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

Nicholas J. Sanders, Ryan Sandler

Abstract: 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. To test for a link between I/M and local air pollution levels, 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 certified after failing an initial emissions inspection as a proxy for emissions-related repairs. Additional passed reinspections of older vehicles with inferior emissions technology (pre-1985 model year) reduce local carbon monoxide, nitrogen oxide, and particulate matter levels, but passed reinspections of newer vehicles with more modern engine technology have no economically significant effect on air pollution. This suggests that biannual emissions inspections as currently implemented will play less of a role in reducing local air pollution as polluting vehicles from the 1970s and 1980s leave the road.

JEL Codes: Q52, Q53, Q58

Keywords: pollution, emissions testing, regulation

DESPITE REGULATORY ADVANCEMENTS and improvements in engine technology, motor vehicles remain responsible for 75% of carbon monoxide (CO) emissions in the

Nicholas J. Sanders is at Cornell University and the National Bureau of Economic Research (njsanders@cornell.edu). Ryan Sandler (corresponding author) is at the Consumer Financial Protection Bureau (ryan.sandler@cfpb.gov). We are grateful to Dan Hosken, Mark Jacobsen, Thomas Koch, Devesh Raval, Joseph Shapiro, and participants in the Economics brown bag at the Federal Trade Commission, the ASSA annual meeting, the University of Chicago Public Policy and Economics Workshop, the University of Illinois at Urbana-Champaign, and Miami University, for helpful comments and suggestions.
Dataverse data: https://doi.org/10.7910/DVN/VZLUG4

Received December 14, 2018; Accepted December 23, 2019;

JAERE, volume 7, number 3. 2333-5955/2020/0703-0006$10.00 https://doi.org/10.1086/707954

United States and over 50% of nitrogen oxide (NO_x) emissions.[1] 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 disproportionately impact low-income households and decrease average road safety (Jacobsen 2013a, 2013b). Driving restriction programs have varied success rates for 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 such programs have substantial problems with adverse selection and may only slightly shift forward the timing of vehicle replacement (Mian and Sufi 2012; Sandler 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 to governments and individuals (Ando et al. 2000) and subject to potential fraud (Oliva 2015). 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, where vehicles recorded as passing government tests actually produced emissions far above allowable levels on the road.[2]

We provide an analysis of the effect of vehicle inspections on local air pollution, using extensive vehicle inspection data from the state of California.[3] We find that California's I/M program, "Smog Check," helps reduce contemporaneous air pollution. However, we only find economically meaningful results from reinspections of older (1985 and prior model year) vehicles that predate advances in engine emissions controls, suggesting that the benefits of inspection-based 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 in an attempt to reduce instances of fraud and ineffective repairs. We find that increasing station quality may help further reduce pollution but, again, only through inspections and repairs of older cars that make up a shrinking share of the automobiles on the road.

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 reinspections following a failed inspection to capture the intensity of I/M-related

---

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

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

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

vehicle repairs. Controlling for local weather effects and a battery of time and region fixed effects, we find that an additional 1,000 passing reinspections of vehicles with older emissions technology decreases ambient CO by 19 parts per billion (ppb) and ambient NO_x by 1.1 ppb, about 5% of a standard deviation for each pollutant. For scale, the average California county passes 1,000 initially failing vehicles of all ages every 12 days. Passed reinspections of vehicles manufactured after 1985 have much smaller effects on air pollution. One possible explanation for our lack of effects is that tests and repairs are poorly executed.

California recently passed substantial reforms of inspection station requirements, hoping to improve the reliability of Smog Check inspections (Bureau of Automotive Repair 2014). Under the new "STAR" system, inspection and repair providers must meet certain quality criteria before the state certifies them to inspect the most high-polluting vehicles.[4] Understanding how inspection station quality impacts air pollution is important given prior findings that I/M programs are subject to gaming (Oliva 2015). Testing the effectiveness of such a quality rating system is challenging due to confounding factors, including strategic customer and station responses to the rating system. To avoid such problems, we construct hypothetical STAR program measures of station quality before the announcement of the policy and test the relationship between ambient air pollution and passed reinspections at what would be considered high-scoring stations under the eventual STAR metrics. We find that passed reinspections of older vehicles at high-quality stations reduce airborne levels of CO and NO_x while passed reinspections at low-quality stations yield no change in local air pollution levels regardless of vehicle age. This result is consistent with low-quality stations allowing vehicles to pass reinspection without appropriate repairs, or that cars passed at such stations received lower-quality, temporary repairs. Much like our findings on the general results of I/M programs, reinspections of newer cars have little impact on air pollution, regardless of station quality. Together, these empirical results suggest that I/M programs like Smog Check may fail to further reduce pollution in the long term, even with improvements to the quality of inspection stations. Our finding that passing reinspections at low-quality inspection stations do nothing to improve air pollution supports prior work showing that gaming is prevalent in I/M programs and that regulation is ineffective at improving environmental quality when enforcement is weak. In the context of I/M programs, Oliva (2015) shows that corruption can be substantial.[5]

---

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

[5] More generally, 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

While the direction of our results is largely robust to different specifications, there are important sensitivities. First, the selection of time controls is important. Our main results use county-by-year fixed effects. Our magnitudes change if we instead opt for more restrictive regional temporal controls, such as linear county trends. Further, a failure to control for any regional time effects finds a positive relationship between passed reinspections and pollution levels, a factor we believe is the result of regional differences in vehicle makeup across time. Alternate model designs (e.g., focusing on inspections per urban square mile rather than total inspections), while largely in agreement with our main results, are noisy. Finally, we explore our results in an event study framework and show that much of the pollution improvement from passed reinspections of older cars is limited to a short time frame. This suggests that there may be greater returns to programs designed to promote less transitory repairs.

As an expansion, in the appendix (available online) we conduct various simulations of air quality changes in California absent the Smog Check program, as well as explore the efficiency of the program and how it has changed over time. Our results suggest that the program as implemented had large returns to both air quality and social welfare but that gains fade with each year while costs remain largely constant.

We next outline the California Smog Check program and the new STAR system in section 1. Section 2 then describes the Smog Check and pollution data. Section 3 describes our identification technique, and section 4 presents estimates of how the Smog Check program changed local pollution levels. Section 5 briefly outlines some of our simulation models, and section 6 concludes. Additional related results may be found in the appendix.[6]

1. BACKGROUND ON CALIFORNIA'S EMISSIONS TESTING PROGRAM

California provides an excellent backdrop for the study of tailpipe emissions inspection programs. Of the approximately 110 million registered automobiles in the United

---

"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 that 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.

[6] Appendix sec. A.1 describes our methods for estimating STAR quality measures. Appendix sec. A.2 discusses various robustness checks on our main results. Appendix sec. A.3 describes the methodology for our simulations. Appendix sec. A.4 uses our results to simulate the impact of Smog Checks on social welfare.

States in 2012, almost 13 million were in California, more than any other single state.[7] 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 2 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 "reinspection" before registering it, thereby removing high-polluting vehicles by inducing repairs or forcing irreparable vehicles off the road. While vehicle inspections in some other states include a safety component, the California system is only about emissions, so failures always indicate above-acceptable levels of emissions. Emissions thresholds vary by the model year of the vehicle in question but within vehicle are consistent over time and across the state.[8] There is also a group of exempt vehicles: vehicles of 1975 model year or older (which represent less than 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 that the first incarnation of the Smog Check program was rife with problems that decreased or eliminated expected 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

---

[7] Bureau of Transportation Statistics, 2014 State Transportation Statistics data, table 5-1. 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.

[8] 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 depend only 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.

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. While the program refers to such vehicles as being "directed," vehicle owners are still able to choose which station inspects their vehicle—directed status simply limits the set of eligible stations. 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 a higher "gross polluter" threshold. The BAR also directs a 2% random sample of all vehicles registered in enhanced areas. In all, 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 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 reinspection at a Smog Check station.[9] A potential implication of the discrepancy is that these cars did not truly pass the reinspections: someone had instead manipulated the testing outcome.[10] The discrepancies were equally common for test-and-repair stations and test-only stations, undermining the logic for sending directed vehicles to the latter.

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 inspections 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

---

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

[10] 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.

the BAR. We use a variation on one of these metrics in our analysis; see appendix section A.1 for details. 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 had to be inspected at qualified STAR stations as of January 1, 2013.

2. DATA

To measure the volume of reinspections 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.[11] While our analysis focuses on the period from 1998 to 2009, our data consist of the population of vehicle inspections conducted in California between 1996 and 2012 and transmitted through the ETS. We use tests from 1996–97 and 2010–12 when constructing leads and lags, and for our estimates of quality measures under the new STAR program.[12]

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 six-digit station identifier, which we join to a crosswalk giving the zip code of each station.[13] We determine model year and vehicle type from the included VIN.[14] We also utilize the provided odometer reading in calculating our station quality scores.[15]

We use pollution data from the California Air Resources Board (CARB) Air Quality Database, a collection of air monitors taking regular pollution readings. We use data from 1998 to 2009, and focus on CO, NO_x, ozone (O_3), and particulate matter less

---

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

[12] Data from 1996 and 1997 are incomplete, as the BAR was phasing in the ETS during these years.

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

[14] Although the Smog Check data contain some direct information on vehicle types, it is at times unreliable when compared to known VIN information. All vehicles manufactured after 1980 have a standardized 17-digit VIN: the first eight digits plus the tenth and eleventh precisely indicate the vehicle type, at the level of make/model/engine/body type/transmission/year/plant. For earlier vehicles, different manufacturers used their own formats. We determine make, model year, and an approximation of the vehicle-identifying prefix for most of the vehicles manufactured 1975–80.

[15] 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 in the data archive for Sandler (2012).

than 10 micrometers (PM_{10}). We aggregate hourly readings for CO, NO_x (including both nitric oxide, NO, and nitrogen dioxide, NO_2), and O_x to the county-day level by averaging individual monitor readings in a given county, and aggregate daily PM_{10} readings to the weekly level (most PM_{10} monitors take measurements once every 6 days). We do not weight monitors by any distance metric and use an unbalanced panel of monitors to maximize available data (balanced panel results are similar). To improve readability of our results, we scale pollution readings for CO, NO_x and O_3 to parts per billion. Our PM_{10} data give the concentration of particles in units of micrograms per cubic meter (μ/m^3).

Of the pollutants we study, only CO and NO_x are emitted directly by gasoline-powered motor vehicles in significant quantities. O_3, 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 O_3, which makes O_3 a difficult pollutant to analyze on a large scale. Regardless, a primary interest of the Smog Check program was to reduce smog, so we test for general impacts on O_3. The link between PM_{10} and automobile use is also largely secondary. The largest direct sources of PM_{10} 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 did not require Smog Checks for diesel vehicles during our sample period, 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_{10}, we include this pollutant as well.[16]

Local weather influences both general air pollution and the contribution of automobile emissions to said air pollution (Knittel et al. 2016). Ambient temperature can also influence inspection results, as emissions control systems generally perform worse when ambient temperatures are cold (Spindt et al. 1979). We control for daily high and low temperatures and daily precipitation at the county level, 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 the geographic information system (GIS), and using historical daily averages to interpolate a measure of localized daily weather.[17] Weather fluctuations should be exogenous to inspection timing, so inclusion does little to change our primary estimates.

---

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

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

3. EMPIRICAL METHODOLOGY

We leverage random variation in the timing and location of repairs of vehicles that fail an emissions inspection. As we cannot observe actual repairs, we focus on the timing of final reinspections following a failed inspection. A final passing reinspection theoretically indicates that a repair took place to reduce a vehicle's emissions below legal thresholds. Thus, we use final reinspections in an inspection cycle as a proxy for repairs. This is not a perfect measure: a final reinspection is a necessary but not sufficient condition for a repair to occur. There will be cases where the final reinspection is passed via owner or station fraud, rather than true repair. This drives our interest in measuring station quality, as final reinspections at high-quality stations are more likely to proxy for a real repair.

It is not possible to estimate the impact of I/M programs in the United States, like Smog Check, using a difference-in-differences or before-after methodology. 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 implementing the Smog Check program, exceeding federal air pollution standards may trigger additional policy responses. For instance, policy makers 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 the effects of Smog Check with the effects of state- and nationwide policy changes, particularly emissions standards on new vehicles.

To avoid such issues, we estimate the effect of Smog Check 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 of failing status but incorrectly passes on an initial inspection, this will not change air pollution levels—emissions will exceed passing levels before and after inspection. However, local air pollution levels will improve if stations conduct repairs as a result of a failed inspection, and correctly verify the effectiveness of the repair by reinspecting the failing vehicle. Station quality is thus important, as a sham initial inspection, or an omitted repair followed by a sham reinspection, will have no more effect on air pollution than no inspection at all.

Inspection timing is exogenous to both levels and 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 2 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. 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 reinspection in the event of failure. None of these sources of variation should correlate with levels of air pollution, except possibly through seasonality, which we control for using fixed effects.[18]

We create a daily panel of reinspection counts aggregated by the county of the station conducting each inspection. Choosing overly fine geography risks attributing inspections to the wrong location and ignoring spillover effects from pollutants blown to neighboring areas, while broad geographic definitions reduce sample size and obscure important variation. We aggregate to the county level as a compromise between these considerations. 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% of California workers live and work in the same county.[19] The EPA determines Clean Air Act attainment status at the county level, making it also a policy-relevant level of aggregation.

Associating air pollution levels on a specific day with reinspections on the same day is problematic due to measurement error. Pollutants like NO_x and CO persist in the air and take time to drift to sensors, and there may be a period of days between a repair and the date the station records the reinspection. We instead use a 90-day rolling total of reinspections as our measure of the intensity of I/M activity. Using a rolling total of passed reinspections to capture recent I/M-related repairs potentially creates its own measurement error to the extent that the effects of repairs are short lived. If repairs deteriorate quickly, our rolling total could conflate recent, still effective repairs with older, now-ineffective repairs, biasing coefficients toward zero. The importance of this source of measurement error depends on whether or not our rolling total window length spans the time it takes for a repair to fully deteriorate. In practice, we obtain qualitatively similar results using a 30-, 60-, or 120-day window (see appendix table A1; tables A1–A8 are available online), suggesting that the effect of repairs is sufficiently persistent to avoid biasing our coefficients.

---

[18] The exogenous nature of timing hinges on owners following the prescribed inspection schedule. If the majority of tests happen outside of the expected timing window, owners may be acting strategically in their choice of timing for Smog Checks. This is problematic if these choices correlate with air quality in ways fixed effects cannot capture. Appendix fig. A1 shows the distribution of timing between Smog Checks for all vehicles in our data located in counties that require biennial inspections and makes clear that the overwhelming majority of Smog Checks occur right around the expected 2-year window.

[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.

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 reinspections of older versus newer vehicles. This requires that 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 on-board monitoring computers (called OBDII), introduced in 1996. Each of these technologies was 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 division for old versus new vehicles, we plot average CO and NO_x emissions measured at passed and failed Smog Checks by model year.[20] Figure 1 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 reinspections of the 1985–95 and 1995+ vehicles, with no change in the effect of 1975–85 vehicles, indicating that the split at 1985 is appropriate (see appendix table A2).

We estimate the effect of passed reinspections (our proxy for repairs) on levels of a pollutant $p \in \{NO_x, CO, O_3, PM_{10}\}$ in county c in time t as:

$$p_{ct} = \left(\sum_{i=0}^{90} R^{\text{old}}_{c(t-i)}\right)\beta_1 + \left(\sum_{i=0}^{90} R^{\text{new}}_{c(t-i)}\right)\beta_2 + \gamma X_{ct} + \varepsilon_{ct}, \tag{1}$$

where R^{old}_{ct} and R^{new}_{ct} denote the number of reinspections of older and newer vehicles, respectively, X_{ct} is a vector of county-level covariates, and ε_{ct} is an error term. We specify passed reinspections in levels. Holding weather, county geography, and related factors constant via fixed effects, to a first approximation repairing and passing 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 reinspections. 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

---

[20] Recall that O_3 and PM_{10} are not tailpipe pollutants and as such are not measured directly by Smog Check inspections.

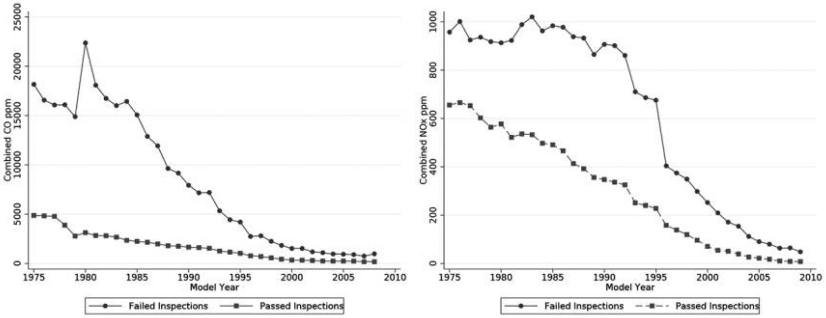

Figure 1. Average Smog Check measured emissions by model year. Graphs show the total combined measured tailpipe emissions at initial test for carbon monoxide (*left*) and nitrogen oxides (*right*) from cars that fail (*circles*) or pass (*squares*) initial inspection by model year. Color version available as an online enhancement.

causes repairs for a large numbers of high-polluting vehicles.[21] The coefficients β_1 and β_2 give the causal effect of additional passed reinspections on air pollution levels. In our preferred specification, X_{ct} includes controls for weather, county fixed effects, statewide calendar-week fixed effects to flexibly control for changes across time, and county-specific year fixed effects. With so many fixed effects, one concern is how much variation remains for identification—appendix figure A2 (figs. A1–A10 are available online) shows the residual variation of the 90-day rolling totals of reinspections for three large urban counties after controlling for the covariates from our preferred specification and demonstrates that there is still substantial high-frequency variation.

4. RESULTS

Panel A of table 1 summarizes pollution and Smog Check variables of interest in our county-day sample. For consistency with our regression results, we show NO_x, O_3, 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.

---

[21] Our results are robust to excluding Los Angeles County. Even if all detectable results existed only in Los Angeles County, the social gains of decreased air pollution would be large; the county has a population of more than 10 million, larger than many European countries and all but the 10 largest US states. For results without Los Angeles County, see appendix table A3, col. 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.

[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.

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)	No. of Inspections	No. of Reinspections
A. All California						
1998	703.0	28.89	28.04	25.50	588.4	39.78
1999	716.2	32.40	28.32	31.00	678.7	54.39
2000	649.6	30.19	26.65	27.16	706.3	61.51
2001	612.1	27.68	27.60	27.79	755.2	69.33
2002	600.8	28.18	29.04	30.10	719.6	75.55
2003	562.6	26.36	28.69	26.78	825.0	86.86
2004	499.6	24.20	27.66	26.38	795.7	82.87
2005	458.4	24.23	26.60	24.05	669.7	71.08
2006	466.8	23.41	28.19	26.14	678.3	66.13
2007	431.7	22.05	27.61	25.90	667.6	63.19
2008	422.3	20.38	28.93	26.75	671.6	64.84
2009	363.5	18.03	27.66	22.18	688.5	67.26
Average	551.9	25.57	27.89	26.61	701.6	66.45
B. Los Angeles County						
1998	1254.5	71.45	21.73		7072.1	526.0
1999	1203.6	79.36	20.78		8324.9	784.2
2000	1029.1	70.98	20.73		8489.9	855.3
2001	936.9	64.25	21.58		9015.2	993.9
2002	869.9	60.89	24.30	37.60	9255.5	1084.6
2003	814.0	58.35	25.27	32.40	9328.7	1127.6
2004	663.9	51.49	26.49	32.41	9418.5	1103.0
2005	580.2	47.09	24.48	30.45	7838.1	928.0
2006	548.8	47.35	25.48	29.98	7880.1	868.2
2007	506.6	43.86	25.05	33.32	7644.9	815.8
2008	470.3	40.49	25.93	30.66	7600.5	809.3
2009	417.4	35.99	26.59	29.86	7781.6	840.2
Average	774.6	55.96	24.04	32.06	8304.3	894.7
C. San Francisco Bay Area						
1998	756.8	33.81	19.76	18.00	5092.4	255.3
1999	769.7	37.48	19.84	21.40	5574.5	291.2
2000	699.7	35.02	18.41	18.45	5946.3	356.7
2001	636.5	31.51	19.86	20.16	6224.7	387.6
2002	599.5	31.08	20.61	21.67	865.1	47.01
2003	569.4	27.56	20.41	16.55	3849.2	282.7
2004	502.9	25.80	20.43	17.32	6820.2	658.4
2005	486.7	25.51	20.21	16.31	5731.4	573.3
2006	466.9	24.93	21.90	18.34	5867.2	512.3

Table 1 (*Continued*)

	CO (ppb)	NO_x (ppb)	Ozone (ppb)	PM_{10} (μ/m^3)	No. of Inspections	No. of Reinspections
2007	418.2	23.18	21.68	16.47	5795.7	481.5
2008	387.5	21.82	22.80	17.44	5820.3	487.5
2009	350.4	21.08	21.24	14.21	5960.8	506.7
Average	553.7	28.23	20.60	18.07	5296.3	403.4

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

O_3 was largely unchanged, remaining within 0.5 ppb of initial 1998 levels. The number of personal automobiles on the road in California increased by almost 1.5 million vehicles across this period, so no change in O_3 levels does not necessarily mean the Smog Check program had no impact on counterfactual O_3 levels. PM_{10} levels fluctuate from year to year with no obvious pattern.

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). NO_x decreased around 40 ppb, just over 50% of 1998 levels, while O_3 levels increased. For northern California, panel C shows averages for the nine 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 O_3 and PM_{10}.

Two general trends appear in the Smog Check summary statistics. First, increases in inspections (and reinspections) correlate with decreases in both CO and NO_x but not with changes in O_3. Second, neither total inspections nor reinspections increase monotonically with time. Total inspections peak in 2004 and then decrease, possibly due to a 2005 Smog Check policy change that exempted pre-1975 vehicles from inspections (state-level week fixed effects control for changes of this nature). In most of the state, reinspections decline around 2002, corresponding to a decrease in the overall failure rate (and thus the need for reinspection/repair). As a caveat, we note that 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 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

---

[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.

stations in the region were upgrading the machines used for Smog Checks. Our results are robust to inclusion or omission of this period for these counties.

4.1. Effect of Reinspections on Primary Pollutants

To test for a link between passed reinspections (our proxy for repairs) and local air pollution, we estimate a series of regression models based on equation (1). The coefficients of interest are the effects of passed reinspections of initially failing cars on local air

Table 2. Reinspections and County-Level Daily Air Quality

	(1)	(2)	(3)	(4)
	A. Outcome Is CO (ppb)			
Thousands of reinspections last 90 days:				
1975–85 vehicles	49.75***	15.11***	−21.34**	−19.00***
	(11.94)	(2.483)	(7.851)	(6.621)
1985+ vehicles	−3.026**	−6.731***	−7.805**	−3.516
	(1.285)	(1.473)	(3.134)	(2.481)
County fixed effects	No	Yes	Yes	Yes
Calendar week fixed effects	No	Yes	Yes	Yes
County-year fixed effects	No	No	Yes	Yes
Weather controls	No	No	No	Yes
Observations	129,260	129,260	129,260	129,260
F-test of equality	16.65	69.07	4.252	5.856
$P > F$	.000279	1.70e-09	.0474	.0214
	B. Outcome Is NO_x (ppb)			
Thousands of reinspections last 90 days:				
1975–85 vehicles	2.558***	.794***	−1.272***	−1.149***
	(.427)	(.107)	(.440)	(.369)
1985+ vehicles	.137***	−.394***	−1.204***	−.793***
	(.0373)	(.0745)	(.367)	(.241)
County fixed effects	No	Yes	Yes	Yes
Calendar week fixed effects	No	Yes	Yes	Yes
County-year fixed effects	No	No	Yes	Yes
Weather controls	No	No	No	Yes
Observations	143,440	143,440	143,440	143,440
F-test of equality	33.75	161.0	.120	2.370
$P > F$	.00000138	1.19e-14	.731	.133

Note. Observations are county-days. Standard errors clustered by county reported in parentheses. *F*-tests of equality report the *F*-statistics from testing the null hypothesis that the coefficient on 1975–85 model year vehicles equals the coefficient on 1985+ model year vehicles.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

pollution. Specifically, in table 2, we estimate the link between the number of county-level reinspections and county-level measures of the direct pollutants CO and NO_x. We scale results to show the effect per 1,000 passed reinspections in the past 90 days. Using values from table 1, the average California county sees 1,000 passed reinspections of all vehicle ages every 12 days, while Los Angeles County has an average of just over 1,000 passed reinspections every day. The nine counties making up the San Francisco Bay Area conduct about 1,000 passed reinspections every 2 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 passed reinspections for older cars: an additional 1,000 passed reinspections correlates with a 49.75 ppb increase in CO (0.09 of a standard deviation) and a 2.5 ppb increase in NO_x. When we focus on newer cars, the sign flips for CO to a 3 ppb decrease per additional 1,000 passed reinspections. The sign on NO_x remains positive but is smaller at 0.14 ppb per 1,000 cars. All results are statistically significant at either the 5% or 1% level.

Adding county fixed effects and calendar week fixed effects (col. 2) increases the negative effect of newer car reinspections on CO to 6.7 ppb per 1,000 passed reinspections, but the impact from older cars remains positive. The sign on NO_x is now negative for newer cars, with an additional 1,000 passed reinspections lowering ambient levels by about 0.4 ppb. The signs on the coefficients for older cars flip once we control for county-specific year effects (col. 3). This sign flip suggests that counties with increasing air quality saw decreases in the number of passed reinspections 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 reinspections of older cars over time due to increased scrappage, and thus would see declining air pollution that is not directly related to Smog Check repairs.[24] Finally, adding controls for weather (col. 4) decreases our coefficients slightly in absolute value, but leaves the essential conclusions unchanged.

Using our preferred specification in column 4, reinspecting and passing an additional 1,000 initially failing older cars decreases ambient CO by 19.0 ppb and ambient NO_x by 1.2 ppb (0.05 of a standard deviation for both pollutants). An additional 1,000 newer

---

[24] Appendix fig. A3 illustrates the trends in reinspections of older cars for three large urban counties. Moreover, there are long-term, county-specific trends in pollutant levels that are unlikely to be related to I/M activity. Appendix fig. A4 demonstrates this. The figure plots the LOWESS-smoothed residuals of CO and NO_x levels for four large counties after controlling for county and (state-level) week fixed effects, and shows substantial trends one can only eliminate by adding county-specific time controls. In appendix table A4, we explore less flexible time controls: county-specific linear, quadratic, and cubic trends. In general, we find that linear and quadratic trends are not sufficient to remove the long-run pollution trends in some counties.

cars passing reinspection decreases ambient CO by 3.5 ppb and NO_x by 0.8 ppb (0.01 and 0.03 of a standard deviation, respectively).

While reinspections immediately before a given day's pollution reading should be associated with decreases in pollution if our assumptions hold, we should not see decreases in pollution coming from passed reinspections further back in time, nor from passed reinspections that will occur in the future. Figure 2 shows coefficients from a regression of pollution levels on in 90-day groupings for lags and leads of reinspections. For older cars, we continue to find a statistically significant negative effect of passed reinspections conducted in the 90 days prior to the pollution reading, but we find no negative effect of leads or lags of passed reinspections on air pollution levels, although we do observe positive and statistically significant coefficients from some bins. For leads of reinspections (reinspections occurring further in the past), this may be spurious correlation but may also be a sign that the effects of repairs are short lived. Our primary results indicate that reductions in emissions occur immediately following the repair and reinspection, but effects may fade over time, consistent with the results of Mérel et al. (2014). This is particularly evident in our results for NO_x. The positive association between pollution levels and future passed reinspections (e.g., a positive coefficient for +180 days) is likely mechanical—if more passed reinspections will happen in the near future, more vehicles are in failing condition today and thus have high emissions. Generally, however, our results support our main findings. For all periods, including lags and leads, effects for newer vehicles are effectively zero.

We present a variety of additional robustness checks for our main results in appendix section A.2, including alternate time controls, additional placebo tests, adjustments for county makeup (urban vs. rural), and testing for possible endogeneity of inspection timing.

As a whole, we find that the Smog Check program's requirements to repair and reinspect high-polluting vehicles moderately improved local CO and NO_x levels, particularly when repairing older model year cars. We find little or no effect from passed reinspections of vehicles manufactured in 1985 or later. To some extent, we expect the small effect of reinspections of newer cars on air pollution. Figure 1 shows that the average differential in tailpipe emissions between a passing and failing vehicle decreases with model year. Bringing the emissions of the average failing 1984 model year vehicle to the level of an average passing vehicle from that year reduces tailpipe CO emissions by around 14,000 parts per million, while fixing an average failing 2001 model year vehicle reduces CO emissions by about 1,300 parts per million, less than one-tenth the gain from older cars. Since repairing one vehicle has a smaller effect on tailpipe emissions, the effect of repairing 1,000 vehicles on ambient pollution is also smaller. The small effect of passed reinspections of newer cars may also relate to 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. The biennial inspection itself then might have little effect; serious emissions

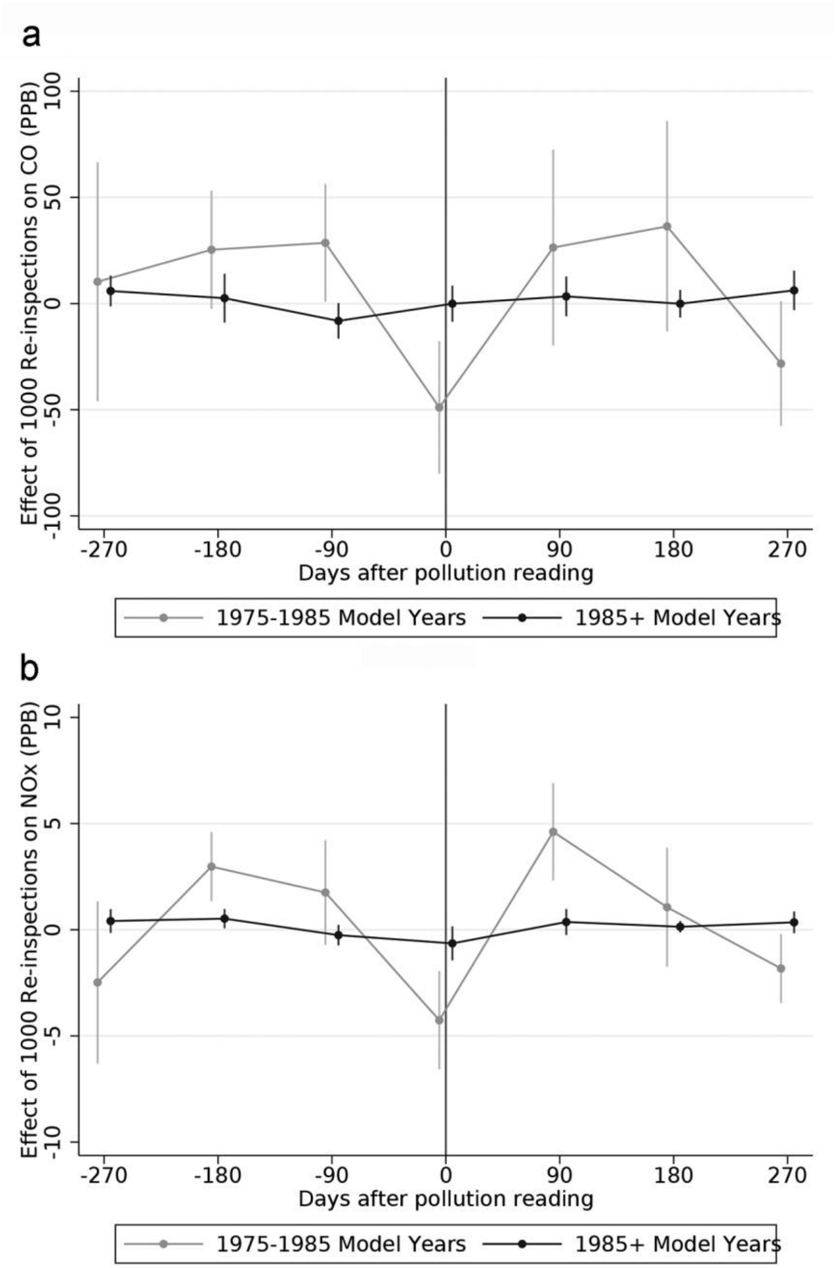

Figure 2. Residual variation in 90-day rolling totals of reinspections after controlling for covariates in preferred regression specification for CO (*a*) and NO_x (*b*). Points report the coefficients from a regression of contemporaneous pollution levels on several leads and lags of a 90-day rolling total of reinspections, controlling for county, calendar week, and county-by-year fixed effects as well as controls for weather. Bars give the uniform 95% sup-t bands for the leads and lags, with critical values obtained via simulation. Color version available as an online enhancement.

failures have already been fixed.[25] Indeed, newer model years, particularly after 1996, have significantly lower failure rates on initial I/M inspections, even though failure levels are more stringent on newer model year vehicles (see appendix fig. A5).[26] OBDII systems may also make it easier to consistently defeat detection of emissions failures, as was the case in the Volkswagen diesel vehicle testing scandal discovered in 2015.

We next consider how station quality controls, as proposed in the new STAR program, affect the benefits of I/M. Our analysis of station quality has the advantage of partially addressing measurement error from using passed reinspections as a proxy for true repairs. Any passed reinspection absent a true, lasting repair induces noise in our treatment. The STAR program quality metrics are designed to rate more highly stations that are less likely to pass initially failing cars absent a true repair. If this measure of quality is valid, systematic differences between repairs and passed reinspections should be lower for inspections at high-quality stations.

4.2. Station Quality

Identification of the effect of station quality follows from the same logic as the effect of passed reinspections. The number of passed reinspections 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. We measure station quality by generating 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. Henceforth, we discuss station "quality" as measured by our reconstructed STAR measures. Our quality metric, which appendix section A.1 describes in detail, is based off the STAR program's quality metrics and runs from 0 to 1, where 1 is the highest possible quality score. As our quality metric was unobservable to consumers and stations during the period covered by our analysis, and the study of roadside inspections leading to the development of the

---

[25] An important related point is that the existence of an I/M program like Smog Check will tend to affect emissions beyond the direct effect of the inspections. The threat of a future failed inspection may lead consumers to fix their vehicles outside of the inspection cycle if problems are detected during routine maintenance. The existence of the I/M program may lead auto mechanics outside the inspection system to flag emissions problems for their customers. Likewise, consumers may decide to scrap older vehicles that require extensive repairs to pass a Smog Check inspection, even if a failed inspection does not itself trigger the scrappage.

[26] As noted above, there are different standards for old and new vehicles—the federal EPA has tightened the standards for new models over time, and the Smog Check emissions standards are broadly based on the emissions standards for a given model year when that model was new. For instance, on the two-speed idle version of the Smog Check, a 1975–80 model year vehicle needs to have CO emissions in the "idle" mode of less than 2% to pass, while a 1993 or newer vehicle has a cut-off that is half that level.

STAR program was not even released until 2010, the distribution of vehicles across high- and low-quality stations should be exogenous as long as, before STAR, consumers had no way to reliably identify low-quality stations willing to pass a failing vehicle. We test this assumption by comparing the average quality of stations inspecting older vehicles to the average quality for stations inspecting newer vehicles. Pre-1985 model year vehicles are more likely to fail and likely more expensive to fix. If consumers had a systematic way to identify "bad" stations willing to give a sham inspection, we expect older vehicles to be disproportionately inspected at lower-quality stations. Instead, the mean quality measure for older and newer vehicles is essentially the same, which shows no active sorting along the age distribution.

Given our argument that only accurate reinspections reduce local air pollution, we expect to see smaller or zero effects from reinspections at low-quality stations and larger effects from reinspections at high-quality stations. As quality is a station-level trait and our estimates are at the county level, we first calculate the quality score for all stations in the county, then aggregate up to the county level, weighting by the number of passed reinspections conducted over the previous 90 days at each station (see appendix sec. A.1 for more details).

Table 3 shows our results. All columns use the controls from column 4 of table 2: weather controls, county fixed effects, county cubic trends, and statewide calendar week fixed effects. Column 1 repeats the model of table 2, column 4, for comparison. Column 2 includes an interaction between the number of passed reinspected vehicles in the last 90 days and the county average quality measure.

Column 2 of table 3 shows that station quality matters in the case of older vehicles. As an extreme example, reinspections in a county with an average quality score of zero would have essentially zero effect on either pollutant, but the effect of reinspections increases with average station quality. For ease of interpretation, we calculate the marginal effect of 1,000 passed reinspection for each vehicle age group at the average county-level quality score in our sample. For older cars, an additional 1,000 passed reinspections in a county of average station quality would decrease ambient CO levels by 57.7 ppb and average NO_x levels by 5.7 ppb, with both average effects statistically significant at 1%.[27]

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 that 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 that 1975–85 model year personal vehicles contribute around 20% of all CO emissions, more so in the earlier years of our sample when these are more common

---

[27] We calculate the marginal effect at the mean quality level using the margins command in Stata 15.

Table 3. Station Quality and County-Level Air Pollution

	(1)	(2)
	A. Outcome Is CO (ppb)	
Thousands of reinspections last 90 days:		
1975–85 vehicles	−19.00***	9.166
	(6.621)	(17.65)
1975–85 vehicles · quality		−120.2*
		(60.56)
Effect at mean station quality		−57.73***
		(19.93)
1985+ vehicles	−3.516	−5.244
	(2.481)	(5.330)
1985+ vehicles · station quality		8.479
		(15.89)
Effect at mean station quality		−.450
		(4.694)
F-test of equality	5.856	6.379
$P > F$	.0214	.0167
	B. Outcome Is NO_x (ppb)	
Thousands of reinspections last 90 days:		
1975–85 vehicles	−1.149***	2.200**
	(.369)	(1.023)
1975–85 vehicles · station quality		−14.31***
		(3.292)
Effect at mean station quality		−5.659***
		(1.124)
1985+ vehicles	−.793***	−.988***
	(.241)	(.320)
1985+ vehicles · station quality		.965
		(1.058)
Effect at mean station quality		−.452
		(.366)
F-test of equality	2.370	21.03
$P > F$	.133	.0000557

Note. Observations are county-days. All regressions control for daily weather, county fixed effects, calendar week fixed effects, and county-year fixed effects. Standard errors clustered by county reported in parentheses. *F*-tests of equality report the *F*-statistic from testing the null hypothesis that either the coefficient (col. 1) or the average marginal effect (col. 2) of 1975–85 model year vehicles equals that of 1985+ model year vehicles.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

cars. Our results in figure 1 indicate that a failing vehicle of these vintages typically has emissions around three times higher than a passing vehicle. Depending on the county, 1,000 passed reinspections of 1975–85 model year vehicles represents the reinspection of 0.2%–20% of all personal automobiles of these vintages. This makes a change of around 58 ppb in CO concentration plausible—this is between 5% and 10% of mean CO levels. Our estimate for NO_x reductions in a county with average station quality is less in line with the models, which predict that older vehicles contribute only around 3%–5% of all NO_x emissions. However, the physical and chemical processes behind NO_x concentrations are more complicated, with nonlinearities such that a modest reduction in NO_x emissions might have a disproportionate effect on NO_x concentrations, depending on atmospheric conditions and concentrations of other pollutants such as hydrocarbons and ozone (see discussion in sec. 4.3).

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 passed reinspections at higher-quality stations reduces the effect of passed reinspections on ambient pollution for both CO and NO_x. These results are precisely estimated but economically insignificant, which is clear when considering the marginal effect at the mean. At the average county quality of around 0.59, an additional 1,000 passed reinspections of newer cars correlates with a decrease in pollution levels of around 0.001 standard deviations for CO and a decrease in pollution levels of about 0.019 standard deviations for NO_x. The marginal effect at the mean is not statistically significant for CO, and average effects on both pollutants are about half the size of the already small overall effects for newer cars from table 2.

To illustrate that the effect of station quality for newer cars is effectively zero, we generate 20 bins of quality scores, in units of 0.05, from 0 to 1 (see appendix sec. A.1 for details). We then estimate regressions including sets of 20 variables each for older and newer cars, giving the counts of reinspections at stations in the appropriate quality bin. Figure 3 plots each coefficient on the bin-specific count variables, and provides a LOWESS fit with bandwidth $N \cdot 0.8$ to illustrate the general patterns across the quality distribution.

The top panel shows results for CO, and the bottom panel shows results for NO_x. For newer cars, the effect of a reinspection is approximately constant at zero across our estimated measure of station quality. Visual analysis shows that the positive results from table 3 are a result of effects for counties with average quality measures very close to 1, which represent a small share of reinspections overall—only 15% of total newer vehicle reinspections across our entire sample occurred at stations with quality levels greater than 0.9. Figure 3 shows that with older cars, there is a clear differential between a passed reinspection at a lower-quality versus a higher-quality station. Increasing passed reinspections in areas with quality scores below approximately 0.3 has no effect on local air pollution, with increasing benefits of passed reinspections at higher-quality stations.

Table 3 and figure 3 jointly suggest that increasing passed reinspections of older cars in areas with higher estimated station quality reduces air pollution in an economically

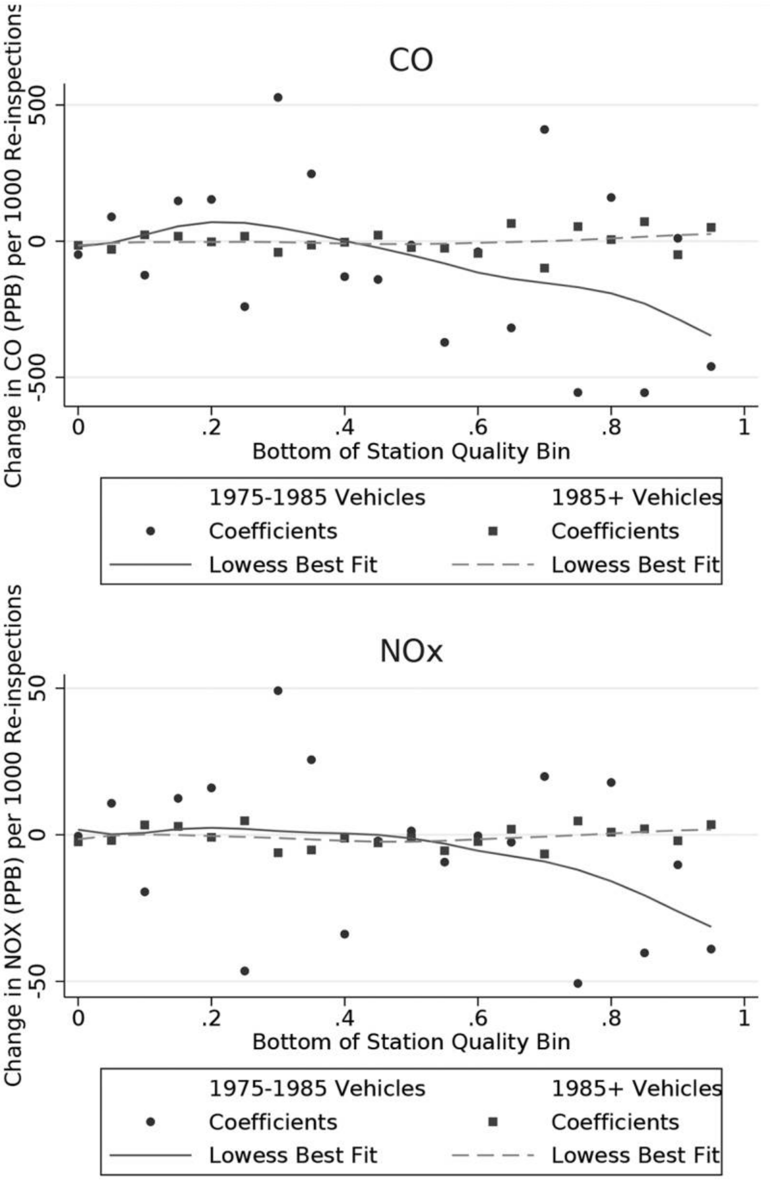

Figure 3. Effect of reinspections on daily air pollution, by follow-up pass rate (FPR) score of test stations. Graphs show the change in measured county-level air quality for carbon monoxide (CO, *top*) and nitrogen oxides (NO_x, *bottom*) resulting from an additional 1,000 nominal repairs (as proxied by a successful reinspection of an initially failed vehicle) done at stations of respective quality as measured across 20 equally spaced bins covering the full range of our contemporaneous FPR (C-FPR) measure (see appendix sec. A.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 $N \cdot 0.8$. Color version available as an online enhancement.

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 passed reinspections of newer cars and the negligible effect of station quality for those vehicles is problematic for the social efficacy of STAR and similar programs in the future. 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 as designed, even enhanced by STAR, is not delivering air pollution benefits from repairing failing newer vehicles, this diminishes the value of the program.

4.3. Effects of Reinspections on Secondary Pollutants

4.3.1. Ozone

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 its control stem from its role in forming secondary pollutants, principally O_3 and PM_{10}. Ozone (O_3) has a complicated formation process, requiring both NO_x and VOCs in a specific ratio. The process by which NO_x and VOCs form O_3 resembles a Leontieff production function, such that when NO_x levels are high, reducing NO_x emissions may have limited effect on O_3 levels. In some cases, reductions in NO_x can increase O_3—excessive NO_x can titrate O_3 by binding with a single O molecule and converting O_3 into basic oxygen (Muller et al. 2009). This makes it a priori difficult to predict program effects on O_3. Moderate reductions in NO_x caused by Smog Check might raise, lower, or have no effect on ambient O_3. Table 4 shows our empirical model using ambient O_3 as the outcome. Columns 1 and 2 replicate the analysis of table 3. The point estimates largely have the "wrong" sign, indicating that passed reinspections increase O_3 levels, but estimates are economically zero. For the average county, passing an additional 1,000 older vehicles increases ambient O_3 by 0.04 of a standard deviation for older cars and decreases ambient O_3 by 0.007 of a standard deviation for newer cars.

While we find effects of passed reinspections on NO_x we expect NO_x to affect O_3 in unpredictable ways. Empirically, in our sample O_3 correlates negatively with both NO_x and CO. To adjust for this, in columns 3 and 4 of table 4 we control for the level of NO_x and NO_x squared. Our coefficients acquire the "right" sign, with passed reinspections of older cars reducing O_3 levels, although the magnitudes are somewhat small—typical O_3 concentrations are on the same scale as those of NO_x. The interaction of station quality with passed reinspections of older cars is noisy but goes in the program-anticipated direction. Our results suggest that passed reinspections of 1,000 older vehicles in a county with average station quality would increase O_3 levels by 0.01 ppb, just 0.001 of a standard deviation. Appendix figure A6 plots results by our quality bins but does not show a clear visual effect of station quality.

Table 4. Station Quality and County-Level Ozone Pollution

	Base		NO_x Controls	
	(1)	(2)	(3)	(4)
Thousands of reinspections last 90 days:				
1975–85 vehicles	.271*	−.500**	.276*	−.0622
	(.156)	(.205)	(.142)	(.213)
1975–85 vehicles · station quality		1.765**		.138
		(.679)		(.700)
Effect at mean station quality		.522		.0136
		(.246)		(.211)
1985+ vehicles	.168*	.736**	.00466	.472*
	(.0960)	(.277)	(.0682)	(.253)
1985+ vehicles · station quality		−1.415***		−1.145**
		(.485)		(.507)
Effect at mean station quality		−.0944*		−.165*
		(.113)		(.0857)
F-test of equality	.715	4.896	3.797	.493
$P > F$	.403	.0323	.0594	.487

Note. Observations are county-days. All regressions control for daily weather, county fixed effects, calendar week fixed effects, and county-year fixed effects. Columns 3 and 4 control for a quadratic in the level of NO_x. Standard errors clustered by county reported in parentheses. *F*-tests of equality report the *F*-statistic from testing the null hypothesis that either the coefficient (cols. 1 and 3) or the average marginal effect (cols. 2 and 4) of 1975–85 model year vehicles equals that of 1985+ model year vehicles.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

4.3.2. Particulate Matter

Because PM_{10} sensors only take readings every 6 days, for our PM_{10} estimates we collapse our data 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 reinspections 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 that passed reinspections of older vehicles have a small negative effect on PM_{10} levels, while passed reinspections of newer cars have a near-zero effect. Adding an interaction with county-level average station quality, on average passed reinspections of older vehicles at poor-quality stations have no effect on PM_{10} levels, while passed reinspections at high-quality stations moderately reduce PM_{10} levels. Our point estimates indicate that at the average station quality, 1,000 passed reinspections of older vehicles lead to a 2.1 μ/m^3 reduction in PM_{10} levels, about 0.13 of a standard deviation.

Table 5. Station Quality and County-Level PM_{10} Pollution

	Outcome Is PM_{10} (μ/m^3)	
	(1)	(2)
Thousands of reinspections last 90 days:		
1975–85 vehicles	−.251	.720
	(1.055)	(2.048)
1975–85 vehicles · station quality		−3.831
		(4.187)
Effect at mean station quality		−1.522
		(1.225)
1985+ vehicles	.0714	.236
	(.226)	(.352)
1985+ vehicles · station quality		−.274
		(.747)
Effect at mean station quality		.0731
		(.261)
F-test of equality	.105	1.613
$P > F$	.747	.204

Note. Observations are county-weeks. All regressions control for daily weather, county fixed effects, calendar week fixed effects, and county-year fixed effects. Standard errors clustered by county reported in parentheses. *F*-tests of equality report the *F*-statistic from testing the null hypothesis that either the coefficient (col.1) or the average marginal effect (col. 2) of 1975–85 model year vehicles equals that of 1985+ model year vehicles.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

We find positive but economically small and statistically noisy results for reinspections of newer vehicles. Appendix figure A7 plots results for PM_{10} by quality bins.

4.4. Mechanisms

To what extent is the failure of reinspections at low-quality stations to reduce pollution driven by consumer behavior? Consumers can potentially influence inspection results by "shopping" for stations willing to pass a failing vehicle without repairs through fraudulent testing behavior.[28] We capture the importance of consumer "shopping"

---

[28] A particularly egregious approach to fraud in I/M is known as "clean piping": entering the information for the failing 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.

behavior by considering 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 pass may be valid or invalid). Table 6 summarizes the experience of consumers who have more than one inspection in a cycle, specifically the proportions that have a passing reinspection within a week, the proportion who switch stations following a failed initial inspection, and the mean change in station quality for switchers. This analysis excludes vehicles designated "gross polluters" due to especially high emissions at the initial inspection. Such vehicles must have follow-up testing at appropriate stations (e.g., test-only stations) for subsequent inspections, and including these vehicles would bias upward the appearance of shopping behavior by showing up as switching across stations.

Table 6 shows that approximately 60% of failed inspections result in a passed final inspection within 1 week of the initial test—this suggests that the majority of failures lead to a (repair and) passed reinspection quickly. But as the number of follow-up inspections prior to passing grows, the likelihood of a passing inspection within 1 week (somewhat mechanically) decreases. Row 2 of the table shows that many of the cars that eventually pass do so at the same station as the initial inspection. Just under 65% of all reinspected cars eventually pass at the same station at which they failed the initial test. As a possible indication of "shopping" for a passing station, switching becomes more common the longer it takes to pass. Almost 70% of cycles with two inspections (an initial and a follow-up) start and end at the same inspection station, compared to just over 20% of cycles with five or more (one initial and at least four follow-up) inspections. Such cycles are rare, making up just 1.5% of all inspection cycles in our data. While this could be the result of consumers abandoning bad testing stations, our data show that those that switch usually end up at stations with lower-quality scores, particularly in cycles with more than two inspections. For example, a car that takes three inspections to pass in a given cycle has a 55% (1 – 0.453 from row 2, col. 3) probability of switching stations

Table 6. Consumers' Shopping Behavior after a Failed Initial Inspection

		Total Number of Inspections in Cycle			
	All	2	3	4	5+
Final inspection within 1 week	.601	.667	.396	.282	.204
Final inspection at same station	.672	.739	.453	.381	.330
Switchers' change in Station Quality Score	–.0588	–.0494	–.0734	–.0788	–.0744
N	13,025,677	10,230,367	2,045,339	510,239	239,732

Note. 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 Station Quality Score" is the difference between the quality score of the first and last station in the inspection cycle.

between the initial inspection and eventual passing, with a decrease of 0.07 in our quality score between stations. This is consistent with a large fraction of consumers putting in some effort to comply with the I/M program but suggests that eventually some give up and seek a less legitimate way to pass.[29]

5. SIMULATION EXERCISES

We use our estimates from the previous section to simulate the impact of California's recently implemented STAR program, which was aimed at improving station quality. We predict what average pollution levels would have been if directed vehicles were inspected at stations with a minimum quality score of 0.4, as STAR would have required had it existed during our sample period. We take the simulated change in pollution 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. Appendix section A.3 provides the details of our simulation methods. In addition, in appendix section A.4, we conduct a rough cost-benefit analysis of both STAR and the Smog Check program as a whole, using estimates from the literature on the economic costs of pollution. Note that these simulations assume no consumer or producer responses to the presence of the Smog Check program, beyond excessively polluting cars receiving repairs. Thus, we are ignoring any producer-driven behaviors designed to improve station quality scores and become eligible to receive business from directed vehicles (including attempts to game the quality score as well as improving true inspection quality).

We simulate the effect of the specific requirements of the STAR program at the beginning and end of our sample on CO and NO_x. 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 then focus on the future impact of more stringent I/M requirements. To predict pollution levels, we use the models from column 2 of table 3, with passed reinspections interacted with county-level average station quality.

Figure 4 presents the results of our simulation as a map of predicted changes in pollution levels 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 pollution in 1998 from placing a floor on station quality, in part because these counties historically had what would be high-quality-score stations to begin with: our theoretical exercise of removing "bad" station reinspections 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.

---

[29] While we cannot observe if there is fraud in the California Smog Check program, we need not do so for our analysis. Our goal is to understand if I/M policies such as Smog Check translate to improved air quality in practice. Because any I/M program will generate both genuine and false inspections, switching behavior speaks to the underlying mechanism behind our results, than being a confounder for those results.

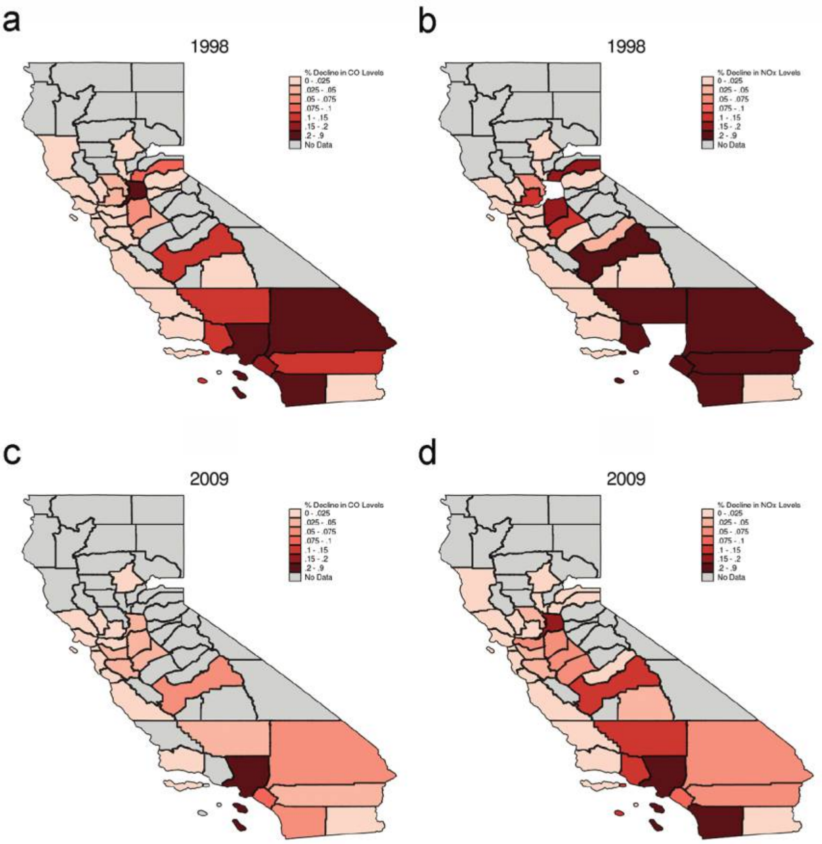

Figure 4. Predicted percentage change in pollution levels from reinspecting all 1975–85 vehicles at STAR stations. Maps show the simulation-based differences in air quality assuming that all 1975–85 model year cars that failed their initial inspections have been reinspected at stations considered of 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 NO_x. 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 sec. 5 for details of our simulation process.

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. 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.

6. CONCLUSION

To reduce public costs of automobile pollution, governments sometimes use inspection and maintenance (I/M) programs. Vehicles are routinely tested for compliance with emissions standards, and must obtain repairs and further inspections in the event of failure. These inspections and repairs are costly to consumers, and while follow-up inspections can show that repairs improve emissions at the tailpipe, there is little causal evidence as to how such programs affect local air quality. We use over a decade of data from the state of California to test whether the California I/M program improved local air pollution in an observable manner on a large scale, beyond laboratory conditions. We find that an increase in passed reinspections (our proxy for emission repairs) corresponds to local improvements in CO, and NO_x levels, with suggestive effects on PM_{10} levels, and essentially no change in local O_3. This relationship persists after controlling for location and time fixed effects and ambient weather controls, though it is sensitive to the omission of region-specific time effects. Overall results show that California's Smog Check program has successfully improved the state's local air pollution.

However, additional gains from the Smog Check program are decreasing with time, as almost all benefits of passed reinspections come from fixing failing older model cars (1985 and prior) with inferior emissions control technology. As vehicles with older emissions control technology disappear, the difference between failing and passing an emissions inspection decreases. Improvements in emissions control technology, potentially driven by the program itself, are reducing the social efficiency of repairing failed vehicles. Moreover, there remains the question of efficiency relative to other policies. I/M programs like California's Smog Check program are intended to reduce O_3 levels by controlling emissions from all vehicles, but our results suggest that the benefits of such programs come only from a shrinking subset of vehicles.

Given fears of false or low-quality repairs of failing vehicles followed by potentially fraudulent reinspections, the new California STAR program uses an estimated station quality measure based on inspection data to determine which stations the state allows to provide reinspections to the dirtiest cars. We consider whether requiring inspections of high-polluting vehicles at high-quality stations can further reduce air pollution. A number of factors, including potential system fraud, complicate measuring the role of station quality after the implementation of the program. We use our pre-STAR program data to construct a modified station-level quality measure that conveys information similar to that of STAR, but without concerns of secondary program effects. When a greater share of I/M stations are of high quality, an additional passed reinspection corresponds to greater marginal improvements in contemporaneous air pollution. Gains from estimated station quality are again limited to reinspections of older model cars. For newer cars, there is no economically meaningful benefit to reinspections from even high-quality stations. Given the decline in the number of older vehicles as they age out of the fleet, it is unlikely that the new STAR program as designed will further improve the effectiveness of the Smog Check program for reducing air pollution.

REFERENCES

Ando, Amy, Virginia McConnel, and Winston Harrington. 2000. Costs, emissions reductions and vehicle repair: Evidence from Arizona. *Journal of the Air Waste Management Association* 50 (4): 509–21.

Bureau of Automotive Repair. 2014. Smog Check performance report (AB2289). Technical report, BAR, Rancho Cordova, CA. https://www.bar.ca.gov/pdf/ab_2289_bill_20100924_chaptered[1].pdf.

Currie, Janet, and Matthew Neidell. 2005. Air pollution and infant health: What can we learn from California's recent experience? *Quarterly Journal of Eocnomics* 120 (3): 1003–30.

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

Dominici, Francesca, Michael Greenstone, and Cass R. Sunstein. 2014. Particulate matter matters. *Science* 344:257–59.

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

Eisinger, Douglas S. 2010. *Smog check: Science, federalism, and the politics of clean air*. New York: Routledge.

Engel, Kirsten H. 2015. The enigma of state climate change policy innovation. In *The law and policy of environmental federalism: A comparative analysis*, ed. Kalyani Robbins and Erin Ryan. Northampton, MA: Elgar.

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

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

Gray, Wayne B., and Jay P. Shimshack. 2011. The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy* 5 (1): 3–24.

Harrington, Winston, Virginia McConnell, and Amy Ando. 2000. Are vehicle emission inspection programs living up to expectations? *Transportation Research, Part D: Transport and Environment* 5 (3): 153–72.

Hoekstra, Mark, Steven L. Puller, and Jeremy West. 2014. Cash for Corollas: When stimulus reduces spending. NBER Working paper 20349, National Bureau of Economic Research, Cambridge, MA.

Hubbard, Thomas N. 1998. An empirical examination of moral hazard in the vehicle inspection market. *RAND Journal of Economics* 29 (2): 406–26.

Jacobsen, Mark R. 2013a. Evaluating US fuel economy standards in a model with producer and household heterogeneity. *American Economic Journal: Economic Policy* 5 (2): 148–87.

———. 2013b. Fuel economy and safety: The influences of vehicle class and driver behavior. *American Economic Journal: Applied Economics* 5 (3): 1–26.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders. 2016. Caution drivers! Children present: Traffic, pollution and infant health. *Review of Economics and Statistics* 98 (2): 350–66.

Li, Shanjun, Joshua Linn, and Elisheba Spiller. 2013. Evaluating "Cash-for-Clunkers": Program effects on auto sales and the environment. *Journal of Environmental Economics and Management* 65 (2): 175–93.

Mérel, Pierre, Aaron Smith, Jeffrey Willams, and Emily Wimberger. 2014. Cars on crutches: How much abatement do smog check repairs actually provide? *Journal of Environmental Economics and Management* 67 (3): 371–95.

Mian, Atif, and Amir Sufi. 2012. The effects of fiscal stimulus: Evidence from the 2009 "Cash for Clunkers" program. *Quarterly Journal of Economics* 1107:1142.

Muehlenbachs, Lucija, Stefan Staubli, and Mark A. Cohen. 2016. The impact of team inspections on enforcement and deterrence. *Journal of the Association of Environmental and Resource Economists* 3 (1): 159–204.

Muller, Nicholas, Daniel Tong, and Robert Mendelsohn. 2009. Regulating NO_x and SO_2 emissions in Atlanta. *B.E. Journal of Economic Analysis and Policy* 9 (2).

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

Sandler, Ryan. 2012. Clunkers or junkers? Adverse selection in a vehicle retirement program. *American Economic Journal: Economic Policy* 4 (4): 253–81.

Schlenker, Wolfram, and Michael J. Roberts. 2006. Nonlinear effects of weather on corn yields. *Review of Agricultural Economics* 28:391–98.

Shimshack, Jay P., and Michael B. Ward. 2005. Regulator reputation, enforcement, and environmental compliance. *Journal of Environmental Economics and Management* 50 (3): 519–40.

Simeonova, Emilia, Janet Currie, Peter Nilsson, and Reed Walker. 2018. Congestion pricing, air pollution and children's health. NBER Working paper 24410, National Bureau of Economic Research, Cambridge, MA.

Spindt, R. S., R. E. Dizak, R. M. Stewart, and W. A. P. Meyer. 1979. Effect of ambient temperature on vehicle emissions and performance factors. Technical report 1979, US Environmental Protection Agency, Ann Arbor, MI.

Wolff, Hendrik. 2014. Keep your clunker in the suburb: Low-emission zones and adoption of green vehicles. *Economic Journal* 124 (578): F481–F512.