Technology and the Effectiveness of Regulatory Programs Over Time: Vehicle Emissions and Smog Checks with a Changing Fleet

Nicholas J. Sanders and Ryan Sandler*

December 14, 2018

Abstract

Personal automobile emissions are a major source of urban air pollution. Many regional and local governments across the United States (US), Europe, and Asia have inspection and maintenance (I/M) programs designed to reduce pollution from personal transportation. But there is little empirical evidence directly linking mandated inspections, maintenance, and local air pollution levels. To test for a link, we estimate the contemporaneous effect of inspections on local air quality in the US state of California. We use day-to-day, within-county variation in the number of vehicles repaired and re-certified after failing an initial emissions inspection, with individual-level data from 1998–2012 from California's inspection program. Additional re-inspections of older vehicles with inferior emissions technology (pre-1985 model year) reduce local carbon monoxide, nitrogen oxide, and particulate matter levels, while re-inspections of newer vehicles with more modern engine technology have no economically significant effect on air pollution. This suggests emissions inspections have become less effective at reducing local air pollution as more high-polluting vehicles from the 1970s and 1980s leave the road, and provides an example of how the social efficiency of programs can change under improving technologies. We also estimate the importance of station quality, using a metric devised for California's new STAR certification program. We show re-inspections of older vehicles conducted by low quality inspection stations do not change air pollution, while inspections at high quality stations have a moderate effect on pollution concentrations, which suggests the potential for ineffective monitoring at low quality inspection stations. We find little effect on ambient ozone levels, regardless of station quality or vehicle age.

*Sanders: Cornell University and NBER, njsanders@cornell.edu. Sandler: Bureau of Consumer Financial Protection, ryan.sandler@cfpb.gov. We are grateful to Dan Hosken, Mark Jacobsen, Thomas Koch, Devesh Raval, Joseph Shapiro and participants in the Economics Brownbag at the Federal Trade Commission for helpful comments and suggestions. Any opinions in this paper are those of the authors and do not necessarily reflect those of the Bureau of Consumer Financial Protection or the United States.

1 Introduction

Automobile pollution has substantial impacts on health, and regulating ambient air pollution from automobile traffic is a public concern throughout the world.[1] Despite regulatory advancements and improvements in engine technology, motor vehicles remain responsible for 75% of carbon monoxide (CO) emissions in the United States, and over 50% of nitrogen oxide (NO_x) emissions.[2] Governments in both developed and developing countries have tried a number of policies to reduce pollution from personal automobiles. Improving fuel standards can decrease emissions per mile driven, but such programs disproportionally impact low-income households and decrease average road safety (Jacobsen, 2013a,b). Driving restriction programs have varied success rates when it comes to actually reducing local pollution (Davis, 2008; Wolff, 2014; Simeonova et al., 2018). Scrappage programs, often referred to as "Cash for Clunkers," can directly remove the dirtiest vehicles from the road, but recent work shows that such programs have substantial problems with adverse selection and may only slightly shift forward the timing of vehicle replacement (Sandler, 2012; Mian and Sufi, 2012; Li et al., 2013; Hoekstra et al., 2014). Inspection and maintenance programs (I/M), the focus of this paper, attempt to limit tailpipe emissions through regular inspections and repairs, without changing driving behavior or fleet composition. Such programs are costly both to governments and individuals (Ando et al., 2000), are subject to potential fraud (Oliva, 2015), and although mandated repairs of high-emitting vehicles often show reduced tailpipe emissions in a testing environment, there exists no large-scale analysis of how I/M programs affect local air pollution. Understanding the effectiveness of I/M in practice is especially important in light of the recent emissions test cheating scandal involving Volkswagen diesel vehicles in U.S. and Europe, where vehicles recorded as passing EPA tests actually produced emissions far above allowable levels on the road.[3]

We provide the first causal analysis of vehicle inspections and local air pollution, using extensive Smog Check data from the state of California.[4] We find an additional inspection-driven repair of a faulty vehicle does reduce contemporaneous local air pollution. However, we only find economically meaningful results from inspections of older (1985 and prior model-year) vehicles that predate advances in emissions control technologies, suggesting

[1]See Currie and Walker (2011) and Knittel et al. (2016).

[2]Emissions data from `http://www.epa.gov/airquality/peg_caa/carstrucks.html` (accessed online June 1, 2015). For discussion on automobile NO_x regulation, see Fowlie et al. (2012).

[3]See `http://www.nytimes.com/2015/09/23/business/international/volkswagen-diesel-car-scandal.html?_r=0` (Accessed September 23, 2015).

[4](Harrington et al., 2000) calculate cost-effectiveness of a similar inspection program in Arizona, but do not link inspections to ambient air pollution levels.

the benefits of I/M programs decline as engine technology improves. Further, we examine a recent reform to California's I/M program that incorporates measures of inspection station quality. We find increasing station quality may help further reduce some pollutants, but again only through inspections and repairs of faulty older cars that are a shrinking share of the automobiles on the road.

To identify the effectiveness of inspection programs, we leverage the fact that although the implementation of California's overall inspection program is endogenous to air pollution, the timing of individual vehicle repairs—the mechanism through which inspections should affect pollution—is random and exogenous. We use counts of final re-inspections following a failed inspection to capture the intensity of I/M-related vehicle repairs. Controlling for local weather effects and a battery of time and region fixed effects, we find repairing cars that failed initial inspections reduces local CO and NO_x levels in a statistically and economically significant fashion in the period following the repairs. Repairing and re-inspecting 1,000 older emissions technology vehicles decreases ambient CO by 26 parts per billion and ambient NO_x by 1.9 parts per billion, about 7% of a standard deviation for each pollutant. For scale, the average California county re-inspects 1,000 failing vehicles of all ages every 12 days. Re-inspections of vehicles manufactured after 1985 have much smaller effects on air pollution. This presents a case where the social efficiency of a program changes as the relevant technology advances — potentially regulation-driven improvements in engine technology are making smog check programs, as currently designed, less socially efficient with time.

California recently passed substantial reforms of inspection station requirements, hoping to improve inspection reliability (Bureau of Automotive Repair, 2014). Under the new "STAR" system, inspection and repair providers must pass certain quality criteria before the state certifies them to inspect the most high-polluting vehicles.[5] Understanding how quality control of testing systems impacts air quality improvements is important given prior findings that I/M programs are subject to gaming (Oliva, 2015). However, testing the effectiveness of quality control programs is challenging due to confounding factors including strategic customer and station responses to the rating system. To avoid such problems, we use historic data to construct hypothetical STAR program measures of station quality before the announcement of the policy, and test the relationship between ambient air pollution and re-inspections at what would be considered high-scoring stations under the eventual STAR metrics. We find re-inspections of older vehicles at

[5] The STAR program also requires that newer vehicles with onboard monitoring computers be tested by computer, rather than by direct tailpipe measurements. In addition, new regulations provide for heavier penalties for stations that are found cheating, as well as for consumers who try to falsify an inspection.

high quality stations reduce airborne levels of CO and NO_x while re-inspections at low quality stations yield no change in local air pollution levels. This result is consistent with the theory that low quality stations allow vehicles to pass re-inspection without appropriate repairs. Much like our findings on the general effectiveness of I/M programs, we find re-inspections of newer cars have little impact on air pollution, regardless of station quality.

Finally, we use our empirical results to conduct two policy simulations. We first simulate the eventual effectiveness of the STAR station quality control program, and show that benefits of the more strict inspection system are likely to fade in the future. We also simulate the impact of removing the Smog Check program entirely. We find that the benefits of Smog Check fell rapidly between 2002 and 2009 as older vehicles left the road, while costs remained relatively constant, being largely a function of the number of inspections per year. Although the program would still pass a simple cost-benefit test in 2009, the trend suggests this will not continue indefinitely.

Beyond our focus on vehicle emissions, this paper relates to the broader literature on the quality of regulatory enforcement. Inconsistent enforcement of regulations can substantially hinder the effectiveness of regulatory programs. In a review of empirical studies on the productivity of environmental monitoring, Gray and Shimshack (2011) find that regular monitoring and enforcement of regulated facilities can reduce violations both through improving regulated areas and deterring future violations in areas that are not directly targeted. But if enforcement is lax, the regulator may appear "toothless," reducing the impact of regulation overall. Shimshack and Ward (2005) show that a regulator having a strong reputation has large positive spillovers, and similarly, weak regulators may have large negative spillovers, undermining compliance overall. Duflo et al. (2013) find honest reporting and actual emissions from factories improve when auditors are assigned at random, rather than being chosen by the factories and subject to potential conflicts of interest. Muehlenbachs et al. (2016) show that, in the context of safety inspections on oil rigs, greater enforcement (as proxied by a greater number of inspectors) improves inspection outcomes and safety. But in the context of I/M programs, Oliva (2015) shows corruption in can be substantial. Our finding that ineffective inspection stations do nothing to improve air pollution supports prior work finding both that gaming is prevalent in I/M programs, and that regulation is ineffective at improving environmental quality when enforcement is weak.

Section 2 outlines the California Smog Check Program and the new STAR system. Section 3 describes the Smog Check and pollution data. Section 4 describes our identification technique and construction of *ex ante* STAR quality measures. Section 5 presents our estimates of how the Smog Check program changed local pollution levels, and section 6 uses these results to simulate the impact of the STAR program. Section 7 concludes.

2 Background on California's Emissions Testing Program

California provides an excellent backdrop for the study of tailpipe emissions programs. Of the approximately 110 million registered automobiles in the United States in 2012, almost 13 million were in California, more than any other single state.[6] California has a history of extensive automobile pollution regulation, and other states often adopt or build off California regulations (Engel, 2015). Prompted by the federal 1977 Clean Air Act Amendments, California began mandating biennial emissions inspections in 1984. Current California law allows the Bureau of Automotive Repair (BAR) to mandate regular measures of tailpipe emissions through "Smog Checks." Most vehicles in California must obtain a Smog Check every two years before renewing their annual vehicle registration. If a vehicle displays emissions levels above the threshold for any regulated pollutant, the owner must repair the vehicle and demonstrate passing levels in a later "re-inspection" before registering it, thereby removing high-polluting vehicles from the fleet by inducing repairs or forcing irreparable vehicles off the road. Emissions thresholds vary by the model year of the vehicle in question, but are consistent over time and across the state.[7] There is also a group of exempt vehicles. These are: vehicles of 1975 model-year or older (which represent less 2% of the 2009 share of vehicles in California according to a California Air Resources Board model), hybrid and electric vehicles, motorcycles, diesel-powered vehicles, and large natural-gas powered trucks.

The California Smog Check program is a decentralized system. Privately-owned repair shops conduct vehicle inspections and, should the vehicle fail initial inspection, these shops make the necessary repairs to bring cars to passing status. Early research found the first incarnation of the Smog Check program was rife with problems that decreased

[6] Bureau of Transportation Statistics, 2014 State Transportation Statistics data, Table 5-1. Available online at `http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/state_transportation_statistics/state_transportation_statistics_2014/index.html/chapter5/table5-1`.

[7] More precisely, there are two different test procedures conducted in the state. Counties under the "enhanced" program regime (discussed below) use the Accelerated Simulation Mode (ASM) test, which uses a dynamometer or rolling road to simulate on-road conditions. Thresholds for the ASM test depend on model year and vehicle weight. Other counties use the simpler Two Stage Idle (TSI) test, which has thresholds that only depend on model year and does not measure NO_x. Regardless of the test, the standards do not change over time, and are the same in any county where the same test is conducted.

or eliminated ambient air pollution benefits (Glazer et al., 1995; Hubbard, 1998), and identified fraud by private station technicians as a major source of problems.

California passed the first major overhaul of the Smog Check program in 1994 in response to the 1990 Clean Air Act Amendments. The state implemented an "Electronic Transmission System" (ETS) to automatically send test results to the BAR, and created an "enhanced" inspection regime for the most polluted areas of the state. In addition to requiring improved testing equipment, the program began directing vehicles in enhanced regions to specially certified stations authorized only to conduct tests but not make repairs. To be clear, while the program refers to such vehicles as being "directed," vehicle owners are still able to choose which station inspects their vehicle. The directed status simply limits the set of stations vehicle owners are permitted to choose. A vehicle is directed for inspections in a testing cycle if a BAR statistical model flags it as meeting a "high emitter profile," and is directed for all follow-up inspections if it fails the initial inspection with emissions greater than or equal to double the legal limits. The BAR also directs a 2% random sample of all vehicles registered in enhanced areas. The BAR directs 30-40% of Smog Checks each year, and directed inspections are a major source of revenue for eligible stations (Eisinger, 2010). The policy of directing vehicles was intended to make California's privately run system more like government-run systems in other states, which were thought to be less prone to fraud. However, test-only stations were still privately run, and lacked the incentive of a test-and-repair station to profit from performing necessary repairs. In 2005, the program allowed a special class of "Gold Shield" test-and-repair stations to inspect directed vehicles as well.

In 2008, the BAR conducted random roadside emissions inspections and compared the roadside results to the same cars' most recent official Smog Check. Many cars listed as passing their last Smog Check failed the equivalent roadside inspection; 19% of older cars passing inspection less than a year prior failed the roadside test. Of the cars that failed roadside testing, approximately half had failed their *initial* official inspection, but then (supposedly) obtained the necessary repairs and passed their final re-inspection at a Smog Check station.[8] A potential implication of the discrepancy is that these cars did not truly pass the re-inspections: someone had instead manipulated the testing outcome.[9] The discrepancies were equally common for test-and-repair stations and test-only stations, undermining the logic for sending directed vehicles to the latter.

[8] "Evaluation of the California Smog Check Program Using Random Roadside Data", 2010 Addendum, California Air Resources Board, February 2010. Available online at http://www.bar.ca.gov/80_BARResources/02_SmogCheck/addendum_with_report.pdf.

[9] An alternative explanation is that the effects of most emissions repairs are short-lived, lasting long enough to pass the follow-up inspection, but degrading to the pre-repair state within a few months.

In response to the roadside inspection study, the California State Legislature further overhauled the Smog Check program. California Assembly Bill AB2289, passed in 2010, directed the BAR to design a new system for certifying stations to inspect directed vehicles, using metrics based on testing results. The system the BAR proposed and eventually implemented was dubbed STAR. Under the new regulations, owners of directed vehicles must obtain checks at STAR-certified locations. STAR stations could be either Test-and-Repair stations or Test-Only, and had to meet specific thresholds on three metrics based on the Smog Check inspection data reported to the BAR. We discuss the three thresholds in detail below.

The BAR finalized regulations for the STAR Program in November 2011, and published STAR scores for all stations in the spring of 2012. The program officially began the next year—all directed vehicles must be inspected at STAR stations as of January 1, 2013.

3 Data

To measure the volume of re-inspections and generate our versions of the STAR quality metrics, we employ inspection-level data from the California Smog Check program. Stations conduct all Smog Check inspections using equipment attached to the ETS that automatically sends results of the test to the California BAR.[10] Our data consist of the population of vehicle inspections conducted in California between 1996 and 2012 and transmitted through ETS.[11]

Each observation in the Smog Check data represents a single inspection, and includes the Vehicle Identification Number (VIN) of the vehicle tested, the date and time of the inspection, the odometer reading, an indicator for the outcome of the test, and emissions readings for hydrocarbons (HCs), NO_x, and CO. Each Smog Check inspection record further contains a 6-digit station identifier, which we join to a crosswalk giving the zipcode of each station.[12]

We determine model year and vehicle type from the included VIN.[13] We also utilize

[10] We obtained access to the Smog Check data via a Public Records Act request, the California equivalent of the Freedom of Information Act.

[11] Data from 1996 and 1997 are incomplete, as the BAR was phasing in the ETS during these years. When available, we use these data to construct any lagged measures for inspections in later years, but our contemporary analysis does not begin until 1998 when the data are more reliable.

[12] We are grateful to Emily Wimberger for providing this crosswalk.

[13] Although the Smog Check data contain some direct information on vehicle types, it is messy and at times unreliable when compared to known VIN information. All vehicles manufactured after 1980 have a standardized 17-digit VIN: the first 8 digits plus the 10th and 11th precisely indicate the vehicle type, at the level of make/model/engine/body type/transmission/year/plant. For earlier vehicles, different

the provided odometer reading in calculating our hypothetical STAR scores.[14]

We use pollution data from the CARB Air Quality Database, a collection of air monitors taking hourly pollution readings. We use data from 1998 to 2009, and focus on CO, NO_x, ozone (O3), and particulate matter (PM). We aggregate hourly readings for CO, NO_x and O3 to the county-day level by averaging individual monitor readings in a given county, and aggregate daily PM readings to the weekly level (most PM monitors take measurements once every six days). We do not weight monitors by any distance metric, and use an unbalanced panel of monitors to maximize available data. To improve readability of our results, we scale pollution readings for CO, NO_x and O3 to parts per billion (PPB).

O3, a major component of the atmospheric condition commonly known as smog, is a secondary pollutant, generated by atmospheric mixing of volatile organic compounds (VOCs) and NO_x. Depending on the current state of local VOC and NO_x levels, additional NO_x can either increase or decrease O3, which makes O3 a difficult pollutant to analyze on a large scale. Regardless, a primary interest of the program was to reduce smog, so we test for general impacts of O3. The link between PM and automobile use is also largely secondary. The largest sources of PM from automobile traffic are combustion of diesel fuel and wear from road and engine friction. We expect little change in these sources from Smog Check—California does not require Smog Checks for diesel vehicle from model years prior to 1997, and I/M should do nothing to change road wear. However, through atmospheric reactions NO_x can form fine particles, providing a vector for an impact, and given the literature on the negative health effects of PM, we include this pollutant as well.[15] Our PM data gives the concentration of particles less than 10 micrometers (PM_{10}) in units of micrograms per cubic meter (μ/m^3). Local weather influences both general air pollution and the types of emissions automobiles generate (Knittel et al., 2016). Ambient temperature can also influence inspection results, as emissions control systems function better when warm. We control for daily high and low temperatures and daily precipitation at the county level. We generate weather data following the methodology of Schlenker and Roberts (2006): taking spatially detailed monthly weather data generated by Oregon State University's PRISM model, aggregating the resulting grid of weather data by county using GIS, and using historical daily averages to interpolate a measure of localized daily

manufacturers used their own formats. We determine make, model year and an approximation of the vehicle-identifying prefix for most of the vehicles manufactured 1975-1980.

[14]We employ an algorithm to "clean" the odometer variable, correcting for rollovers, typos and other glitches that produce unbelievable values for miles traveled between inspections. Specifics of our algorithm are available upon request, or see the data archive for Sandler (2012).

[15]See `http://www3.epa.gov/airtrends/aqtrnd95/pm10.html`, (accessed October 30th, 2015). See also Dominici et al. (2014) for a review of recent literature on particulates.

weather.[16] Weather fluctuations should be exogenous to inspection timing, and their inclusion should do little to change our primary estimates.

Finally, for our simulations we use county-level annual birth rates from the California Department of Finance.

4 Empirical Methodology

We estimate the impact of the Smog Check program, and the projected impact the new STAR program in particular, by leveraging random variation in the timing and location of repairs of failing vehicles. We cannot observe actual repairs, and we instead focus on the timing of final re-inspections following a failed inspection. A final passing re-inspection theoretically indicates a repair took place to reduce a vehicle's emissions below the legal thresholds. Thus, we use final re-inspections in an inspection cycle as a proxy for repairs. Of course, this is not a perfect measure: a final re-inspection is a necessary but not sufficient condition for a repair to occur. There will be cases where the final re-inspection is passed via owner or station fraud, rather than true repair. This drives our interest in measuring station quality, as final re-inspections at high quality stations are much more likely to proxy for a real repair.

The effectiveness of I/M depends on inspections accurately assessing vehicle-level emissions, and both inspectors and vehicle owners can influence test results through dishonest behavior (Oliva, 2015). Evaluating the environmental benefits of programs to increase I/M station quality using a simple pre/post examination of air pollution levels is problematic due to the possibility of stations attempting to game the rating system. Drivers of marginal cars have an incentive to seek test providers that are more lax in testing or willing to falsify tests, which in turn provides an incentive for test providers to engage in such duplicitous behavior to draw business. To avoid such problems, we use the California STAR quality-control program metrics to examine the role of "better" stations, as measured by the STAR metrics, *before* the development of the program. We generate retrospective STAR scores for the period 1998–2009 based on the metrics the BAR eventually used. This allows us to establish the link between station quality and local air pollution using what *would be* "better" stations by STAR standards before the state even proposed the program, such that no gaming behavior is possible. This section describes our methodology in detail, including construction of our retrospective station quality metric and our identification of the effects of the Smog Check I/M program and

[16] We are grateful to Wolfram Schlenker for providing code to create the interpolated daily weather series.

STAR.

4.1 STAR-based Station Quality Metrics

The STAR program certifies stations based on past inspection results. To qualify for the program and receive business from "directed" vehicles as defined in section 2, stations must have a passing grade on three metrics, based on the results of their inspections in the Smog Check database. These include a short-term measure called the Similar Vehicle Failure Rate (SVFR), using inspections taking place in the most recent calendar quarter, and a longer-term measure called the Follow-up Pass Rate (FPR), based on *current* inspection results of vehicles that a station inspected and passed on the *previous* inspection. The other measure involves deviations from standard Smog Check test procedure.[17]

All STAR metrics compare test outcomes at the station in question to average rates for similar vehicles statewide. The BAR defines "similar" vehicles as having the same make, model, year, engine displacement, transmission type, and body style. Both the long-term FPR and the key short-term measure, the SVFR, construct expected total failure rates at initial inspections based on similar vehicles. The BAR compares these measures to the actual failure rate for vehicles inspected at a given station. Although the BAR did not start calculating these measures until after our sample period, we have all available information the STAR program uses and can construct our own measures of both the FPR and SVFR using observed historical inspection results.

Vehicles registered in basic and enhanced Smog Check areas must be inspected every two years. We use the term "cycle" to refer to each time the vehicle is up for inspection before registration. Each cycle may involve multiple inspections if the initial inspection is failed, culminating with a final re-inspection where the vehicle eventually passes. Let s index inspection station, n index vehicle, m vehicle type, c inspection cycle, and q calendar quarter. We index inspections within cycles with $i \in \{1, ..., I\}$. If a vehicle passes the first inspection of a cycle, $I = 1$. Let $\eta_s(c,i)$ denote the set of vehicles that receive their ith inspection at station s during cycle c . We define the general expected failure rate for station s as:

$$\Theta(\eta_s(c,i), c', i') = \frac{1}{|\eta_s(c,i)|} \sum_{n \in \eta_s(c,i)} P(fail_{nc'i'} = 1 | m_n, q, X_{nmqc'}), \tag{1}$$

[17] If a station deviates from procedure more than the average for similar vehicles, the station can be ineligible for STAR. The STAR program separates out one of these deviations, selecting the incorrect gear during the test, from seven other deviations. Stations can fail on up to one of the set of seven deviations and still be eligible for STAR, but cannot fail the incorrect gear selection metric.

where $fail_{nc'i'}$ is an indicator equal to one if vehicle n fails the i' inspection of cycle c', and $X_{nmqc'}$ is a vector of time-varying vehicle characteristics (e.g., mileage). We calculate $P(fail_{nc'i'} = 1|m_n, q, X_{nmc'})$ as:

$$P(fail_{nc'i'} = 1|m_n, q, X_{nmc'}) = F(m, q) + (X_{nmc'} - \bar{X}_{mc'})\beta, \tag{2}$$

where $F(m, q)$ is the proportion of type m vehicles that fail during quarter q, $\bar{X}_{mc'}$ denotes the type-specific mean of $X_{nmc'}$ during cycle c' and β is derived from the following linear probability model using only initial inspections ($i = 1$):

$$P(fail_{nc1} = 1) = \alpha_m + \gamma_q + X_{nmc}\beta + \varepsilon_m, \tag{3}$$

where α_m and γ_q are vehicle type and quarter fixed effects, respectively. Following the BAR procedure, X includes odometer reading at the time of the inspection and days since last inspection. In words, the expected failure rate is the mean of the predicted failure probability for vehicles inspected during the relevant cycle at the station in question.

Using the notation above, we define the SVFR for station s in quarter q as:

$$SVFR_{sq} = \frac{\frac{1}{|\eta_s(c,1)|}\sum_{n\in\eta_s(c,1)} fail_{nc1}}{\Theta(\eta_s(c,1), c, 1)}. \tag{4}$$

The SVFR is the ratio of the actual failure rate for initial inspections at the station during the current period to the expected failure rate for those inspections.

The BAR calculates the FPR score used by the STAR program as the p-value of the following hypothesis test:

$$h_0 : \frac{1}{|\eta_s(c-1, I)|} \sum_{n\in\eta_s(c-1,I)} fail_{nc1} \leq \Theta(\eta_s(c-1, I), c, 1) \tag{5}$$

$$h_A : \frac{1}{|\eta_s(c-1, I)|} \sum_{n\in\eta_s(c-1,I)} fail_{nc1} > \Theta(\eta_s(c-1, I), c, 1) \tag{6}$$

The FPR tests whether vehicles given final inspections by station s on the previous cycle fail more than expected during the current cycle. Note the vehicles in $\eta_s(c-1, I)$ need not be inspected at station s during cycle c — this tracks the outcome of a given automobile across all stations.[18]

We use the Smog Check inspection data to test how well these measures reflect the

[18] The STAR program actually calculates FPR at the technician level, and assigns stations the lowest score of all their registered technicians. Because we calculate our version of the FPR at the station level, we abstract from this distinction for clarity of notation.

ability of stations to reduce tailpipe measured vehicle emissions. We calculate our STAR metrics for the period before the program was in place, and thus our metrics are not subject to gaming or any other responses to implementation of STAR. We regress emissions at a vehicle's initial inspection of the current cycle on the SVFR or FPR of the station that conducted the final, *passing* inspection of the previous inspection cycle. If STAR metrics capture station "quality" (propensity to accurately test and pass vehicles), vehicles passed by higher scoring stations should be less likely to fail, and thus have lower emissions, on their initial inspection of the next cycle. Specifically, we model Y_{ic1}, the CO reading of vehicle n during its initial inspection in cycle c as:

$$Y_{nc1} = \alpha_m + \gamma_q + X_{nmqc}\beta + \sum_{b=1}^{10} Q_{n(c-1)I} + \mu SVFR_{nc1} + \varepsilon n, \tag{7}$$

Where again, m indexes vehicle types and q calendar quarters. X_{nmqc} is a vector of potentially time-varying covariates, which includes fixed effects for vehicle age, inspection type, and county of inspection. Our variable of interest is a set of indicators for 10 bins of each quality score Q, as measured at the last inspection of the previous inspection cycle. We also control for the SVFR of the station conducting the *current* initial inspection, to hold constant the probability of failing independent of underlying emissions, and include controls for the type of inspection and vehicle type, vehicle age, year, and county fixed effects.

Figure 1 plots the coefficients on the STAR score indicators for CO emissions—results for NO_x emissions are similar, and versions of figure 1 are available in the appendix as figure A1. Panel (A) shows results for bins of SVFR, and panel (B) shows results for bins of FPR. For both measures, vehicles that fail the initial inspection on the previous cycle are more likely to fail (have higher emissions) on the initial test of the next inspection cycle. While we do see lower emissions for vehicles passed by stations with higher SVFR scores, the effect flattens out at higher levels, reflecting an undesirable feature of the SVFR as a quality metric: a very high SVFR does not necessarily mean a station is high quality. Low failure rates may indicate fraud against the Smog Check program, but unusually high failure rates may indicate fraud against consumers rather than rigorous inspections (Schneider, 2012). For the FPR, the relationship is the opposite of that intended. For vehicles that failed the initial inspection on the previous cycle, the relationship between emissions and FPR is essentially flat. For vehicles that pass the initial inspection, higher FPR correlates with *higher* CO emissions at the next initial inspection. This counter-intuitive result suggests vehicles inspected and passed at nominally higher quality stations, as measured by the FPR, have greater emissions in the future.

What explains this odd result? The theoretical appeal of the FPR for measuring station quality is that it captures whether or not failing vehicles inspected at a station are correctly repaired before eventually being certified. As we discuss in the next section, this assumption is critical for any I/M program to reduce pollution levels—inspections only reduce pollution if they lead to repairs that reduce emissions. The downside of the FPR score used by the STAR program is that it is necessarily retrospective, capturing a station's performance generally two years in the past. This lagged performance metric may not correlate with current performance, which presents a potential problem for implementation and effectiveness of the current STAR program. We discuss this issue in detail in section 5.

As we use retrospective data, and have the entire history of each station's inspections up through 2012, we can calculate an alternate version of the FPR capturing *contemporaneous* final inspections, including re-inspections. We refer to this contemporaneous FPR as the C-FPR. This score is the p-value from the following hypothesis test:

$$h_0 : \frac{1}{|\eta_s(c,I)|} \sum_{n \in \eta_s(c,I)} fail_{n(c+1)1} \leq \Theta(s, \eta(c+1, I), c, 1) \tag{8}$$

$$h_A : \frac{1}{|\eta_s(c,I)|} \sum_{n \in \eta_s(c,I)} fail_{n(c+1)1} > \Theta(s, \eta(c+1, I), c, 1) \tag{9}$$

Our C-FPR tests whether vehicles given their final inspections by station s in the current cycle fail more than average on their initial inspections in the next cycle (which may be conducted at other stations). This is a modification from the STAR-measured FPR, which uses cycle $c-1$ in place of the $c+1$ in the above metrics. An important feature of this measure is that it is independent of station incentives. If a station has a high C-FPR in a given quarter, that means that, for whatever reason, the vehicles it re-inspected that quarter were more likely to be repaired and thus pass again two years later. Whether the station actually had good practices or not, a high C-FPR means that, empirically, good inspections were done that quarter.

Panel (C) of figure 1 shows the relationship between the C-FPR score and future emissions rates. Not only are higher C-FPR scores associated with lower CO emissions on the next inspection, the relationship is nearly linear. We focus on measuring station quality using the C-FPR for our air pollution results in the section 5.

4.2 The Effect of I/M Programs on Local Air Pollution

Counties and air quality management districts must implement the basic or enhanced Smog Check program when pollution levels cross thresholds set by the federal Clean Air Act. As a result, not only will the pre-trend of increasing emissions lead to bias in estimating the effect of the Smog Check program, exceeding the federal air pollution standards may trigger additional policy responses. For instance, policymakers may respond to high pollution levels by increasing mass transit options, or subsidizing engine modifications on commercial trucks. Further, any kind of before/after comparison risks confounding effects of Smog Check with effects of state- and nation-wide policy changes, particularly emissions standards on new vehicles.

To avoid such issues, we estimate the effect of Smog Check and the new STAR program by exploiting a key realization about the nature of vehicle inspection programs. Initial inspections, passed or failed, correct or not, cannot directly impact air pollution. That is, to a first approximation, *inspections do not affect local air pollution levels*; detection of faulty engines and subsequent repairs do. The only way inspections affect air pollution is through inducing repairs or scrappage. If a vehicle is actually in a failing status but is incorrectly passed on an initial inspection, this should not change air pollution levels—emissions will be high both before and after the inspection. However, local air pollution levels would improve if stations conduct repairs as a result of a failed initial inspection, and correctly verify the effectiveness of the repair by re-inspecting the failing vehicle. Station quality is thus important, as a sham initial inspection, or an omitted repair followed by a sham re-inspection, should have no more effect on air pollution than no inspection program at all.

Conveniently, the timing of inspections is exogenous to both levels and (non-Smog Check) policy-driven changes in local air pollution. An annual vehicle registration notice from the California DMV prompts vehicle owners to obtain a Smog Check every two years. Vehicle registrations are due on the anniversary of the date the vehicle was initially registered in California, and the registration notice comes in the mail around 60 days before the vehicle's registration expires. In California, the expiration date for a vehicle's registration stays constant even if the vehicle is sold. Thus, if a vehicle ever changes owners, the timing of the biennial Smog Check is unrelated to any choice on the part of the current vehicle owner. This provides variation in registration dates, variation in when the vehicle owner obtains the initial inspection, variation in whether a vehicle fails the initial inspection, and variation in how quickly the owner repairs the vehicle and schedules the re-inspection in the event of an inspection failure. None of these sources of variation should correlate with levels of air pollution, except possibly through seasonality, for which we control using fixed effects.

We create a daily panel of re-inspection counts aggregated by county of the station conducting each inspection. Choosing an overly fine geography risks attributing inspections to the wrong location and ignoring spillover effects from pollutants blown to neighboring areas. At the same time, broad geographic definitions reduce our sample size and obscure important variation. We aggregate at the county level as a compromise between these constraints. Counties in California are large relative to other parts of the United States, but a county is a reasonable approximation of the area in which a vehicle owner does most of their commute driving; data on county-to-county migration flows from the 2000 Census show that 82% California workers live and work in the same county.[19] Further, the EPA determines Clean Air Act attainment status at the county level, making it a policy-relevant level of aggregation.

Although daily variation in re-inspections is useful for identification, there is a lag between the repair of a failed vehicle and when we expect to see a change in air pollution levels. Pollutants like NO_x and CO persist in the air and may take time to drift to a sensor, and there may be a period of days between a repair and the date the station records the re-inspection. Thus, associating air pollution levels on a specific day with re-inspections on the same day is overly restrictive. To address this issue, we use a 90-day rolling total of re-inspections as our measure of the intensity of re-inspections. We obtain qualitatively similar results using a 30-, 60- or 120-day window (see online appendix table A1). We also obtain similar results aggregating pollution levels at the weekly level (see online appendix table A3).

Because older vehicles are more polluting on average and thus more likely to be targeted by California's policy of directing vehicles, we split results by vehicle age to estimate separate effects for re-inspections of older vs. newer vehicles. This requires we formally define "older" in our context. Beginning in 1980 three major improvements in engine technology led to significant reductions in vehicle emissions: the three-way catalytic converter, introduced in 1980; fuel-injection systems, introduced in 1985; and second-generation onboard monitoring computers (called OBDII), introduced in 1996. Each of these technologies were adopted rapidly, and affected emissions both under normal operating conditions and potentially under "failing" conditions as measured by Smog Checks.

To help determine a reasonable cutpoint for old versus new vehicles, we plot average CO and NO_x emissions measured at passed and failed Smog Checks by model year.[20]

[19]See `http://www.census.gov/population/www/cen2000/commuting/index.html`, accessed August 1, 2015. We obtain qualitatively similar results when we aggregate at the air basin level.

[20]O3 and PM_{10} are not tailpipe pollutants and as such are not measured directly by Smog Check inspections.

Figure 2 shows that, for both pollutants, emissions at passing inspections are decreasing across model years in a largely continuous fashion as general automobile technology improves. For failed inspections, average Smog Check test results exhibit a strong decline in CO levels beginning with the 1985 model year, and a strong decline in NO_x levels beginning with the 1994 model year. To be more conservative in what we call "older" cars, we opt for 1985 as the cut-off year between old and new vehicles. If we separate out effects into three groups, one for each "era" of pollution control technology, we obtain essentially identical effects from re-inspections of the 1985–1995 and 1995+ vehicles, with no change in the effect of 1975–1985 vehicles, indicating that the split at 1985 is appropriate. See appendix table A2.

We estimate the effect of re-inspections on levels of a pollutant $p \in \{NO_x, CO, O3, PM_{10}\}$ in county g in time t as:

$$p_{tg} = \left(\sum_{i=0}^{90} R^{old}_{g(t-i)} \right) \beta_1 + \left(\sum_{i=0}^{90} R^{new}_{g(t-i)} \right) \beta_2 + \gamma X_{gt} + \varepsilon_{gt}, \tag{10}$$

where R^{old}_{ct} and R^{new}_{ct} denote the number of re-inspections of older and newer vehicles, respectively, X_{gt} is a vector of county-level covariates, and ε_{gt} is an error term. We specify re-inspections in levels. Holding weather, county geography and related factors constant via fixed effects, to a first approximation repairing one failing vehicle should remove the same quantity of emissions, and by extension reduce air pollution levels by the same amount, regardless of whether this represents a 1% change or a 0.001% change in the level of re-inspections. This will cause the model to predict much larger effects for densely populated areas such as Los Angeles County, a desirable feature of our specification. Los Angeles County is heavily polluted in part because it has a very large number of cars on the road, and we expect it to experience large reductions in pollution relative to the counterfactual if the Smog Check program causes repairs for a large numbers of high-polluting vehicles. Our results are robust to excluding Los Angeles County, but even if this county were driving our results, it is worth noting that Los Angeles County has a population of more than 10 million, larger than many European countries and all but the 10 largest U.S. states.[21] β_1 and β_2 give the causal effect of additional re-inspections on air pollution levels. In our preferred specification, X_{gt} includes controls for weather, county fixed effects, calendar-week fixed effects to flexibly control for statewide changes across time, and county-specific cubic time trends. Online Appendix Figure A2 shows

[21] For results without Los Angeles County, see Online Appendix Table A3, column 2. Because observations from Los Angeles County substantially increase the variance of our right hand side variables, omitting them naturally increases our standard errors substantially, but our point estimates stay relatively consistent.

the residual variation for three large urban counties of the 90-day rolling totals of re-inspections after controlling for the covariates from our preferred specification.

Identification of the effect of station quality follows from the same logic as the effect of re-inspections. The number of re-inspections at higher- and lower-quality stations will fluctuate over time with the variation in the timing of the initial Smog Checks of vehicles eventually taken to those stations. As our quality metric was unobservable during the period covered by our analysis, and the study of roadside inspections leading to the development of the STAR program was not even released until 2010, the distribution of vehicles across high- and low-quality stations is exogenous. This further assumes consumers had no other way to identify low-quality stations willing to pass a failing vehicle before STAR. One way to test this assumption is to compare the average C-FPR of stations inspecting older vehicles to the average for stations inspecting newer vehicles. Pre-1985 model year vehicles are more likely to fail and likely more expensive to fix. If consumers had any systematic way of identifying "bad" stations willing to give a sham inspection, we would expect to see older vehicles disproportionately inspected at lower quality stations. We observe the mean C-FPR for older and newer vehicles is essentially the same. Given our argument that only accurate re-inspections reduce local air pollution, we expect to see smaller or zero effects from re-inspections at low-quality stations, and larger effects from re-inspections at high-quality stations.

Conducting analysis at the county rather than station level requires we construct a county-level measure of station quality. We estimate the interaction of station quality and re-inspections using two different approaches. First, we calculate a 90-day rolling average of the C-FPR score at the county level, weighting our individual station C-FPR scores by the number of re-inspections at each station. We interact this county-level average with the count of re-inspections:

$$\begin{aligned} p_{tg} = \left(\sum_{i=0}^{90} R^{old}_{g(t-i)}\right)\beta_1 + \left(\frac{1}{91}\sum_{i=0}^{90} CFPR^{old}_{g(t-i)} \cdot \sum_{i=0}^{90} R^{old}_{g(t-i)}\right)\delta_1 + \left(\sum_{i=0}^{90} R^{new}_{g(t-i)}\right)\beta_2 \\ + \left(\frac{1}{91}\sum_{i=0}^{90} CFPR^{new}_{g(t-i)} \cdot \sum_{i=0}^{90} R^{new}_{g(t-i)}\right)\delta_2 + \gamma X_{gt} + \varepsilon_g, \end{aligned} \quad (11)$$

where $CFPR^{old}_{gt}$ and $CFPR^{new}_{gt}$ denote the county average C-FPR of stations conducting re-inspections of older and newer cars, respectively.

We next take a more flexible semi-parametric approach, equivalent to interacting re-inspections with bins of average county-level C-FPR scores. Because our air pollution analysis uses aggregated data, this takes the form of counting re-inspections of older and newer cars at stations falling into each of 20 equal-size C-FPR bins:

$$p_{tg} = \sum_{b=1}^{20} \beta_1^b \left(\sum_{i=0}^{90} R_{bg(t-i)}^{old} \right) + \sum_{b=1}^{20} \beta_2^b \left(\sum_{i=0}^{90} R_{bg(t-i)}^{new} \right) \beta_2 + \gamma X_{gt} + \varepsilon_g, \tag{12}$$

where $R_{bg(t-i)}^{old}$ counts total re-inspections of older vehicles conducted at stations with C-FPR scores falling in bin b within a county in period $(t-i)$, and $R_{bg(t-i)}^{new}$ is the corresponding count for newer vehicles.

5 Results

Panel A of table 1 summarizes our variables of interest. For consistency with our regression results, we show NO_x, O3 and CO in parts per billion (ppb).[22] Average levels of both NO_x and CO fell across our sample period, dropping almost 350 ppb (0.35 ppm) for CO and 10 ppb for NO_x. O3 was largely unchanged, remaining within 0.5 ppb of initial 1998 levels. We note the number of personal automobiles on the road in California increased by almost 1.5 million vehicles across this period, so no change in O3 levels does not necessarily mean that the Smog Check program had no impact on O3. PM_{10} levels fluctuate from year to year with no obvious pattern. To give a sense of the change in vehicle technology, we also show the average NO_x emissions reading at I/M emissions tests in each year. NO_x emissions largely fell over the sample period.

California is a large and geographically diverse state, with very different climate, topology and population density in the northern and southern regions. For southern California, panel B shows averages for Los Angeles County, the county with the most annual inspections. Improvements in CO were more drastic in Los Angeles County than the state as a whole, decreasing by over 800 ppb (0.8 ppm). NOx decreased around 40 ppb, just over 50% of 1998 levels, while O3 levels increased. For northern California, panel C shows averages for the 9 counties that make up the San Francisco Bay Area.[23] The San Francisco Bay Area follows the general pattern of California, with similar improvements in both CO and NO_x and little change in O3 and PM_{10}.

[22]Researchers often show CO in parts per million (ppm), so at first glance our CO numbers appear larger than prior studies looking at CO in California. For example, Currie and Neidell (2005) show that in 2000, the average California 8-hour high CO level was 1.3 ppm (1,300 ppb), compared to a full day average of 649 ppb for the same period in our data.

[23]The counties in the San Francisco Bay Area are: Alameda County, Contra Costa County, Marin County, Napa County, San Francisco County, San Mateo County, Santa Clara County, Solano County and Sonoma County.

Turning to the data on Smog Check, two general trends appear. First, increases in inspections (and re-inspections) correlate with decreases in both CO and NO_x but with little change in O3. Second, neither total inspections nor re-inspections increase monotonically with time. Total inspections peak in 2004 and then drop, likely due to a 2005 Smog Check policy change that exempted pre-1975 vehicles from inspections. As we control flexibly for such state-level changes across time in our regressions, such trends should not bias our coefficient estimates. In most of the state, re-inspections decline around 2002, corresponding to a decrease in the overall failure rate (and thus the need for re-inspection/repair). As a caveat, we note the San Francisco Bay Area seems to be missing a large number of inspections from 2002 through mid-2003—we exclude these years for these counties from our empirical analysis. We are not aware as to why these counties are "missing" inspections for this period, though the timing corresponds to a shift in the Smog Check regime in these counties, during which stations in the region were upgrading the machines used for Smog Checks. Results are robust to inclusion of this period for these counties.

5.1 Effect of Re-inspections

To establish a causal link between re-inspections (our proxy for repairs) and local air pollution, we estimate a series of regression models based on equation (10), where the coefficient of interest is the effect of re-inspections of failing cars on local air pollution. Specifically, in table 2 we estimate the link between the number of county-level re-inspections and county-level measures of the direct pollutants CO and NO_x. We scale our results to show the effect per 1,000 re-inspections in the past 90 days. Using values from table 1, the average California county sees 1,000 re-inspections of all vehicle ages every 12 days, while Los Angeles County has an average of just over 1,000 re-inspections every day. The nine counties making up the San Francisco Bay Area conduct about 1,000 re-inspections every two days.

Column 1, the most basic model, includes no controls. Both CO and NO_x show a positive correlation between county pollution levels and the number of re-inspections for older cars: an additional 1,000 re-inspections correlates with a 42.5 ppb increase in CO (0.09 of a standard deviation) and a 2.0 ppb increase in NO_x. When we focus on newer cars, the sign flips for CO to a 2.4 ppb decrease per additional 1,000 re-inspections. The sign on NO_x remains positive, but is smaller at 0.15 ppb per 1,000 cars. All results are statistically significant at either the 5% or 1% level.

Adding county fixed effects (Column 2) increases the negative effect of newer car re-inspections on CO to 13.3 ppb per 1,000 re-inspections, but the impact from older cars remains similar. The sign on NO_x is now negative for newer cars, with an additional 1,000 re-inspections lowering ambient levels by 0.6 ppb. Adding controls for weather (Column 3) does little to change the size or sign of our results—while weather can influence the relationship between emissions and ambient pollution levels, we expect it to be exogenous to the number of re-inspections. Thus, controlling for weather serves largely to increase precision of our estimates. Column 4 adds calendar week effects to flexibly control for state-wide trends and within-year seasonality.[24] Controlling for state-wide time effects pushes coefficients toward zero for both vehicle age groups and both pollutants, but the coefficients on re-inspections of older cars remain positive.

Finally, Column 5 controls for local trends using a third-order polynomial time trend for each county, in addition to the county fixed effects, weather controls, and calendar week fixed effects of other models. After controlling for differential trends across counties, increasing re-inspections (repairs) of both older and newer cars *reduces* ambient CO and NO_x. The sign flip associated with the inclusion of county-specific trends suggests counties with increasing air quality saw decreases in the number of re-inspections of older cars over time. Older cars are more polluting than newer cars, but also more likely to be scrapped or sold out of state, especially if these vehicles cannot pass a Smog Check without significant repairs. As a result, counties with many older cars early in our sample period would see fewer re-inspections of older cars over time due to increased scrappage and thus declining air pollution that is not directly related to Smog Check inspections and repairs. Figure A3 in the Online Appendix illustrates these trends for three large urban counties. Failing to account for such trends biases the estimated impact of repairing older cars. In our preferred specification, re-inspecting an additional 1,000 older cars *decreases* ambient CO by 14.4 ppb and ambient NO_x by 0.7 ppb (0.05 and 0.03 of a standard deviation, respectively). An additional 1,000 newer car re-inspections decreases ambient CO by 4.2 ppb and NOx by 0.8 ppb (0.01 and 0.03 of a standard deviation, respectively). Our basic results are robust to several alternate specifications of time controls, the results of which are presenting in online appendix Table A3. Specifically, we obtain similar results dropping Los Angeles county, including-day of-week and month-of-year fixed effects, including county-by-year fixed effects, and aggregating our data at the weekly level.

As a whole, we find the Smog Check I/M program's requirements to repair and re-

[24] While weather is exogenous to the number of inspections, time of year is not: inspection timing corresponds with time of initial purchase, and the number of new cars sold varies systematically across time of year.

inspect high-polluting vehicles moderately improved local CO and NO_x levels, particularly when repairing older model-year cars. We next consider how station quality controls, as proposed in the new STAR program, affect the benefits of I/M.

5.2 Station Quality

To test for the role of station quality, we examine how our C-FPR score correlates with the effect of re-inspections. We first calculate the C-FPR score for all stations in the county, then aggregate up to the county level, weighting by the number of re-inspections conducted over the previous 90 days at each station (see section 4).

Table 3 shows our results. All columns use the controls from column 5 of table 2: weather controls, county fixed effects, county cubic trends, and calendar week fixed effects. Column 1 repeats the model of table 2 column 5 for comparison. Column 2 includes an additional interaction between the number of re-inspected vehicles in the last 90 days with the continuous C-FPR measure. Higher C-FPR corresponds to the county having a greater share of re-inspections at stations the STAR program would consider to be of better quality, had STAR existed at the time.

Column 2 shows station quality matters in the case of older vehicles. As an extreme example, re-inspections in a county with an average C-FPR of zero would have zero effect on either pollutant, but the effect of re-inspections increases with average station quality. For ease of interpretation, we present the estimated effect for each vehicle age calculated at the average county-level C-FPR in our sample. For older cars, an additional 1,000 re-inspections in a county of average C-FPR would decrease ambient CO levels by 46.64 ppb and average NO_x levels by 3.7 ppb, with both average effects statistically significant at 1%.[25] These effects are significantly larger than estimates that do not account for station quality. For CO, the average effect is roughly consistent with older vehicles' share of all CO emissions, and engineering models predict successful repairs of most failing older vehicles would create an effect this large. The California Air Resource Board's criteria emissions (CEPAM) and emissions factor (EMFAC) emissions inventory models indicate 1975–1985 model year personal vehicles contribute around 20 percent of all CO emissions, more so in the earlier years of our sample when these are more common cars. Our results in Figure2 indicate a vehicle of these vintages that fails a Smog Check typically has emissions around 3 times higher than a vehicle that passes inspection. Depending on the county, 1,000 re-inspections of 1975–1985 model year vehicles represents the re-inspection of 0.2 to 20 percent of all personal automobiles of these vintages. This makes a change of

[25]We calculate the effect at the average level using the *margins* command in Stata 15.

around 47 ppb in CO concentrations, between 5–10 percent of mean CO levels, plausible. Our estimate for NOx reductions in a county with average station quality is less in line with the models, which predict older vehicles contribute only around 3–5 percent of all NOx emissions. However, the physical and chemical processes behind NOx concentrations are more complicated, with non-linearities such that a modest reduction in NOx emissions might have a disproportionate effect on NOx concentrations, depending on atmospheric conditions and concentrations of other pollutants such as hydrocarbons and ozone.

The case for station quality when testing newer vehicles is less clear. The signs of both the baseline effect and the interaction effect are reversed from that of older vehicles: a greater share of re-tests at higher quality stations reduces the effect of re-inspections on ambient pollution for both CO and NO_x. Importantly, these results are precisely estimated but economically insignificant, which is clear when considering the marginal effect at the mean. At the average county C-FPR of around 0.59, an additional 1,000 re-inspected newer cars would increase pollution levels by around 0.0002 standard deviations for CO and 0.014 standard deviations for NO_x. The result is not statistically significant for CO, and the average effects on both pollutants are about half the size of the already small overall effects for newer cars from table 2.

Figure 3 clearly illustrates the effect of station quality for newer cars is effectively zero. Following equation (12), we generate 20 bins of C-FPR scores, in units of 0.05, from 0 to 1. We include sets of 20 variables each for older and newer cars, giving the counts of re-inspections at stations in the appropriate C-FPR bin. Figure 3 plots each coefficient on the bin-specific count variables, and provides a LOWESS fit with bandwidth $N * 0.8$ to illustrate the general patterns across the C-FPR distribution.

Panel A shows results for CO, and panel B shows results for NO_x. For newer cars, the effect of a re-inspection is approximately constant at zero across our estimated measure of station quality. Visual analysis shows the puzzling positive results from table 3 are a result of effects for counties with average FPRs very close to 1, which represent a small share of re-inspections overall. For example, only 24% of total newer vehicle re-inspections across our entire sample occurred at stations with C-FPR levels greater than 0.9. Panel B shows that with older cars, there is a clear differential between a re-inspection at a lower-quality vs. a higher-quality station. Increasing re-inspections in areas with C-FPR scores below approximately 0.3 has no effect on local air pollution, with increasing benefits of re-inspections for higher-quality stations.

Table 3 and figure 3 jointly suggest increasing re-inspections of older cars in areas with higher estimated station quality reduces air pollution in an economically and statistically significant manner. However, there is no effect for areas with lower station quality, and economically insignificant effects regardless of station quality when considering newer model-year cars.

The small effect of re-inspections of newer cars on air pollution is to be expected, to some extent. As figure 2 shows, the average differential in tailpipe emissions between a passing and a failing vehicle decreases with model year. Bringing an average failing 1984 model year vehicle up to passing standard reduces the CO concentration in its tailpipe emissions by around 14,000 parts per million, while fixing an average 2001 model year vehicle reduces CO emissions by about 1,300 parts per million, more than 10 times less. With a smaller effect on emissions per vehicle of conducting repairs, it should not be surprising that the effect of repairing 1,000 vehicles on ambient pollution is smaller. The small effect of re-inspections and the zero effect of station quality may also have to do with the OBDII computers installed in vehicles manufactured after 1996. These computers trigger the familiar "check engine" light when they detect an emissions failure, possibly leading to repairs outside the Smog Check inspection cycle. Station quality at the biennial inspection would then have little effect, as most serious emissions failures would have already been fixed. Indeed, as we show in appendix figure A4, newer model years, particularly after 1996, have significantly lower failure rates on initial I/M inspections. More troublingly, OBDII systems may make it easier to consistently defeat detection of emissions failures, as was the case in the scandal Volkswagen diesel vehicle testing scandal discovered in 2015. The STAR quality metrics, including our C-FPR variant, are designed to catch cheating stations by comparing their results to others statewide. However, if a vehicle passes incorrectly at every inspection, this will lead to scores that are uncorrelated with actual inspection results, especially if the cheating behavior is on the part of the vehicle owner or manufacturer rather than the inspection station. A related issue with inspection programs, identified by Mérel et al. (2014), is that vehicle repairs may not last for the full period between inspections. Depending on the rate of repair decay, the benefits of passed re-inspections could be highly transient.

Regardless of the reason, the small effect of re-inspections of newer cars and the negligible effect of station quality for those vehicles is problematic for the social efficacy of such I/M programs in general, and STAR in particular going forward. Newer vehicles make up the majority of inspections, and the majority of the costs of both inspections and repairs come from these vehicles. If the Smog Check program, even enhanced by STAR, is not delivering air pollution benefits from newer vehicles, this diminishes the value of the program. Importantly, this does not necessarily imply that out-of-cycle repairs do not raise social welfare by reducing emissions. However, one cannot attribute such repairs to the I/M program itself.

5.3 Effects of Re-inspections on Secondary Pollutants

We focus on the effect of the Smog Check I/M program on CO and NO_x because these pollutants are directly emitted by motor vehicles. However, the main social harms from NO_x and the policy interest in controlling it stem from its role in forming secondary pollutants, principally O3 and PM. O3 in particular has a complicated formation process, and it is possible that the moderate reductions in NO_x caused by Smog Check might not translate into reductions in O3. In table 4 we estimate our empirical model using ambient O3 as the outcome. Columns 1 and 2 replicate the analysis of tables 2 and 3 using O3 levels in parts per billion as the outcome variable. The point estimates largely have the wrong sign, indicating re-inspections increase O3 levels, but although some estimates are statistically significant, they are economically zero. For the average county, re-testing an additional 1,000 cars increases ambient O3 by 0.0019 of a standard deviation for older cars and increases ambient O3 by 0.008 of a standard deviation for newer cars.

O3 levels are negatively correlated with CO and NO_x levels in our sample, and so the small effect and "wrong sign" for our O3 results could simply reflect that correlation. The process by which NO_x and VOCs form O3 resembles a Leontieff production function, such that when NO_x levels are high, reducing NO_x emissions may have limited effect on O3 levels.[26] To test this, in columns 3 and 4 we control for the level of NO_x and NO_x squared. Our coefficients acquire the "right" sign, with re-inspections of older cars reducing O3 levels, but the magnitudes are still very small. In appendix table A4 we allow the coefficients on re-inspections to vary depending on whether the NO_x level is high or low on a given day. We get similar results as those in table 4. It does not appear our zero findings for O3 are a result of the availability of NO_x. Finally, figure 4 plots results by C-FPR bins, and we see no visual effect.

We next test for the effect of re-inspections on levels of PM_{10}. Because PM_{10} sensors only take readings every 6 days, we collapse to the county-week level. For counties with more than one sensor and thus more than one observation per week, we take the average PM_{10} reading for each week and use the weather controls and the rolling total of re-inspections for the day of the last reading. Table 5 repeats the analysis of table 3 with PM_{10} as the dependent variable. Without controlling for station quality, our estimates are imprecise, but the point estimates indicate re-inspections of older vehicles have a small negative effect on PM_{10} levels, while re-inspections of newer cars have a near-zero effect. Adding an interaction with county-average C-FPR, we find that re-inspections of older vehicles have no effect on PM_{10} levels when conducted at poor quality stations,

[26] There is an additional process by which certain forms of NO_x can actually reduce O3 when NO_x is sufficiently abundant. See Muller et al. (2009).

while re-inspections at high quality stations moderately reduce PM_{10} levels, a result that is statistically significant at 5%. At the average C-FPR, 1,000 re-inspections of older vehicles leads to a 1.97 μ/m^3 reduction in PM_{10} levels, about 0.08 of a standard deviation. As with CO and NO_x, re-inspections of newer cars have the wrong relationship with station quality, although the effect at the mean C-FPR is economically insignificant. Figure 5 plots results for PM_{10} by C-FPR bins, and shows that the effect of re-inspections of older cars increases with the C-FPR score of the stations doing the inspections, while the effect re-inspections of newer cars is effectively zero at all station quality levels.

5.4 Mechanisms

Before turning to our simulation exercise, we examine how much of the failure of low-quality stations is driven by consumer behavior, as this may be relevant for policymakers. In particular, consumers may "shop" for stations that are willing to pass a failing vehicle without repairs by engaging in fraudulent testing behavior.[27] One way to capture the importance of consumer "shopping" behavior is to look at how many consumers switch inspection stations after a failed initial inspection, and the profile of the stations at which they have their final passing test performed (where the passing may be valid or invalid). Table 6 summarizes the likelihood of two more more re-tests and how often consumers switch following a failed initial I/M inspection. This analysis excludes vehicles whose emissions during testing exceeded twice the Smog Check thresholds and were designated "gross polluters." As noted in section 2, such vehicles must have follow-up testing at appropriate stations for subsequent inspections, and including these vehicles would bias upward the appearance of shopping behavior by showing up as switching across stations.

Row 1, column 1 of table 6 shows that approximately 60% of failed inspections result in a passed final inspection within 1 week of the initial test. As the number of inspections prior to passing grows, the likelihood of a quick final passing decreases. Row 2 shows that many of the cars that eventually pass do so at the same station as the initial inspection. Just under 65% of all re-tested cars eventually pass at the same station at which they failed the initial test. Switching becomes more common the longer it takes to pass. Almost 70% of cycles with two inspections start and end at the same inspection station, compared to just over 20% of cycles with five or more inspections. Note that such cycles

[27]The most egregious approach to fraud in I/M is known as "clean piping": entering the information for the car to be tested, but connecting the testing apparatus to a known passing vehicle. For tests done by plugging into the vehicle's onboard diagnostic system rather than directly measuring tailpipe emissions (something done in other I/M programs and in California after the STAR program), the equivalent practice is "clean plugging." There are a variety of less extreme tricks, generally involving deviations from the approved test procedure.

are considerably rarer, making up just 1.5% of all inspection cycles in our data. We also observe that consumers that do switch end up at stations with lower C-FPR scores, particularly in cycles with more than two inspections. For example, a car that takes 3 inspections to pass in a given cycle has a 56% (1 - 0.438 from row 2, column 3) probability of switching stations between initial test and passing, with a decrease of 0.07 in our C-FPR score between stations. Recall the C-FPR measure is an accurate measure of station quality in terms of later passage by vehicles re-inspected at said station. Thus, "switchers" that move stations tend to move to stations of lower contemporaneous quality as measured by their staying-power of repairs and re-inspections. This is consistent with a large fraction of consumers putting in some effort to comply with the I/M program, but eventually some give up and seeking a less legitimate way to pass. No testing data can truly separate all fraud from genuine repairs. But we do not need to do so from the perspective of policy evaluation. Our goal is to understand if I/M policies such as STAR translate to improved air quality. Both genuine and false repairs are a byproduct of I/M programs, making such switching an outcome of interest rather than a confounder.

6 Simulation Exercises

We can use our estimates to simulate the impact of California's recently implemented STAR program, aimed at improving station quality, and the impact of the Smog Check program as a whole. We first predict what average pollution levels would have been under STAR during our sample period. We take our estimate for the last year of our sample, 2009, as a rough estimate of the predicted impact of STAR at the time of implementation in 2013. We then predict what pollution levels would have been throughout our sample absent any I/M program. In addition to predicting pollution levels, we conduct a rough cost-benefit analysis considering the benefits of abating the pollutants we study and the costs of Smog Check and STAR.

The benefits of pollution reduction predominantly arise from reduced infant and elderly mortality—although pollution impacts other health outcomes, the value of a statistical life (VSL) is generally high enough that it dwarfs morbidity effects, which are also more difficult to measure. We use estimates from prior literature in economics on the impacts of CO and PM_{10} on mortality, which largely focus on infants. As such, these benefit estimates ignore mortality effects for those age 1 and older and morbidity effects for all ages. CO and PM_{10} have documented direct effects on infant mortality, while the harms of NO_x derive from its role in forming O3 and particulate matter, which also have documented health effects. As we find no economically significant effect of the Smog Check program on O3, we focus on benefits from reducing PM_{10} and CO.

Instrumental variables estimates from Knittel et al. (2016), Chay and Greenstone (2003), and Arceo et al. (2015) all find that a one unit reduction in ambient PM_{10} results in approximately 10 fewer infant deaths per 100,000 births. However, these studies do not provide useful estimates of the harms of CO: Chay and Greenstone (2003) do not estimate CO effects, CO results from Knittel et al. (2016) are too imprecise to draw conclusions, and estimates from Arceo et al. (2015), which use data from Mexico, are less applicable to our setting given nonlinearities in the effect of CO and the higher CO levels in Mexico. Instead, for the benefits of CO reduction we use estimates Currie and Neidell (2005) and Currie et al. (2009), which both estimate fixed effects models with U.S. data and find a 1 part per million (1,000 ppb) reduction in CO lowers the infant mortality rate by 18.1 and 17.6 deaths per 100,000 live births, respectively. Currie and Neidell (2005) and Currie et al. (2009) find essentially zero effect of PM_{10} on infant mortality, which suggests the possibility of co-pollutant effects. That is, the instrumental variables studies mentioned above place all mortality effects on PM, while fixed effects models place it all on CO, when in fact the total weight likely is shared between the two. By drawing results from different estimation strategies, we over-attribute effects to any single pollutant, making our benefits an over-estimate. With this in mind, we use the 17.6 deaths per 100,000 live births estimate from Currie et al. (2009) to estimate the benefits of CO reductions. We calculate these benefits for each county separately, using county level birthrate data.

6.1 Predicting the Impact of STAR

We first use our regression results to simulate the effect of the specific requirements of the STAR program at the beginning and end of our sample on CO and PM_{10}. Knowing the theoretical impact of STAR in 1998 is interesting, but tells us little about what to expect as the policy goes forward. Thus, we focus on the *future* impact of more stringent I/M requirements. To predict pollution levels, we use the models whose results we present in the second columns of tables 3 and 5 for CO and PM_{10}, respectively.

To project future impacts of STAR, we ask how increasing mean station quality would change the impact of the existing level of re-inspections. The minimum acceptable FPR for STAR certification is 0.4. Thus, we estimate changes in air pollution if every re-inspection of an older vehicle were conducted at a station with our C-FPR score of 0.4 or higher. We take every station with a C-FPR below 0.4, and re-assign re-inspections of older vehicles from those stations to stations in the same county with C-FPR greater than or equal to 0.4. Re-inspections are redistributed to "good" stations in proportion to each station's share of re-inspections in same county and the current quarter. Then, given the simulated distribution of re-inspections over inspection stations, we recalculate the 90-day county-level moving average C-FPR, and predict counterfactual air pollution levels using the model and results from column 2 of table 3. One can think of this exercise as simulating the effect of an idealized certification program targeting true station quality, as distinct from observable quality. Recall our C-FPR measure is not feasible for use in certifying stations. The actual STAR program uses the FPR, which uses contemporaneous inspections to evaluate lagged station performance, and as we show in section 4 the FPR metric poorly predicts future vehicle emissions.

Figure 6 maps our predicted results by county for the years 1998 (top) and 2009 (bottom), with results for CO and NO_x on the left and right, respectively. The counties making up the San Francisco Bay Area see relatively small changes in 1998, in part because these counties historically had what would be high C-FPR score stations to begin with: our theoretical exercise of removing "bad" station re-inspections thus has little bite. But improving station quality in 1998 would have substantially reduced CO and NO_x levels in a number of California's urban areas, with Los Angeles County and nearby portions of southern California receiving the greatest benefit. Although relatively few counties see a large change in pollution levels, in 1998 more than 15 million people lived in counties with counterfactual decreases of 20% or more, so welfare effects would have been large. We discuss such welfare effects further below.

The case for improving inspection station quality in 2009 is less clear. As a result of older vehicles aging out of circulation, the effect of moving older vehicles to higher quality stations is limited. Many counties, including the most populous in the state, see reductions of less than 1% from already lower baseline pollution levels.

These estimates likely overstate the potential impact of STAR, as our measure of station quality, while based on the STAR program methodology, uses future inspection results to rate current stations. This improves accuracy in measuring low- vs. high-quality stations, but is impossible to use in practice. To predict the effect of the STAR program using the standard STAR measures, we conduct another simulation, this time reassigning re-inspections of pre-1985 vehicles from stations with STAR metrics below the cut-offs (FPR below 0.4 and SVFR below 0.75) to stations with STAR scores above those cut-offs.

Table 7 shows the results of simulating implementation of the STAR program in 2009. Assigning all older vehicles to stations with our C-FPR of 0.4 or higher would raise the average C-FPR of stations inspecting older vehicles from 0.6 to 0.77. This would cause the average 2009 CO level to decrease by 11 ppb, the NO_x level by almost 0.9 ppb, and the PM_{10} level by 0.3 μ/m^3. The statewide effect of implementing the STAR program is much smaller, with a 3.4 ppb reduction in CO, a 0.29 ppb reduction in NO_x, and a 0.1 μ/m^3 reduction in PM_{10}. The statewide effect hides substantial variation across the state's most polluted areas. The average C-FPR in Los Angeles County was 0.39 in 2009. Moving re-inspections at low quality stations to stations with C-FPR of 0.4 or higher would raise the average C-FPR to 0.65, lowering CO levels by 75 ppb, lowering NO_x levels by 7.4 ppb, and lowering PM_{10} by 3.25 μ/m^3. Implementing the actual STAR program in Los Angeles County in 2009 would raise the C-FPR to just 0.46, with an effect on all three pollutants roughly one fourth as large as the idealized program. The San Francisco Bay Area had higher than average C-FPR in 2009, and either of our simulations has only a small effect on both average station quality and air pollution. In contrast, San Diego County had an average C-FPR score of 0.33 in 2009, and as in Los Angeles County, implementing the actual STAR program would have about one fourth the effect of using the C-FPR to certify stations.

The effect of increasing station quality is quite small in either simulation for 2009. Even this overestimates the *future* effect of STAR, as the number of older model year vehicles on the road falls with time.

6.1.1 Welfare Effects of STAR

Our simulations suggest the current incarnation of the STAR program will do little to further improve local air pollution in California. Because the gains from inspections mostly derive from older vehicles that are becoming scarce on the road, even an inspection program using our contemporaneous quality measure would have a small impact by 2013. Further, the STAR quality metrics available in practice appear to do a poor job of measuring current station quality. As a result, moving vehicle inspections to STAR-approved stations is unlikely to meaningfully reduce air pollution. However, while STAR is unlikely to have a substantial impact on air pollution levels, it is a relatively inexpensive policy, and may have net positive effects on welfare.

Using 2009 county-level birth rates, our predicted effect of STAR on CO would have prevented an expected 0.64 infant deaths over the year statewide. Using the U.S. EPA's preferred estimate of the VSL, $7.6 million, this translates to annual benefits of about $4.9 million. If PM_{10} affects infant mortality, STAR would have prevented an additional 16.5 infant deaths in California, with welfare benefits of about $125 million. Note that these results assume a *lasting* decrease in pollution levels following faulty car repairs. If the benefits of repairs degrade with time, the health gains will be more short lived and thus the social benefits lower. As an important caveat, our estimated pollution reductions are on a large geographic scale, effectively estimating the average reductions across a county.

PM in particular can travel large distances, so this may be the relevant geographic scale. But if reductions are more local and the health gains of reduction are non-linear, then health benefits could be much higher.

While the benefits of STAR for CO in particular are modest, STAR is a fairly inexpensive program. The California State Legislature estimated additional administrative costs of $350,000-$450,000 per year.[28] The more significant source of costs is in time—directing vehicles to STAR stations instead of the more abundant test-only stations might lead to increased driving times. In 2009, there were about 3.6 million directed inspections in the state of California. At the 2009 average hourly wage for California ($23.82),[29] if STAR caused owners of directed vehicles to drive an extra 10 minutes round trip to their Smog Checks, the time costs would total more than $14.5 million that year.

6.2 Do "Smog Checks" Affect Smog?

We next estimate the effect of the Smog Check Program as a whole on California pollution levels. The methodology for this simulation is simple–we predict pollution levels in each year of our sample with the number of re-inspections set to zero for all days and counties. That is, we estimate pollution levels for the counterfactual world where California had no Smog Check Program at all. We again use our regression models presented in the second columns of 3 and 5. However, we assume for purposes of our simulations that re-inspections of newer vehicles have zero effect on PM_{10}. Our regression results have a positive effect at the mean C-FPR level, but both the base effect and interaction term are noisily estimated and statistically indistinguishable from zero, and the marginal effect at the mean C-FPR is only marginally significant. Taken literally, our point estimates would imply that eliminating the Smog Check Program would drastically *reduce* PM_{10} levels, which seems unlikely.

For the costs of Smog Check, we assume that each inspection has a social cost of $20. There is no database of Smog Check station prices that we are aware of. However, anecdotal evidence indicates that inspection stations typically charge at between $30 and $70 for an inspection (see, e.g. `http://www.smogtips.com/smog_program.cfm`). We use $20 as an estimate of cost both to be conservative and to account for the fact that some portion of the price of a Smog Check inspection is a transfer from the consumer to either the station or the state, rather than a true social cost. We further assume that each inspection takes 20 minutes of consumer time, which we again value at the 2009 average

[28]See `http://www.leginfo.ca.gov/pub/09-10/bill/asm/ab_2251-2300/ab_2289_cfa_20100617_172946_sen_comm.html` (Accessed September 23, 2015).

[29]See `http://www.bls.gov/oes/2009/may/oes_ca.htm`, (accessed November 9, 2015)

hourly wage for California, \$23.82. An additional cost of the Smog Check Program is the required repairs when vehicles fail an inspection. To estimate this cost, we employ an additional set of data from the Smog Check program, containing records on repairs logged by Smog Check stations. We assume that the number of Smog Check-related repairs each year are equal to the number of official Smog Check repair records — i.e., there are no additional repairs that are not logged. A subset of the repair records indicate the cost of repairs,[30] and we use these records to calculate the average repair cost by year. We only have repair data beginning in 2002, and so our welfare estimates will begin in that year. Our estimate of the total cost of Smog Check in a given year is then the total number of inspections, multiplied by \$20 plus 0.33 times \$23.82, plus the total number of repairs times the average repair cost.

We note that this exercise will to some extent understate the benefits of Smog Check, as it implicitly assumes away long-term effects of regular inspections on vehicle emissions. Vehicle emissions controls deteriorate over time. Even if many inspections are low quality, without any inspection the average emissions rate might climb above what we observe in our data, as the emissions controls fail to an extent that even an ineffective inspection system would not permit. We see some evidence of this in the Smog Check data. California requires vehicles that are imported from other states to receive a Smog Check before getting an initial registration, and the data indicates which inspections are "initial registration inspections" of this nature. As not all states have I/M programs, any difference between emissions at initial registration inspections and normal biennial inspections will partially capture the long-term effects of having no I/M program. As we show in appendix table A5, newer vehicles imported from out of state have emissions readings as much as 20% higher than in-state vehicles, even after controlling for vehicle type, age, and odometer reading (however, we find much smaller differences among the oldest and dirtiest vehicles). Note also that we are predicting significantly out of sample in this simulation. In contrast to our STAR simulation, where we observe a range of station quality in our data, we do not observe a period with no Smog Checks. Thus, while we take the trends in the welfare analysis to be informative, the levels may be off significantly in either direction.

We display the results of our simulation in Figure 7, which plots by year our estimates of the health benefits of CO reductions, the health benefits of PM_{10} reductions, and the costs, of the Smog Check program. Two things stand out. First, despite the relatively small effect of re-inspections on PM_{10}, particulate matter is so harmful that the aggregate

[30] Specifically, actual repair costs are reported for repairs where the consumer applied to the Consumer Assistance Program, or CAP, a program that subsidizes emissions repairs for low-income consumers.

benefits were quite large, particularly earlier in our sample. Second, the benefits of the Smog Check Program declined rapidly due to older cars disappearing from the fleet, while the costs remained relatively constant. We predict that the program still had positive net benefits in 2009, but the ratio of benefits to costs fell precipitously between 2002 and 2009. In 2002, the benefits from CO reductions alone exceeded the costs of the program, and the total benefits of pollution reduction were more than 10 times the costs. By 2009, the benefits of the much smaller reductions in CO were less than a quarter of the cost of the program, and the combined benefits of CO and PM_{10} reductions are only about twice the costs. Given this trend, it seems likely that the benefits fell below the costs between 2009 and the present.

7 Conclusion

Motor vehicles generate a number of pollutants, many of which have an established negative impact on human health. To reduce the public damages of automobile use, governments may use inspection and maintenance (I/M) programs, where stations routinely test engines for compliance with emissions standards, and require repairs and further tests in the event of failure. These inspections and repairs are costly to consumers, and while follow-up tests can show repairs improve emissions at the tailpipe, there is little causal evidence of how such programs change local air quality. We test two important questions related to such programs. First, do I/M programs improve local air pollution in an observable manner on a large scale, beyond laboratory conditions? Using over a decade of data from the state of California, we show an increase in emissions-related repairs, as proxied by passing post-repair inspections, corresponds to local improvement in CO, NO_x and PM_{10} levels, but with little change in local O3. This relationship persists after controlling for a wide variety of location and time fixed effects and ambient weather controls, and shows California's Smog Check program has successfully improved local air pollution. However, *additional* gains from the Smog Check program are decreasing with time, as almost all benefits of repairs and re-inspections come from fixing failing older model cars (1985 and prior) with inferior emissions control technology. As older technology cars disappear from the road, the differential between failing and repaired emissions decreases. This is a case where growth of regulated technology, potentially driven by the program itself, is reducing the social efficiency of the regulatory program over time.

We combine information on estimated costs of emissions inspections, our estimated regional pollution improvements, and pollution mortality effects from prior economics literature to perform a back-of-the-envelope cost/benefit analysis of the California I/M program. Our estimates suggest that, while the costs of the program remained largely constant during our sample, the benefits declined sharply with each passing year as the older-model cars that benefit most from Smog Checks disappear from the road. Even so, as of the end of our data sample in 2009, the benefits of the program still exceeded the costs by approximately 2 to 1, suggesting it would have passed a basic cost/benefit test at that time, although given the trend it is questionable whether the program would pass such a test today. Moreover, there remains the question of efficiency relative to other policies. I/M programs like California's Smog Check Program are intended to reduce O3 levels by controlling emissions from all vehicles. Given that we find the primary benefit of Smog Check is from inspections of a shrinking number of very old cars, and manifests primarily through reductions in ambient particulate matter we must ask—is there a better policy to achieve these pollution reductions?

We partially address this issue in our second question as we consider whether using a certifying high quality stations using inspection results can further reduce air pollution. Given fears of false or low-quality repairs of failing vehicles followed by potentially fraudulent re-inspections, the new California STAR program uses a "Follow-up Pass Rate" (FPR) measure based on inspection data to determine which stations the state allows to provide repairs and re-inspections to the dirtiest cars. A number of factors, including potential system fraud, complicate measuring the role of the FPR after the implementation of the program, making estimation of program effects difficult. We use the California Smog Check data from 1998-2009 to construct a modified station-level FPR that conveys information similar to that of the STAR program, but without concerns of secondary program effects. We show how local air pollution changes with an increase in average quality of local stations allowed to administer inspections and repairs. When a greater share I/M stations are of high quality, an additional repair corresponds to a greater marginal improvement in contemporaneous air pollution. However, the gains of estimated station quality are again limited to repairs and re-inspections of older model cars. For newer cars with modern emissions control technology, there is no economically meaningful benefit to restricting repairs to high-quality stations. Between a decline in effectiveness of the overall Smog Check program as older vehicles age out of the fleet and poor predictive power of the actual STAR quality metrics, it is unlikely the new STAR program as designed will further improve the effectiveness of Smog Check for reducing air pollution.

References

Ando, Amy, Virginia McConnel, and Winston Harrington, "Costs, Emissions

Reductions and Vehicle Repair: Evidence from Arizona," *Journal of the Air Waste Management Association*, 2000, *50* (4), 509–521.

Arceo, Eva, Rema Hanna, and Paulina Oliva, "Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City," *The Economic Journal*, 2015.

Bureau of Automotive Repair, "Smog Check Performance Report (AB2289)," Technical Report 2014. Accessed from `https://www.bar.ca.gov/pdf/ab_2289_bill_20100924_chaptered[1].pdf`.

Chay, Kenneth Y. and Michael Greenstone, "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *The Quarterly Journal of Economics*, 2003, *118* (3), 1121–1167.

Currie, Janet and Matthew Neidell, "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience," *Quarterly Journal of Eocnomics*, 2005, *120* (3), 1003–1030.

_ **and Reed Walker**, "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 2011, *3* (1), 65–90.

_ **, Mathew Neidell, and Johannes F. Schmieder**, "Air Pollution and Infant Health: Lessons from New Jersey," *Journal of Health Economics*, 2009, *28*, 688–703.

Davis, Lucas W, "The effect of driving restrictions on air quality in Mexico City," *Journal of Political Economy*, 2008, *116* (1), 38–81.

Dominici, Francesca, Michael Greenstone, and Cass R. Sunstein, "Particulate Matter Matters," *Science*, 2014, *344*, 257–259.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India," *The Quarterly Journal of Economics*, 2013, *128* (4), 1499–1545.

Eisinger, Douglas S, *Smog Check: Science, Federalism, and the Politics of Clean Air*, Routledge, 2010.

Engel, Kirsten H., "The Enigma of State Climate Change Policy Innovation," in Kalyani Robbins and Erin Ryan, eds., *The Law and Policy of Environmental Federalim: A Comparative Analysis*, Edward Elgar Publishing Ltd., 2015.

Fowlie, Meredith, Christopher R Knittel, and Catherine Wolfram, "Sacred cars? Cost-effective regulation of stationary and nonstationary pollution sources," *American Economic Journal: Economic Policy*, 2012, *4* (1), 98–126.

Glazer, Amihai, Daniel B. Klein, and Charles Lave, "Clean on Paper, Dirty on the Road: Troubles with California's Smog Check," *Journal of Transport Economics and Policy*, 1995, *29* (1), 85–92.

Gray, Wayne B. and Jay P. Shimshack, "The Effectiveness of Environmental Monitoring and Enforcement: A Review of the Empirical Evidence," *Review of Environmental Economics and Policy*, 2011, *5* (1), 3–24.

Harrington, Winston, Virginia McConnell, and Amy Ando, "Are vehicle emission inspection programs living up to expectations?," *Transportation Research Part D: Transport and Environment*, 2000, *5* (3), 153–172.

Hoekstra, Mark, Steven L. Puller, and Jeremy West, "Cash for Corollas: When stimulus reduces spending," 2014. NBER Working Paper 20349.

Hubbard, Thomas N., "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market," *RAND Journal of Economics*, 1998, *29* (2), 406–426.

Jacobsen, Mark R, "Evaluating US fuel economy standards in a model with producer and household heterogeneity," *American Economic Journal: Economic Policy*, 2013, *5* (2), 148–187.

_ , "Fuel Economy and Safety: The Influences of Vehicle Class and Driver Behavior," *American Economic Journal: Applied Economics*, 2013, *5* (3), 1–26.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders, "Caution Drivers! Children Present: Traffic, pollution and infant health," *Review of Economics and Statistics*, 2016, *98* (2), 350–366.

Li, Shanjun, Joshua Linn, and Elisheba Spiller, "Evaluating "Cash-for-Clunkers": Program Effects on Auto Sales and the Environment," *Journal of Environmental Economics and Management*, 2013, *65* (2), 175–193.

Mérel, Pierre, Aaron Smith, Jeffrey Willams, and Emily Wimberger, "Cars on Crutches: How Much Abatement Do Smog Check Repairs Actually Provide?," *Journal of Environmental Economics and Management*, 2014, *67* (3), 371–395.

Mian, Atif and Amir Sufi, "The Effects of Fiscal Stimulus: Evidence from the 2009 'Cash for Clunkers' Program," *The Quarterly Journal of Economics*, 2012, *1107*, 1142.

Muehlenbachs, Lucija, Stefan Staubli, and Mark A. Cohen, "The Impact of Team Inspections on Enforcement and Deterrence," *Journal of the Association of Environmental and Resource Economists*, 2016, *3* (1), 159–204.

Muller, Nicholas, Daniel Tong, and Robert Mendelsohn, "Regulating NO_x and SO_2 Emissions in Atlanta," *The B.E. Journal of Economic Analysis and Policy*, 2009.

Oliva, Paulina, "Environmental regulations and corruption: Automobile emissions in Mexico City.," *Journal of Political Economy*, 2015, *123* (3), 686–724.

Sandler, Ryan, "Clunkers or Junkers? Adverse selection in a vehicle retirement program," *American Economic Journal: Economic Policy*, 2012, *4* (4), 253–281.

Schlenker, Wolfram and Michael J. Roberts, "Nonlinear Effects of Weather on Corn Yields," *Review of Agricultural Economics*, 2006, *28*, 391–398.

Schneider, Henry S., "Agency Problems and Reputation in Expert Services: Evidence from Auto Repair," *The Journal of Industrial Economics*, 2012, *60* (3), 406–433.

Shimshack, Jay P. and Michael B. Ward, "Regulator reputation, enforcement, and environmental compliance," *Journal of Environmental Economics and Management*, 2005, *50* (3), 519–540.

Simeonova, Emilia, Janet Currie, Peter Nilsson, and Reed Walker, "Congestion Pricing, Air Pollution and Children's Health," 2018. NBER Working Paper 24410.

Wolff, Hendrik, "Keep Your Clunker in the Suburb: Low-emission Zones and Adoption of Green Vehicles," *The Economic Journal*, 2014, *124* (578), F481–F512.

Table 1: Average Daily Air Quality and Smog Check Inspections in California Counties, 1998-2009

	CO (PPB)	NO_x (PPB)	Ozone (PPB)	PM_{10} (μ/m^3)	# Inspections	# Re-Inspections	Mean NO_x I/M Reading (PPM)
Panel A: All California							
1998	703.0	28.89	28.04	25.50	587.5	63.07	4.775
1999	716.2	32.40	28.32	31.00	678.5	74.12	5.226
2000	649.6	30.19	26.65	27.16	706.2	85.29	5.172
2001	612.1	27.68	27.60	27.79	755.2	93.99	5.012
2002	600.8	28.18	29.04	30.10	719.6	96.61	4.796
2003	562.6	26.36	28.69	26.78	825.0	107.2	4.688
2004	499.6	24.20	27.66	26.38	795.7	101.7	4.494
2005	458.4	24.23	26.60	24.05	669.7	84.99	4.684
2006	466.8	23.41	28.19	26.14	678.3	78.04	4.448
2007	431.7	22.05	27.61	25.90	667.6	73.65	4.265
2008	422.3	20.38	28.93	26.75	671.6	72.64	4.098
2009	363.5	18.03	27.66	22.18	688.5	73.90	3.958
Average	551.9	25.57	27.89	26.61	701.5	83.19	4.593
Panel B: Los Angeles County							
	CO (PPB)	NO_x (PPB)	Ozone (PPB)	PM_10 (μ/m^3)	# Inspections	# Re-Inspections	Mean NO_x I/M Reading (PPM)
1998	1254.5	71.45	21.73		7064.1	814.8	4.915
1999	1203.6	79.36	20.78		8324.6	1023.0	5.305
2000	1029.1	70.98	20.73		8489.6	1121.0	5.274
2001	936.9	64.25	21.58		9015.2	1253.4	5.121
2002	869.9	60.89	24.30	37.60	9255.0	1318.1	4.931
2003	814.0	58.35	25.27	32.40	9328.7	1316.7	4.810
2004	663.9	51.49	26.49	32.41	9418.5	1271.9	4.621
2005	580.2	47.09	24.48	30.45	7838.1	1042.3	4.773
2006	548.8	47.35	25.48	29.98	7880.1	962.3	4.561
2007	506.6	43.86	25.05	33.32	7644.7	900.2	4.382
2008	470.3	40.49	25.93	30.66	7600.5	871.8	4.175
2009	417.4	35.99	26.59	29.86	7781.6	895.2	4.074
Average	774.6	55.96	24.04	32.06	8303.5	1065.9	4.757
Panel C: San Francisco Bay Area							
	CO (PPB)	NO_x (PPB)	Ozone (PPB)	PM_{10} (μ/m^3)	# Inspections	# Re-Inspections	Mean NO_x I/M Reading (PPM)
1998	756.8	33.81	19.76	18.00	5078.1	422.3	4.872
1999	769.7	37.48	19.84	21.40	5571.8	403.6	4.534
2000	699.7	35.02	18.41	18.45	5941.3	514.9	4.100
2001	636.5	31.51	19.86	20.16	6223.4	567.5	3.955
2002	599.5	31.08	20.61	21.67	865.1	69.89	4.126
2003	569.4	27.56	20.41	16.55	3849.2	379.1	4.565
2004	502.9	25.80	20.43	17.32	6820.2	831.2	4.558
2005	486.7	25.51	20.21	16.31	5731.4	693.5	4.738
2006	466.9	24.93	21.90	18.34	5867.2	608.5	4.474
2007	418.2	23.18	21.68	16.47	5795.3	561.4	4.274
2008	387.5	21.82	22.80	17.44	5820.1	545.0	4.131
2009	350.4	21.08	21.24	14.21	5960.5	553.0	3.984
Average	553.7	28.23	20.60	18.07	5294.2	512.6	4.373

Note: Excludes counties and years where biennial inspections are not required

Table 2: Re-Inspections and County-Level Daily Air Quality

	A: Outcome is Carbon Monoxide (PPB)				
	(1)	(2)	(3)	(4)	(5)
000s of Re-Inspections Last 90 Days					
1975-1985 Vehicles	42.50***	34.87***	36.38***	15.81***	-14.43***
	(10.10)	(8.450)	(9.153)	(2.168)	(3.972)
1985+ Vehicles	-2.390**	-13.34***	-10.55***	-4.639***	-4.226*
	(1.125)	(4.708)	(3.836)	(1.104)	(2.222)
County FE	No	Yes	Yes	Yes	Yes
Weather Controls	No	No	Yes	Yes	Yes
Calendar Week FE	No	No	No	Yes	Yes
County Time Trends	No	No	No	No	Yes
	B: Outcome is NO_x (PPB)				
	(1)	(2)	(3)	(4)	(5)
000s of Re-Inspections Last 90 Days					
1975-1985 Vehicles	2.033***	1.454***	1.549***	0.863***	-0.680**
	(0.318)	(0.237)	(0.289)	(0.0917)	(0.263)
1985+ Vehicles	0.150***	-0.639***	-0.456***	-0.244***	-0.788***
	(0.0382)	(0.204)	(0.146)	(0.0451)	(0.148)
County FE	No	Yes	Yes	Yes	Yes
Weather Controls	No	No	Yes	Yes	Yes
Calendar Week FE	No	No	No	Yes	Yes
County Time Trends	No	No	No	No	Yes
Observations	143440	143440	143440	143440	143440

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Observations are county-days. Standard errors clustered by county reported in parentheses.

Table 3: Station Quality and County-Level Air Pollution

	A: Outcome is CO (PPB) (1)	(2)
000s of Re-Inspections Last 90 Days		
1975-1985 Vehicles	-14.43***	8.287*
	(3.972)	(4.576)
1975-1985 Vehicles · C-FPR		-98.07***
		(22.09)
Effect at Mean C-FPR		*-46.64****
		(9.769)
1985+ Vehicles	-4.226*	-13.17***
	(2.222)	(3.991)
1985+ Vehicles · C-FPR		23.32***
		(8.462)
Effect at Mean C-FPR		*0.111*
		(2.611)

	B: Outcome is NO_x (PPB) (1)	(2)
000s of Re-Inspections Last 90 Days		
1975-1985 Vehicles	-0.680**	1.628***
	(0.263)	(0.281)
1975-1985 Vehicles · C-FPR		-9.623***
		(1.340)
Effect at Mean C-FPR		*-3.695****
		(0.678)
1985+ Vehicles	-0.788***	-1.741***
	(0.148)	(0.276)
1985+ Vehicles · C-FPR		2.457***
		(0.453)
Effect at Mean C-FPR		*-0.366***
		(0.168)

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Observations are county-days. All regressions control for daily weather, county fixed effects, calendar week fixed effects, and county-specific cubic time trends. Standard errors clusted by county reported in parentheses.

Table 4: Station Quality and County-Level Ozone Pollution

	Base		NO_x Controls	
	(1)	(2)	(3)	(4)
000s of Re-Inspections Last 90 Days				
1975-1985 Vehicles	-0.121	-0.485***	-0.195***	-0.259**
	(0.0809)	(0.152)	(0.0700)	(0.127)
1975-1985 Vehicles · C-FPR		0.873**		-0.401
		(0.419)		(0.405)
Effect at Mean C-FPR		*0.0228*		*-0.480*
		(0.166)		*(0.177)*
1985+ Vehicles	0.176***	0.472**	0.0918**	0.248
	(0.0490)	(0.185)	(0.0425)	(0.154)
1985+ Vehicles · C-FPR		-0.635**		-0.307
		(0.279)		(0.244)
Effect at Mean C-FPR		*0.0981**		*0.0762**
		(0.0523)		*(0.0481)*

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Observations are county-days. All regressions control for daily weather, county fixed effects, calendar week fixed effects, and county-specific cubic time trends. Columns (3) and (4) control for a quadratic in the level of NO_x. Standard errors clusted by county reported in parentheses.

Table 5: Station Quality and County-Level PM10 Pollution

	Outcome is PM_{10} (μ/m^3) (1)	(2)
000s of Re-Inspections Last 90 Days		
1975-1985 Vehicles	-0.857	0.527
	(0.605)	(0.815)
1975-1985 Vehicles · C-FPR		-4.236**
		(1.800)
Effect at Mean C-FPR		*-1.966***
		(0.861)
1985+ Vehicles	0.132	-0.0936
	(0.170)	(0.392)
1985+ Vehicles · C-FPR		0.634
		(0.503)
Effect at Mean C-FPR		*0.283**
		(0.166)

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Observations are county-weeks. All regressions control for daily weather, county fixed effects, calendar week fixed effects, and county-specific cubic time trends. Standard errors clusted by county reported in parentheses.

Table 6: Consumers' Shopping Behavior After a Failed Initial Inspection

	All	Total Number of Inspections in Cycle 2	3	4	5+
Final Inspection Within 1 week	0.603	0.659	0.388	0.281	0.207
Final Inspection At Same Station	0.644	0.695	0.438	0.372	0.323
Switchers' Change in C-FPR	-0.0442	-0.0332	-0.0681	-0.0746	-0.0706
N	16317721	13311589	2219152	537029	249951

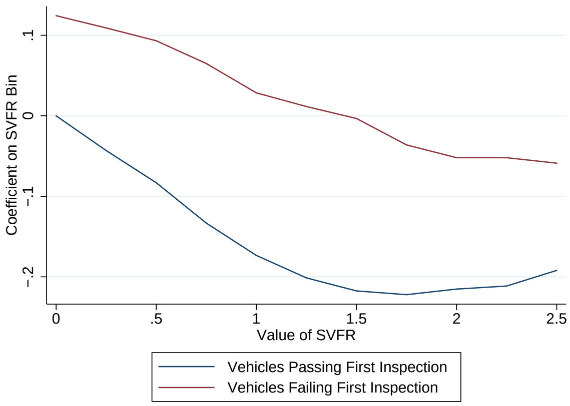
Statistics are means. A unit of observation is an inspection cycle, limited to cycles with more than one inspection, with no "gross" failures. "Final Inspection at Same Station" is equal to 1 if the first and last inspection in the cycle are at the same station. "Switchers' change in SVFR" is the difference between the SVFR score of the first and last station in the inspection cycle; "Switchers' change in C-FPR" is calculated similarly.

Table 7: Predicted 2009 Station Quality and Pollution, Simulating an Ideal and Real STAR Program

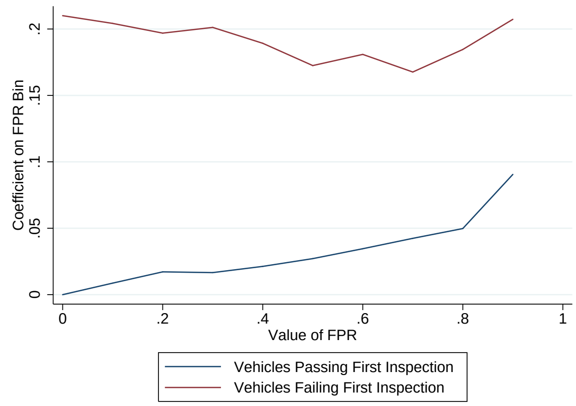
	All CA	LA County	Bay Area	San Diego
C-FPR Score				
Baseline	0.597	0.393	0.632	0.328
Re-assign using C-FPR	0.772	0.651	0.787	0.698
Reassign using SVFR & FPR	0.664	0.458	0.695	0.416
CO Level (PPB)				
Baseline	363.9	453.3	341.3	611.9
Re-assign using C-FPR	352.9	378.1	337.5	577.7
Reassign using SVFR & FPR	360.5	434.1	340.2	603.9
NO_x Level (PPB)				
Baseline	17.54	38.43	20.95	23.94
Re-assign using C-FPR	16.67	31.04	20.58	20.59
Reassign using SVFR & FPR	17.25	36.54	20.84	23.16
PM_{10} Level (μ/m^3)				
Baseline	20.92	28.63	16.73	28.70
Re-assign using C-FPR	20.59	25.37	16.57	27.23
Reassign using SVFR & FPR	20.81	27.79	16.67	28.36

Notes: Baseline scenario shows actual average C-FPR and predicted CO and NO_x from the model in column 2 of table 3. "Reassign using C-FPR" redistributes inspections at stations with C-FPR below 0.4 to stations which pass these measures, recalculates the county average C-FPR accordingly, and predicts CO and NO_x. "Reassign to STAR Stations" redistributes inspections at stations failing the SVFR and FPR threshholds to stations which pass these measures.

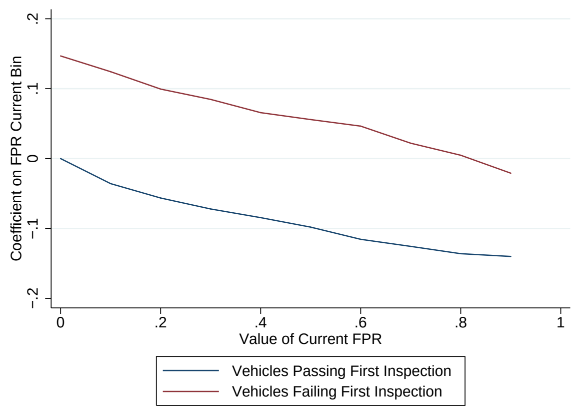
Figure 1: STAR Metrics of Re-inspection Station and Log CO Emissions at Next Inspection



(A) STAR SVFR Score



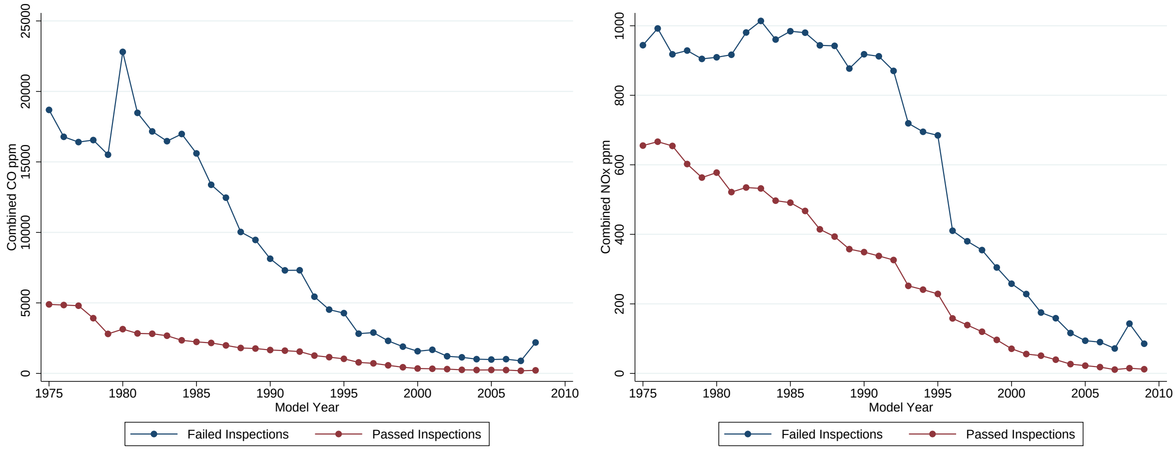
(B) STAR FPR Score



(C) Contemporaneous FPR (C-FPR)

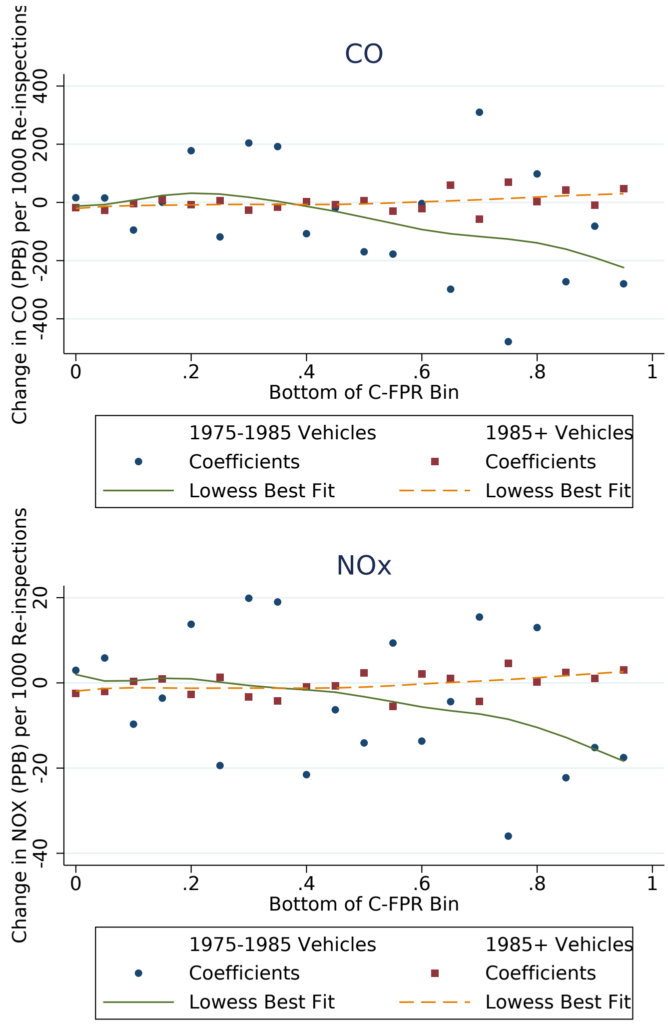
Notes: Graphs show the link between three different possible measures of I/M station quality and future carbon monoxide emissions rates of vehicles at later tests. We obtain the results by regressing changes in emissions levels between the previous and current initial tests (y-axis) on a series of 10 indicators equal to 1 if a tested car falls into one of the 10 equally-spaced bins that cover the distribution of the quality measure, and 0 otherwise. In the case of all three metrics, an increase in the measure should indicate an increase in station quality *as determined by the specific measure.* Panels (A) and (B) show the STAR program measures of station quality. Panel (C) shows our alternate measure of station quality, which uses observed future tests to *ex post* evaluate the effectiveness of a nominal repair to a failing vehicle. We split results by initial inspection outcomes (pass vs. fail).

Figure 2: Average Smog Check Measured Emissions by Model Year



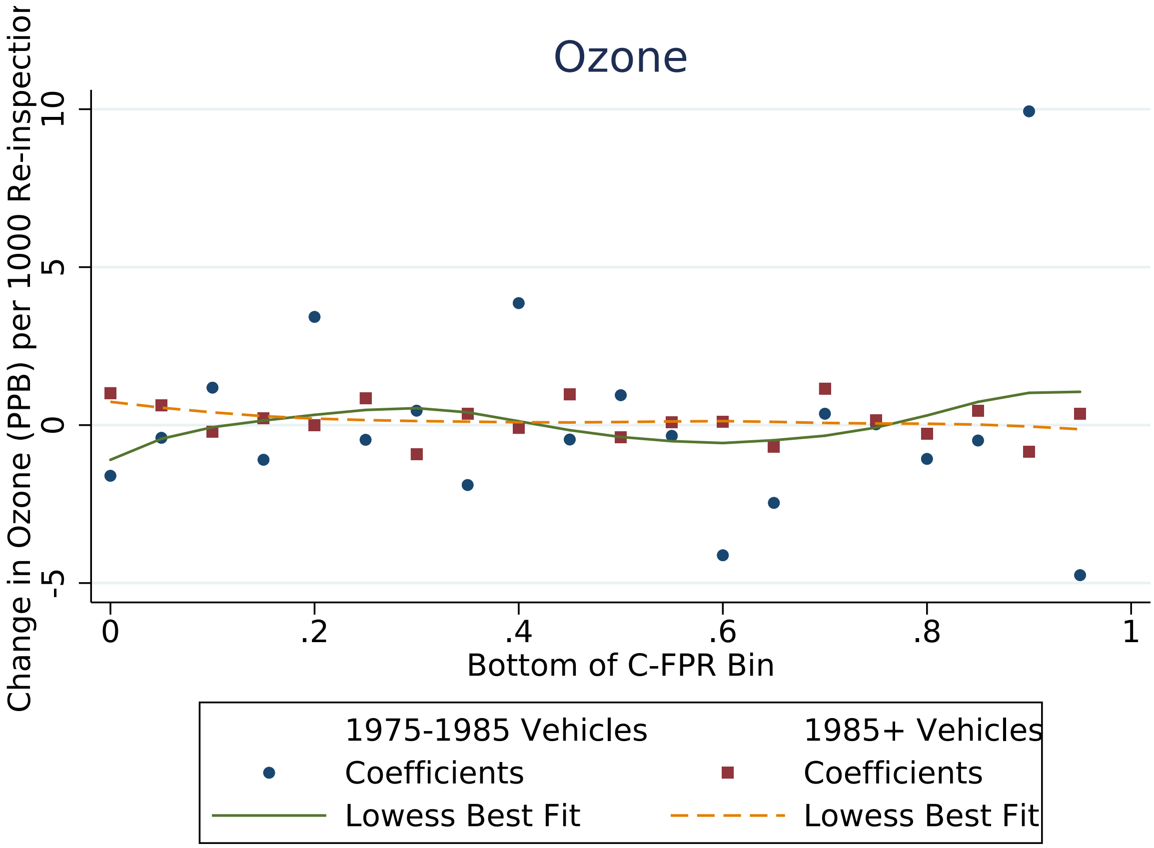
Notes: Graphs show the total combined measured tailpipe emissions at initial test for carbon monoxide (left) and nitrogen oxides (right) from cars that fail (blue line) or pass (red line) initial inspection by model year.

Figure 3: Effect of re-inspections on Daily Air pollution, by FPR Score of Test Stations



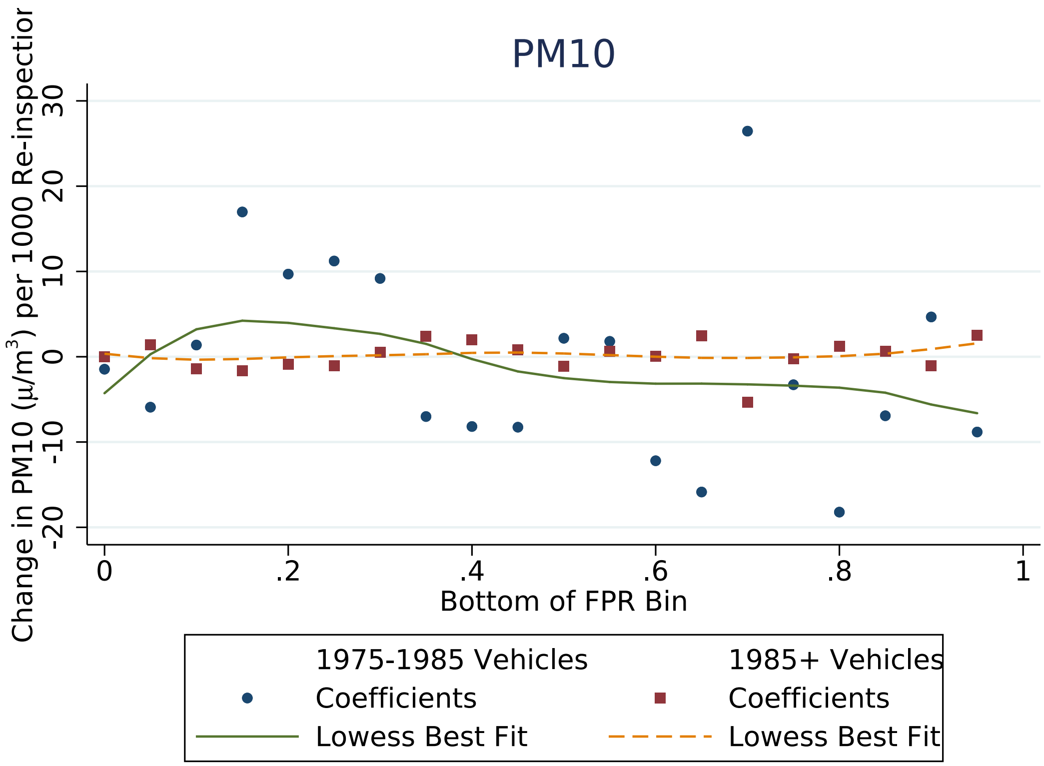
Notes: Graphs show the change in measured county-level air quality for carbon monoxide (CO, top) and nitrogen oxides (NO_x, bottom) resulting from an additional 1,000 nominal repairs (as proxied by a successful re-inspection of an initially failed vehicle) done at stations of respective quality as measured across 20 equally-spaced bins covering the full range of our C-FPR measure (see section 4.1). We split results by older (pre-1985) and newer (1985 and onward) model-year vehicles. LOWESS best-fit lines use a bandwidth of of $N * 0.8$.

Figure 4: Effect of re-inspections on Daily Ozone Levels, by FPR Score of Test Stations



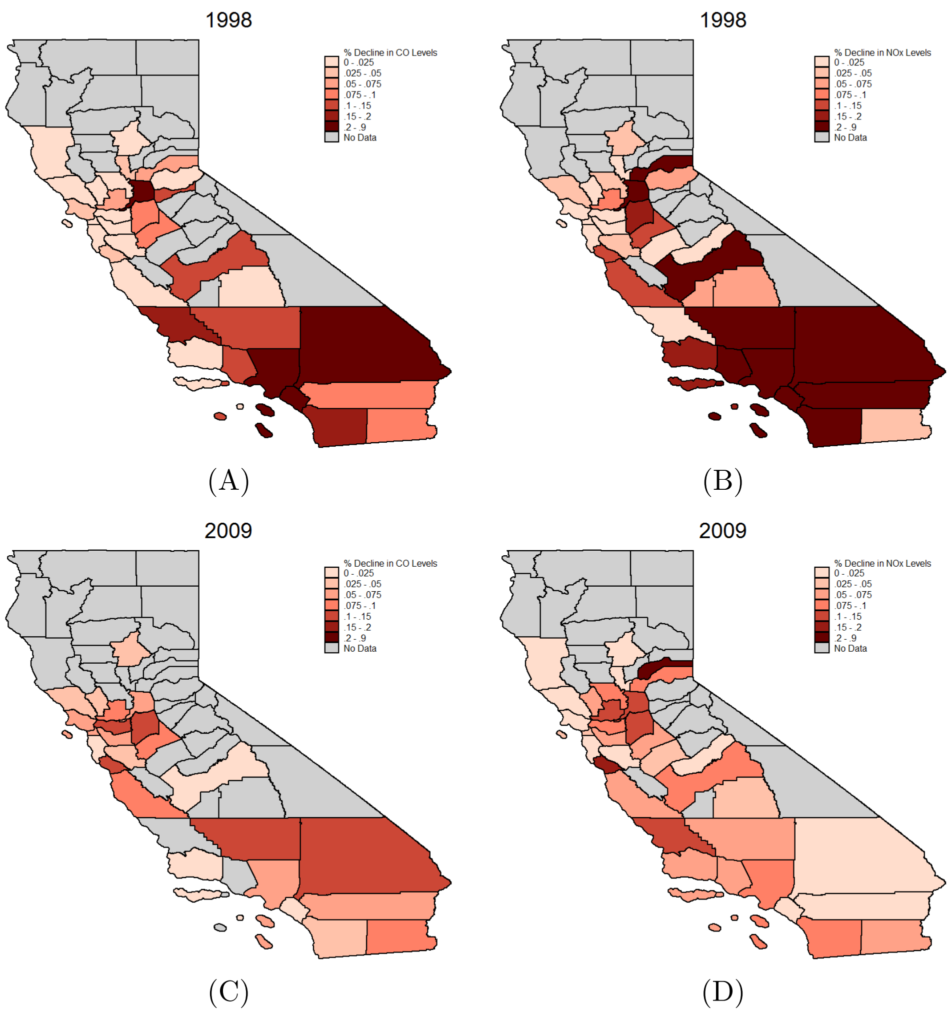
Notes: Graphs show the change in measured county-level air quality for ozone (O3) resulting from an additional nominal 1,000 repairs (as proxied by a successful re-inspection of an initially failed vehicle) done at stations of respective quality as measured across 20 equally-spaced bins covering the full range of our C-FPR measure (see section 4.1). We split results by older (pre-1985) and newer (1985 and onward) model-year vehicles. LOWESS best-fit lines use a bandwidth of of $N * 0.8$.

Figure 5: Effect of re-inspections on Weekly PM10 Levels, by FPR Score of Test Stations



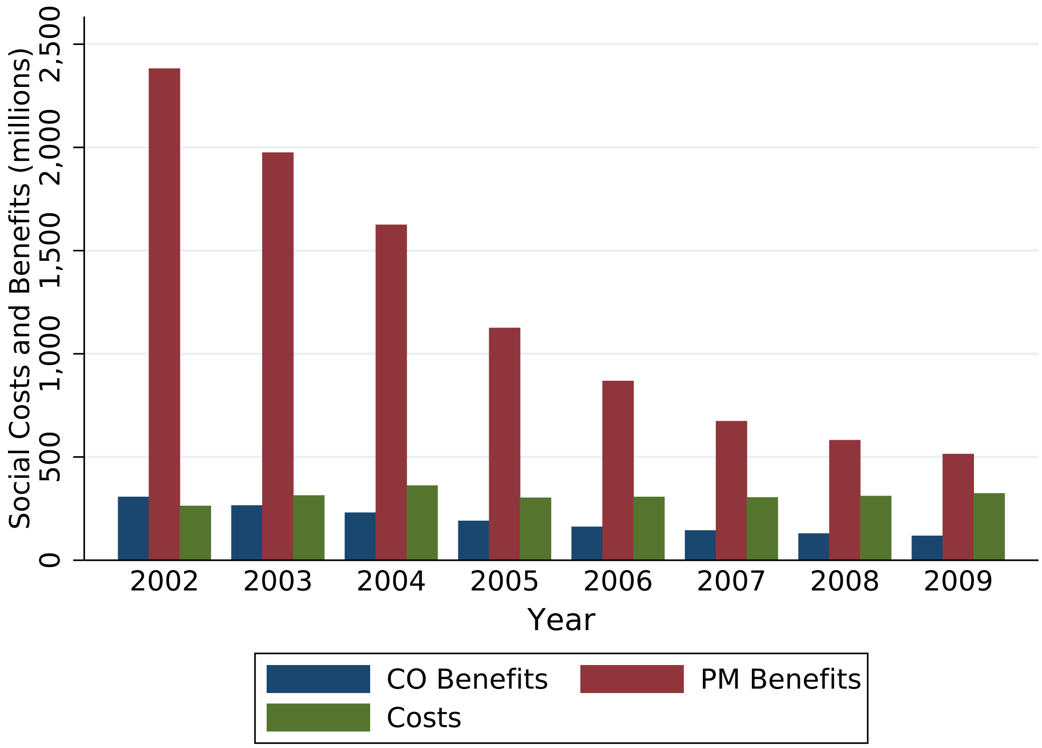
Notes: Graph shows the change in measured county-level air quality resulting from an additional 1,000 nominal repairs (as proxied by a successful re-inspection of an initially failed vehicle) done at stations of respective quality as measured across 20 equally-spaced bins covering the full range of our C-FPR measure (see section 4.1). We split results by older (pre-1985) and newer (1985 and onward) model-year vehicles. LOWESS best-fit lines use a bandwidth of of $N * 0.8$.

Figure 6: Predicted Percent Change in Pollution Levels From Re-inspecting All 1975–1985 vehicles at STAR stations



Notes: Maps show the simulation-based differences in air quality assuming that all 1975-1985 model year cars that failed their initial inspections been re-inspected at stations considered of high quality based on the current STAR metrics. Panels (A) and (C) show simulated differences in ambient CO, while panels (B) and (D) show simulated differences in ambient NOx. Darker shades indicate larger improvements in air quality under our simulation. Gray counties are those for which we have either no emissions data or limited inspection data (i.e. biennial inspections are not required). See section 6 for details of our simulation process.

Figure 7: Costs and Benefits of the Smog Check Program, by Year



Notes: Graph shows the estimated costs and benefits of the Smog Check program, in millions of dollars, across time. We derive costs using estimated cost-of-test and cost-of-time measures, which generate a total cost of $27.86 per test. We derive benefits using estimates of pollution on mortality from the economics literature. See section 6.2 for calculation details.

A Appendix

Table A1: Re-Inspections and County-Level Daily Air Quality: Robustness Checks

	A: Outcome is Carbon Monoxide (PPB)			
	30 Days	60 Days	90 days	120 days
000s of Re-Inspections in Window				
1975-1985 Vehicles	-44.79***	-24.09***	-14.43***	-8.260***
	(10.57)	(5.995)	(3.972)	(2.947)
1985+ Vehicles	-11.73**	-7.259**	-4.226*	-1.779
	(5.256)	(3.178)	(2.222)	(1.647)
County FE	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes
Calendar Week FE	Yes	Yes	Yes	Yes
County Time Trends	Yes	Yes	Yes	Yes
	B: Outcome is NOx (PPB)			
	30 Days	60 Days	90 days	120 days
000s of Re-Inspections in Window				
1975-1985 Vehicles	-2.243***	-1.308***	-0.680**	-0.206
	(0.667)	(0.397)	(0.263)	(0.179)
1985+ Vehicles	-1.860***	-1.257***	-0.788***	-0.423***
	(0.365)	(0.225)	(0.148)	(0.0981)
County FE	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes
Calendar Week FE	Yes	Yes	Yes	Yes
County Time Trends	Yes	Yes	Yes	Yes
Observations	143440	143440	143440	143440

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Observations are county-days. Standard errors clustered by county reported in parentheses.

Table A2: Re-Inspections and County-Level Daily Air Quality: Robustness Checks

	CO	NO_x
000s of Re-Inspections Last 90 Days		
1975–1985 Vehicles	-15.41***	-0.737**
	(4.001)	(0.277)
1985-1995 Vehicles	-3.059	-0.720***
	(2.425)	(0.207)
1995+ Vehicles	-6.377*	-0.910***
	(3.370)	(0.210)
County FE	Yes	Yes
Weather Controls	Yes	Yes
Calendar Week FE	Yes	Yes
County Time Trends	Yes	Yes
Observations	129260	143440

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table A3: Re-Inspections and County-Level Daily Air Quality: Robustness Checks

	A: Outcome is Carbon Monoxide (PPB)				
	(1)	(2)	(3)	(4)	(5)
000s of Re-Inspections Last 90 Days					
1975-1985 Vehicles	-14.43***	-28.22	-14.42***	-9.600**	-15.56***
	(3.972)	(18.19)	(3.971)	(4.562)	(4.223)
1985+ Vehicles	-4.226*	-2.986	-4.205*	-4.068	-5.135**
	(2.222)	(5.920)	(2.223)	(2.443)	(2.301)
County FE	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
County Time Trends	Yes	Yes	Yes	No	No
Exclude LA	No	Yes	No	No	No
Month-of-Year and Day-of Week FE	No	No	Yes	No	No
County-Year FE	No	No	No	Yes	No
Weekly Data	No	No	No	No	Yes
	B: Outcome is NOx (PPB)				
	(1)	(2)	(3)	(4)	(5)
000s of Re-Inspections Last 90 Days					
1975-1985 Vehicles	-0.680**	-0.951	-0.678**	-0.258	-0.760***
	(0.263)	(1.010)	(0.262)	(0.209)	(0.255)
1985+ Vehicles	-0.788***	-0.716**	-0.786***	-0.801***	-0.839***
	(0.148)	(0.328)	(0.148)	(0.211)	(0.143)
County FE	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
County Time Trends	Yes	Yes	Yes	No	No
Exclude LA	No	Yes	No	No	No
Month-of-Year and Day-of Week FE	No	No	Yes	No	No
County-Year FE	No	No	No	Yes	No
Weekly Data	No	No	No	No	Yes
Observations	143440	139057	143440	143440	20505

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Observations are county-days. Standard errors clustered by county reported in parentheses.

Table A4: Station Quality and County-Level Ozone

	(1)	(2)
Low NO_x Days		
1975-1985 Vehicles	0.205	0.335
	(0.176)	(0.470)
1975-1985 Vehicles · C-FPR		-0.670
		(1.219)
1985+ Vehicles Vehicles	0.211***	0.202
	(0.0624)	(0.233)
1985+ Vehicles · C-FPR		0.106
		(0.458)
High NO_x Days		
1975-1985 Vehicles	-0.188**	-0.553***
	(0.0709)	(0.149)
1975-1985 Vehicles · C-FPR		0.556
		(0.427)
1985+ Vehicles Vehicles	0.0657	0.411***
	(0.0412)	(0.130)
1985+ Vehicles · C-FPR		-0.770***
		(0.267)
Observations	143049	143027

Note: Observations are county-days. All regressions control for daily weather, county fixed effects, calendar week fixed effects, and county-specific cubic time trends. Low and high NO_x days are days with NO_x levels below (above) the median NO_x level of 18 ppb. Standard errors clusted by county reported in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

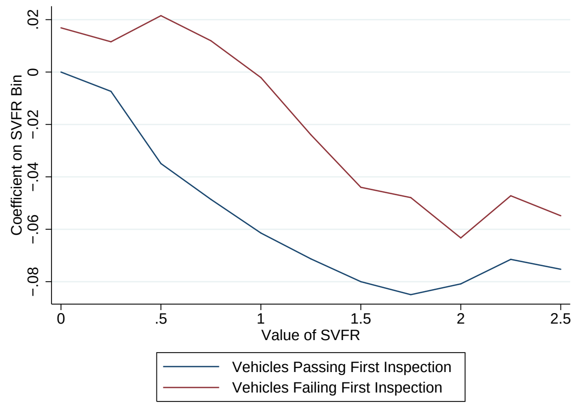
Table A5: Emissions levels by Inspection Reason

	(1) Ln(HC)	(2) Ln(NO_x)	(3) Ln(CO)	(4) P(Fail)
Change of Ownership				
Model Years 1976-1985	0.0304***	-0.00708**	0.0270***	-0.00610***
	(0.00159)	(0.00258)	(0.00285)	(0.000557)
Model Years 1985+	0.0704***	0.0297***	0.0617***	0.00176***
	(0.000696)	(0.00127)	(0.000790)	(0.000197)
Initial Registration				
Model Years 1976-1985	0.0473***	-0.101***	0.0465***	-0.00195
	(0.00259)	(0.00395)	(0.00488)	(0.00104)
Model Years 1985+	0.201***	0.171***	0.202***	0.0149***
	(0.00212)	(0.00354)	(0.00243)	(0.000254)
Odometer (00000s)	0.340***	0.480***	0.378***	0.0628***
	(1.05e-08)	(2.67e-08)	(1.27e-08)	(3.25e-10)
Years Since Last Inspection	0.0109***	0.0135***	0.0213***	0.0112***
	(0.000270)	(0.000458)	(0.000369)	(0.0000982)
ASM Test	0.392***	0.253***	0.478***	0.0339***
	(0.00337)	(0.00425)	(0.00474)	(0.000519)
Vehicle Age FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
VIN Prefix FE	Yes	Yes	Yes	Yes
N	124575462	95543435	124575462	124586809

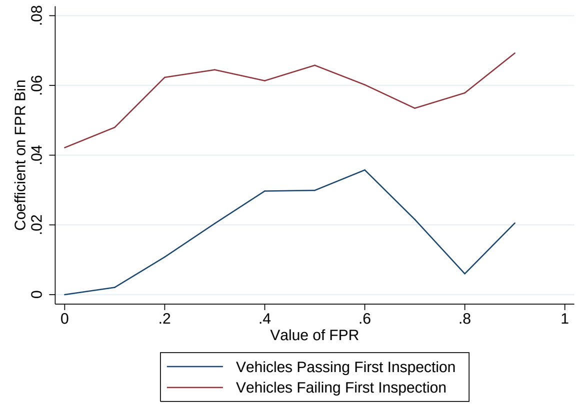
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

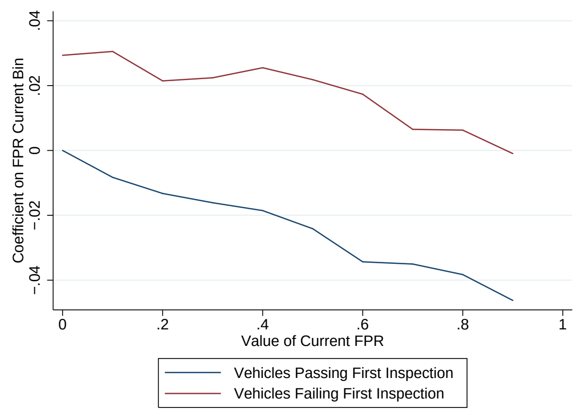
Figure A1: STAR Metrics of Re-inspection Station and Log NO_x Emissions at Next Inspection



(A) STAR SVFR Score



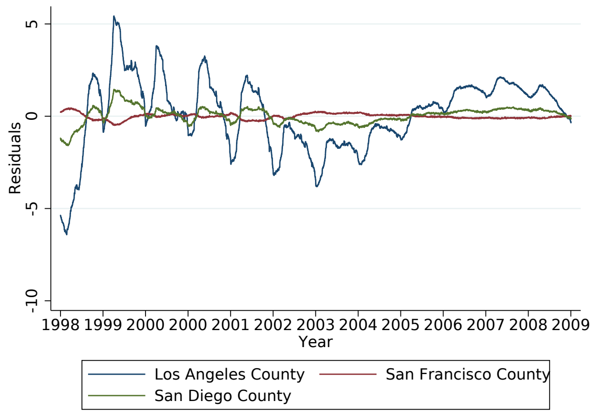
(B) STAR FPR Score



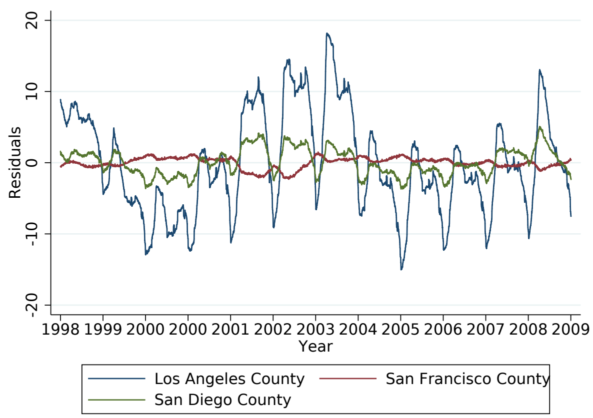
(C) Contemporaneous FPR (C-FPR)

Notes: Graphs show the link between three different possible measures of I/M station quality and future NO_xemissions rates of vehicles at later tests. We obtain the results by regressing changes in emissions levels between the previous and current initial tests (y-axis) on a series of 10 indicators equal to 1 if a tested car falls into one of the 10 equally-spaced bins that cover the distribution of the quality measure, and 0 otherwise. In the case of all three metrics, an increase in the measure should indicate an increase in station quality *as determined by the specific measure*. Panels (A) and (B) show the STAR program measures of station quality. Panel (C) shows our alternate measure of station quality, which uses observed future tests to *ex post* evaluate the effectiveness of a stated repair to a failing vehicle. We split results by initial inspection outcomes (pass vs. fail).

Figure A2: Residual variation in 90-day rolling totals of re-inspections after controlling for covariates in preferred regression specification.



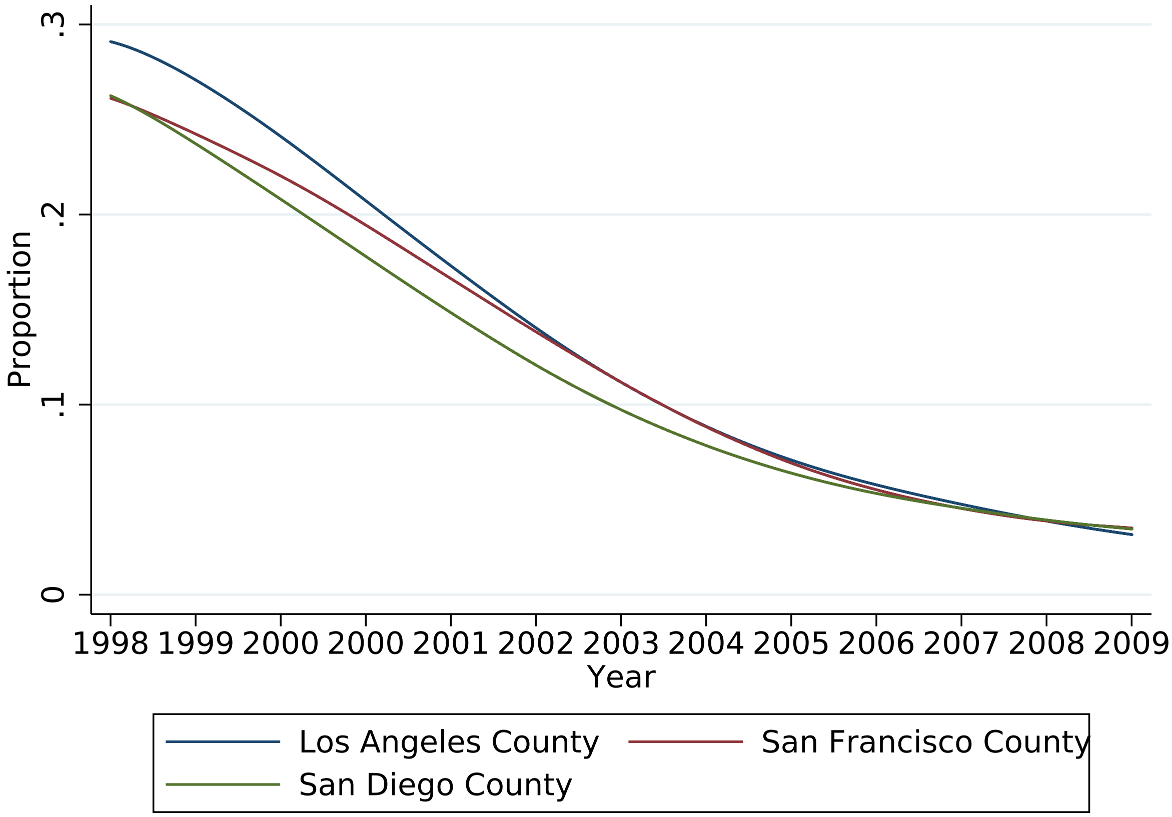
(A) 1975–1985 Model Year Vehicles



(B) 1985+ Model Year Vehicles

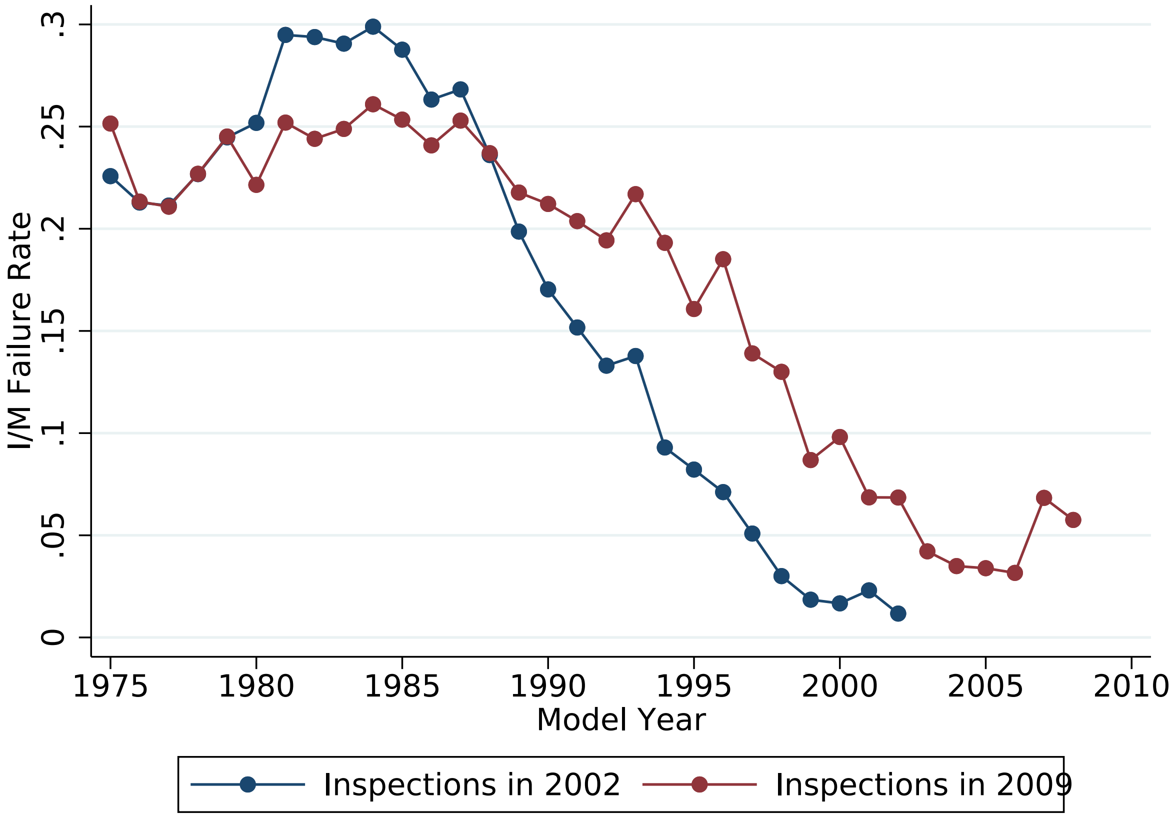
Notes: Residual results split by county.

Figure A3: Share of Daily Total Re-Inspections that are 1975–1985 Model Year (LOWESS Smoothed)



Notes: Graph shows the share of re-inspections, by year and region, that are for cars of the 1975—1985 model year range. Vertical axis represents share of total (e.g., 0.1 is 10%).

Figure A4: I/M Emissions Inspection Failure Rates By Model Year, 2002 and 2009



Notes: Lines show the initial failure rate for inspections done in the year 2002 (blue line) and the year 2009 (red line). Vertical axis shows the initial failure rate, while the horizontal axis shows the relevant model year of the tested vehicle.