Estimating Full IM240 Emissions from Partial Test Results: Evidence from Arizona

Amy W. Ando
Winston Harrington
Virginia McConnell

Discussion Paper 98-24

March 1998

RESOURCES FOR THE FUTURE
1616 P Street, NW
Washington, DC 20036
Telephone 202-328-5000
Fax 202-939-3460

Estimating Full IM240 Emissions from Partial Test Results: Evidence from Arizona

Amy W. Ando, Winston Harrington, and Virginia McConnell

Abstract

The expense and inconvenience of enhanced vehicle emissions testing using the full 240-second dynamometer test has led states to search for ways to shorten the test process. In fact, all states that currently use the IM240 allow some type of fast-pass, usually as early in the test as second 31, and Arizona allows vehicles to fast-fail after second 93. While these shorter tests save states millions of dollars in inspection lanes and driver costs, there is a loss in information since test results are no longer comparable across vehicles. This paper presents a methodology for estimating full 240 second results from partial-test results for three pollutants: HC, CO and NOx. Using random sample of vehicles in Arizona which received full 240 second tests, we use regression analysis to estimate the relationship between emissions at second 240 and emissions at earlier seconds in the test. We examine the influence of other variables such as age, model-year group, and the pollution level itself on this relationship. We then use the estimated coefficients in several applications. First, we attempt to shed light on the frequent assertion that the results of the dynamometer test provide guidance for vehicle repair of failing vehicles. Using a probit analysis, we find that the probability that a failing vehicle will passing the test on the first retest *is* greater the longer the test has progressed. Second, we test the accuracy of our estimates for forecasting fleet emissions from partial test emissions results in Arizona. We find that forecast fleet average emissions are very close to the actual fleet averages.

Key Words: inspection and maintenance, mobile source, fast pass

JEL Classification No.: Q25

Table of Contents

List of Figures and Tables

Estimating Full IM240 Emissions from Partial Test Results: Evidence from Arizona

Amy W. Ando, Winston Harrington, and Virginia McConnell[1]

I. INTRODUCTION

Four states -- Arizona, Colorado, Ohio and Wisconsin -- now use IM240 tests for vehicle emission measurements in their state Enhanced I/M programs, and other states are expected to do so shortly. Each of these states has, or is expected to adopt, a fast-pass algorithm, which allows emission tests for most vehicles to be truncated before the test has run its full 240 seconds. In addition, Arizona has adopted a fast-fail algorithm. The use of fast-pass and fast-fail algorithms generates raw test results that are not comparable to the results of full-length IM240 tests. In this paper we describe a method for estimating full-length test results from the results of tests that were truncated by these fast-pass and fast-fail options. We also examine whether the use of fast-fail interferes with the main objective of I/M programs: the diagnosis and repair of high-emitting vehicles.

The principal advantage of fast-pass and fast-fail is that their use can substantially reduce both the time motorists spend waiting in testing queues and the costs of vehicle emission testing itself. Also, short tests avoid the high-speed phase of the IM240 trace, which some critics claim has been responsible for vehicular damage. These advantages may be achieved with only a small reduction in the ability to discriminate between normal vehicles and vehicles with high emissions. It was the aim of the fast-pass and fast-fail algorithms recommended by the EPA (and adopted with some minor modifications by the various states so far) to be sufficiently conservative that only a small number of fast-passing vehicles would fail the full test, and few fast-failing vehicles would pass the full test. Although no rigorous test of these algorithms has been completed, most observers are confident that they have largely succeeded in these objectives.[2]

---

[1] Amy W. Ando, Fellow, and Winston Harrington and Virginia McConnell, Senior Fellows, Resources for the Future. Portions of this work were funded by a grant from the EPA. We would like to thank Phil Enns, Ed Glover, and Phil Lorang of EPA's Office of Mobile Sources, Phil Heirigs of Sierra Research, and employees at both the Arizona Department of Environmental Quality, and the I/M contractor, Gordon Darby, Inc. Through their efforts and assistance, we were able to obtain and make sense of the data for this project. The EPA staff and Tom Wenzel of Lawrence Berkeley Labs also provided us with helpful comments at various stages of the effort. All errors remain our responsibility

[2] We used the extrapolation method described in this paper to extrapolate from fast-pass and fast-fail test results to full emission tests, including a random disturbance term. When we compare the fast-pass and fast-fail results to these simulated full-test results, we find very few false positives or false negatives. This leads us to a tentative conclusion that the fast-pass and fast-fail algorithms do reasonably well at discriminating high-emitting vehicles. However, a more conclusive test would be the reverse of our procedure: begin with a sample of complete tests and use the algorithm to simulate fast-pass and fast-fail results. We have not been able to perform this analysis because we have been unable to get a reliable definition of the algorithms actually in use.

Despite the apparent conservatism of the algorithms, the great majority of vehicles fast-pass; in Arizona, for example, over half the vehicles fast-pass at the earliest second allowed by the test guidelines (31 seconds), and for only about 3.7 percent of vehicles is the IM240 test run to completion. Truncation of this magnitude may lead states to need many fewer test lanes, saving millions of dollars annually in inspection costs and driver "convenience" costs.[3]

Unfortunately, allowing fast-passes and fast-fails also has two potential disadvantages. There is some question whether truncation of failing tests interferes with the diagnostic value of the test -- i.e., its ability to help mechanics identify and repair emission malfunctions. In addition, allowing fast-passes and -fails in I/M programs makes it more difficult to use the results of IM240 tests for other purposes. The results of emission tests terminated at different times are not directly comparable because the test conditions vary over the test cycle. Thus, fast-pass and fast-fail procedures could undermine the integrity of a valuable source of data on real-world vehicle emissions. Despite the apparent success of fast-pass and fast-fail in identifying high emitters, it is important for several reasons to have a way of extrapolating to full-test estimates from partial IM240 tests. Some of the reasons are:

i. To construct, prior to the start of an I/M program, a baseline emission inventory for comparison to the emission inventory generated by computer models (EPA's MOBILE5 in all states except California, which has its own model).
ii. To estimate the extent of emission reductions from various vehicle emission reduction programs, including I/M itself.
iii. To estimate relationships between real-world vehicle emissions and vehicle characteristics, such as age, model year, engine size and type, and manufacturer.
iv. To provide the owners of vehicles that fast-pass or -fail with test results that are consistent with the results of full-length tests.

Although EPA regulations permit states to use fast-pass and fast-fail algorithms, they attempt to preserve some of the measurement potential of the test by encouraging programs to include a two percent random sample of vehicles that complete the entire IM240 test. However, not all states plan to subject a random sample of vehicles to the full test; indeed, of the first four states adopting the IM240 test, Ohio does not. Furthermore, a two percent random sample may not provide a sufficiently large sample for some purposes. For example, all states allow or plan to allow failing vehicles to fast-pass on re-test, which means that the number of failing vehicles receiving both complete initial tests and complete re-tests is potentially quite small (especially if the fast-fail algorithm is in use).[4] Thus, the sample of vehicles with comparable emission test results before and after emission repair could be too

[3] Using a cost model of I/M described in McConnell and Harrington (1992).

[4] Arizona forces failing random-sample vehicles to take full-length re-tests, but the number of such vehicles is quite small.

small for meaningful statistical analysis. This is all the more true if it is further desired to subset the data by age, vehicle type, manufacturer, or emitter class.

Nonetheless, it is the two percent random sample that permits the analysis discussed in this paper. We use such a random sample of second-by-second test results collected in Arizona in 1996 to relate the complete-test results to emission levels from truncated tests. We use ordinary least squares (OLS) linear regression to estimate the link between partial- and full-test results, for each partial-test length (from 31 to 239 seconds) and for three pollutants (HC, CO, NOx). The method is designed to be a convenient way for state officials to estimate full tests from partial results, because the independent variables used in these regressions are routinely reported by emission-testing contractors. The matrix of regression coefficients for all regression equations is available on request from the authors.

II. THE ARIZONA IM240 PROGRAM AND EMISSION DATA

Arizona was the first state to implement the IM240 dynamometer test as part of its vehicle inspection program. Light duty cars and trucks from model years 1981 and later that are registered in the Phoenix metropolitan area must be tested every two years with the IM240 test in centralized stations. Figure 1 shows the test trace, i.e. the prescribed vehicle speed over the entire 240 second test cycle. Vehicles can fast-pass the test starting at second 31 (after the initial acceleration), and they can fast-fail any time after second 93 (once the first "warm-up" phase of the test is completed). To fast-pass, vehicles must have emissions below second-specific cutpoints for all three pollutants. A vehicle can be identified as failing for any of the pollutants after second 93, but cannot fast-fail the test until all three pollutants have either fast-passed or fast-failed. This means that the second at which a vehicle fails may not be clearly indicative of the vehicle's pollution levels. The decision rules for the fast-pass and fast-fail in Arizona are given in the Appendix.

Figure 2 shows the proportion of vehicles that have fast-passed or fast-failed either before or on each second of the test. This figure is based on data from all IM240 tests conducted in Arizona between January 1 and May 31, 1996. The graphs show that over half of all vehicles fast-pass at second 31. Of the vehicles that fast-fail, a cluster fail at second 94 (the first second at which a vehicle is eligible to fail), but significant numbers of additional failures do not occur at any particular second until after second 150. From the trace in Figure 1, we see that the majority of failures do not occur until the last major acceleration phase of the test.

III. ESTIMATION METHODOLOGY

To determine the relationship between full test emissions and truncated emissions for vehicles that either fast-pass or fast-fail we use regression analysis on data from the Arizona IM240 test program, using tests done January-June 1996 for the random sample of vehicles. There were 6,803 vehicles in this sample, and the analysis uses the entire second-by-second trace for every vehicle. Many approaches could be taken in developing this methodology. In general, the choices involve tradeoffs between accuracy (on the one hand) and simplicity and

Figure 1

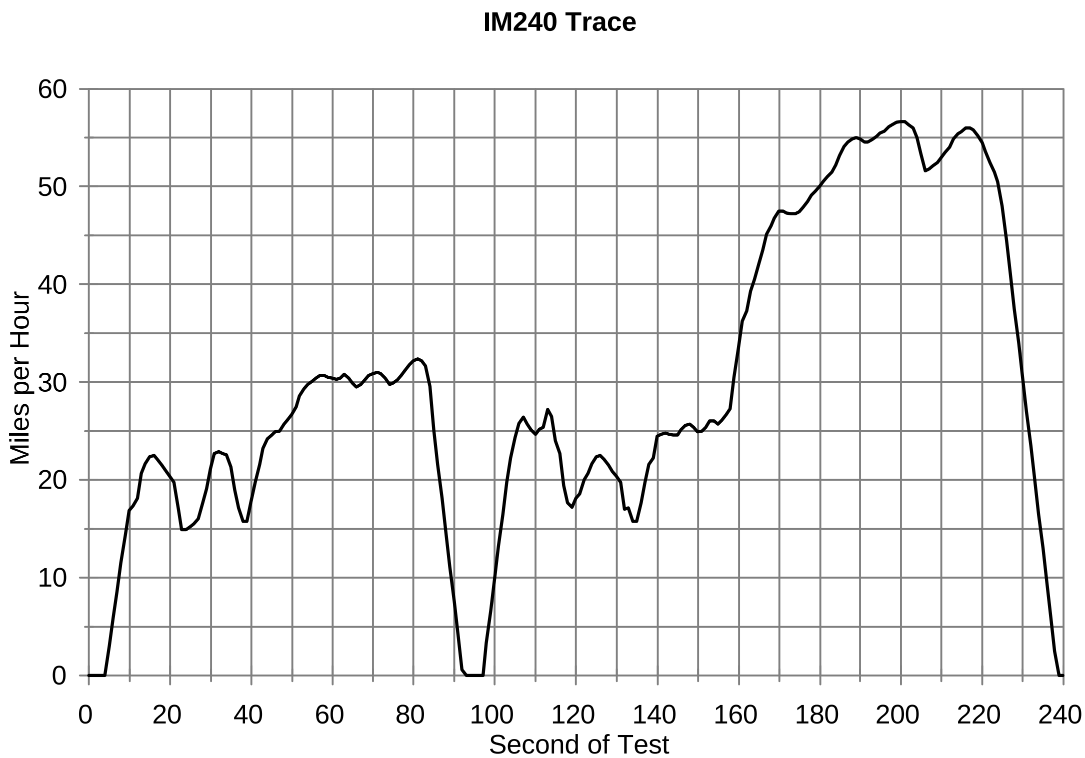
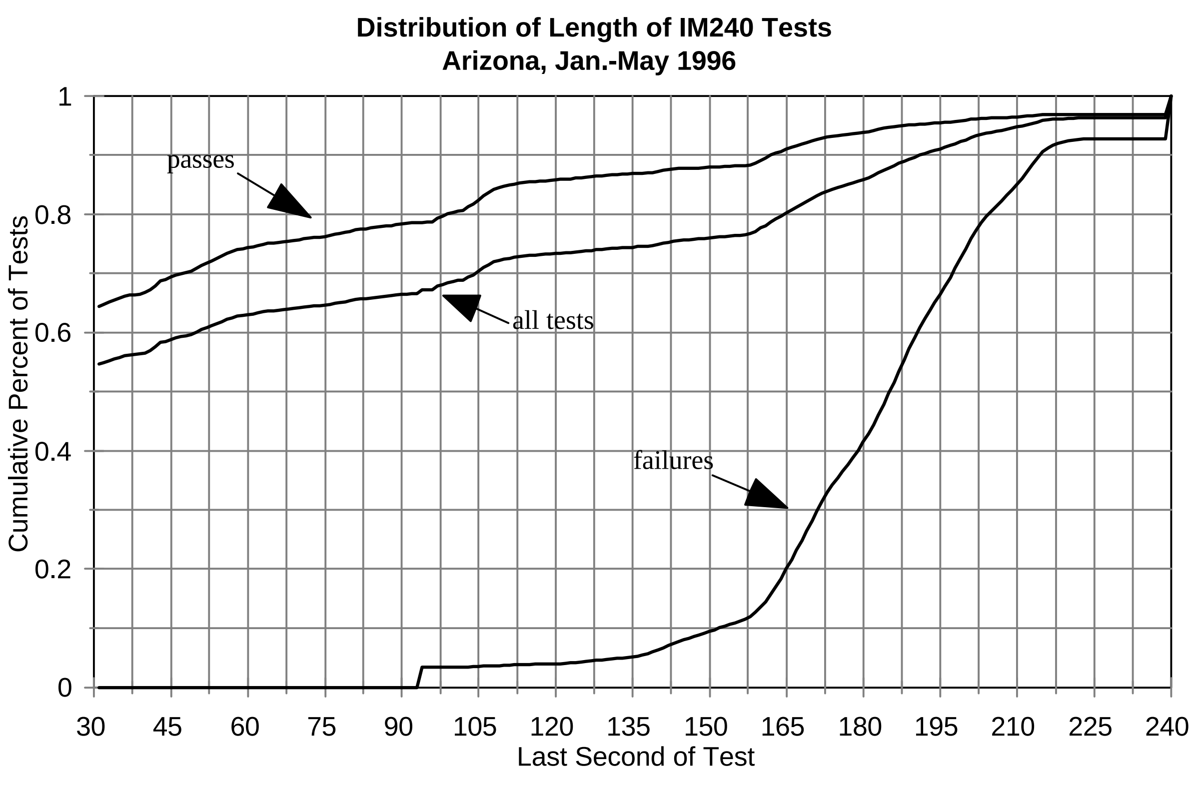


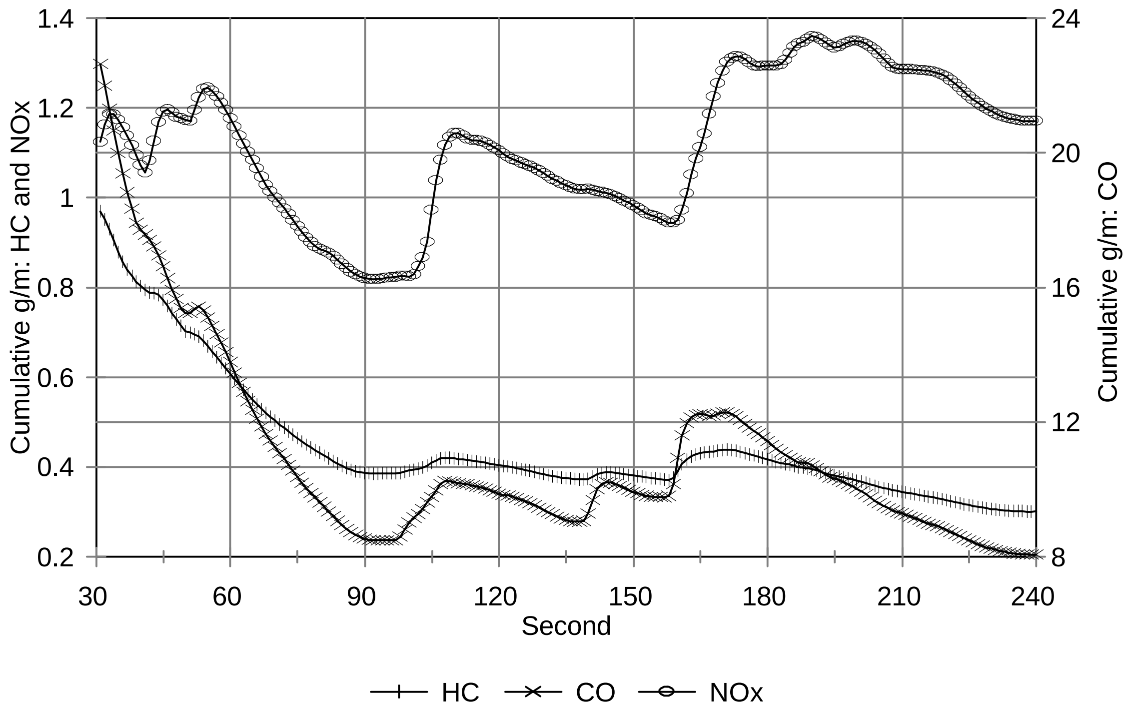
Figure 2



generalizability (on the other). We have designed our methodology to yield a good fit between predicted and actual full-test emissions, but not to be so linked to the Arizona program as to be inapplicable to other testing programs, and not to use data which may not be widely available.

Figure 3

Cumulative Grams per Mile: Sample Car



We estimate the relationships between partial and full-test emissions with separate equations for each of the three pollutants at each second at which a vehicle can fast-pass or fast-fail.[5] Other analysts have estimated similar relationships, but specify only one equation for each of a relatively small number of groups of seconds.[6] If there is much variation in emissions from one second to the next, however, the accuracy of this method's predictions of full-test emissions is inevitably challenged. Figure 3 shows the cumulative grams per mile of HC, CO, and NOx for one car in the Arizona program from seconds 31 to 240. For NOx and CO in particular, there are few groups of seconds in the test for which emission rates are relatively constant. Furthermore, although the patterns are often similar, emissions traces

[5] Although this produces many equations and coefficients, they are quite easy to transfer among analysts and to use for forecasting.

[6] An analysis of the Colorado fast-pass breaks the test into 10 second segments to reduce the number of equations for forecasting the full 240 second reading from short test readings (Colorado Department of Public Health and Environment, 1998). New York's analysis breaks the full 240 second trace into 12 different driving modes and uses those to predict the full IM240 reading (Shih and Whitby, 1996).

vary greatly among vehicles. Although grouping seconds does produce an apparently simpler procedure for converting fast-pass data to full-test estimates, that advantage is rendered almost meaningless by the easy availability of computers. Taken as a whole, this evidence leads us to conclude that estimating over groups of seconds could compromise the forecasting power of the methodology with no compensating advantage in simplicity.

Some analysts have suggested that different groups of vehicles should be used to estimate each equation. For example, the relationship between emissions at second 31 and second 240 might be estimated using only vehicles that would have passed at second 31, since the resulting coefficient is to be applied only to test results for vehicles that do pass at that point in the test. This approach probably would yield more accurate predictions when the resulting coefficients were applied to short-test results from the current Arizona program. However, this method has a serious drawback. The estimated coefficients would be an implicit function of the cutpoints and the fast-pass/fast-fail algorithm used in Arizona during the time period of the sample. Thus, they would be of questionable validity in estimating full-test results from truncated tests in other states or regions, or even in Arizona itself were the cutpoints or algorithm to be changed.

Having weighed these concerns, we use all of the vehicles in the sample to estimate each equation. However, we do explore the possibility that clean and dirty vehicles may have different relationships between short- and full-test results. In one version of our estimates, we allow the estimated relationships to vary with the emissions rate of the vehicle (see the discussion of the spline function below).

Choosing Explanatory Variables

Our approach, then, is to perform a linear regression, for each second and for each pollutant. With one exception to be discussed below, all specifications we examine can be nested in the following:

$$P_j^{i,240} = \alpha^{i,t} + \beta^{i,t} P_j^{i,t} + \delta^{i,t} Z_j + \gamma^{i,t} P_j^{i,t} Z_j + \varepsilon_j^{i,t} \qquad (t = 31,...,239) \tag{1}$$

where j indexes the particular vehicle, $P^{i,240}$ is the cumulative emissions in grams for pollutant i (i =HC, CO and NOx) at second 240, $P^{i,t}$ is cumulative emissions in grams for pollutant i at second t, and the Z variables are other factors that may have an impact on full-test emissions.[7] We explore a number of Z variables that might influence relationships between the partial and full IM240 test results. These variables include: model-year-group dummies (as proxies for technological variation), vehicle age, the type of vehicle (car or

[7] It is worth emphasizing that we estimate the relationship between partial- and full-test cumulative grams of a given pollutant. We do not perform the estimation in terms of grams per mile because our data do not contain a transparent measure of miles traveled at each second. If such data were available, our approach could easily be used to estimate grams per mile at second 240 as a function of grams per mile at second t.

truck), and the partial-test emission results for the other two pollutants (to exploit correlations among them).

The bar charts in Figures 4, 5 and 6 show the R^2 statistics[8] for the estimation of various equations for the three pollutants at a sampling of three different seconds: 31, 94 and 155. (Also see Table A1 in the Appendix for the exact R^2 statistics.) These illustrate the effects of the various Z variables on the explanatory power of the equation. The basic equation is from estimation of Equation (1) above with no *Z* variables. The other estimated equations include additional explanatory variables as labeled in the figures. The "interact" terminology in the labels means that the Z factors enter the equation multiplied by $P^{i,t}$ as well as separately. At all seconds of the test, the estimated equations have the highest R^2 for HC, and the lowest for NOx predictions.

Engine technology could affect relative emissions levels at various stages of acceleration in the test, and therefore the magnitude of the coefficients on emission readings at seconds less than 240. We try to capture the effects of engine technology with model-year groups. These groupings are unlikely to capture technology differences perfectly, but they are easily specified from available data. The model-year groups are:

Group 1: pre-1983
Group 2: 1983-1991
Group 3: 1992 +

Model-year groups are included first just as dummy variables (R^2 results shown as the second set of bars in Figures 4-6).[9] Then they are included both as separate variables and multiplied by the emissions reading at the second for which the equation is being estimated, $P^{i,t}$ (R^2 results shown in the third set of bars in Figures 4-6). Including model-year dummies adds to the explanatory power, particularly for the NOx equations, but allowing for interaction seems to have little impact for any second or for any pollutant.

A more precise specification of engine technology would include information about whether each vehicle is carbureted or has a fuel injection system. Unfortunately, this information was unavailable for all vehicles in the Arizona data set. We were able to decode the VIN to obtain such information for a small number of vehicles in our sample. However, including the carbureted/fuel injected variable for the resulting limited sample did not add more explanatory power compared to model year or age. Given that, and our limited information on the fuel systems of individual vehicles, we do not report the results here.

Including age by itself seems to add slightly more to the explanatory power of most equations compared to model-year groups alone. In fact, including age and age interacted with the *P* terms results in the largest R^2 for the pollutants at all seconds (except for the

---

[8] The R-squared is a common test statistic; for OLS regressions like those used here, it ranges from 0 to 1 and measures the fraction of the total variation in the dependent variable that is explained by the regression.

[9] Since the equation contains a constant, only dummies for the latter two groups are included; the first group is the implicit default.

Figure 4

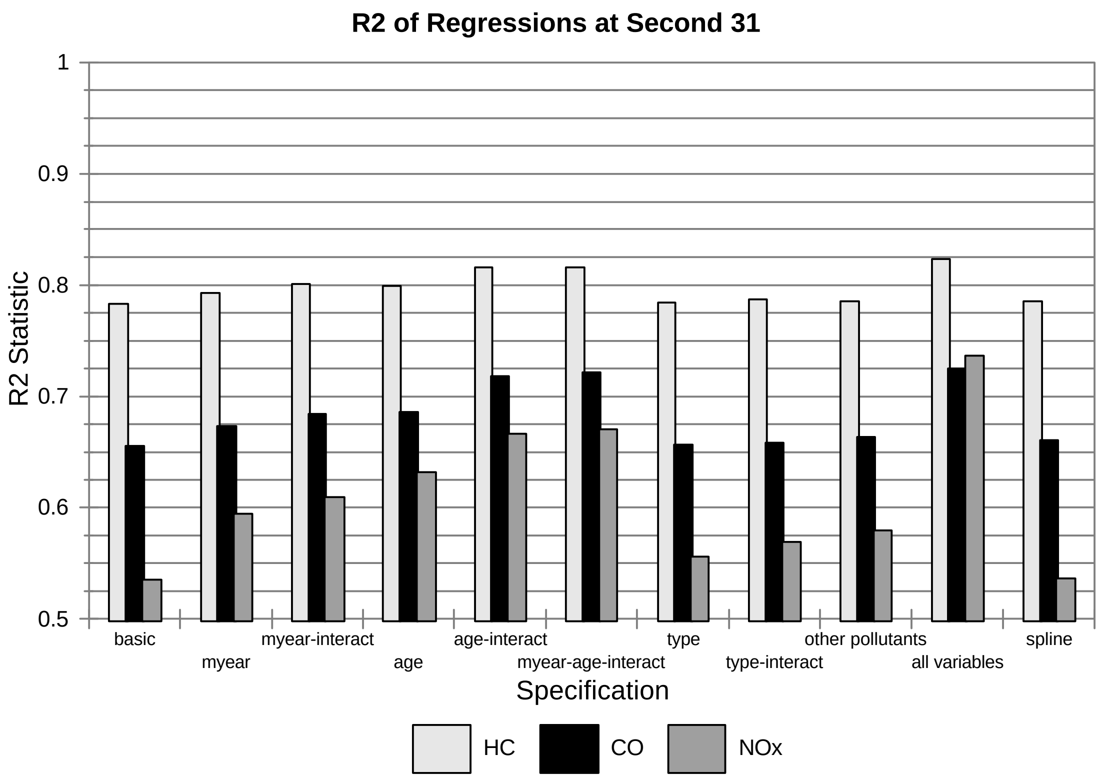


Figure 5

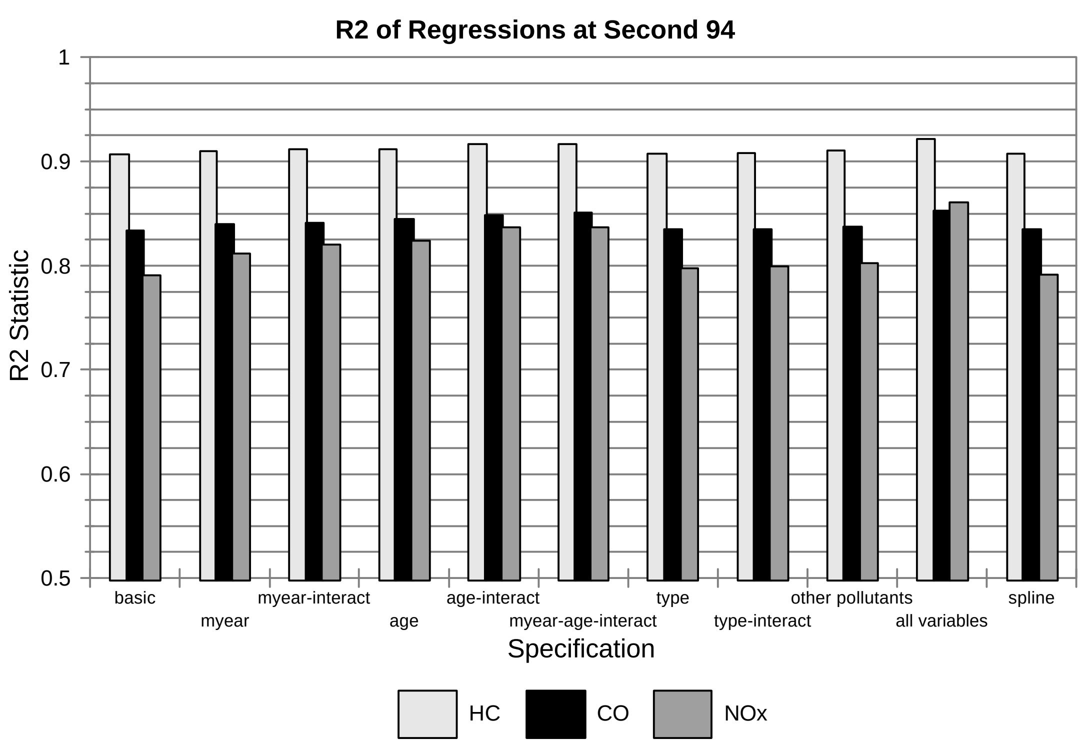
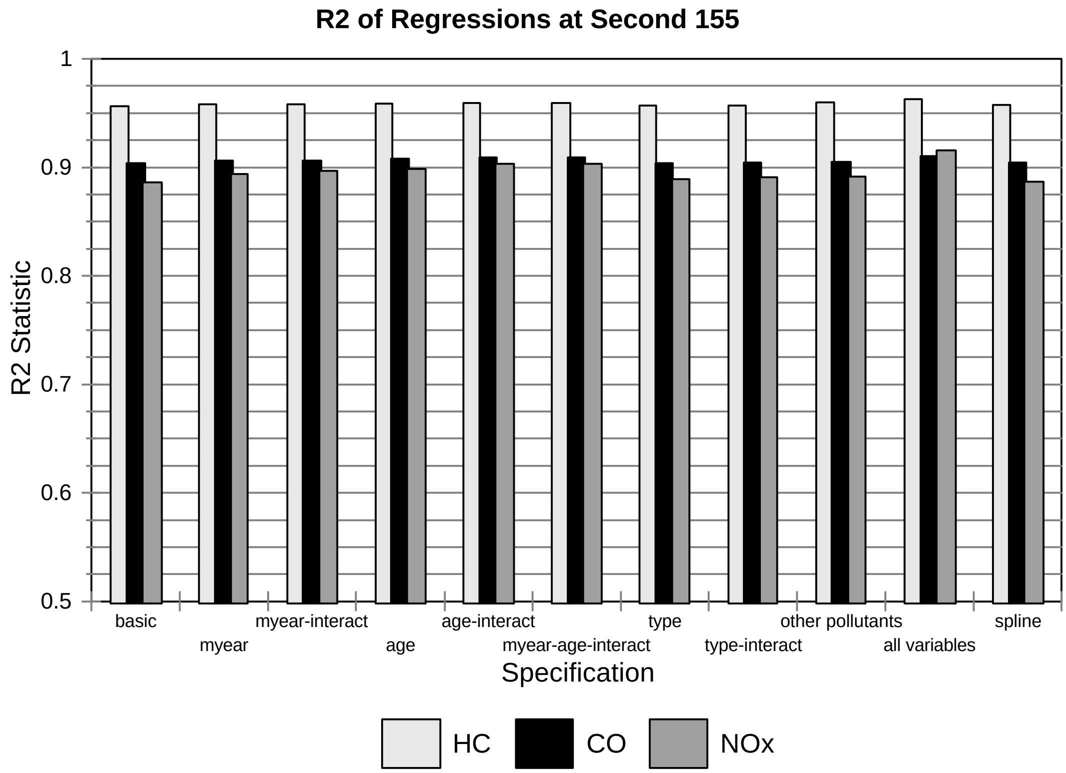


Figure 6



equation that includes all the variables, which has a cumbersome number of coefficients). Adding the model-year-group dummies and their interaction terms in addition to age does not seem to improve the fit (as shown in Figures 4-6). The equation that includes age and age interacted with partial-test result is simple and has high explanatory power. Therefore, we use this equation below to forecast the full test results.

The type of vehicle, i.e. whether it is a car, light truck, or medium-duty truck, seems to add little to the explanatory power of the equations, even when also interacted with own-pollution levels. The partial-test cumulative readings of the other two pollutants also add little to the model.

Finally, we examined an alternative functional form for equation (1) by allowing the *B* coefficients to vary with the emissions of the vehicle. It has been suggested that dirtier cars may have a different relationship between emission levels early in the test and the final 240 second reading. To test for this possibility, we estimated a four part spline function which allows both the constant term and the slope term to vary over the range of emissions for the second being estimated, but constrains the linear segments to intersect at the points where the variations are allowed to occur. This is essentially a piecewise linear function.[10]

Figures 4, 5 and 6 show that the spline specification adds little to the estimated results. The R^2 is not much larger than the basic model, and most of the estimated coefficients for the four-part spline were not significantly different from each other. This result supports our decision to estimate the equation for each second using the entire sample of vehicles.

Regression Analyses Using Chosen Model

We estimate the most promising of the models described above for each of the 209 seconds and each of the three pollutants, yielding 627 equations. The chosen model includes a constant and *t*-second pollution term. Age is the only *Z* variable, and is included alone and interacted with the *t*-second cumulative grams of pollution. The age variable is defined as age relative to the 1996 model year.

$$P_j^{i,240} = \alpha^{i,t} + \beta^{i,t} P_j^{i,t} + \delta^{i,t} Age_j + \gamma^{i,t} P_j^{i,t} Age_j + \varepsilon_j^{i,t} \quad (t = 31,\ldots,239) \tag{2}$$

Table 1 shows detailed regression results for each of the three pollutants at seconds 31, 94 and 155 of the test. Almost all terms are significantly different from zero at the 5 percent level. Especially in the early seconds, the overall fit of the HC equation is the best and the NOx equation is not nearly as good (R^2 of 0.67 for the NOx equation compared to 0.82 for the HC equation at second 31). As we would expect, the coefficients on cumulative pollution at second *t* fall toward one as *t* increases towards 240. The coefficient of age interacted with cumulative grams is positive, large and highly significant, especially for HC and NOx at early seconds in the test. Total 240-second emissions relative to cumulative emissions at 31 seconds are larger for older cars than for newer cars. This difference diminishes as the test goes closer to the full 240 seconds.

Figure 7 illustrates how the R^2 statistics from estimating Equation (2) vary between pollutants and as the test progresses through time; the statistics are superimposed on the MPH trace of the IM240 test. As we have shown above for three sample seconds, the predictive power of the regressions improves as data from later in the test are used to predict full-test emissions. The predictive power of the HC equations is better than either CO or NOx in the early seconds, and by second 185, the predictive power for all three pollutants is close to 100 percent. The figure shows that the fit improves most rapidly during or immediately after periods of acceleration, particularly for CO and NOx. There is a large improvement in the fit for all three pollutants between seconds 31 and 80, and by second 80, the R^2 is above .80 for CO and NOx and above .90 for HC.

---

[10] For more detail on the spline function, see Greene (1990).

Table 1: Regression Coefficients

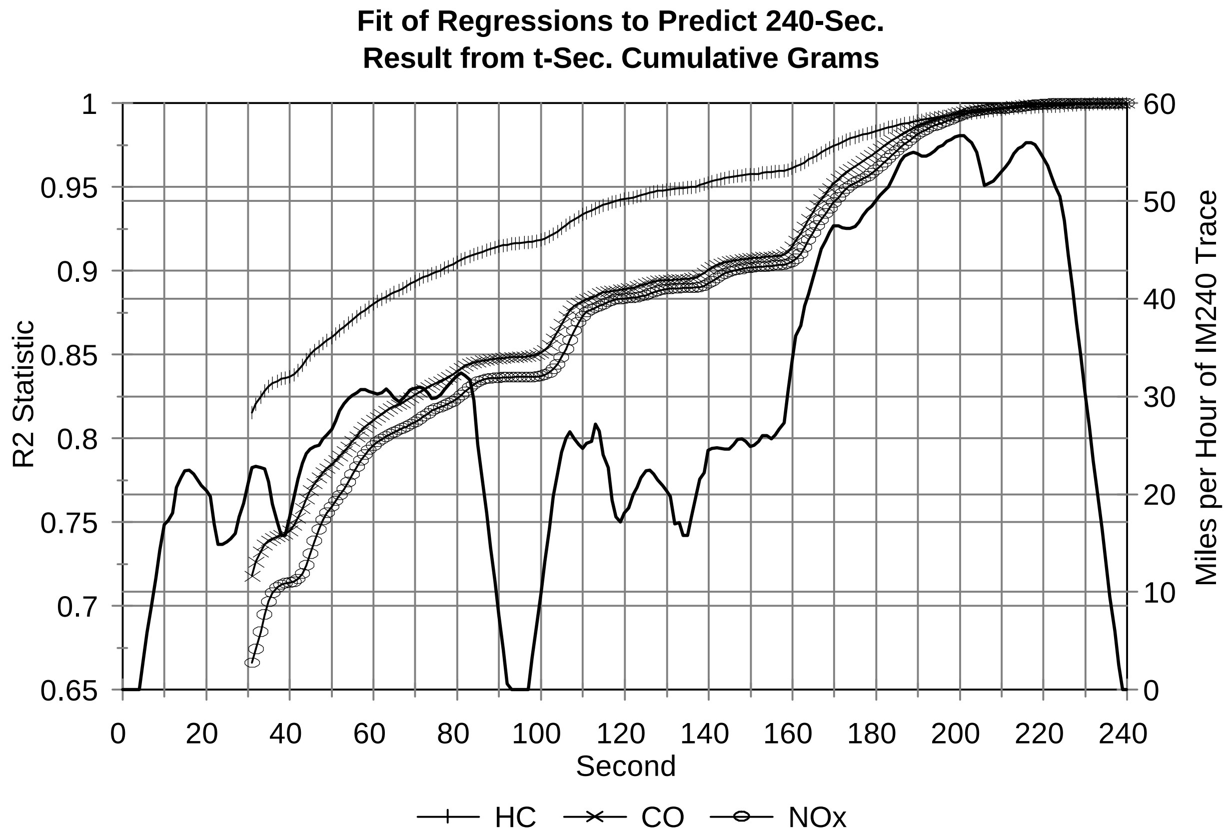
Pollutant	HC			CO			NOx		
Second	31	94	155	31	94	155	31	94	155
Cumulative grams (at second *t*)	3.56 (0.35)*	1.90 (0.089)*	1.41 (0.37)*	2.27 (0.64)*	2.04 (0.20)*	1.55 (0.08)*	3.44 (0.24)*	2.01 (0.06)*	1.62 (0.03)*
constant	-0.11 (0.041)*	-0.094 (0.023)*	-0.065 (0.015)*	0.57 (1.09)	-1.53 (0.70)*	-1.50 (0.42)*	0.45 (0.04)*	0.27 (0.03)*	0.14 (0.02)*
Age	0.030 (0.005)*	0.022 (0.0031)*	0.019 (0.0024)*	0.73 (0.12)*	0.80 (0.07)*	0.54 (0.05)*	0.08 (0.008)*	0.06 (0.006)*	0.037 (0.005)*
(Cumulative grams)*Age	0.27 (0.032)*	0.053 (0.008)*	0.015 (0.003)	0.40 (0.06)*	0.055 (0.017)*	0.017 (0.0067)*	0.55 (0.03)*	0.10 (0.008)*	0.040 (0.004)*
R^2	0.8157	0.9165	0.9591	0.7181	0.8486	0.9087	0.6662	0.8366	0.9031

Notes:

(1) Heteroskedasticity-robust standard errors are in parentheses

(2) * indicates significant at 5% level, ** indicates significant at 10% level.

Figure 7



It is important to note that including the age variable alone in the analysis is probably capturing at least two effects: vehicle aging, and technological differences (such as the shift to fuel injection) among vehicles of different vintages. In the analysis above we tried to include both age and model-year dummy variables in an attempt to differentiate between these two factors. However, we really do not have the ability to separate these two effects in this dataset. We only have six months of data, so we do not observe vehicles of the same model year at substantially different ages. As more data become available, it may be possible to disentangle the two effects. In the meantime, the precise coefficients presented here are should be used only for analyses of data from around 1996, since a five-year-old car is likely to have very different technology in 2001 than in 1996. Analysis of vehicle-emission tests from another time period should use coefficients from equations estimated on test data from that period.

IV. FLEET-AVERAGE EMISSION-RATE PREDICTIONS

The regression coefficients can be applied to data on vehicle age and partial test cumulative emission results to predict full-test cumulative emissions for a given vehicle. That prediction can easily be converted into a prediction of the test result in terms of grams per mile by dividing predicted grams by the average length of the IM240 test in miles.[11]

Policy makers and administrators are often interested in average fleet emissions, in order to track changes over time and to evaluate any impact that policy may have had on those averages. Here, we try to evaluate the accuracy of average emission predictions made by applying our technique to a sample of vehicles that took the IM240 under a fast-pass/fast-fail regime.

First, we take our 6,803 random-sample vehicles, all of which went the full 240 seconds, and subject them to a hypothetical test with the fast-pass/fast-fail algorithm turned on. In particular, we estimate when those vehicles would have passed or failed the IM240 test in Arizona using decision rules and cutpoints much like those used in the Arizona program (see Appendix), and note what the cumulative emissions of HC, CO and NOx would have been for each vehicle at the end of its test. Then we use the coefficients estimated by applying Equation (2) to the same sample of 6,803 vehicles, and calculate the predicted full-test emissions for each vehicle.

Figures 8 and 9 show how the average predicted cumulative emissions in grams compare to the actual average emission levels, when averages are taken first by model-year and then by vehicle type. Looking first at Figure 8, we find that the predicted averages are very close to the actual averages. The largest errors occur in pre-1986 NOx estimates; however, even those differences are small in absolute terms. Of the three vehicle types shown in Figure 9, cars dominate the sample (comprising 65 percent of all the tests). Thus, it is especially good news that our predictions of average emission levels are extremely accurate for that group. The predictions are a little further off for light trucks, and further still for medium-duty trucks. Our model performs best when predicting average emissions for cars, the most common type of vehicle in the sample; conversely, errors are highest for the least numerous vehicle type. If one were truly interested in accurate predictions of emissions from medium-duty trucks, it would be worth including dummies for vehicle type as explanatory variables in the regressions and interacting them with partial-test emission levels. However, precisely because cars so dominate the fleet, the overall accuracy of the methodology is little compromised by not controlling for vehicle type.

V. TEST LENGTH AND EASE OF REPAIR

Some enhanced I/M programs have touted the potential usefulness of the 240-second trace as a diagnostic tool for repairing vehicles that fail the test. Owners of Arizona vehicles that trigger a fast-fail have relatively short traces available for diagnostic use. At the same time, owners of similar vehicles may have full-length traces available if those vehicles happen to fall into the random sample. We exploit that random variation in test length in an analysis of the diagnostic usefulness of the IM240 trace.

[11] In our sample, average miles-traveled for the 240-second test was 1.96.

Figure 8

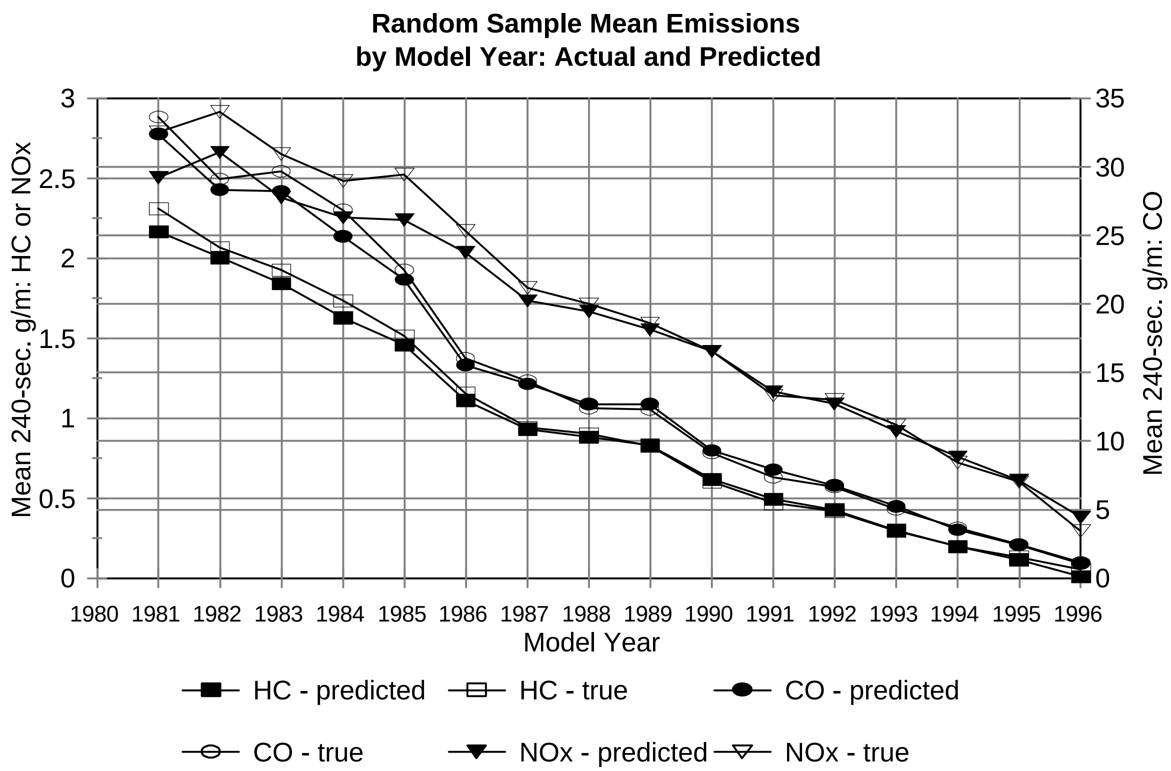
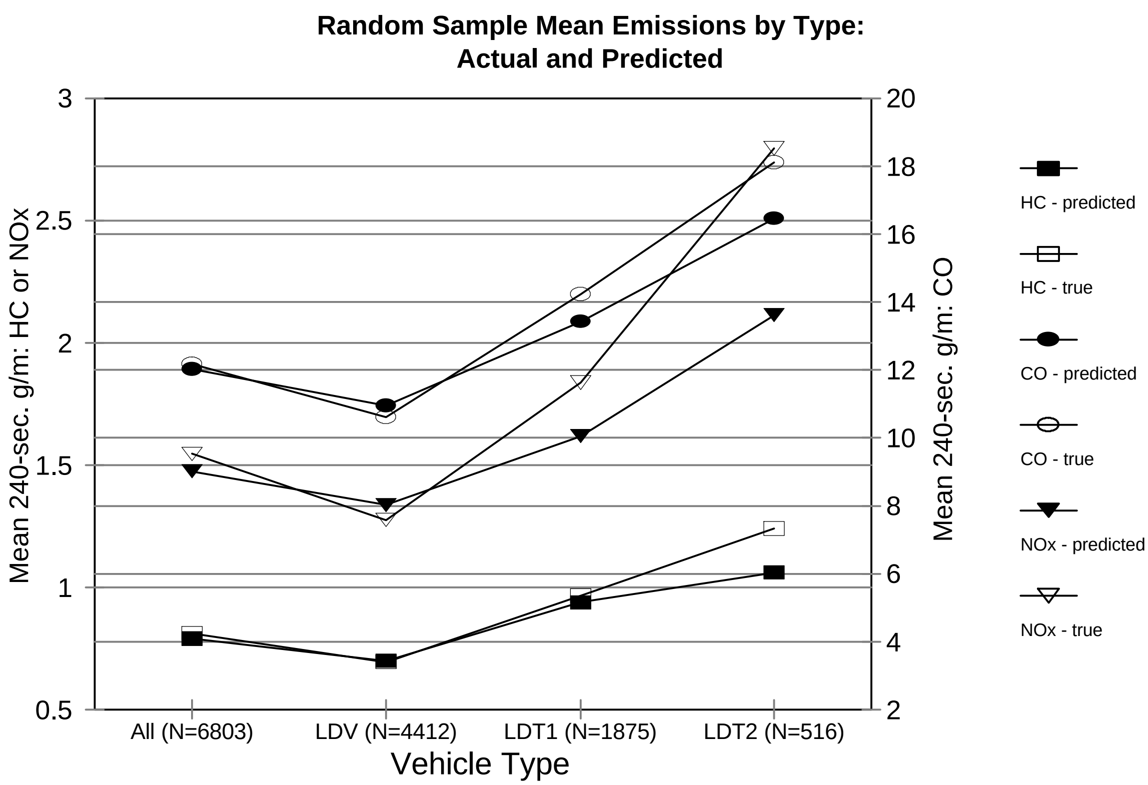


Figure 9



We construct a sample for the analysis beginning with all vehicles in Arizona that failed their initial tests between January 1996 and May 1996. We then keep only those vehicles for which at least one re-test was recorded before the end of May 1996 (this yields a sample size of 31,163). The dependent variable in the analysis is a dummy variable equal to 1 if a vehicle passed its first re-test, and 0 if it did not. This serves as a simple indicator of the ease with which a vehicle is able to be diagnosed and repaired; about 63 percent of the sample had a successful first re-test.

We perform a simple probit analysis[12] of a variety of factors that may influence the probability that a given vehicle passes its first re-test. The first independent variable is the second at which the vehicle's initial test ended. If the trace is a useful diagnostic tool, then longer tests should yield better diagnostic information and increase the probability of a successful re-test. Second, we include variables that capture, for each of HC, CO, and NOx, the difference between the vehicle's estimated initial pollutant levels in grams per mile[13] and the emission standards to which that vehicle is held. It may be more difficult to repair a vehicle successfully with emission levels that greatly exceed the standards. Third, we include the age of the vehicle, in case (for example) older vehicles, which may have had time to accumulate numerous unrepaired malfunctions, are trickier to diagnose. Fourth, we control for variations in the ease of repair by vehicle type (dummies for light- and medium-duty trucks are included; vehicles are the excluded base category). Fifth, we include a dummy for whether the vehicle failed only a single pollutant.

Interestingly, a full 24 percent of the vehicles in the main sample do not have recorded failures for HC, CO, or NOx even though they failed the overall test. Presumably, these vehicles were picked out for tampering. Since diagnosis is less of an issue when a vehicle fails only due to tampering, we include a dummy variable that captures that status. We also perform the analysis using a sub-sample which excludes vehicles that passed all three pollutants.

Table 2 contains the probit results for the full sample. Increased test length does appear to improve the owner's chances of successfully repairing his vehicle in the first round. This supports the hypothesis that the IM240 trace is a valuable diagnostic tool in emission-related repair. Many of the other variables in the analysis are significant as well. Vehicles with emission rates that far exceed the standards they must meet are harder to repair; that effect is most pronounced for HC, and more for NOx than for CO. Old cars seem more difficult to diagnose and repair in a single round. This may be because of differences in the technologies found in vehicles of different vintages, or because old cars tend to be plagued with a number of problems, making the culprit responsible for high emission levels difficult to

---

[12] Probit analysis is a commonly-used statistical method for estimating regression models where the dependent variable can only take on two values – in our case, pass or failure of the re-test. The predicted value can be interpreted as the probability that the re-test is passed given the values of the independent variables. For more information see, for example, Greene (1990).

[13] All tests have emissions in grams reported at the last second of the test. For partial tests, full-test emissions are estimated using the estimation method outlined in Section III. Then g/m are derived for all tests by dividing total cumulative grams by 1.96 miles.

identify. Light- (but not medium-) duty trucks seem more difficult to repair. Finally, vehicles that failed the test only because of tampering have a much higher rate of first-round success. The average probability that a vehicle passes its first re-test is fully 0.37 greater if its only problem is tampering.

Table 2: Probit Results of Whether Failed Vehicle Passed First Re-Test (N=31163)

Variable	Coef.	S.E.	Sig?	dP/dx
Second initial test ended	0.0018	0.00026	*	0.00067
Estimated HC g/m - HC standard g/m	-0.17	0.0075	*	-0.060
Estimated CO g/m - CO standard g/m	-0.0015	0.00047	*	-0.00053
Estimated NOx g/m - NOx standard g/m	-0.057	0.0065	*	-0.021
Age of vehicle	-0.07	0.0030	*	-0.026
LDT1 (light-duty truck)	-0.19	0.033	*	-0.070
LDT2 (medium-duty truck)	-0.06	0.041		-0.024
Failed only one pollutant in initial test	0.24	0.024	*	0.087
Failed no pollutants in initial test	1.29	0.046	*	0.37
constant	85.05	5.54	*	

Table 3 shows the results of the same analysis performed on a sub-sample of vehicles that includes only those that failed at least one of the three pollutants. The findings are qualitatively similar, but the coefficients tend to be larger, and thus the estimated effects of variables on the probability of successful first-round repair are often larger. Recall that this probit analysis was designed primarily to capture the importance of the trace to diagnosing vehicles with high emission levels. Hence, we use the results of the sub-sample analysis in the calculations of Table 4. Those calculations provide a sense of how much the fast-fail part of the test algorithm reduces owners' chances of being able to successfully repair their failing vehicles in just a single round of repair.

Second 94 is the first at which a vehicle is allowed to fail. Only 3.5 percent of all failures occur then, but those unlucky car owners would have had their probability of success in the first round of repair raised by 0.26 if the tests had been allowed to run the full 240 seconds. In general, the loss of diagnostic ability gets smaller as the fast-fail occurs later in the test. However, even by second 165, at which point 20 percent of all failures have occurred, the reduction in success probability associated with the fast-fail is still .14.

Each round of repair entail substantial transaction costs for a vehicle's owner. For many vehicles, the fast-fail algorithm does seem to substantially increase the chances that the owner will have to engage in more than one round of time-consuming, inconvenient emission repair. States must balance that cost against the benefits of fast-fail that come from faster-moving queues at the testing stations.

Table 3: Probit Results of Whether Failed Vehicle Passed First Re-Test
(Exclude Vehicles that Failed No Pollutants, N=23772)

Variable	Coef.	S.E.	Sig?	dP/dx
Second initial test ended	0.0045	0.00032	*	0.0018
Estimated HC g/m - HC standard g/m	-0.18	0.0083	*	-0.070
Estimated CO g/m - CO standard g/m	-0.0020	0.00050	*	-0.00078
Estimated NOx g/m - NOx standard g/m	-0.074	0.0071	*	-0.030
Age of vehicle	-0.064	0.0032	*	-0.025
LDT1 (light-duty truck)	-0.20	0.034	*	-0.080
LDT2 (medium-duty truck)	-0.062	0.043		-0.025
Failed only one pollutant in initial test	0.21	0.025	*	0.084
constant	120.74	6.05	*	

Note: dP/dx gives the marginal effect of each variable on the probability a vehicle passes its first re-test, given other variables at their mean values. This is evaluated at the mean of the variable (if continuous). For a dummy variable, this compares the probability given dummy=1 to the probability given dummy=0.

Note: * indicates statistically significant at the 5% level.

Table 4: Quantifying Impact of Fast Fail on Probability that First Re-Test is Successful

Second test ends	Fraction of all failures occurring exactly then	Cumulative fraction of failures occurring by then	Increased probability if wait until second 240
94	0.035	0.035	0.26
165	0.017	0.20	0.14
213	0.012	0.88	0.049

Note: Uses coefficient from sub-sample analysis in Table 3.

Note: The numbers in the second and third columns are from the distribution shown in Figure 2.

VI. CONCLUSIONS AND RECOMMENDATIONS

Truncated emission tests -- "fast-pass" and "fast-fail" tests -- are nearly universal in states using or planning to use the IM240 emission test in their Enhanced I/M programs. Such tests appear to do a good job of distinguishing vehicles that do and do not comply with applicable emission standards, although to our knowledge there has been no conclusive test of this assertion. Also, we are not aware that the savings in time and equipment that serve as the main justification for such tests have been quantified using empirical data, at least in the open literature. We believe that empirical investigations of the benefits in time and equipment savings as well as the costs of errors in classification that result from use of truncated tests should therefore be completed as soon as possible.

In this paper we have examined two other potential disadvantages of the use of truncated tests: (i) whether the use of a fast-fail procedure interferes with vehicle repair and (ii) how to extrapolate the truncated test results to estimates of full IM240 test results, for use in making quantitative estimates of vehicle emissions.

In order to examine the determinants of success in passing the re-test, we used a probit model applied to all initial emission-test failures during the first five months of 1996. We found that after controlling for the age, type of vehicle, initial pollutant levels and other variables, the duration of the initial test had a positive and significant effect on the probability of a successful re-test. Moreover, the result was not merely significant in a statistical sense; it also suggested a quantitatively important effect. We estimate that had a vehicle fast-failing at second 94 been allowed to complete the test, the probability of a successful first-round repair would have been increased by an increment of 0.26. Considering the high cost to motorists of repeat vehicle repair and emission tests, these results call into question the continued use of the fast-fail algorithm in Arizona.

Notwithstanding this interesting result, the principal achievement of the paper has been the development of a method of extrapolating the fast-pass and fast-fail test results to completed IM240 test estimates. Our method applies OLS regression to second-by-second emission data on a random sample of vehicles required to have a full IM240 test.

Unlike some other methods that have been proposed, the method described here is independent of the emission standards in use and the fast-pass and fast-fail algorithms employed. The predictive ability of the methodology is improved somewhat by the inclusion of vehicle age, but since we only had test data from six-month period, we could not distinguish the effects of age from those of model year. It may therefore be useful to repeat this analysis in a couple of years to see if separate age and model-year effects can be identified and if so, whether those effects are important. Caution should be used in using the coefficients generated here in extrapolating partial- to full-test results in future years.

We find that the correspondence between predicted and actual full test results is fairly good for HC even for tests that are truncated at 31 seconds $(R^2 = 0.80)$, though somewhat less so for CO $(R^2 = 0.68)$ and NOx $(R^2 = 0.61)$. For longer tests, the agreement between predicted and actual is even better, so that, for example, for tests truncated at 94 seconds the R^2 exceeds 0.8 for all three pollutants. In view of this improvement in statistical performance, it would be very useful to know the costs, in terms of waiting time and additional equipment requirements, of raising the minimum test length.

We investigated how well this procedure predicts emissions and conclude tentatively that it predicts average fleet emission levels quite well, both for all vehicles and within model years. We found that the errors in predicted average emissions were very small for cars and somewhat larger for trucks. As the fraction of trucks in the fleet increases, it may be worthwhile to account for vehicle type more fully in the estimation.

APPENDIX

Decision Rule Used for Fast-pass:

See EPA Guidance Document rules (USEPA, 1996).

The rules are:

The vehicle passes if, for any second between seconds **31 and 93**:
$E_{HC} < S_{HC}$ for the composite emissions, <u>and</u>
$E_{CO} < S_{CO}$ for the composite emissions, <u>and</u>
$E_{NOX} < S_{NOX}$ for the composite emissions.

The vehicle passes if, for any second between seconds **94 and 239**:
$E_{HC} < S_{HC}$ for the composite emissions or $E_{HC} < S_{HC}$ for the Phase II emissions, <u>and</u>
$E_{CO} < S_{CO}$ for the composite emissions or $E_{CO} < S_{CO}$ for the Phase II emissions, <u>and</u>
$E_{NOX} < S_{NOX}$ for the composite emissions or $E_{NOX} < S_{HC}$ for the Phase II emissions (there are no Phase II standards for NOx so we did not include a Phase II check for NOx).

Where,
E_{HC} , E_{CO} , E_{NOX} are cumulative emissions for each pollutant up to any second,
S_{HC} , S_{CO} , S_{NOX} are the standards for each second (from EPA Guidance document (1996)). Composite emissions are cumulative emissions over the entire test up to any given second. Phase II emissions are cumulative emissions over the second phase of the test which starts at second 94.

Decision Rules for Fast-fail:

See Gordon/Darby memo titled "Arizona Fast-pass and Fast-fail Guidelines."

The vehicle can only fast-fail after second 93. The decision rules are:

The vehicle fails if, for any second between seconds **94 and 239**:
Fast-fails HC: $E_{HC} > S_{HC}$ for the composite emissions and $E_{HC} > S_{HC}$ for the Phase II emissions,
Fast-fails CO: $E_{CO} > S_{CO}$ for the composite emissions and $E_{CO} > S_{CO}$ for the Phase II emissions,
Fast-fails NOx: $E_{NOX} > S_{NOX}$ for the composite emissions and $E_{NOX} > S_{HC}$ for the Phase II emissions

However, the vehicle does not get a fast-fail until all three of the pollutants have been found to pass the fast-pass or fast-fail criteria.

Note: Although we have been told that these are the decision rules being used in Arizona, we have not been entirely successful in using these to correctly predict when vehicles actually do pass or fail. For example, we were able to predict the correct second at which the test ended for about 80% of both the fast-passes and fast-fails for 10,000 vehicles test in January 1996. Gordon-Darby, the contractor for Arizona, was unable to help us resolve this discrepancy.

Table A1. R^2 Statistics for Regression of Cumulative Pollutant Grams at Second 240 on Cumulative Grams at Second *t* and Other Variables

Variables included **in addition to** "own grams" and "constant"	HC at second:			CO at second:			NOx at second:		
	31	94	155	31	94	155	31	94	155
none	0.783	0.907	0.956	0.656	0.834	0.904	0.535	0.791	0.886
model-year group dummies	0.793	0.910	0.958	0.673	0.840	0.906	0.595	0.812	0.894
model-year-group dummies interacted with own grams	0.801	0.912	0.958	0.684	0.841	0.906	0.610	0.820	0.897
age	0.799	0.912	0.958	0.686	0.845	0.908	0.632	0.824	0.899
age, and age interacted with own grams	0.816	0.917	0.959	0.718	0.849	0.909	0.666	0.837	0.903
model-year-group dummies, age, and dummies and age interacted with own grams	0.816	0.917	0.959	0.722	0.851	0.909	0.671	0.837	0.903
vehicle-type dummies (LDT1 and LDT2 vs. LDV)	0.784	0.908	0.957	0.657	0.835	0.904	0.556	0.798	0.889
vehicle-type dummies interacted with own grams	0.787	0.908	0.957	0.658	0.835	0.904	0.569	0.799	0.890
other pollutants' cumulative grams	0.785	0.911	0.960	0.663	0.837	0.905	0.580	0.802	0.891
all variables	0.824	0.921	0.963	0.725	0.853	0.910	0.736	0.861	0.915
components to specify own grams in a four-part spline	0.785	0.908	0.957	0.661	0.835	0.904	0.537	0.791	0.887

References

Colorado Department of Public Health and Environment. 1998. "The Colorado Enhanced I/M Program 0.5% Sample Annual Report," Remote Sensing Technologies, Inc., Tucson, Arizona, January 27.

Greene, William H. 1990. *Econometric Analysis* (New York: MacMillan Publishers, Inc.).

McConnell, Virginia, and Winston Harrington. 1992. "Cost-Effectiveness of Enhanced Motor Vehicle Inspection and Maintenance Programs," Discussion Paper QE92-18-REV, Resources for the Future, Washington, D.C., September.

Shih, Celia, and Robert Whitby. 1996. "A Predictive Modal Regression Methodology for IM240 Fast-Pass/Fail Decisions," Task 1, New York State Department of Environmental Conservation, Division of Air Resources, Bureau of Mobile Sources.

U.S. Environmental Protection Agency (USEPA). 1996. "7/96 Draft Revised High Tech I/M Guidance," Office Of Mobile Sources, U.S. EPA, Ann Arbor.
Available at: http://www.epa.gov/OMSWWW/im.htm#guidance