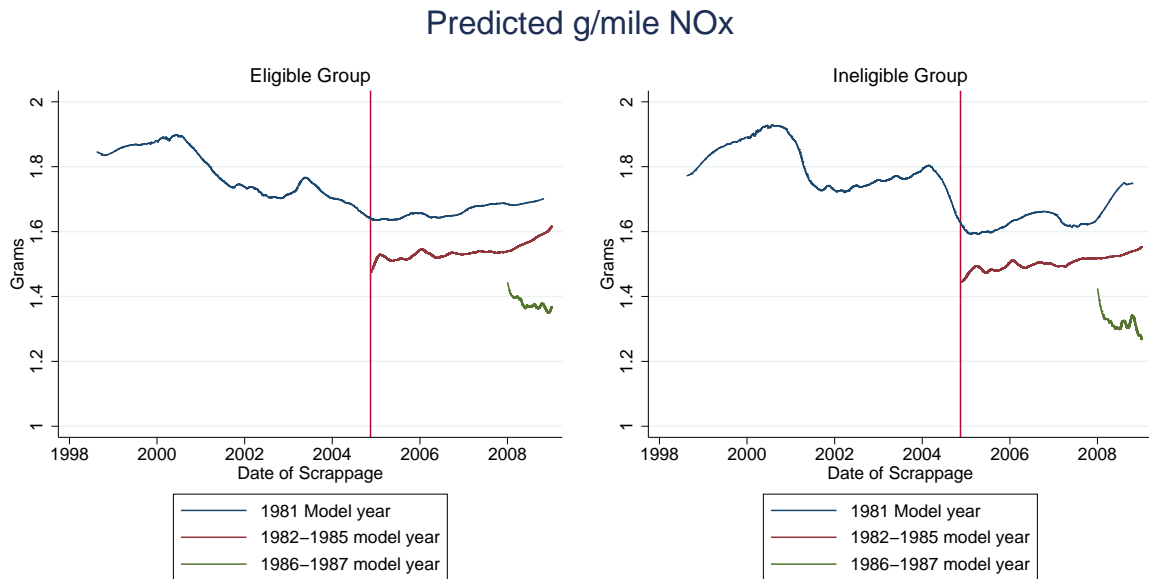
# Figures

Figure A.1: Projected Pollution Levels



LOWESS plot of estimated average g/mile HC on date of purchase. Centered subsets of  $N^*.1$  observations are used to calculate each smoothed point for model year 1981. Model years 1982-1985 and 1986-1987 use subsets of  $N^*.15$  and  $N^*.25$ , respectively.

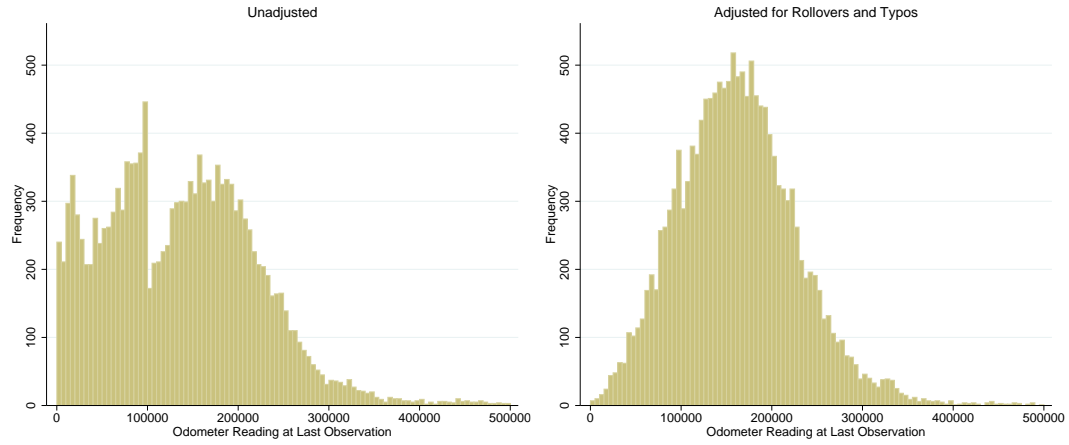
Figure A.2: Projected Pollution Levels



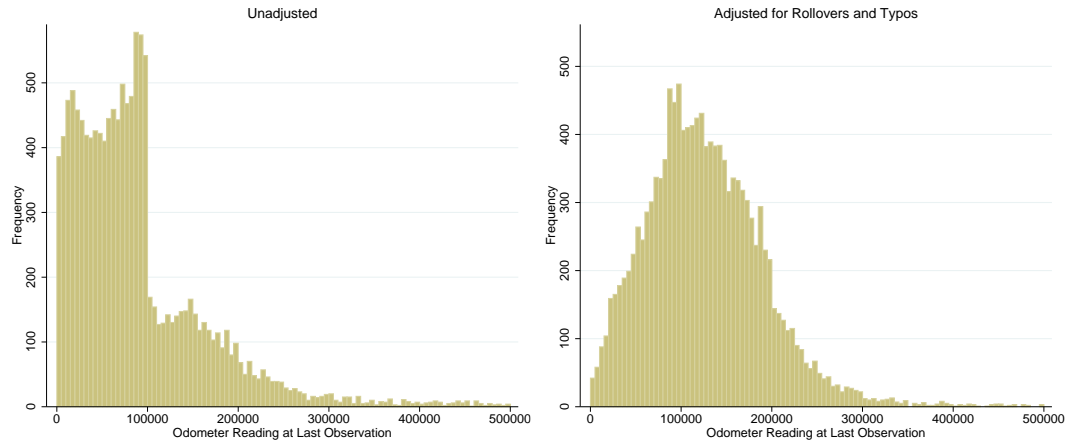
LOWESS plot of estimated average g/mile HC on date of purchase. Centered subsets of  $N^*.1$  observations are used to calculate each smoothed point for model year 1981. Model years 1982-1985 and 1986-1987 use subsets of  $N^*.15$  and  $N^*.25$ , respectively.

Figure A.3: Distribution of Odometer Readings Before and After Correction

(a) Buyback Vehicles, 1981-1987 Model Year



(b) Buyback Vehicles, 1980 and Earlier Model Year



(c) Control Vehicles

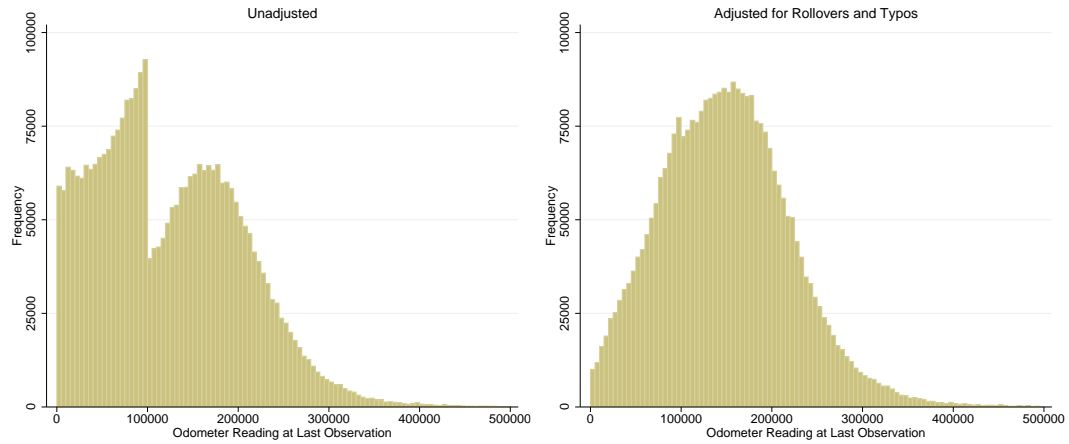
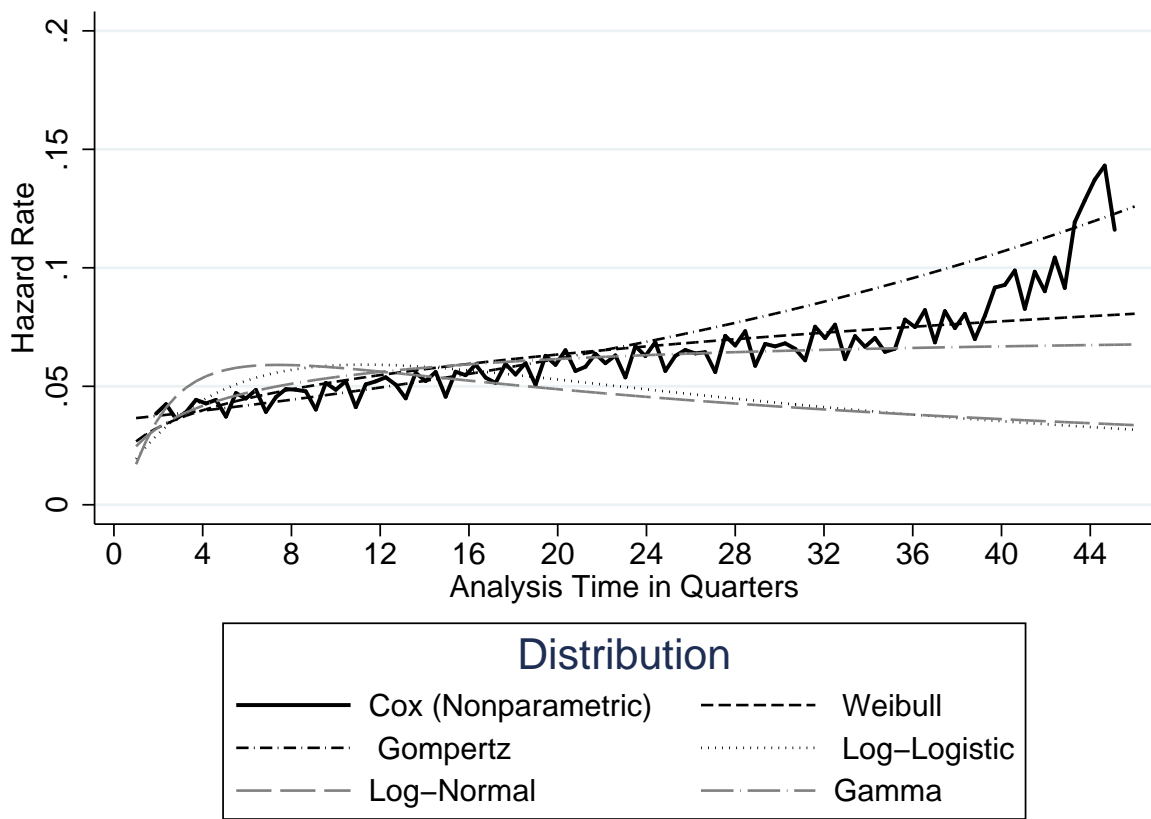


Figure A.4: Quarterly Baseline Hazard Functions of Parametric and Non-Parametric Models of Hazard of Scrappage for Control Vehicles Over 18 Years (72 quarters) Old



# Tables

Table A.1: Marginal Benefit of reducing a kilogram of NOx and HC (year 2000 dollars)

Location	MM 2009 Baseline		MM 2009 EPA		CMP
	NOx	HC	NOx	HC	NOx+HC
Alameda County	-0.29	2.94	1.65	19.28	14.10
Contra Costa County	0.42	3.97	4.72	26.42	14.10
Marin County	0.83	2.18	5.00	14.90	14.10
Napa County	1.23	1.16	7.70	7.36	14.10
San Francisco County	0.37	4.32	2.52	30.03	14.10
San Mateo County	0.24	2.05	1.55	13.59	14.10
Santa Clara County	0.18	2.71	1.44	17.56	14.10
Solano County	1.31	2.17	8.51	14.12	14.10
Sonoma County	1.23	0.96	8.09	6.39	14.10
BAAQMD Average	0.61	2.50	4.57	16.63	-
National Average	0.50	0.51	3.53	3.28	-

Table A.2: Estimated Counterfactual Outcomes of Scrapped Vehicles (Ineligible Control Group)

	Quartile of VMT Remaining			
	1st	2nd	3rd	4th
VMT Remaining	3,184.5 (56.7)	8,947.7 (89.5)	15,505 (123.3)	32,579 (313.4)
Days of driving remaining	1,247.4 (10.74)	1,341.8 (9.49)	1,403.9 (8.84)	1,544.9 (9.35)
Total HC + NOx reduction per car (kg)	14.1771 (0.275)	39.3807 (0.497)	66.1223 (0.769)	128.601 (1.467)
Mean Propensity Score Distance	.061287 (0.0010)	.052265 (0.0009)	.047349 (0.0009)	.043788 (0.0009)
Number of Control Vehicles Used	3.5003 (0.0156)	3.73264 (0.0121)	3.80478 (0.0108)	3.73444 (0.0131)
<b>Net Pollution Benefits</b>				
MM 2009 Baseline	-704.565 (0.586)	-663.744 (0.992)	-618.372 (1.616)	-510.826 (3.042)
MM 2009 USEPA	-563.615 (3.468)	-279.014 (6.578)	38.5451 (10.816)	780.764 (20.832)
CMP Guidelines	-528.878 (3.90)	-174.058 (6.97)	203.153 (10.91)	1,088.1 (20.94)
<b>Proportion Cost Effective</b>				
MM 2009 Baseline	0 (0.0000)	.000596 (0.0004)	.003284 (0.0010)	.017267 (0.0024)
MM 2009 USEPA	.010149 (0.0021)	.123398 (0.0062)	.408358 (0.0097)	.778803 (0.0075)
CMP Guidelines	.012239 (0.0023)	.215499 (0.0092)	.612537 (0.0099)	.897291 (0.0052)
Observations	14033			

Statistics are means of estimated outcomes. Bootstrapped standard errors are reported below in parentheses. Counterfactual estimated from average actual outcomes of 4 closest matches by VIN prefix and propensity score.

Table A.3: Estimated Outcomes by Make

	CMP Benefit	VMT	Emissions (kg)	HC g/mile	NOx g/mile	N
GMC	1,211	17480	85.911	2.25	1.8573	110
Chevrolet	1,059	16721	75.1	2.0527	1.7674	805
Toyota	995.09	18470	70.593	1.7183	1.4596	3,198
Mercury	924.47	15533	65.243	1.7157	1.4994	219
Nissan	915.62	16343	64.96	1.7996	1.5553	1,021
Volvo	884.45	14825	62.749	1.6456	1.7118	1,066
Ford	882.84	14144	62.598	2.0312	1.6811	1,121
BMW	867.4	16232	61.539	1.5832	1.5838	584
Dodge	840.82	12172	59.709	2.1541	1.7679	393
Buick	835.14	14958	59.382	1.7108	1.5898	302
Oldsmobile	831.96	15144	59.139	1.7153	1.597	457
Lincoln	821.48	14309	58.282	1.8196	1.6812	108
Pontiac	821.41	16250	58.276	1.5903	1.4981	112
Cadillac	785.6	12750	55.736	1.7795	1.7955	329
Honda	784.96	16887	55.682	1.4125	1.3373	1,770
Mercedes-Benz	697.49	13783	49.485	1.5627	1.6053	134
Mazda	675.07	13907	47.894	1.6707	1.3355	310
Chrysler	583.19	12459	41.375	1.4378	1.4366	144
Volkswagen	568.38	12287	40.315	1.4523	1.42	375
Total	887.18	16090	62.935	1.724	1.5469	1,490
Observations	12134	12147	12147	12558	12558	12558

Table A.4: Average Difference in Emissions Readings Between ASM and TSI tests

	HC (ppm)	CO (%)
Correction	50.41	0.07
SE	0.07	0.0007
$R^2$	0.73	0.59



Table A.5: Propensity to Participate in Vehicle Buyback Program

	(1)	(2)	(3)	(4)	(5)	(6)
Vehicle Characteristics	✓	✓	✓	✓	✓	✓
Subsidy Level	✓	✓	✓	✓	✓	✓
Days until Next Registration	✓	✓	✓	✓	✓	✓
Station ZIP Code Characteristics		✓	✓	✓	✓	✓
Subsidy-Vehicle Characteristic interactions			✓	✓	✓	✓
Polynomial Terms				✓	✓	✓
Model Year Fixed Effects					✓	✓
Make Fixed Effects					✓	
VIN Prefix Fixed Effects						✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	745142	736590	736590	736590	736585	736394
Pseudo $R^2$	0.076	0.081	0.089	0.111	0.178	0.184

Table A.6: Robustness Checks - Propensity Score Specifications

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
VMT Remaining	+1539	+1554	+1159	+603	+245	16028
Days of driving remaining	+27	+35	+24	+43	+13	1394
Total HC reduction per car (kg)	+4.024	+3.883	+2.497	+1.514	+0.215	34.859
Total NOx reduction per car (kg)	+2.639	+2.62	+1.621	+0.796	*	28.453
Total CO reduction per car (kg)	+38.47	+39.37	+27.77	+16.91	+3.28	355.89
Total HC + NOx reduction per car (kg)	+6.663	+6.503	+4.117	+2.309	+0.268	63.312
Estimated average grams of HC per VMT	*	*	*	*	*	1.7341
Estimated average grams of NOx per VMT	*	*	*	*	*	1.5543
Estimated average grams of CO per VMT	+0.156	+0.077	+0.116	+0.102	*	17.18
Control Vehicles	+0.1504	+0.1412	+0.128	+0.1169	+0.0675	3.6098
<b>Social Benefit of Scrapping</b>						
MM 2009 Baseline	+12.72	+12.15	+7.81	+4.9	+0.74	104.42
MM 2009 USEPA	+87.1	+83.78	+53.53	+33.46	+4.78	723.47
CMP guidelines	+97.59	+95.33	+61.17	+33.21	+4.01	889.33
<b>Net Welfare Gain/Loss</b>						
MM 2009 Baseline	+12.7	+12.18	+7.8	+5.05	+0.92	-625.13
MM 2009 USEPA	+87.09	+83.81	+53.53	+33.62	+4.97	-6.0877
CMP Guidelines	+97.58	+95.36	+61.17	+33.37	+4.2	159.78
<b>Proportion Cost Effective</b>						
MM 2009 Baseline	*	*	*	*	*	.00331
MM 2009 USEPA	+0.04554	+0.04343	+0.03698	+0.01821	*	.32951
CMP Guidelines	+0.05326	+0.0482	+0.03771	+0.0173	*	.43886
N	13129	13257	13272	13390	13560	13599

\* : Cannot reject equal to base specification,  $p > .05$

Table A.7: Robustness Checks - Matching Estimators

	(1) Base	(2) Bias-Adjusted	(3) No Miles	(4) No Outliers	(5) Top 4	(6) All Matches
VMT Remaining	16028	-157	-701	+820	+201	+1580
Days of driving remaining	1394	-38	-14	+9	+13	+32
Total HC reduction per car (kg)	34.859	-1.381	-2.762	+1.91	+0.902	+4.439
Total NOx reduction per car (kg)	28.453	-0.557	-1.691	+1.209	+0.318	+2.43
Total CO reduction per car (kg)	355.89	-14.18	-28.98	+25.04	+9.14	+54.05
Total HC + NOx reduction per car (kg)	63.312	-1.938	-4.453	+3.118	+1.22	+6.869
Estimated average grams of HC per VMT	1.7341	-0.0673	-0.0696	*	*	*
Estimated average grams of NOx per VMT	1.5543	-0.0508	-0.0522	*	*	-0.0164
Estimated average grams of CO per VMT	17.18	-0.726	-0.754	+0.209	+0.183	+0.479
Control Vehicles	3.6098	+0.0286	+0.0286	+43.31	+0.1164	+47.1
<b>Social Benefit of Scrapping</b>						
MM 2009 Baseline	104.42	-2.09	-6.641	+6.01	+2.72	+13.9
MM 2009 USEPA	723.47	-13.17	-45.55	+41.21	+18.07	+94.53
CMP guidelines	889.33	-11.63	-48.81	+46.3	+18.11	+100.8
<b>Net Welfare Gain/Loss</b>						
MM 2009 Baseline	-625.13	-2.08	-6.63	+6.05	+3.47	+14.7
MM 2009 USEPA	-6.0877	-13.14	-45.53	+41.26	+18.82	+95.34
CMP Guidelines	159.78	-11.62	-48.8	+46.33	+18.85	+101.6
<b>Proportion Cost Effective</b>						
MM 2009 Baseline	.00331	*	*	*	*	*
MM 2009 USEPA	.32951	*	-0.02294	+0.06272	*	+0.1041
CMP Guidelines	.43886	*	-0.02206	+0.144	*	+0.1826
N	13599	13599	13599	13599	14636	14703

\* : Cannot reject equal to base specification,  $p > .05$

Table A.8: Robustness Checks - Bootstrap Procedures

	Bootstrapped SE Clustered By:					Naive SE
	Mean	VIN Prefix	Prefix + Wild	Make	VIN	
VMT Remaining	16028	(161.02)	(173.55)	(501.16)	(118.06)	(116.78)
Days of driving remaining	1394.2	(9.5085)	(8.8795)	(37.170)	(5.0884)	(5.0599)
Total HC reduction per car (kg)	34.86	(0.5227)	(0.5162)	(1.5380)	(0.3216)	(0.3213)
Total NOx reduction per car (kg)	28.45	(0.3418)	(0.3615)	(0.9514)	(0.2359)	(0.2297)
Total CO reduction per car (kg)	355.9	(6.2751)	(6.4517)	(22.047)	(3.5866)	(3.6435)
Total HC + NOx reduction per car (kg)	63.31	(0.7980)	(0.8147)	(2.3110)	(0.5311)	(0.5216)
Estimated average grams of HC per VMT	1.734	(0.0153)	(0.0119)	(0.0646)	(0.0044)	(0.0044)
Estimated average grams of NOx per VMT	1.554	(0.0092)	(0.0069)	(0.0471)	(0.0028)	(0.0028)
Estimated average grams of CO per VMT	17.18	(0.1841)	(0.1527)	(0.9215)	(0.0572)	(0.0586)
Number of Control Vehicles Used	3.610	(0.0153)	(0.0127)	(0.0634)	(0.0077)	(0.0078)
Mean Propensity Score Distance	0.037	(0.0010)	(0.0007)	(0.0030)	(0.0004)	(0.0004)
<b>Social Benefit of Scrapping</b>						
MM 2009 Baseline	104.4	(1.6290)	(1.5519)	(4.8427)	(1.0376)	(1.0554)
MM 2009 USEPA	723.5	(11.143)	(10.633)	(32.895)	(7.1888)	(7.2857)
CMP Guidelines	889.3	(11.277)	(11.433)	(32.671)	(7.5167)	(7.3958)
<b>Net Welfare Gain/Loss</b>						
MM 2009 Baseline	-625.1	(1.7435)	(1.6431)	(5.0326)	(1.0284)	(1.0559)
MM 2009 USEPA	-6.088	(11.201)	(10.687)	(33.050)	(7.1474)	(7.2557)
CMP Guidelines	159.8	(11.319)	(11.473)	(32.895)	(7.4717)	(7.3586)
<b>Proportion Cost Effective</b>						
MM 2009 Baseline	0.003	(0.0006)	(0.0007)	(0.0011)	(0.0005)	(0.0005)
MM 2009 USEPA	0.330	(0.0058)	(0.0057)	(0.0163)	(0.0039)	(0.0040)
CMP Guidelines	0.439	(0.0061)	(0.0059)	(0.0188)	(0.0043)	(0.0043)

Table A.9: Hazard of Scrappage

	(1)	(2)	(3)	(4)	(5)
Failed last smog check	1.030*	1.023	1.023	1.014	1.010
	(0.011)	(0.012)	(0.013)	(0.011)	(0.011)
Gross polluter last smog check	0.993	0.995	0.990	0.986	0.976
	(0.027)	(0.026)	(0.027)	(0.017)	(0.016)
Mileage	1.769*	1.632*	1.605*	1.948*	1.962*
	(0.213)	(0.182)	(0.179)	(0.186)	(0.187)
Mileage <sup>2</sup>	0.909*	0.921*	0.923*	0.904*	0.903*
	(0.018)	(0.016)	(0.017)	(0.017)	(0.017)
Cylinders		1.107	1.101		0.975
		(0.087)	(0.088)		(0.033)
Engine Displacement (liters)		0.810*	0.813*		0.870*
		(0.067)	(0.068)		(0.027)
Fuel Injection		0.907	0.899		0.955
		(0.085)	(0.088)		(0.037)
Stratified on Make	No	No	No	Yes	Yes
Model Year Fixed Effects	No	No	Yes	Yes	Yes
N	588068	585267	585267	588068	585267
Predicted Remaining Survival Time:					
Mean	10.217	10.521	10.411	10.606	10.651
Standard Deviation	2.714	2.991	2.936	4.446	4.579
Max	26.910	37.241	34.831	73.011	72.567
Min	0.920	0.909	0.905	0.890	0.892

Exponentiated Coefficients, reported as hazard ratios

Standard errors clustered by 3-digit VIN prefix reported in parantheses