Fraud and Enforcement in Equilibrium: Evidence from Environmental Inspections

Jacob Bradt Jackson Dorsey*

November 6, 2026

Abstract

Many regulations rely on inspectors to certify compliance. When compliance is costly, regulated parties often demand fraudulent certifications and corrupt inspectors supply them. Regulators typically respond by cracking down on corrupt inspectors. We study how enforcement crackdowns affect the market for fraudulent certifications in equilibrium, using nearly 100 million vehicle emissions inspections in Texas from 2016–2024. First, we develop a new method to classify fraudulent inspections and find that 9.5% of passing inspections are fraudulent. On-road remote pollution sensors show that the excess emissions of each fraudulently passed vehicle cause about $251 of health damages a year. Second, we estimate an equilibrium model of supply and demand for fraud and use it to study a 2022 crackdown that shut down almost every major fraud seller. Despite more than 1,500 enforcement actions against stations, about half of the volume of fraudulent inspections persisted. Relative to a market in which enforcement stayed at its pre-crackdown level, the crackdown reduced fraudulent inspections by only 48%, as displaced buyers switched to the remaining sellers and the market concentrated among a few large "fraud mills." Drivers lost surplus, while the public gained from lower air pollution and repair shops gained profits, leaving the crackdown approximately welfare neutral on net. Finally, penalties on drivers who buy fraudulent inspections would complement the supply-side crackdown, further reducing fraud and raising net welfare relative to supply-side enforcement alone.

JEL: K42, Q58, Q53, L51, L84, C38

Keywords: fraud, enforcement, third-party certification, illicit markets, vehicle emissions inspections, air pollution

*Bradt: University of Texas at Austin. Dorsey: University of Texas at Austin. The views expressed are those of the authors and do not represent the views of The University of Texas at Austin or The State of Texas. All fraud classification results are statistical estimates and should not be interpreted as definitive findings about any individual or station. We thank Dan Ackerberg, Mary Evans, and Ashley Langer for helpful comments as well as participants at the UT–ITAM TexMex Conference, NYU Stern, Stanford SITE, and Rice University for constructive feedback. We thank Yiqun Zhao and Josh Dawe for excellent research assistance; Lawrence Cortez and Lisa Cargill at the Texas Department of Public Safety for data access and help understanding the I/M Program; Les Carnes, Jose Escribano, and Mike Bradburn at the Travis County Clean Air Task Force and Scott Friedman and Eva Parks of NBC DFW for helpful conversations. This research was supported in part by a Faculty Research Excellence Grant from the McCombs School of Business. All errors are our own.

1 Introduction

In many settings, regulators rely on inspectors to certify that regulated parties comply. Auditors certify that industrial plants meet pollution limits (Duflo et al., 2013), accounting firms certify that companies' financial statements are accurate (Dyck et al., 2010), and licensing examiners certify that applicants are qualified to drive (Bertrand et al., 2007). When compliance requires costly action, regulated parties may seek to avoid paying its full cost. Inspectors, in turn, can profit by selling false certifications. The result is a market for fraudulent certification, with corrupt inspectors on the supply side and regulated parties on the demand side. While fraudulent certification is privately beneficial, it often imposes social costs. False environmental certifications, for instance, cause health damages through excess emissions. To limit these costs, regulators often respond by targeting the supply side with enforcement campaigns that shut down corrupt inspectors.

Such crackdowns have targeted building inspectors who took bribes for inspection approvals,[1] physicians who certified excessive sick leave (Celhay et al., 2025), and pain clinics whose physicians prescribed opioids without medical need (Soliman, 2025). The appeal of these crackdowns is simple: if fewer inspectors are willing to fake certifications, fewer false certifications will be sold. However, the effectiveness of a crackdown depends on how both sides of the market respond. On the demand side, regulated parties can often choose their inspector (Bennett et al., 2013; Hubbard, 1998). If a regulator removes corrupt auditors, for example, the crackdown's effect depends in part on whether the firms they audited turn to honest auditors or to the corrupt ones who remain. On the supply side, the remaining corrupt inspectors may expand and raise their prices, and new ones may enter. Theoretical work on illicit markets makes the same point. Enforcement against drug traffickers raises their costs and the street price of drugs, but how much drug use falls depends on the elasticities of demand and supply (Becker et al., 2006). Measuring these responses is difficult, because false certifications, like other illicit transactions, are inherently hard to observe.

Our setting is vehicle emissions inspections in Texas from 2016 to 2024. Under the Clean Air Act, metropolitan areas that fail to meet federal air-quality standards must require regular vehicle emissions inspections. Most states delegate these inspections to privately operated auto shops. Texas runs one of the largest such programs: we observe nearly 100 million emissions inspections at more than 5,000 stations over our sample period. The driver of a vehicle that fails an inspection faces a repair bill that can run from hundreds to thousands of dollars. The driver can pay for the repair and pass the inspection honestly, or buy a fraudulent pass, known as a "clean scan," often for a fraction of the repair bill. In 2022, the program's regulator launched a major supply-side crackdown aimed at shutting down stations that fraudulently passed vehicles.

These records, together with the crackdown, let us ask how much supply-side enforcement reduces fraud in equilibrium, and who gains and who loses. We find that fraud is common: 9.5% of passing inspections are fraudulent. The supply-side crackdown shut almost every major seller, yet cut fraud by only about half. It was roughly welfare neutral, moving surplus from drivers to the public and to repair shops. A penalty on drivers who buy fraudulent

[1] In 2002, federal prosecutors charged 15 of New York City's 24 plumbing inspectors with taking bribes in exchange for inspection approvals.

inspections would remove the same fraud at a net welfare gain, and it complements the supply-side crackdown.

Our analysis proceeds in three steps. First, we develop a method that estimates the probability that each inspection is fraudulent from the administrative records the program already collects. Administrative data often contain fraud indicators, red flags that mark a record as suspicious. On its own, an indicator cannot identify which records are fraudulent, because indicators produce both false positives and false negatives. Our method combines information from a set of multiple indicators. Under our identifying assumption, the covariance between the indicators identifies the share of each station's inspections that are fraudulent and the probability that each inspection is fraudulent. We find that a substantial share of passing inspections in our sample, 9.5%, are fraudulent: about 1.0 million fraudulently passed inspections a year. Using roadside remote sensors, we show that fraudulently passed vehicles emit 216% more nitric oxide than similar vehicles that passed honestly, and that each fraudulently passed vehicle causes about $251 of health damages a year.

Second, having measured fraud, we document how the market responded to the crackdown. Fraud was rising in the years before the crackdown, peaking statewide at 11.6% of passing inspections in 2022. The regulator's campaign suspended and revoked the licenses of stations selling fraudulent passes. By the end of 2024, it had taken 1,545 enforcement actions against stations in the Dallas–Fort Worth and Houston metro areas. Despite the scale of the crackdown, fraud fell much less than one for one: the number of major fraud stations, those selling more than 100 fraudulent passes a month, fell by 92%, but the share of passing inspections that were fraudulent fell by only 37%. New stations entered to replace many of the sellers that were shut down, and a few large stations captured most of the fraud market. As a result, the market became much more concentrated: by 2024, just 5 stations sold half of the two metros' fraudulent passes. Rather than repair their vehicles, some drivers skipped the inspection and drove unregistered, which kept polluting vehicles on the road.

Third, to measure the crackdown's effects on fraud and welfare and to compare it with other enforcement policies, we build a structural model of both sides of the market for fraudulent inspections. In the model, the regulator moves first and sets an enforcement policy: how intensively it audits inspection stations and how it penalizes the inspection stations it catches. We measure this policy from the regulator's actions but take it as given, without modeling the regulator's objective. On the demand side, each driver of a malfunctioning vehicle makes a discrete choice: repair the vehicle and pass honestly, pay an inspection station to pass it fraudulently, or let its registration lapse. Two parameters of the demand model are particularly important for our counterfactuals. The first measures how readily drivers substitute among sellers of fraudulent passes. When the regulator shuts down a driver's preferred inspection station, this parameter determines whether the driver turns to another corrupt inspector or repairs the vehicle. The second is the distribution of drivers' price sensitivity, which determines how large a repair bill must be before a driver chooses not to repair. We estimate it by comparing the choices of drivers who face higher and lower repair bills to bring their vehicles into compliance. On the supply side, inspection stations decide whether to enter the market for fraudulent passes, how long to operate, and what price to charge, knowing that the regulator may detect and penalize them. Together, these pieces let us measure how counterfactual enforcement policies affect the equilibrium quantity of fraudulent inspections and the welfare of each side of the market.

Our estimates show that drivers treat sellers of fraudulent passes as substitutes: when an inspection station is shut down, the most common response of its buyers is to buy from another corrupt station. Drivers are also sensitive to the cost of repair: for the median driver whose vehicle has just failed an inspection, a repair bill above $640 makes a fraudulent pass or a lapsed registration more attractive than repair. We use the estimated model to simulate the market with and without the crackdown, and find that the crackdown reduced fraudulent inspections in 2024 by 48%. This is larger than the observed fall in the fraud share because fraud would have kept rising without the crackdown, and the model measures the effect against that market. Two responses offset part of the fraud the closures removed directly: displaced buyers switched to the remaining sellers, including stations outside their metropolitan area, and the surviving sellers expanded. Fraud still fell by half, because the higher chance of being caught made selling worth only a few months before detection, so far fewer stations chose to operate at any one time. An accounting based only on the stations the crackdown shut down, which assumes that their buyers stopped buying fraud, would overstate the crackdown's effect on fraud, implying a fall of 63%.

We next use the model to measure the crackdown's effects on welfare. We estimate that the crackdown cost drivers $407 million of consumer surplus in 2024. Many drivers had benefited from fraudulent passes, which let them keep their vehicles registered while avoiding costly repairs. By shutting down many of the inspection stations these drivers used, the crackdown forced them to pay for costly repairs, let their registrations lapse, or switch to a less preferred inspection station. Importantly, these losses fall on the drivers who bought fraudulent passes, who buy mainly from stations in the lower-income parts of each metropolitan area (Section 3). On the other hand, we estimate that the crackdown reduced health damages by $278 million through lower on-road emissions and increased the profits of local repair shops by $131 million.[2] The crackdown left total welfare nearly unchanged. Thus, the case for the crackdown as implemented hinges more on the distribution of welfare than on efficiency: how hard a regulator cracks down depends on how it weighs pollution damages against the surplus of the drivers who bought fraudulent passes.

Finally, we ask whether the regulator could do better by also targeting the demand side of the market. In Texas, as in many settings, the regulator targets corrupt suppliers, but in principle it could also penalize the regulated parties, here the drivers who buy fraudulent passes.[3] To understand how enforcement on both sides of the market interacts, we first simulate a penalty on drivers who buy fraudulent passes, set so that it reduces fraud by as much as the supply-side crackdown.[4] We then combine the two policies. We find that the two instruments are complements. Together, they leave 13% of the fraud in the market without the crackdown, against the 27% that would remain if the two effects simply multiplied. With fewer sellers left and less demand for their passes, almost no seller can cover the cost of entering, so enforcement on each side makes enforcement on the other more effective. Adding the penalty to the crackdown also raises the net welfare gain from $4 million to $180 million in 2024.

Related literature. Our paper relates to several strands of literature. First, we con-

[2]Free entry keeps the profits of corrupt stations small, so the crackdown changes them little.

[3]Texas law already makes it a misdemeanor to knowingly present a fraudulent inspection report (Texas Transportation Code §548.603), but we find no record of drivers being charged.

[4]The penalty is the expected cost to a buyer: the fine times the chance of being caught.

tribute to work that detects misconduct in administrative data. This literature uses statistical red flags to detect cheating by teachers, tax evasion, and illegal arms trade (DellaVigna and La Ferrara, 2010; Fisman and Wei, 2004; Jacob and Levitt, 2003). We develop a method that classifies which observations are fraudulent using the covariance of several imperfect indicators. It formalizes the intuition of Jacob and Levitt (2003) that cheating sets off several independent red flags at once. We show that, if the indicators are independent conditional on fraud status, three indicators are enough to point identify fraud rates (Allman et al., 2009). Furthermore, we show that with a fourth indicator the model is overidentified, and we derive a test of the independence assumption. Importantly, our method is useful across a wide range of settings, when researchers observe several imperfect indicators of misconduct.

Second, we contribute to work on the impacts of enforcement in illicit markets. Theory shows that enforcement against suppliers can reshape an illicit market without eliminating it (Becker et al., 2006), and empirical work documents how crackdowns divert trade in illegal drugs (Dell, 2015; Soliman, 2025). Galenianos and Gavazza (2017) estimate a model of both sides of the retail drug market but study penalties without a crackdown in the data. To our knowledge, ours is the first paper to use an actual crackdown on sellers to estimate an equilibrium model of an illicit market in which both buyers' choices and sellers' entry respond to enforcement, and the first to measure whether penalties on buyers and on sellers are complements.

Third, we contribute to work on third-party certification, in which inspectors may be captured by the parties they inspect. Third-party auditors misreport (Duflo et al., 2013), and competition for customers pushes vehicle inspectors and physicians toward leniency (Bennett et al., 2013; Hubbard, 1998; Markussen and Røed, 2017). In concurrent work, Lugthart (2026) uses the same Texas inspection records to study how competition among stations affects cheating. Earlier work by Oliva (2015) measures the extent of cheating in Mexico City's vehicle emissions inspection program and simulates policies that raise the price of a bribe and the fine for driving without a valid certificate. We instead study how both sides of a market for fraud respond when the regulator ramps up enforcement.

Finally, we contribute to broader work on environmental enforcement. Existing work studies how enforcement intensity and monitoring affect compliance (Blundell et al., 2020; Greenstone et al., 2022; Shimshack and Ward, 2005; Zou, 2021). In vehicle markets, related work documents how manufacturers evade emissions rules and the health costs of their cheating (Ale-Chilet et al., 2025; Alexander and Schwandt, 2022; Reynaert and Sallee, 2021), and how effective inspection programs are at reducing emissions (Sanders and Sandler, 2020). Theory shows that the second-best policy toward pollution depends on how closely illegal disposal substitutes for the legal options (Fullerton and Stechuk, 2026). We document noncompliance by the inspectors themselves and estimate these substitution patterns for vehicle emissions, where drivers choose among repair, fraud, and letting their registration lapse.

The rest of the paper proceeds as follows. Section 2 describes the inspection program and our data. Section 3 develops our method for measuring fraud, and Section 4 documents how the market responded to the crackdown. Section 5 presents the model, Section 6 describes its estimation, and Section 7 reports the estimates and counterfactuals. Section 8 concludes.

2 Setting and Data

2.1 Texas Vehicle Emissions Inspection Program

The Clean Air Act requires regions that fail to meet federal air quality standards to operate vehicle inspection and maintenance (I/M) programs.[5] Many such programs are decentralized: rather than running government testing facilities, states delegate inspections to privately operated auto service stations. We study Texas, which operates one of the largest I/M programs in the country. Under the state's Clean Air Act Implementation Plan, 17 counties in the Dallas–Fort Worth, Houston–Galveston–Brazoria, Austin, and El Paso metropolitan areas require annual emissions inspections.[6] Owners of vehicles from model year 1996 and later must pass an annual test of the vehicle's on-board computer to renew their registration; a vehicle that fails must be repaired and retested before registration can be renewed.

Two state agencies oversee the program. The Texas Commission on Environmental Quality (TCEQ) sets policy and maintains the central inspection database (TIMS). The Texas Department of Public Safety (DPS) licenses stations and inspectors, audits them, and enforces compliance through administrative and criminal sanctions. Violations of Texas Transportation Code Chapter 548 carry both: DPS may suspend or revoke a station or inspector license and levy civil penalties.[7]

Inspections themselves are conducted by privately licensed stations, and the inspection fee is capped by regulation at $18.50 for an emissions test.[8] Given the low inspection fees, stations typically enter the inspection market not to profit from the tests themselves but to attract customers for repair and service work.[9]

2.2 Driver Compliance Choice: Repair, Fraud, or Lapse

Each year a driver must pass an emissions inspection to register their vehicle within any program county. In most cases, a driver's vehicle passes an honest emissions inspection on the first try, and the driver simply renews the registration. These vehicles already have a functioning emissions system, and hence, do not face a meaningful compliance choice.[10] In this paper, our main focus is on *malfunctioning vehicles*: vehicles with an emissions malfunction at the start of the compliance period that a legitimate inspection would detect, whether or not an inspection is ever performed. Importantly, we use the *malfunctioning*

[5] About 28 U.S. states and the District of Columbia operate I/M programs (U.S. Environmental Protection Agency, 2026b).

[6] The 17 emissions counties are Brazoria, Collin, Dallas, Denton, Ellis, Fort Bend, Galveston, Harris, Johnson, Kaufman, Montgomery, Parker, Rockwall, Tarrant, Travis, Williamson, and El Paso.

[7] In addition, falsifying a government record—such as an inspection report—is a felony and can be punished by criminal charges.

[8] The $18.50 cap applies in the Dallas–Fort Worth and Houston–Galveston–Brazoria program areas, the two major markets in the state (30 TAC §114.53); the cap is lower in Austin and El Paso.

[9] To become licensed, a station applies through a local DPS Vehicle Inspection Office, meets facility and equipment standards, purchases an approved On-Board Diagnostic (OBD) analyzer, and establishes a data-sharing agreement with TCEQ. Station-type composition and summary statistics are in Appendix B.4.

[10] Section 3 describes how we separate honest passes from fraudulent ones; a vehicle that passes an inspection we do not classify as fraudulent is honest.They could choose to drive unregistered. Our counterfactuals in Section 7 focus on enforcement policies that are unlikely to change these drivers' choices.

vehicle label to describe the vehicle's true condition at the start of the compliance period, which is distinct from its recorded inspection result.[11] There are several ways a driver can learn their vehicle is malfunctioning: they could notice that their check-engine light is on, be told by a mechanic during a service visit, or fail a previous inspection.

The owner of a malfunctioning vehicle faces three options. The first option is **repair**: pay to fix the malfunction, retest, and renew the registration. The second option is **fraud**: obtain a passing certificate from a station willing to fake the inspection result for a fee (a "clean scan"), then renew the registration. The third option is **lapse**: skip passing the inspection altogether and forgo registration. The lapse option further splits into three terminal actions: *drive unregistered*, let the vehicle *sit idle* (e.g., not driven but kept in the garage), or *scrap* it (or sell it outside the regulated area). Hence, the five terminal outcomes—Repair, Fraud, Unregistered, Idle, Scrap—are the drivers' choice set in the structural model (Section 5).

Drivers who choose the fraud route face a further choice about which inspection station to use. Importantly, solicitations and advertisements for illegal clean scans circulate openly online, on Reddit, Craigslist, and Facebook Marketplace.[12] As we discuss in the next subsection, common clean scanning methods do not require the target vehicle to be physically present at the station. This fact, along with the ubiquity of online advertisements, means that drivers have many options for where to purchase fraudulent passes, and the station they ultimately choose could be far away from where they live.

2.3 On-Board Diagnostic Inspections and Fraud Indicators

Modern emissions inspections read the vehicle's On-Board Diagnostics (OBD-II) computer through an analyzer plugged into its diagnostic port. During an honest test, the analyzer downloads the vehicle's *electronic identity*: the electronic VIN ("eVIN") stored in the engine control module and an *electronic profile* that identifies the computer and its software. The analyzer also downloads the status of the *readiness monitors* (the computer's self-checks of the emissions controls), any stored diagnostic trouble codes, and whether the check-engine light is on (Appendix B.2 gives the pass rule and which vehicles must be tested). Most of our fraud indicators come from these data, which a fraudulent pass must fake.

An inspector cannot manually enter a pass: faking a pass requires feeding the analyzer data that looks like a passing vehicle. There are two widespread ways to fake an inspection result. The first uses an *OBD simulator*, a device that plugs into the analyzer in place of a vehicle and mimics a passing OBD system; simulators frequently report no valid eVIN. The second uses a *surrogate vehicle*: the analyzer is connected to a different, well functioning vehicle while the inspector types in the VIN of the malfunctioning vehicle, so the reported eVIN does not match the entered VIN. Because neither method requires the vehicle to be at the station, a station can fabricate inspections in volume: one analyzer can process many fraudulent inspections a day. Clean scanning is not unique to Texas: it has been documented across the U.S. (California Bureau of Automotive Repair, 2020; Georgia Office of the Attorney General, 2015; New Jersey Office of the Attorney General, 2016; U.S. Department of

[11] For example, a malfunctioning vehicle may still pass an inspection fraudulently.

[12] For example, in a January 2023 thread on the Plano subreddit a driver asks where to pass an inspection with the check-engine light on, and another user replies, "DM me I know a guy that will pass anytime." See Reddit, r/plano (2023).

Justice, Eastern District of California, 2024).

Both OBD simulators and surrogate vehicles leave traces in the inspection database, and we use four of them as *fraud indicators*. (1) An *eVIN anomaly* flags a test whose downloaded eVIN is missing or does not match the VIN the inspector entered: a simulator often reports no valid eVIN, and a surrogate reports its own. (2) A *profile double match* flags consecutive tests on the same analyzer that report identical electronic profiles for nominally different vehicles, as happens when one simulator or surrogate is used for several cars in a row. (3) A *communication-protocol mismatch* flags a test whose OBD communication protocol is not the one the vehicle's make, model, and year should use, as happens when the device connected to the analyzer is not the car on the record. (4) A *round odometer* flags a passing test whose recorded mileage is an exact multiple of 1,000, which suggests that the inspector typed an estimate rather than reading the car's odometer.

No single indicator measures fraud on its own, because each mixes honest tests that trip it (false positives) with fraudulent tests it misses (false negatives). Section 3 combines them into a measure of fraud; Appendix C.1 details their construction and the candidates we considered and excluded.

2.4 Enforcement and the 2022 Crackdown

DPS monitors inspection stations in two ways. The first is overt audits: announced visits, which federal rules require at least twice a year at each station, in which DPS checks the station's records, licenses, equipment, and inspector proficiency (40 CFR §51.363; 37 TAC §23.51). The second is covert audits, in which DPS sends an undercover vehicle to a station to see whether it is inspected properly. How DPS chooses which stations to audit covertly is not public, but it could target them using patterns in the inspection data. When an audit uncovers a violation, DPS can suspend or revoke the station's or the inspector's license.

The state also monitors vehicles on the road. Under the Texas On-Road Vehicle Emissions Testing (TORVET) program, run for DPS and TCEQ by a contractor since 2001, roadside remote sensors mounted in vans are placed where traffic passes in a single lane while accelerating, at sites in the Dallas–Fort Worth and Houston program areas.[13] According to the program's contractor, sites are chosen to obtain a large number of readings of distinct, representative vehicles within its budget.[14] DPS can also order the owner of a high-emitting vehicle to pass an inspection, but rarely does so (Appendix B.5), so enforcement against clean scanning operated almost entirely through monitoring and sanctioning stations.

In August 2022, DPS quietly launched "Operation Cinderblock," a coordinated regulatory campaign against stations engaged in clean scanning. It did not announce the operation publicly until March 2023 (Texas Department of Public Safety, 2023).

During Cinderblock, enforcement rose sharply (Figure 1(a)). Before 2022, DPS ran about 3,400 covert audits a year but took few enforcement actions: 18 in 2019 and 4 in 2021. Actions then rose to 174 in 2022, 842 in 2023, and 609 in 2024, roughly a 6-fold increase. Covert

[13] Environmental Systems Products, "Texas" (program description), https://esp-global.com/en_US/RSD/Program-Descriptions/TexasRSD; TCEQ, "Vehicle Emissions Inspection Program Enforcement," https://www.tceq.texas.gov/airquality/mobilesource/vim/program_enforcement.html.

[14] Conversations with staff at Opus Inspection, the TORVET contractor, January 2026.

audits rose as well, peaking at 6,996 in 2023 before falling to 4,339 in 2024. By March 2023 DPS had suspended or revoked the licenses of hundreds of stations and inspectors.

Figure 1(b) shows that fraud indicator rates rose steadily from 2016 until 2022. eVIN anomalies, profile double matches and protocol mismatches flagged 5.6, 7.6 and 7.9% of passing inspections in 2016 and all peaked in 2022 at 9.9–10.5%, while round odometers flagged under 1% of passes throughout (0.4–0.8%). In mid-2022 the program's independent evaluation reported clear signs of clean scanning (Eastern Research Group, 2022). After the crackdown began, the indicator rates fell. Operation Cinderblock is a key source of variation we use to estimate the equilibrium effects of enforcement in the following sections.

2.5 Data

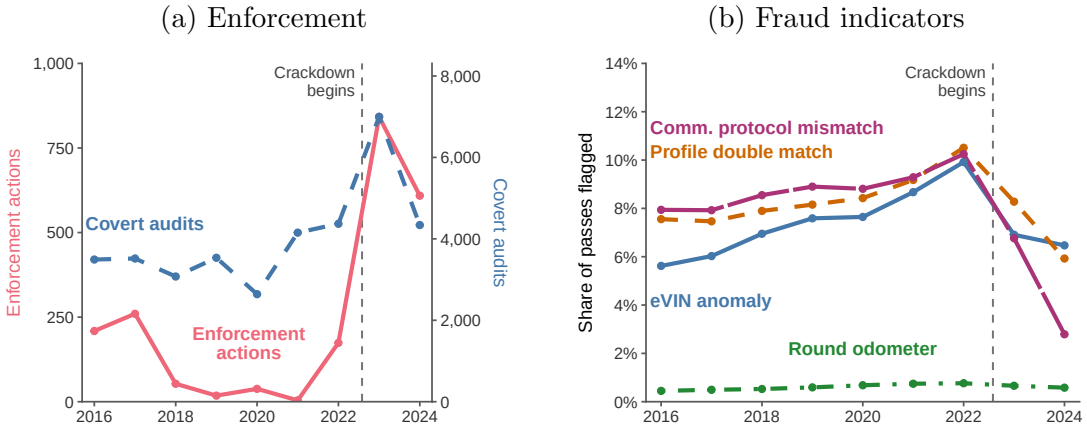
We use three main sources of data: inspection records, roadside remote sensing, and DPS enforcement records. Appendix B.1 gives further detail on each, along with the other minor supporting data sources. We organize the inspection records into *inspection cycles*. A pass ends a cycle, because the driver now has a certificate and can renew the registration. A fail does not: the driver still has a decision to make. An initial fail therefore tells the driver, and us, that the vehicle has a malfunction and which diagnostics failed, which we use to predict the cost of repair. The last test in a cycle records the decision itself: an honest pass, a fraudulent pass, or, if the vehicle never passes, a lapse. Section 3 explains how we detect whether a final pass was fraudulent. If a vehicle has not passed within 300 days of its first test, we treat the cycle as a lapse, and any later test starts a new cycle. Appendix B.3 describes the cycle construction.

Our main analysis covers 2016 to 2024. We estimate the fraud detection model of Section 3 on all 95,470,830 passing inspections in 2016–2024 in the 17 program counties, which span the Dallas–Fort Worth, Houston, Austin, and El Paso areas. The rest of the analysis, including the model of the market, uses only the Dallas–Fort Worth and Houston areas. These two areas account for most of the program, 84.0% of passing inspections (80,238,309), but we restrict the analysis to these metros because they are the only areas with roadside remote-sensing data, which we need to tell vehicles driven unregistered from those that sit idle or are scrapped, and to measure on-road emissions.

3 Measuring Fraud

Section 2 described how an inspector can fraudulently pass a malfunctioning vehicle, and how this leaves *fraud indicators* in the inspection data. We define a fraud indicator as a binary variable that is more likely to fire on a fraudulent inspection than on an honest one. An indicator marks inspections as suspicious, but on its own it can identify neither which inspections are fraudulent nor how many, because indicators produce false positives and false negatives. Two cars of the same popular model tested back to back share an electronic profile and set off the profile double match. In the other direction, most vehicles built before 2005 report no electronic VIN, so the eVIN indicator cannot be applied to them and misses any fraud on older cars. An indicator's firing rate therefore mixes three unknowns: how much fraud there is, how often fraud trips the indicator, and how often honest inspections do.

Figure 1: Enforcement and Fraud Indicators, 2016–2024



Note: Panel (a): DPS enforcement actions against stations (suspensions, revocations, and closures; left axis) and covert audits (right axis), by year. Panel (b): share of passing OBD inspections in the program's counties flagged by each of the four fraud indicators, by year; the sample is the one the detection model of Section 3 is estimated on (95,470,830 inspections). The vertical line marks the launch of Operation Cinderblock in August 2022.

This section develops a model that uses a set of fraud indicators to infer which passing inspections were fraudulent. We show that if the indicators are independent of one another conditional on an inspection's fraud status, the way the indicators co-move identifies fraud rates. Our model is general and can be used in many settings where a researcher observes several imperfect indicators of hidden misconduct. We present the model in Section 3.1 and its estimates in Section 3.2. We then compare the on-road emissions of vehicles that passed fraudulently with those of similar vehicles that were repaired, which both validates the measure and quantifies the pollution damages of fraud (Section 3.3). Finally, we test the identifying assumption and run several further checks (Section 3.4).

3.1 Fraud Detection Model

Each passing inspection i takes place at a station s, in a given year, on a vehicle of type v.[15] We fit the model separately by year and suppress the year index throughout. Inspection i is either fraudulent, $F_i = 1$, or honest, $F_i = 0$, and we never observe F_i directly. Instead, we observe K fraud indicators $Y_{ij} \in \{0, 1\}$, $j = 1, \ldots, K$, each equal to one when indicator j fires on inspection i. We group inspections into cells (s, v), and the model has three sets of primitives. The *fraud share* $\rho_{sv} = \Pr(F_i = 1 \mid s, v)$ is the share of a cell's inspections that are fraudulent. The *detection rate* $\beta_{jv} = \Pr(Y_{ij} = 1 \mid F_i = 1, v)$ is the probability that indicator j fires on a fraudulent inspection (one minus its false-negative rate), and the *false-positive rate* $\Lambda_{jv} = \Pr(Y_{ij} = 1 \mid F_i = 0, v)$ is the probability that it fires on an honest one. Both rates vary with vehicle type, because whether an indicator fires can depend on the vehicle's electronics, but we assume they are common to all stations.[16] We likewise treat

[15] A vehicle type can contain any observed vehicle characteristic, such as make, model, age, or mileage. In our application it is a make and a three-year model-year bin.

[16] The assumption is reasonable for the false-positive rate because nearly all inspections run on analyzers from two manufacturers (three before 2020), so the equipment that produces false positives is the same across

all fraudulent inspections of a vehicle type as sharing the same detection rates, whatever method produced them, and all honest inspections as sharing the same false-positive rates. The probability that indicator j fires on an inspection in cell (s, v), $\bar{Y}_{j,sv} = \Pr(Y_{ij} = 1 \mid s, v)$, is then

$$\underbrace{\bar{Y}_{j,sv}}_{\text{observed firing rate}} = \underbrace{\rho_{sv}}_{\text{fraud share}} \cdot \underbrace{\beta_{jv}}_{\text{detection rate}} + \underbrace{(1-\rho_{sv})}_{\text{honest share}} \cdot \underbrace{\Lambda_{jv}}_{\text{false-positive rate}} . \tag{1}$$

A simple example shows why observed firing rates alone cannot identify fraud. Consider a station at which indicator 1 fires on 8% and indicator 2 fires on 6% of inspections, and two scenarios that could produce these rates. In scenario A, no inspection is fraudulent ($\rho = 0$), so every firing is a false positive: 8% and 6% are the indicators' false-positive rates. In scenario B, 10% of inspections are fraudulent, indicators 1 and 2 detect 80% and 60% of them, and neither fires on an honest inspection. Equation (1) then gives the same firing rates, $0.1 \times 0.8 = 8\%$ and $0.1 \times 0.6 = 6\%$. From the firing rates alone, the two scenarios are indistinguishable, as are many other combinations of fraud share, detection rates and false-positive rates. Without further restrictions, or auxiliary data on the detection rates or the false-positive rates, the fraud share is not identified. Such data are rarely available, because measuring either rate requires knowing which inspections are fraudulent in the first place. Instead, we show that the model is identified under the following assumption, whose key part restricts the relationship among the fraud indicators.

ASSUMPTION 1 (Identification). (A) Conditional independence. *Within a cell (s, v), the fraud indicators are independent of one another given an inspection's fraud status:*

$$\Pr(Y_{i1}, \ldots, Y_{iK} \mid F_i, s, v) = \prod_{j=1}^{K} \Pr(Y_{ij} \mid F_i, s, v).$$

(B) Regularity. *(i) There are at least three fraud indicators. (ii) Each indicator is more likely to fire on a fraudulent inspection than on an honest one, $\beta_{jv} > \Lambda_{jv}$. (iii) Within each vehicle type, at least one cell contains both fraudulent and honest inspections, $0 < \rho_{sv} < 1$.*

To understand Assumption 1A, suppose an inspection is honest. Then each indicator fires only by chance, with probability equal to its false-positive rate, and the probability that two indicators both fire is the product of their false-positive rates, $\Pr(Y_{ij} = 1, Y_{ik} = 1 \mid F_i = 0, s, v) = \Lambda_{jv}\Lambda_{kv}$. Likewise, on a fraudulent inspection the probability that two indicators both fire is the product of their detection rates, $\beta_{jv}\beta_{kv}$. Importantly, the assumption restricts only how the indicators fire together given an inspection's fraud status. It allows fraud to be more common at some stations or on older vehicles, and the indicators to fire more often on some vehicle types than on others.

The assumption is likely to hold when the indicators pick up unique and unrelated signals of fraudulent activity. For example, on an honest inspection the profile double match fires when the previous car tested on the same analyzer happened to be the same model, while the protocol check fires only if the car's computer transmits a protocol other than the one expected for its VIN. The order in which cars arrive at the analyzer does not change

stations.

which protocol a car transmits, so the two false positives should arise independently. The assumption is most at risk for the eVIN and protocol indicators. Both are signals sent by the vehicle's computer that describe the vehicle's identity. Suppose an honest inspection is performed on a car whose engine and computer were replaced with used parts from a different model. Both the eVIN anomaly and the protocol indicators could then fire together, which would violate Assumption 1A. We nonetheless use both, because each covers inspections the other cannot.[17] Section 3.4 shows that this co-firing barely moves our estimates.

Assumption 1A pins down how often any two indicators fire together. Define the covariance of indicators j and k in a cell as the share of inspections on which both fire minus the product of their firing rates, $\Pr(Y_{ij} = 1, Y_{ik} = 1 \mid s, v) - \bar{Y}_{j,sv}\,\bar{Y}_{k,sv}$. Under the assumption, this covariance factors into three terms (Appendix C.3):

$$\operatorname{Cov}(Y_{ij}, Y_{ik} \mid s, v) \;=\; \underbrace{\rho_{sv}(1-\rho_{sv})}_{\text{fraud mix}}\;\underbrace{(\beta_{jv}-\Lambda_{jv})}_{\substack{\text{informativeness}\\ \text{of } j}}\;\underbrace{(\beta_{kv}-\Lambda_{kv})}_{\substack{\text{informativeness}\\ \text{of } k}}. \tag{2}$$

The first term, the variance of fraud status in the cell, is zero at a station that performs only honest or only fraudulent inspections and positive where the two mix. The other two terms measure how much more often each indicator fires on fraudulent than on honest inspections. The covariances between indicators are therefore the key evidence of fraud. The covariance separates scenarios A and B in our earlier example. In scenario A, where there is no fraud, the two indicators fire together on $8\% \times 6\% \approx 0.5\%$ of inspections, exactly what chance predicts, so their covariance is zero. In scenario B, a fraudulent inspection can set off both indicators at once, so they fire together on $0.1 \times 0.8 \times 0.6 = 4.8\%$ of inspections. A single covariance, however, cannot distinguish the amount of fraud from how informative the indicators are. A little bit of fraud caught by sharp indicators produces the same covariance as substantial fraud caught by weak ones. Under Assumption 1, the fraud share in every cell and the detection and false-positive rates of every indicator are identified (Appendix C.3 gives a constructive proof).[18] A third indicator identifies the model. The model has seven unknowns: the fraud share and a detection and a false-positive rate for each indicator. With three indicators, we obtain seven moments: three firing rates, three pairwise covariances, and how the three indicators move together. That three-way co-firing provides the additional equation that separates the fraud share from the informativeness of the indicators (Equation (22) in Appendix C.3).

In our preferred specification we use four indicators, the ones described in Section 2.3: eVIN anomaly, profile double match, communication protocol and round odometer. A fourth indicator adds eight moments but only two unknowns. The extra moments are overidentifying restrictions (Equation (23) in Appendix C.4.2), which let us test Assumption 1A directly in Section 3.4.

We estimate the model by maximum likelihood with the EM algorithm, using the probability of every firing pattern implied by Assumption 1A (Appendix C.2). The estimation returns the fraud share $\hat{\rho}_{sv}$ of each cell and the detection and false-positive rates of each

[17] The eVIN indicator cannot be applied to most vehicles built before 2005, which report no electronic VIN, while the protocol check applies to every model year. Conversely, the protocol check loses about half of its signal in 2024 (Appendix C.5.1).

[18] See Allman et al. (2009, Corollary 2).

Table 1: Indicator Detection Rates and the Fraud Rate by Vehicle Age

	Indicator				
	eVIN	Profile	Protocol	Odometer	
Panel A. Indicators, 2016–2024					Fraud rate (%)
Observed firing rate (%)	7.36	8.18	7.92	0.61	9.52
Detection rate (%)	65.4	73.8	80.0	3.2	
False-positive rate (%)	1.10	1.18	0.11	0.33	
Share of firings due to fraud (%)	86	87	99	51	
Panel B. Detection rate by vehicle age (%)					Share of passes (%)
Age 2–15	85.7	73.4	81.3	2.9	86.3
Age 16–19	30.4	74.5	74.5	3.8	9.6
Age 20–24	–	76.8	85.9	4.5	4.1
All ages 2–24	65.3	73.9	80.2	3.2	100.0
Panel C. False-positive rate by vehicle age (%)					Fraud rate (%)
Age 2–15	1.16	0.75	0.05	0.28	7.6
Age 16–19	1.14	3.43	0.46	0.65	22.2
Age 20–24	–	6.71	0.77	1.00	22.1
All ages 2–24	1.12	1.18	0.11	0.34	9.6

Note: Passing inspections in the 17 program counties, 2016–2024 (95,470,830 inspections). Detection and false-positive rates are averaged over vehicle-type-years, weighted by each inspection's posterior probability of fraud and of honesty, respectively. The fraud rate is the share of inspections classified as fraudulent ($\hat{\tau}_i > 0.5$). Panels B and C drop the 1.59% of inspections of vehicles aged below 2 or above 24; their last columns give each age bin's share of inspections and its fraud rate. The eVIN indicator is not defined before model year 2005, so its 20–24 cells show a dash and its other rates include model years it cannot fire on.

indicator. From these we compute, by Bayes' rule, the posterior probability $\hat{\tau}_i$ that each inspection was fraudulent given which indicators fired on it (Equation (18) in Appendix C.2.3). The posterior combines how common fraud is at the station with how much more likely the inspection's firing pattern is under fraud than under honesty. At a station with substantial fraud, an inspection on which the eVIN anomaly and the profile double match both fire has a posterior near one, and an inspection on which no indicator fires has one near zero. We classify an inspection as fraudulent when $\hat{\tau}_i > 0.5$, that is, when fraud is more likely than not.

3.2 Fraud Detection Estimates

Table 1 summarizes the estimates. We classify 9.5% of passing inspections in the program counties as fraudulent between 2016 and 2024 (standard error 0.46 percentage points; Appendix Table A.2). The classification barely depends on the cutoff of one half, because only 0.4% of inspections have a posterior probability of fraud between 0.2 and 0.8 (Appendix Table A.3).

No single indicator captures all fraud. Each indicator fires on fewer inspections than we

classify as fraudulent (Panel A), because none of them detects every fraudulent inspection. Their false-positive rates are low, between 0.11% and 1.18% of honest inspections, but not negligible. The indicators differ in the ways their mechanics predict. The protocol indicator detects the most fraud, 80.0% of fraudulent inspections, and almost never fires on an honest one. The profile double match detects less, 73.8%, because it misses the first fraudulent test after an inspector switches surrogate vehicles or runs an honest test.

Detection also differs across vehicles, and the model accounts for it. The eVIN indicator detects 85.7% of fraudulent inspections of vehicles aged 2–15 years and cannot be applied at ages 20–24, because it cannot fire on vehicles built before 2005 (Panel B). Detecting fraud with eVIN anomalies alone would therefore miss most fraud on older vehicles. The model instead learns their detection rates from the other indicators, and we find that fraud is most concentrated on exactly those older vehicles: 22.2% and 22.1% of inspections of vehicles aged 16–19 and 20–24 years are fraudulent, against 7.6% at ages 2–15.

Fraud is also concentrated in a small number of stations (Appendix Figure A.2). In 29.9% of station-years we classify no inspection as fraudulent, and the median station-year has a fraud rate of 0.2%. At the other extreme, the 5.4% of station-years in which at least 50% of passes are fraudulent perform 77.8% of all fraudulent inspections.

Many of the stations selling most of the fraudulent inspections are newly licensed: stations in their first two years of operation perform 21.8% of passing inspections but 55.4% of fraudulent ones (Appendix Figure A.3, with more detail in Appendix A.2). Stations that sell fraud also sit in lower-income, central parts of each metropolitan area: stations in the poorest quarter of ZIP codes perform a quarter of inspections but 63.1% of fraud (Appendix Figures A.3c and A.4).

The fraud rate differs across the four metropolitan areas. Over 2016–2024 it is 10.4% in Dallas–Fort Worth, 11.2% in Houston, 9.1% in El Paso and only 0.3% in Austin (Appendix Figure A.5).[19] Dallas–Fort Worth and Houston account for 94.7% of fraudulent inspections.

3.3 Pollution and Damages

We compare the on-road emissions of vehicles whose last inspection was a fraudulent pass with those of similar vehicles that passed after a repair, using the remote-sensing data (Section 2.5). The remote-sensing data measure emissions independently of the inspections, so these regressions check our fraud classification. Vehicles we classify as fraudulently passed should emit more on the road than repaired vehicles or vehicles that passed honestly at their first test. Our regressions compare vehicles of the same make and model year, observed on the same day and under the same driving conditions (Appendix D.1). A vehicle that passed honestly at its first test emits on average 1.09 grams of nitric oxide (NO) and 0.48 grams of hydrocarbons (HC) per kilogram of fuel. Table 2 shows that a fraudulently passed vehicle emits 2.37 grams more NO (+216%) and 1.03 grams more HC (+214%). A repaired vehicle emits about as much as an honestly passed one, so a repair returns a malfunctioning vehicle to the honest level. We call the difference between fraudulently passed and repaired vehicles the excess emissions of a fraudulent pass. The main threat to giving these comparisons a

[19] Since 2017 a Clean Air Task Force in the Travis County Constable's office has investigated and prosecuted inspection fraud (KXAN Austin, 2023), which likely explains the low rate in Austin.

Table 2: Roadside Emissions after Fraudulent and Repaired Passes

	NO (g per kg of fuel)		HC (g per kg of fuel)	
	(1)	(2)	(3)	(4)
Fraudulent pass	2.37	1.74	1.03	0.77
	(0.03)	(0.09)	(0.02)	(0.07)
Repaired pass	0.09		−0.01	
	(0.02)		(0.02)	
Failed, not yet passed	1.14		0.39	
	(0.08)		(0.06)	
Fraudulent minus repaired	2.28	1.74	1.04	0.77
	(0.03)	(0.09)	(0.03)	(0.07)
Mean, honest first-time passes	1.09		0.48	
Driving conditions	✓	✓	✓	✓
Make, model-year and date fixed effects	✓	✓	✓	✓
Failed-test diagnostic fixed effects		✓		✓
Sample	All	Failed a genuine test	All	Failed a genuine test
Readings	2,199,150	86,815	2,199,150	86,815
Damages per fraudulent vehicle-year (2024 $)	174		77	
Total, NO_x and VOC	251			

Note: Each column regresses roadside emissions on the vehicle's last inspection outcome within 365 days, Dallas–Fort Worth and Houston, 2016–2024. Columns 1 and 3 use all readings, relative to honest first-time passes. Columns 2 and 4 keep vehicles that failed a genuine test, compare fraudulent passes after the failure with repaired passes, and hold the defect fixed with failed-test diagnostic fixed effects; their ratio to the same sample without those controls (0.85 NO, 0.82 HC) scales the damages. Readings are weighted by the inverse of the vehicle's readings that year; standard errors are clustered two ways, by vehicle and by day. Damages value adjusted excess emissions at county pollution damages (Appendix D).

causal interpretation is that drivers choose fraud when a repair is expensive, so fraudulently passed vehicles may have worse defects than repaired ones. Among vehicles that failed a genuine test, holding fixed the diagnostic profile of that test, recorded before the driver chooses between repair and fraud, lowers excess emissions by 15% for NO and 18% for HC (Appendix Table D.1, columns 7 and 8). Thus, the bias from worse defects appears modest. Appendix D.5 reports further robustness checks.

The excess emissions also give the environmental cost of fraud, an input to the counterfactuals in Section 7. Our estimates imply that about 1.0 million malfunctioning vehicles a year pass their emissions inspection fraudulently. We convert excess emissions into pollution damages per year of a fraudulent pass with the fuel a vehicle burns per mile, the miles it drives in a year and the marginal damage of each ton of pollution in the vehicle's county from the AP4 model (Muller and Mendelsohn, 2009; Muller et al., 2011). To be conservative, we lower the excess emissions of all fraudulently passed vehicles by the 15% (NO) and 18% (HC) that holding the defect fixed removes. We find that a fraudulently passed vehicle causes $251 of damages a year on average (Table 2, last row). Appendices D.3–D.4 describe the adjustment and how damages vary across vehicles.

3.4 Validation and Robustness

We next test the identifying assumption (Assumption 1A). It fails if two indicators share a cause on honest inspections, so that they fire together more often than independence implies even when no fraud takes place. We classify 22.6% of passing inspections at stations eventually sanctioned for clean scanning as fraudulent, against 0.36% at outlets of national auto repair chains that were never sanctioned. We also find little fraud among inspections of nearly new vehicles (1.9%), which rarely have a defect to hide (Appendix Figure C.1). The national chains should therefore provide a useful benchmark for how the indicators behave on honest inspections. There, the round odometer fires together with each of the other indicators about as often as independence predicts, while the eVIN anomaly and the protocol indicator, which both read the vehicle's computer, fire together more often (Appendix Table C.3). With four indicators, the data must also satisfy over-identifying restrictions on how the pairwise covariances relate. Testing each year separately and correcting for multiple testing, we reject none of the restrictions; pooled across years, the tests detect a small departure. These departures barely move our estimates. Re-estimating the model on simulated data with the strongest dependence that the indicators' firing rates allow changes the fraud rate by at most +0.51 percentage points. Merging the eVIN anomaly and the protocol indicator into a single indicator changes it by +0.38 points, against yearly fraud rates of 7.0–11.6% (Appendix C.4.3).

We run several further checks and find that our estimates barely depend on which indicators we use: replacing the round odometer with high station volume leaves the estimated fraud rate essentially unchanged (9.48% against 9.52%; Appendix C.1.4). Nor do the estimates depend on how we group vehicles into types, or on the 2024 change in how analyzers record the communication protocol (Appendix C.5). A limitation of the model is that it can detect fraud only if fraud leaves a signal in the indicators, but we find no evidence that stations got better at hiding fraud after the crackdown (Appendix Figure A.7; Appendix D.5). Finally, fraudulent failures, which a station could use to sell the driver a repair or a fraudulent pass, are rare: we classify only 1.3% of failed inspections as fraudulent, against 9.8% of passing inspections (Appendix C.6). We therefore focus the rest of our analysis on clean scans, the fraudulent passes of malfunctioning vehicles.

4 The Enforcement Crackdown: Descriptive Evidence

Section 3 measured which inspections are fraudulent. We now turn to the paper's main question: how does enforcement change the market for fraudulent inspections? Enforcement against illicit markets often targets prominent suppliers. Police dismantle a regional drug-trafficking network, only to see the same product reappear nearby from another seller. If supply responds this way, enforcement becomes a game of whack-a-mole. Such responses are hard to see in most illicit markets, where sales go unrecorded. In our setting, every inspection is recorded, and the detection estimates of Section 3 tell us which stations sold fraud and how much, before and after the crackdown that began in 2022 (Section 2.4).

Four facts stand out. First, fraud fell, but by much less than the scale of the crackdown would suggest: the regulator took hundreds of enforcement actions and eliminated nearly

all of the largest sellers, yet most of the fraud market survived. Second, new fraud stations, many of them under new licenses operating at the same locations, stepped in to replace the ones that were shut down. Third, the fraud that remained was sold by far fewer, much larger stations. Fourth, some drivers who lost access to their incumbent fraud supplier let their registrations lapse rather than repair their cars, keeping polluting vehicles on the road. Together, these facts show why removing sellers did not translate one for one into less fraud or more repairs.

4.1 Enforcement, Station Closures, and Fraud

We focus on Dallas–Fort Worth (DFW) and Houston, the two metropolitan areas where roadside sensors also observe vehicles on the road. Figure 2 traces six monthly measures from January 2019–December 2024.

DPS took almost no actions against inspection stations before the crackdown, even as fraud was rising in DFW and high in Houston (panels a and c). Its actions rose abruptly once the crackdown began in August 2022 and continued at a rapid pace after DPS announced the campaign publicly in March 2023 (the second dashed line). From the start of the crackdown to the end of 2024, DPS took 670 actions against stations in DFW and 875 in Houston.

Panel (b) shows how these actions translated into the number of major fraud stations in each market, which we define as stations selling more than 100 fraudulent passes in a month. Their number fell from 71 to 8 in DFW (89% fewer) and from 144 to 10 in Houston (93% fewer).[20] In Houston, where DPS took more actions and started them slightly sooner, major fraud stations had largely disappeared by mid-2023, while in DFW they dwindled more gradually over the following year. Because DPS shut these stations down, the crackdown cut the supply of fraud by different amounts in the two markets and at different times. This variation lets us see how demand for fraud responds when supply is cut, and we use it to estimate our model in Section 6.

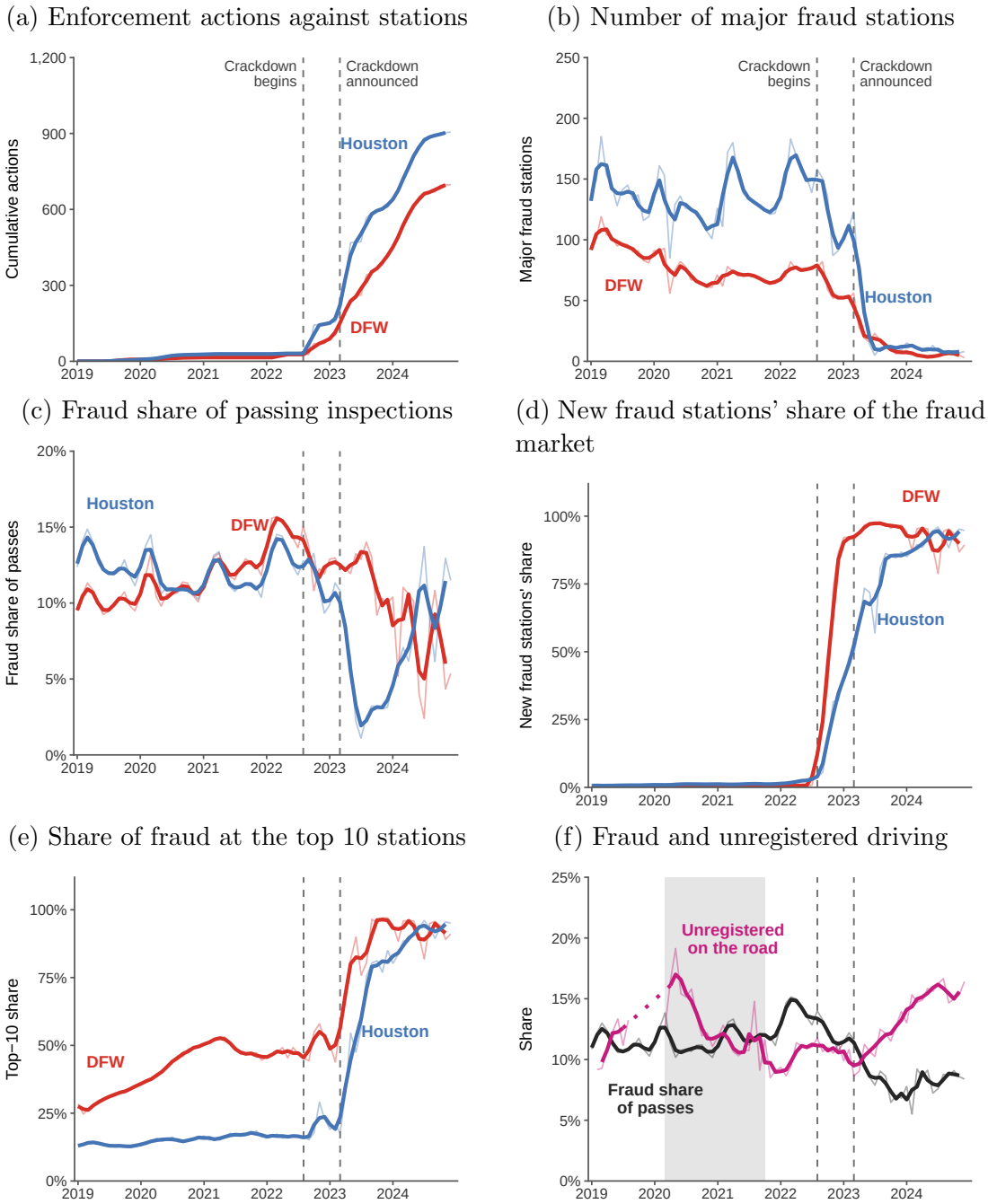
Despite the removal of so many fraud suppliers, fraud fell by far less (panel c). The share of passing inspections we classify as fraudulent fell from 13.2% to 9.1% in DFW (31% lower) and from 12.3% to 6.8% in Houston (44% lower). The fraud share collapsed in Houston in 2023 and then climbed back above DFW's by late 2024, while in DFW it held steady through most of 2023 and then fell only gradually.

4.2 Entry and Concentration

Panel (d) shows how quickly stations that had not sold fraud before the crackdown, most of them newly licensed, took over the market. After the crackdown, these new fraud stations sold 94.9% of the fraud in DFW and 91.2% in Houston (Appendix Figure A.8). Here, as throughout, a station is an inspection license, because DPS records inspections and takes action by license. We call a station a new fraud station if it first became a major fraud station in or after August 2022.

[20] Throughout this section, before-and-after numbers are averages over two 18-month windows, January 2021–June 2022 and July 2023–December 2024. The after window skips the crackdown's first year, while DPS was still closing stations, so that it reflects the market that emerged.

Figure 2: The Enforcement Crackdown and the Market for Fraud



Note: By station metro. Fraudulent pass: a vehicle's final passing inspection with posterior fraud probability above 0.5. (a) Cumulative DPS actions against stations since January 2019. (b) Stations selling over 100 fraudulent passes monthly. (c) Fraud share of passes. (d) Share at new fraud stations (first month as a major fraud station on/after August 2022). (e) Share at the top 10 stations. (f) Metros pooled; unregistered: roadside readings of Texas-registered vehicles expired over a month. Dotted where registration records were not updated (September 2019–February 2020). Grey band: COVID registration waiver and renewal backlog (March 2020–September 2021). Thick: 3-month centered average; thin: monthly. Text numbers: January 2021–June 2022 vs July 2023–December 2024. Dashed lines: crackdown begins (August 2022), announced (March 2023).

Many of the stations that sprang up to supply fraud during the crackdown had visible links to incumbent fraud stations that had just closed. A business that loses its license can reopen under a new one, often at the same address and with a new name on the paperwork. Of the new fraud stations that appeared in the year from August 2022, 66% began selling fraud at the same address, or with the exact same analyzer, as a fraud station that had just closed (Appendix E).

Turnover surged during the crackdown (Appendix Figure A.9 and Appendix A.3). In DFW, entry of major fraud stations rose from about 12 a quarter before the crackdown to 45 in its first year, and exit rose from 14 to 60. Before the crackdown, a typical major fraud station sold inspections for 31 months, whereas similar stations that started after the crackdown began lasted only 3 months.

Panel (e) shows that because station exit outpaced entry during the crackdown, the market also became much more concentrated. The share of fraud sold by the 10 largest fraud stations rose from 48.6% to 91.6% in DFW and from 16.6% to 85.8% in Houston. By 2024, just 3 stations in DFW and 3 in Houston sold half of each metro's fraudulent inspections. The Herfindahl–Hirschman index of station shares of fraud shows the same shift (Appendix Figure A.10).

The largest of the new fraud stations illustrates the scale at which some suppliers operated after the crackdown. Wynnewood Inspections, in DFW, opened on March 1, 2024 and over the rest of the year sold 137,657 fraudulent passes, about 450 for every day it was open. DPS revoked its license on March 7, 2025.

4.3 Unregistered Driving

While many drivers responded to the loss of their incumbent fraud supplier by moving to another seller, others responded by skipping the inspection altogether and letting their registration lapse. Panel (f) shows this response. It compares the fraud share of passing inspections with the share of vehicles seen on the road whose registration had expired more than a month earlier, measured by the roadside remote sensors (Section 2.4). Once the crackdown is announced, the unregistered share climbs as the fraud share falls. Pooled across both metropolitan areas, the fraud share fell by 4.7 percentage points, while the unregistered share rose from 10.7% to 14.6%. Notably, the two shares have different bases, so the rise does not tell us how many displaced fraud buyers let their registration lapse.[21]

The descriptive evidence cannot tell us how many fraud buyers the crackdown pushed into repair rather than unregistered driving, or how other enforcement policies would compare. Section 5 develops a model of both sides of the market to answer these questions.

5 A Model of the Fraud Market

The patterns in Section 4 suggest several margins that shape the equilibrium effects of crackdowns on inspectors who sell false certifications. The first is how readily regulated entities (drivers) can switch to another inspector (station) when the regulator shuts down the one

[21] Appendix Figure A.11 shows a smaller and later rise for registrations expired a year or more, which by construction cannot respond until a year after the crackdown.

they use. The second is whether regulated entities that stop buying false certifications come into compliance (repair their vehicles) or forgo certification altogether (let their registration lapse). This margin matters for the harm the inspections exist to prevent, because some entities that forgo certification keep causing that harm anyway (drivers who keep driving, and polluting, without a registration). The third is how readily new inspectors begin selling false certifications to fill the hole left by those the regulator shuts down. Each margin can undermine a crackdown aimed at reducing fraud and other noncompliance. To quantify these margins, and to predict what other enforcement policies would achieve, we develop a model of both sides of the market.

5.1 Environment and Timing

The market has three kinds of agents: the regulator, the licensed inspection stations, and the drivers of malfunctioning vehicles. A market is a metropolitan area in a year, indexed by m and t. Within a market, three things are fixed for the year: the population of drivers whose registrations come due, the regulator's policy, and the stations' plans and the qualities they draw. Within a year, two arrival processes govern how the market evolves. Each month, indexed by ℓ, new drivers arrive as their registrations come due, and the regulator's detections arrive, removing sellers from the market.

The regulator's policy has two levers (Section 5.4): its intensity, how likely it is to detect a station that sells fraud, and its severity, the penalty a detected station pays. Following Blundell et al. (2020), we take this policy as given, measure it from the regulator's actions (Section 6.2), and change it in the counterfactuals.[22] Our counterfactuals can therefore evaluate how stations and drivers respond to a policy the regulator puts in place, but not how the regulator itself would adjust its policy to a change in the environment.

At the start of the year, each station chooses whether to enter the market for fraudulent passes, knowing the regulator's current policy. Its risk of being caught depends on the policy and on how much fraud it sells, so the policy can deter fraud before any audit takes place. Each month, drivers arrive at random as their registrations come due, each with their own repair bill and tastes. Each driver faces one compliance choice a year, in the month the registration comes due: repair, lapse, or buy from one of the active sellers. Meanwhile, sellers exit: each month, the regulator detects some stations, which ends their operation for the year, and others stop at the end of the time they planned to operate. Because no one revisits a strategic choice between months, each month differs from the last only in which drivers arrive and which sellers remain, and Section 6.1 uses this variation to measure how readily drivers substitute among stations.

> Assumption 2 (Annual choices). *Stations make their strategic choices (whether to enter, how long to operate, and how much to sell each month) once, at the start of the year, to maximize that year's expected profit, and do not revise them as the year unfolds.*

[22] Blundell et al. (2020) take this approach to the EPA's enforcement of air-pollution rules at industrial plants. Duflo et al. (2018), by contrast, model the objective of an environmental regulator in India, which lets them derive its best policy but requires assumptions about what it values. The 2022 crackdown itself responded to rising fraud (Section 4); we take its timing as given (Section 6.2).

The assumption has two parts. The first is that stations look only one year ahead, so we model their choices as a static entry game played separately each year. We believe this is reasonable in our context because a license is cheap to replace. A new license requires a $100 fee, an approved analyzer, and premises that can be as modest as a storage unit or a strip-mall suite, and a station that is shut down can reopen under a new license (Appendix E). Many new fraud stations opened at the site of one that had just closed (Appendix E). A license that is easy to replace carries little value beyond the current year, so a static entry model is a reasonable approximation of the incentives sellers face.

The second part is that stations do not revise these choices within the year. In many markets this would be hard to defend: when a rival closes, the remaining sellers raise their prices or expand. Here, stations learn only slowly which rivals have closed. The regulator publicizes its enforcement actions only in monthly reports, which can arrive months late: during the crackdown, the reports covering February through July 2023 were all produced on a single day in late September, seven months after the previous report.[23] The market also has many sellers, and many of them operate from inconspicuous storefronts with no online presence. During the crackdown, when this matters most, closures came quickly and in large numbers (Section 4), so stations had little way to track which rivals remained. The assumption therefore rules out surviving sellers raising their prices or expanding within the year after their rivals close. Stations do respond to the closures they expect: a station's choice at the start of the year accounts for the risk that it and its rivals will be caught during the year. The model is therefore suited to measuring how behavior changes from one year to the next, not responses to short-run changes within a year.

5.2 Demand

Each year, every vehicle's registration comes due. If the vehicle is malfunctioning at that time, its driver chooses among three options: repair the vehicle, buy a fraudulent pass, or let the registration lapse. A driver who lets the registration lapse drives the vehicle unregistered, keeps it idle, or scraps it.[24] The distinction between the three lapse options is relevant for pollution and welfare because a vehicle driven unregistered keeps polluting, while an idle or scrapped vehicle does not. Driver i's utility from option k is $u_{ik} = v_{ik} + \varepsilon_{ik}$, where ε_{ik} is an idiosyncratic preference shock, distributed type-I extreme value, and

$$v_{ik} = \begin{cases} -\alpha_i\, r_i + \mathbf{x}_i'\boldsymbol{\beta}_{\text{repair}} - \eta_R\, \mathbf{1}\{o_i \in \{\text{fraud}, \text{lapse}\}\} & \text{if } k = \text{repair}, \\ \mathbf{x}_i'\boldsymbol{\beta}_{\text{fraud}} + \eta_F\, \mathbf{1}\{o_i = \text{fraud}\} + \lambda_F \log S_i & \text{if } k = \text{fraud}, \\ \mathbf{x}_i'\boldsymbol{\beta}_k + \eta_L\, \mathbf{1}\{o_i = \text{lapse}\} & \text{if } k \in \{\text{unregistered}, \text{idle}, \text{scrap}\}. \end{cases} \tag{3}$$

The term S_i in the fraud utility is the inclusive value of the stations selling fraudulent passes, which we define in equation (4) below.

[23] We date each report by the production date recorded in its file, since a report cannot be posted before it is produced. In 26 of the 69 months in our sample, the report was produced more than a month after the month it covers.

[24] Drivers of vehicles that are not malfunctioning may also scrap, idle, or drive unregistered, but we assume that their choices do not respond to the enforcement policies we consider.

The first term of the repair utility is the cost of coming into compliance: the repair bill r_i that the driver faces to pass an honest inspection, weighted by the driver's price sensitivity α_i. Repair costs vary widely across drivers because vehicles malfunction for different reasons. One vehicle may need a new oxygen sensor, which is cheap, while another needs a catalytic converter replacement, which is expensive. Hence, the cost of repair depends on which diagnostics the vehicle fails. Together with differences in price sensitivity across drivers, the distribution of repair costs determines the share of drivers who repair.

The next term in each utility, $\mathbf{x}_i'\boldsymbol{\beta}_k$, is the mean utility of option k for a driver with characteristics $\mathbf{x}_i$. For repair it is the value of complying with the law and of a repaired vehicle; for fraud, any distaste for breaking the rules and any penalty the driver expects;[25] and for the lapse options, the cost of driving unregistered or of losing the vehicle's use. Each utility is measured relative to scrapping the vehicle ($\boldsymbol{\beta}_{\text{scrap}} = \mathbf{0}$).

The next set of terms captures inertia in drivers' choices. Many drivers make the same choice year after year. Although our model is static, we let a driver's previous choice shift the current one in reduced form, as in the literature on consumer inertia (Abaluck and Gruber, 2016; Dubé et al., 2010; Handel, 2013). The variable o_i records how the vehicle's previous inspection cycle ended, with an honest pass, a fraudulent pass, or a lapse. The terms η_R, η_F, and η_L shift the utility of repair, fraud, and lapse for a driver whose previous cycle ended without a repair, in fraud, or in a lapse. The specification is agnostic about whether the inertia comes from learning, search costs, or persistent preferences.

The last term in the fraud utility of equation (3), $\lambda_F \log S_i$, reflects the driver's choice of which station to buy from. It is the inclusive value of a nest that contains every station selling fraudulent passes when the driver makes the choice. As in a standard nested logit (Berry, 1994; McFadden, 1978), $\lambda_F \log S_i$ is the expected utility of the best station available to the driver, where

$$S_i \;=\; \sum_{s \in \mathcal{A}_{m\ell t}} \exp\!\left(\frac{v_{is}}{\lambda_F}\right) \qquad \text{and} \qquad v_{is} \;=\; \delta_{smt} + \zeta_{is} - \alpha_i\, p_{smt}. \tag{4}$$

The set $\mathcal{A}_{m\ell t}$ contains the stations selling in market m in month ℓ of year t, the month in which driver i's registration comes due. It includes the large stations, each with its own quality; the smaller stations, which together form a competitive fringe, a single alternative with a common quality; and a composite option that encompasses all stations outside the driver's home metro area. The utility of buying from station s has three parts. The first, δ_{smt}, is the station's quality. A fraudulent pass is close to a homogeneous good, but stations differ in how easily buyers find them: some advertise on Facebook Marketplace, and some may receive referrals from used-car dealers, which could give a few stations far higher volumes than others. We think of quality as mostly capturing this visibility. The second, ζ_{is}, is a match value between the driver and the station, which can capture the travel cost of reaching a station (Appendix H.3). The third is the price p_{smt} the station charges for a pass, weighted by the driver's price sensitivity α_i, the same sensitivity that applies to the repair bill.

The nesting parameter λ_F governs how much the inclusive value grows as more stations, or more visible ones, sell fraudulent passes. When λ_F is near one, drivers' tastes for different stations are nearly independent. Drivers then value variety: each additional station gives

[25] The regulator rarely penalizes drivers, and the risk changes little over our sample (Appendix B.5).

some drivers a better match, so more stations make fraud more attractive. When λ_F is near zero, drivers' tastes for different stations are highly correlated, so a market with few stations offers drivers similar convenience to a market with many. This parameter is crucial for the counterfactuals, because it determines how many drivers substitute to another seller, rather than repair or lapse, when the regulator shuts down stations.

5.3 Supply

We model stations' decisions as a two-stage game played each year, following entry models in other settings (Berry, 1992; Bresnahan and Reiss, 1991; Jia, 2008; Seim, 2006). The two stages correspond to the two margins on which the crackdown changed supply, which stations sold fraud and for how long (Section 4). In the first stage, at the start of the year, each potential seller decides whether to enter the market for fraudulent passes. It knows its type, the enforcement policy, and its own cost of entering, but not its quality or rivals' decisions. In the second stage, each station that entered learns its quality and plans its year: how many months to operate and how much to sell in each. Detection by DPS, with a probability set by its enforcement regime, can end a station's operations before it completes its plan.

Potential sellers are of two fixed types, $\omega_s \in \{\text{large}, \text{fringe}\}$. A large seller runs a high-volume operation and is a separate alternative for drivers. Fringe sellers sell on a small scale and together form the competitive fringe of Section 5.2, a single alternative. The fringe sells at a going rate, the price that a small seller sets against the demand for its own passes when it takes its rivals' prices as given. It divides its sales equally among the fringe sellers operating, so a fringe seller does not plan quantities. Each market m has a set of potential sellers of each type in year t, $\mathcal{S}_{mt}^{\text{large}}$ and $\mathcal{S}_{mt}^{\text{fringe}}$. The stations drivers can buy from in month ℓ, $\mathcal{A}_{m\ell t}$ in equation (4), are therefore the large sellers operating that month and the fringe, with $N_{m\ell t}^{f}$ fringe sellers operating. The composite option of Section 5.2, the stations outside the driver's home metro area, also sells at the going rate, and we do not model their entry.

> Assumption 3 (Capacity and a posted price). *A large seller plans the quantity it will sell in each month it operates and posts one price for the year, the price at which it expects those quantities to sell. Each month it sells what drivers demand at its posted price, up to the quantity it planned.*

The assumption reflects how these operations work: a station commits its capacity in advance, through its analyzers and its hours of operation, and posts a price for its passes. Capacity chosen in advance, with prices set afterward, is the classic rationale for quantity competition (Kreps and Scheinkman, 1983). Within a year a seller cannot expand beyond its plan to serve the drivers whose seller DPS closed, so the restriction matters for the transition rather than for the 2024 comparisons. We maintain the same conduct in estimation and in every counterfactual.

We derive stations' profits from the competition among the stations that enter. A station's quality, θ_{smt}, is the value drivers place on buying from it apart from its price. It equals the station's own δ_{smt} in equation (4) plus the taste for fraud common to all drivers in the market and year, the part of $\mathbf{x}_i'\boldsymbol{\beta}_{\text{fraud}}$ in equation (3) that every seller shares. For a fringe seller it is the fringe's common quality, θ_{mt}^{f}. A station of type ω and quality θ that operates

for T months and plans to sell q_ℓ passes in month ℓ expects to earn

$$\Pi_{mt}(T, q; \omega, \theta) = \mathbb{E}\Big[\big(1 - \chi\, \mathbf{1}\{d \le T\}\big) \sum_{\ell=1}^{\min(T,d)} p_{mt}(q; \omega, \theta)\, q_\ell\Big]. \tag{5}$$

The station collects revenue in each month of its plan until DPS detects it, and detection costs it a penalty of χ per dollar it has earned that year. Here, $q = (q_1, \ldots, q_T)$ lists the station's planned quantities, and $p_{mt}(q; \omega, \theta)$ is the price it posts for the year (Assumption 3). The month in which DPS detects the station is d, with $d > 12$ if DPS does not detect it in year t. We treat revenue as profit before any penalty: once a station holds a license and an analyzer, the marginal cost of performing an additional fraudulent inspection is likely negligible. The expectation is over what the station does not know when it plans: which rivals enter, their qualities, and when DPS detects each of them and the station itself.

Once a station has entered and learned its quality, as entrants learn their productivity in Hopenhayn (1992), it chooses the plan that maximizes its expected payoff,

$$\big(T^*_{mt}(\omega, \theta),\, q^*_{mt}(\omega, \theta)\big) \;=\; \arg \max_{T \in \{1,\ldots,12\},\, q \in \mathbb{R}^T_+} \Pi_{mt}(T, q; \omega, \theta). \tag{6}$$

The plan weighs revenue against the risk of detection on two margins, months and passes (Appendix J.2 gives the first-order conditions). Each further month adds that month's revenue, but it also raises the penalty the station pays if DPS detects it. Each further pass in a month adds revenue, but it lowers the price at which the station expects its planned passes to sell. It also raises the station's chance of detection (Section 5.4).

At entry, a station weighs the payoff of its best plan, averaged over the qualities it might draw, against the cost of entering:

$$\Pi^*_{mt}(\omega) = \mathbb{E}_\theta\Big[\,\Pi_{mt}\big(T^*_{mt}(\omega, \theta), q^*_{mt}(\omega, \theta); \omega, \theta\big)\Big] - c_\omega, \qquad a_{st} = \mathbf{1}\big\{\Pi^*_{mt}(\omega_s) + \varepsilon_{st} \ge 0\big\}. \tag{7}$$

Here, $\mathbb{E}_\theta$ averages over the distribution of quality among stations of type ω in the market, and c_ω is the cost of entering for a station of that type. The indicator a_{st} equals one if station s enters, and ε_{st} is a shock to the payoff from entering that only the station observes and that does not depend on its plan. Stations of the same type are alike, apart from their payoff shocks, when they decide whether to enter, so Π^*_{mt} depends on a station only through its type. Through the plan and the entry decision, enforcement deters on three margins: fewer stations enter, those that enter operate for fewer months, and each sells less in the months it operates.

5.4 The Regulator

The regulator enters the model through the two levers of its enforcement policy (Section 5.1), the probability of detection and the severity of the penalty (Becker, 1968; Polinsky and Shavell, 2000). Each lever enters the station's payoff, equation (5), through a different term: intensity through the month of detection d, and severity through the penalty rate χ.

The first lever is intensity: how likely the regulator is to detect a station that is selling fraud. We summarize it by a monthly detection hazard $h_{s\ell t}$, the probability that the regulator

detects station s in month ℓ of year t if the station is operating. The hazard depends on the station's type and rises with its own fraud rate, its fraud volume relative to all the inspections it performs. For a seller of a given type and plan, it therefore sets the distribution of the detection month d in equation (5). A station that sells more raises its own risk, which is the marginal cost of volume in Section 5.3. We estimate the hazard separately from the rest of the model, from the regulator's enforcement actions at the station-month (Section 6.2).

The crackdown raised intensity: the regulator detected stations far more often after it began (Section 4; Appendix I). We hold severity and stations' own primitives fixed across the crackdown, so the model attributes the change in supply to the change in intensity and to the shifts in demand that Section 5.2 measures. The counterfactuals vary intensity and severity separately. Each acts through equations (6) and (7), on the three margins of Section 5.3.

5.5 Information and Equilibrium

Each station knows the enforcement policy, the demand it faces, its type, and its own payoff shock ε_{st}, but not its rivals' shocks. It therefore forms beliefs about how likely potential sellers of each type are to enter, and it plans against the same beliefs once it knows its quality (Section 5.3).

Let $\phi_{mt}(\omega)$ denote the probability that each potential seller in $\mathcal{S}^{\omega}_{mt}$ enters market m in year t, and let $\phi_{mt} = \big(\phi_{mt}(\text{large}), \phi_{mt}(\text{fringe})\big)$. The payoff in equation (5) depends on these probabilities through the expectation over which rivals will operate and how much they plan to sell, so we write the value of a station's best plan as $\Pi^{*}_{mt}(\omega; \phi_{mt})$. In equilibrium, each rival's plan is T^{*}_{mt}, q^{*}_{mt} of Section 5.3 for a seller of its type and quality, the price it posts follows from that plan, and stations' beliefs about entry are correct: for each type,

$$\phi_{mt}(\omega) = \Pr\big\{\Pi^{*}_{mt}(\omega; \phi_{mt}) + \varepsilon_{st} \geq 0\big\}, \qquad \omega \in \{\text{large}, \text{fringe}\}. \tag{8}$$

Each type's entry probability is thus the probability that its best plan beats staying out, given the entry probabilities and the plans of all types. This is the Bayesian Nash equilibrium of entry models with private information (Bajari et al., 2010; Seim, 2006).

The model has three sets of primitives: drivers' preferences, including the distribution of sellers' quality; stations' entry costs, the distribution of their payoff shocks, and the sets of potential sellers; and DPS's two enforcement levers, the detection hazard and the penalty rate χ. Section 6 describes how we obtain each of them. Given the primitives, the model determines in equilibrium which stations enter and how long they operate, the quantities and prices of fraudulent passes each month, and drivers' choices among repairing, buying a fraudulent pass, and letting the registration lapse.

6 Estimation and Identification

In this section, we discuss how we take the model of Section 5 to the data: our specification choices, the details of implementation, and the key assumptions. We discuss the components of the model in turn: Section 6.1 covers demand, Section 6.2 the regulator's detection technology, and Section 6.3 supply.

6.1 Demand

The choice data and specification. We estimate demand using inspection records from two geographic markets, Dallas–Fort Worth and Houston, over 2016 to 2024. We treat each metro-year as a market, because only 6.1% of fraudulent passes are bought outside the buyer's home metro. Each driver makes approximately one compliance choice per year, one per inspection cycle as defined in Section 2.5. Although every driver in these metros must comply with the program, we model only the drivers of vehicles that are malfunctioning at the start of an inspection cycle. A driver whose vehicle is already compliant can simply pass an honest inspection, at the regulated fee, and register it. That option costs less than a fraudulent pass, so the driver has no meaningful choice to make.

A major challenge is that we do not directly observe which vehicles are malfunctioning. We identify malfunctioning vehicles in two ways. First, we classify a vehicle as malfunctioning if it fails the first inspection of an inspection cycle. By failing that inspection, the vehicle reveals that it was malfunctioning when the cycle began, and we follow what its driver does next. A vehicle also reveals a malfunction if its driver skips the failed inspection and buys a fraudulent pass directly, which we observe through that pass.[26] Second, some drivers already know that their vehicle is malfunctioning, from a failed test in a previous year, a visit to a mechanic, or a check-engine light. These vehicles are harder to identify in the data, so we rely on each vehicle's inspection history. We classify a vehicle as malfunctioning if its previous inspection cycle ended without an honest pass, in a failure that was never cleared or in a fraudulent pass. We call these carryover-malfunction vehicles, and all other malfunctioning vehicles new-malfunction vehicles. A vehicle whose previous cycle ended without an honest pass counts as carryover-malfunction even if it also fails its first inspection in the current cycle. Thus, a malfunctioning vehicle remains in the malfunctioning population until it passes an honest inspection. Appendix H.1 gives the full definitions and the resulting sample.

From the inspection data, we infer one of three outcomes for each malfunctioning vehicle in each inspection cycle: repair, fraud, or lapse. For a new-malfunction vehicle that fails an initial inspection, we look at what happens next. If the vehicle returns and passes an honest inspection, we infer that its driver repaired it. If it returns and passes an inspection the detection model of Section 3 flags as fraudulent, we infer that its driver bought a fraudulent pass. If it does not pass within 300 days of the failed inspection, we infer that its driver let the registration lapse. For a carryover-malfunction vehicle, an honest pass is a repair and a fraudulent pass is fraud, whether or not a failed inspection comes first. A carryover-malfunction vehicle that does not pass in a given year remains lapsed in that year.[27]

These choices depend on the covariates in the utility of Section 5.2 (equation (3)). The vector $\mathbf{x}_i$ contains the vehicle's age group, an indicator for mileage above the median for its age group, and metro-by-year fixed effects, with 2016–2018 pooled into a single period in each metro.[28] Each enters the utility of every alternative, relative to scrapping, with its

[26] We treat every fraudulent pass as evidence of a malfunction: the driver of a compliant vehicle has no reason to pay more for a pass it would obtain honestly. New-malfunction drivers who never test and never buy fraud leave no record; we return to them under Estimation below.

[27] A vehicle that is lapsed for only part of a year, for example because it passes partway through the year, counts as lapsed in proportion to the share of the year it spends lapsed (Appendix G.5).

[28] The remote-sensing data, which we use to split lapses into scrapping, idling and driving unregistered, are sparse in 2016–2018, when the sensors operated at a few sites for a few weeks a year (Appendix G.2).

own coefficient, so the attractiveness of repairing, buying fraud or lapsing can vary with a vehicle's age and mileage and across metros and years (Appendix H.2). The fraud alternative also contains fixed effects for the driver's home county (Appendix H.1).

We assume that price sensitivity is lognormal:

$$\alpha_i = \exp\left(\bar{a} + \sigma_\alpha \nu_i\right), \qquad \nu_i \sim N(0,1). \tag{9}$$

The lognormal form keeps every driver's sensitivity positive, so a higher repair bill always makes repair less attractive, and σ_α measures how much drivers differ in how much a given repair bill deters them.

The price of repair. Section 5.2 treats the repair bill r_i as the price of compliance. We observe the repair bill for a subset of drivers: inspectors can record repair costs when a repaired vehicle passes its retest (Appendix F.1). We use these reported costs to estimate a mapping from a vehicle's diagnostic readings at the failed inspection to its repair bill. The readings record what went wrong with the vehicle: whether the check-engine light was on, how many diagnostic trouble codes the onboard computer had stored, and which of its emissions self-tests had not completed. With this mapping, we can predict the repair bill $\hat{r}_i$ of any vehicle that we observe failing an honest inspection, from its own diagnostics, even when no cost was recorded for its repair. We call drivers who take their vehicle for an initial inspection and fail it *testers*. Drivers who never fail an honest inspection during the inspection cycle are, analogously, *non-testers*.[29]

For non-testers, we observe no diagnostics, so we cannot predict the bill in the same way. We integrate a non-tester's choice over the distribution of predicted bills among tested vehicles with the same age group, mileage, metro and year. This requires an assumption.

> Assumption 4 (Non-testers' repair costs). *Conditional on the covariates $\mathbf{x}_i$, testers and non-testers draw their repair bills from a common distribution, and they draw their price sensitivities from a common distribution.*

We need Assumption 4 for the model to capture the equilibrium and the supply-side incentives of the entire market for fraudulent passes. Stations compete to sell fraudulent passes to testers and non-testers alike, and non-testers make up 76% of the choices in our sample.

The assumption has two parts. The first is that testers and non-testers with the same characteristics face the same distribution of repair costs. We cannot test this part directly, but repair costs rise with a vehicle's on-road emissions, and untested fraud buyers emit about as much on the road as tested ones with similar characteristics (Appendix F.4).

The second part of the assumption restricts the shape of the heterogeneity in price sensitivity. For non-testers, we never observe repair costs, only their distribution given the drivers' characteristics, so their choices alone cannot reveal how price sensitivity varies across them. Importantly, a common distribution of price sensitivity does not mean that the two groups have the same preferences. The inertia terms of Section 5.2 let drivers' past choices shift

[29] The distinction is related to, but not the same as, the one between new- and carryover-malfunction vehicles. Most carryover-malfunction vehicles never take an honest test, but some return and fail one again, which makes them testers; and a new-malfunction driver who buys a fraudulent pass without first failing a test is not a tester.

their current choices, so testers and non-testers who face the same repair cost can make very different choices, and Section 7.1 shows that they do.

Fraud sellers. The utility of fraud depends in part on which stations are actively selling fraudulent passes. Each market has many inspection stations, and most sell only a small number of fraudulent passes or none at all. We therefore treat each large station as a separate option within the fraud nest, group the smaller stations in a metro into a competitive fringe, and include a composite option that encompasses all stations outside the driver's home metro area (Section 5.2). We classify a station as *large* in a year if it sells at least 1,000 fraudulent passes to vehicles in our sample.[30] We treat the stations in the fringe as identical, with a common quality and a common price that together enter through a single term, δ^f_{mt}, but the value of the fringe option grows with the number of fringe stations that are active. We allow for a simple driver-specific match: drivers can value a large station in their home county more highly than stations outside it (Appendix H.3). The inclusive value we estimate is therefore

$$S_i = \underbrace{\sum_{s \in \mathcal{A}^L_{m\ell t}} \exp\left(\frac{\delta_{smt} + \mu\, \bar{w}_{is}}{\lambda_F}\right)}_{\text{large stations active in the month}} + \underbrace{N^f_{m\ell t} \exp\left(\frac{\delta^f_{mt}}{\lambda_F}\right)}_{\text{fringe}} + \underbrace{\exp\left(\frac{\delta^{\text{out}}_{mt}}{\lambda_F}\right)}_{\text{out-of-metro composite}}, \tag{10}$$

where $\mathcal{A}^L_{m\ell t}$ is the set of large stations active in the month of the driver's choice, $N^f_{m\ell t}$ is the number of active fringe stations, and $\bar{w}_{is}$ indicates that station s is in driver i's home county, so that the match value of Section 5.2 is $\zeta_{is} = \mu\, \bar{w}_{is}$.

Since we do not observe the prices of fraudulent passes, each large station enters equation (10) through δ_{smt}, which combines the station's quality and its price, and we rely on the supply model to separate the two (Section 6.3). Each station charges one price for the year (Assumption 3), and its quality is fixed for the year, so δ_{smt} is fixed within each station-year. What changes from month to month is which stations are active, $\mathcal{A}^L_{m\ell t}$ and $N^f_{m\ell t}$, and this variation is what pins down λ_F, as we discuss below. We treat a station as active from the month of its first fraudulent sale to the month of its last, and each driver faces the stations active in the month of the driver's compliance choice (Appendix H.3).[31]

Scrapping, idling, and driving unregistered. The inspection data alone cannot tell which of the three lapse alternatives a driver chose, because a vehicle that is scrapped, parked, or driven unregistered simply stops appearing in the inspection records. We therefore combine two other pieces of information: whether a lapsed vehicle returns for an inspection in a later year, which shows that it was not scrapped, and the roadside remote-sensing data of Section 2.5, which show whether it is still being driven.

A vehicle that does not return, however, may have been scrapped, or it may still exist and remain lapsed. We assume that vehicles the roadside sensors record on the road, which

[30] Section 4 defines major fraud stations by monthly volume; the two groups largely overlap but are not identical.

[31] We therefore treat the month of the inspection as exogenous. This is reasonable because most drivers return for an inspection when their registration expires, and Appendix H.7 finds no evidence that drivers delayed their inspections to wait for a seller during the crackdown.

cannot have been scrapped, return for an inspection at the same rate as other surviving lapsed vehicles with the same characteristics. Only surviving vehicles can return, so the share that survive is the return rate of all lapsed vehicles divided by the return rate of those seen on the road. For example, if 20 percent of all lapsed vehicles return within a year but 40 percent of those seen on the road do, this implies that half of the lapsed vehicles survive. Among the survivors, we separate the vehicles driven unregistered from the idle ones by comparing how often the sensors record lapsed vehicles with how often they record comparable fraud-passed vehicles, which are registered and so on the road (Appendices G.2–G.3). Each lapse then enters the likelihood as a weighted combination of the three alternatives, with weights equal to the estimated probability that a vehicle with characteristics $\mathbf{x}_i$ is scrapped, idle, or driven unregistered (Appendix H.2).

Estimation. We estimate the parameters of demand by maximum likelihood, and our standard errors treat the first-stage estimates as known (Appendix H.2). For testers, we evaluate the probability of the driver's observed choice at the predicted repair cost, averaging over the distribution of price sensitivity. For non-testers whose vehicle was already malfunctioning, we also average over the distribution of repair costs given their characteristics. Among drivers of newly malfunctioning vehicles, we observe the untested fraud buyers, but drivers who repair or let the registration lapse without ever taking a test leave no record of a malfunction. We therefore assume that, among these drivers, non-testers choose similarly to testers with the same characteristics, so that the number of untested fraud buyers implies how many repaired or lapsed without testing (Appendix H.2).

Identification. Two parameters matter most for our results, and much of our empirical strategy is built to pin them down. The first is the distribution of price sensitivity, which measures drivers' willingness to pay to avoid a repair. It matters for welfare, and it determines how responsive buyers are to the price of a fraudulent pass and the incentives of the stations that sell them. The second is the nesting parameter λ_F, which governs what drivers do when the regulator shuts down stations, the central question of our counterfactuals.

We predict each tester's repair cost from the vehicle's diagnostic readings at the failed inspection, which record what is wrong with the vehicle. Our key identifying assumption is that, conditional on the covariates $\mathbf{x}_i$, a vehicle's diagnostic readings, and hence its predicted repair cost, are unrelated to its driver's unobserved tastes for the alternatives. The main threat is that repair costs are correlated with drivers' preferences. Older vehicles, for example, tend to need more expensive repairs, and their drivers may also have less income to spend on them, so comparing drivers of old and new vehicles would confound price sensitivity with income. We address this threat in two ways. First, we condition on the vehicle's age and mileage, so we compare drivers of similar vehicles. Second, we predict the repair cost from the diagnostics alone, for a typical vehicle at a typical station, so it varies only with the specific defect that caused the vehicle to malfunction (Appendix F.1).

Consider two drivers with similar vehicles, alike in age, mileage, and location. One vehicle fails because of a faulty oxygen sensor, the other because of a broken catalytic converter.[32]

[32]In independent retail quotes, an oxygen sensor replacement averages \$469 against \$2,347 for a catalytic converter. Inspection fees, by contrast, are fixed by regulation, unlike in Oliva (2015), who uses variation in

The second driver faces a much higher repair cost, and we compare the choices of these two similar drivers, who differ only in their cost of coming into compliance. How fast the share who repair falls as the repair cost rises pins down the mean of price sensitivity, and how widely that decline is spread over repair costs pins down its variance. Appendix Figure A.12 shows nonparametric fits of drivers' choice probabilities by predicted repair cost. At a predicted repair cost of \$159, 92.5% of drivers repair; at \$589, only 42.0% do.

The nesting parameter λ_F is identified from changes in the set of active stations within a year. Under Assumptions 2 and 3, each station's quality and price are fixed for the year, while the crackdown closed stations at different times. Consider two drivers with similar vehicles in the same county, one whose registration comes due early in 2023 and one whose registration comes due late in 2023. The crackdown shut down stations in between, so the two drivers face different sets of available stations, and our identification strategy is to compare choices across such drivers. If closing stations leads buyers to substitute toward buying fraud from other stations, λ_F will tend to be closer to zero; if they substitute toward repair or lapsing, λ_F will tend to be closer to one. Our key identifying assumption is that, conditional on the metro-year and home-county fixed effects, month-to-month changes in the set of active stations are unrelated to month-to-month shifts in the demand for fraud. The estimate of λ_F is robust when we control for a direct deterrent effect of the crackdown on the demand for fraud (Appendix H.4), and we find no evidence that drivers delayed their inspections to wait for new sellers (Appendix H.7).

6.2 Detection

The regulator enters the model through the monthly detection hazard $h_{s\ell t}$, the probability that DPS detects station s in month ℓ of year t while it sells fraud (Section 5.4). We estimate the hazard from DPS's record of enforcement actions at the station-month over 2016 to 2024, with each station's conduct measured by its monthly fraud rate from Section 3. The hazard is estimated before, and separately from, the supply model. It describes the regulator's behavior given a station's conduct, so the enforcement record identifies it on its own. The supply model then takes it as a primitive, and the counterfactuals vary it.

The hazard rises with a station's fraud rate through the two ways in which DPS enforces, covert audits and a screen of stations' inspection records (Section 2.4), and its parameters differ between the enforcement regimes before and after the crackdown (Appendix I.1).

Three sources of variation identify the parameters. Differences in fraud rates among stations in the same county and month identify the hazard's slope in the fraud rate, γ_r, and the difference between large sellers and other stations at the same fraud rate identifies λ_r. Differences in audit rates across counties and years separate the audit channel from the records screen. The change across regimes is identified by the onset of the crackdown, which we date from the enforcement record (Appendix I.4). We treat the onset as exogenous to any one station's conduct: DPS launched the campaign in response to the rise in fraud across the market as a whole (Section 4), so the regime change shifts every station's hazard at the same date, while a station's own fraud rate moves its hazard only within a regime, through the slope. Appendix Figure I.1 plots the observed and fitted detection rates against

inspection fees to identify demand.

the fraud rate before and after the crackdown; after it, the hazard is higher at every fraud rate in the data and rises more steeply in it.

6.3 Supply

Parameters and estimators. The supply side has seven parameters: the penalty rate χ of equation (5), an entry cost $c_\omega(m)$ for each type and market, and a scale σ_ω of each type's payoff shocks ε_{st} in equation (7). Free entry (Assumption 6) sets σ_{large} to zero, which leaves six parameters, including the fringe's scale σ_f. Both stages take demand (Section 6.1) and the detection hazard (Section 6.2) as given, with large sellers' post-crackdown hazard scaled to the detections DPS recorded (Appendix J.1). The second stage then has one parameter, χ, which we estimate by the generalized method of moments from how long large sellers operated after the crackdown. Given $\hat{\chi}$, we estimate the first stage's entry costs and σ_f by maximum likelihood from the number of entrants of each type. We recover each large seller's unobserved price from its first-order condition for quantity at its observed sales, and the rest of δ_{smt} is its quality, which the counterfactuals hold fixed (Appendix J.2).

The penalty rate. In the second stage, how long stations operate pins down the penalty rate. A station's spell is the run of months it operates after entering. Consider a station that sells q passes at price p in every month it operates and faces a constant monthly detection rate $\bar{h}$. If it plans a spell of T months, it collects pq in each month until DPS detects it or it stops. It pays the penalty χ on what it has collected if detection comes first. Its expected payoff in equation (5) reduces to

$$\Pi(T) \;=\; pq\Big[(1-\chi)\,\frac{1-e^{-\bar{h}T}}{\bar{h}} \;+\; \chi\,T e^{-\bar{h}T}\Big]. \tag{11}$$

One more month adds pq if the station is still operating, and it exposes the penalty on what the station has earned, χpqT, to the hazard $\bar{h}$. The station lengthens its spell until the two balance:

$$\frac{d\Pi}{dT} \;=\; pq\,e^{-\bar{h}T}\big(1-\chi\bar{h}T\big) \;=\; 0 \qquad\Longrightarrow\qquad T^* \;=\; \min\Big\{12,\;\frac{1}{\chi\bar{h}}\Big\}. \tag{12}$$

A station that faces twice the detection rate therefore plans a spell half as long, and the probability that DPS detects it before it stops,

$$1-e^{-\bar{h}T^*} \;=\; 1-e^{-1/\chi} \qquad \text{for } T^*<12, \tag{13}$$

depends on χ alone. In the model, a station's hazard varies across months with its sales, and its profit with demand. The plan maximizes equation (5) over the profits and hazards the station expects in each month (Appendix J.2), and equations (11) to (13) are that problem at a constant profit and hazard. We estimate χ from the spells of large sellers that began in 2023. It was the first year in which spells were shorter than the year and the last whose spells had mostly closed when our data end.[33] The stopping rule depends on a station's

[33] Appendix J.3 describes how we treat the spells still open when our data end.

hazard and not directly on the number of its rivals, so the 2023 spells inform χ even though the number of sellers was still adjusting that year. Two moments compare these spells with the model's entrants of 2023, simulated in that year's market with the sellers that operated in it,

$$g(\chi) \;=\; \begin{pmatrix} \hat{s} - \mathbb{E}\big[\,1 - e^{-\bar{h}\,T^*(\chi)}\big] \\ \hat{T} - \mathbb{E}\big[\,T^*(\chi) \;\big|\; \text{undetected}\big] \end{pmatrix}, \qquad \hat{\chi} = \arg\min_{\chi}\; g(\chi)'\, W\, g(\chi). \tag{14}$$

Here, $\hat{s}$ is the share of the spells that ended in detection and $\hat{T}$ is the average number of months open among those that did not. The expectations run over these entrants and the qualities they draw, each with its average hazard for the year, $\bar{h}_{mt}(\omega, \theta)$, and its plan T^* at penalty rate χ. A higher penalty rate shortens the planned spells, so fewer end in detection and the rest close sooner.

The estimation sample. We estimate the first stage on two steady states of each market, the five years before the crackdown and 2024. We leave out 2022 and 2023, when the market moved from one regime to the other. Appendix J.4 explains why neither year is a steady state.

> Assumption 5 (Steady states). *In each market, the years 2017–2021 and the year 2024 are stationary equilibria of the model of Section 5 under the enforcement regimes before and after the crackdown.*

The record of 2024 supports treating it as a steady state: DPS's recorded enforcement actions fell sharply in the second half of the year, yet the number of large sellers open did not change (Appendix J.4).

Entry costs and scale parameters. In the first stage, we estimate the entry costs and the scale parameters by maximum likelihood, given $\hat{\chi}$. A potential large seller learns its quality only after it enters, drawing it from the qualities that demand estimates for large sellers in its market (Section 6.1). Fringe sellers share a common quality. Potential sellers of the same type are therefore alike when they decide whether to enter. With payoff shocks that are logistic with scale σ_ω, each potential seller of type ω enters market m in a year of window τ with probability

$$\phi_{m\tau}(\omega) \;=\; \frac{\exp\big\{\big[\bar{\Pi}_\omega(m,\tau) - c_\omega(m)\big]/\sigma_\omega\big\}}{1 + \exp\big\{\big[\bar{\Pi}_\omega(m,\tau) - c_\omega(m)\big]/\sigma_\omega\big\}}. \tag{15}$$

Here, the window τ is either the years before the crackdown or 2024, and $\bar{\Pi}_\omega(m,\tau)$ is the expected payoff of entering before the entry cost, averaged over the window's years. We compute the expectation in equation (7) from the second stage at the equilibrium beliefs of equation (8) (Appendix J.3). The payoff falls as more sellers of a type enter, because they compete for the same drivers. We fix the sets of potential sellers of Section 5.3 by a convention, the same in every year and every counterfactual (Appendix J.1). Appendix J.3 gives the likelihood of the observed counts of entrants of each type (equation (42)).

> Assumption 6 (Free entry of large sellers). *In each market and year, large sellers enter until the expected value of a spell, net of the expected penalty, equals the entry cost:* $\Pi^*_{mt}(\textit{large}) = 0$.

Free entry is the limit of equation (15) for large sellers as σ_{large} goes to zero, and the entry of large sellers becomes the zero-profit count of Bresnahan and Reiss (1991). The full model, with a free scale for large sellers' payoff shocks, places that scale near zero (Appendix J.3). Under free entry, we estimate $c_{\text{large}}(m)$ from the number of large sellers before the crackdown (Appendix J.3). The number of large sellers in 2024 then plays no part in the estimate, and Section 7.3 compares the model's prediction of it with the data. For the counterfactuals, free entry means that the number of large sellers adjusts fully to any change in the value of a spell. If large sellers' shocks matter, entry responds less than free entry implies, and the model overstates how many large sellers a change in enforcement adds or removes.

Identification. The spells pin down χ through the share that ends in detection and the months that undetected stations stay open. The share depends on χ alone and is therefore robust to an error in the separately estimated hazard. The months depend on both, so they test the stopping rule. Given $\hat{\chi}$, demand and the hazard give the value of entering at any number of rivals, and the counts of entrants pin down the entry costs and σ_f. Free entry makes $c_{\text{large}}(m)$ the value of a spell at the number of large sellers before the crackdown. For the fringe, $c_{\text{fringe}}(m)$ absorbs each market's level of participation, and the change in participation between the two windows is informative about σ_f. Finally, intensity and severity lower the value of a spell, so the counts alone cannot distinguish them. The spells distinguish them, which lets the counterfactuals vary the two levers of Section 5.4 separately.

7 Estimates, Fit and Counterfactuals

In this section, we report the estimates of the demand and supply models and use the estimated model to quantify the equilibrium effects of the crackdown and of alternative enforcement policies. Section 7.1 reports the demand estimates and Section 7.2 the supply estimates. Section 7.3 compares the estimated model with the market before and after the crackdown on the outcomes that the counterfactuals rest on: drivers' compliance choices and large sellers' share of the market, volume, and number. Section 7.4 then simulates the market in 2024, the post-crackdown steady state (Section 6.3), with and without the crackdown. It decomposes the crackdown's effect into the contributions of drivers' substitution among sellers, the adjustment of the sellers that remain, and entry; varies the intensity and the severity of enforcement separately; compares the crackdown with a penalty on the drivers who buy fraudulent passes; and reports the welfare effects.

7.1 Demand Estimates

Table 3 reports the demand estimates. The nesting parameter $\lambda_F = 0.514$ measures how different sellers of fraudulent passes are in drivers' eyes. At one, sellers would be as different from one another as fraud is from repair, and a driver whose seller leaves would exit the fraud market in proportion to that seller's share of fraud. At zero, sellers would be perfect substitutes. The estimate sits in between. When we remove from drivers' choice sets, all at once, the stations that DPS's enforcement actions closed in the crackdown (Section 4), holding prices and the remaining sellers fixed, 48.8% of the closed stations' buyers still buy

a fraudulent pass, 23.5% repair, and 27.7% let the registration lapse. About half of the drivers who lose their seller stay in the fraud market, and of the rest more lapse than repair. Appendix H.5 reports the alternative specifications.

The second parameter that shapes the counterfactuals is cost sensitivity, which governs how the cost of repair affects the decision to comply. Its mean is 0.545 per \$100 of the repair bill, and its dispersion on the log scale, $\sigma_\alpha = 0.404$, implies a standard deviation of 0.23 across drivers, so there is substantial heterogeneity in cost sensitivity. A helpful way to interpret these estimates is through the implied indifference repair cost, the bill at which a driver is as likely to repair as to take the best of the other options. For new-malfunction drivers, 45% of the population, the median indifference cost is \$640, and it ranges from \$370 at the tenth percentile to \$1,110 at the ninetieth. The median predicted repair bill for these drivers is \$248, so most of them repair, but a bill of several hundred dollars turns a large share of them away from repair, while the price of a fraudulent pass is on the order of a hundred dollars (Table 5). The home-county match value $\mu = 0.118$ is worth about \$22 at the mean cost sensitivity, a modest advantage for a seller in the driver's own county.

Panel B reports the inertia terms, which let a driver's previous choice shift the current one. A driver whose previous cycle ended in a fraudulent pass or a lapse rather than an honest pass values repair less by $\eta_R = 2.357$, which cuts the odds of repairing to 0.09 of a new-malfunction driver's. On top of that common shift, a previous fraudulent pass raises the utility of fraud by $\eta_F = 0.487$ and a previous lapse raises the utility of lapsing again by $\eta_L = 1.114$. In dollars, the median indifference repair cost is \$75 for a driver whose previous cycle ended in fraud and \$25 for one whose previous cycle ended in a lapse. Drivers whose previous cycle ended in fraud are 52% of the population, so fraud is persistent, whatever mix of inertia and persistent tastes lies behind it (Section 5.2): half of the malfunctioning vehicles on the road arrive with a fraudulent pass from the previous cycle, and half of their drivers would not repair a bill above \$75. These are the drivers a crackdown has to move. The remaining parameter, τ_C, scales the direct purchases in the likelihood and does not enter drivers' utility (Appendix H.2).

7.2 Supply Estimates

Table 4 reports the supply estimates, and three results carry into the counterfactuals. First, entering the market as a large seller is expensive. Under free entry (Assumption 6), a large seller's entry cost equals what a spell, a run of consecutive months selling fraudulent passes at scale, was worth before the crackdown: \$349.8 thousand in DFW and \$196.9 thousand in Houston, about a year of a large seller's pre-crackdown profit (0.90 and 0.96 years). A fringe seller's entry cost is far smaller, \$20.0 and \$22.0 thousand, and the scale of fringe sellers' payoff shocks, \$15.4 thousand, is large relative to it, so fringe participation responds smoothly to the value of selling: the elasticity of the number of fringe sellers with respect to the value of a fringe spell is 0.6.[34]

Second, detection is costly. The penalty rate χ is what a detected station loses per dollar of the fraud revenue it earned during its spell, and we estimate $\chi = 1.188$: a detected station

[34]The main specification sets the scale of large sellers' payoff shocks to zero; Appendix J.3 reports the model with it estimated.

Table 3: Demand Estimates

	Estimate	Std. error
Panel A. Substitution among sellers and cost sensitivity		
Nesting parameter for the fraud alternatives, λ_F	0.514	(0.003)
Home-county match value, μ	0.118	(0.011)
Cost sensitivity, mean $E[\alpha_i]$ (per \$100 of the repair bill)	0.545	(0.002)
Cost sensitivity, median	0.502	(0.002)
Cost sensitivity, dispersion σ_α	0.404	(0.003)
Panel B. Inspection history and direct purchases		
Carryover-malfunction vehicles: shift away from repair	2.357	(0.004)
Previous cycle ended in fraud: further shift toward fraud	0.487	(0.003)
Previous cycle ended in a lapse: further shift toward lapse	1.114	(0.004)
Probability that a new-malfunction driver who comes to the market is tested, τ_C	0.0725	(0.0001)
Log-likelihood	−18,796,902	
Decision cells	9,072	
Parameters, including fixed effects	88	

Note: Maximum likelihood estimates on the inspection cycles of malfunctioning vehicles in DFW and Houston, 2016–2024, in 9,072 decision cells by metro, year, vehicle group (age and mileage) and inspection history (Section 6.1). Drivers choose among repair, a fraudulent pass from one of the available sellers, and the three lapse alternatives (scrap, idle, unregistered), whose split the roadside remote-sensing data pin down (Appendix G). Metro-by-year fixed effects on every alternative (2016–2018 pooled) and metro-by-home-county effects on fraud are not reported. Panel A: λ_F governs how readily drivers substitute among sellers; μ is the match value of a seller in the driver's home county; cost sensitivity α_i is lognormal (equation (9)), in utility per \$100 of the repair bill. Panel B: shifts in utility for vehicles still malfunctioning from the previous cycle, relative to new-malfunction vehicles, and the probability that a new-malfunction driver who comes to the market takes an inspection rather than buying a fraudulent pass directly. Standard errors, in parentheses, are unclustered inverse-information standard errors with the fixed effects re-solved at every parameter value; they condition on the detection model and the predicted repair bills. The likelihood is over decision cells, and part of their lapse mass is carried from earlier cycles rather than observed, so a vehicle-level observation is not defined for that mass and we report unclustered standard errors.

loses more than its fraud revenue to date. A rate above one is plausible because χ captures, in reduced form, everything a detected station loses in proportion to its fraud business, the fines and restitution, its legal costs, and the honest inspection business and license value that a revocation takes with it (Section 4). Because the penalty grows with what the station has earned, each additional month of selling raises what it stands to lose, so stations plan short spells when detection is likely.[35]

Third, the penalty rate and the crackdown's detection hazard together explain why the number of large sellers fell so far. The detection hazard facing a large seller with a fraud rate

[35]The two moments that identify χ do not agree exactly: in the model, 61% of the spells of 2023 entrants end in detection, against 68% and 79% of the spells that began in 2023 in DFW and Houston, and the months that undetected stations stayed open point to a higher penalty rate than the share of spells ending in detection does (Appendix J.5).

Table 4: Supply Estimates

	DFW	Houston
Panel A. Parameters		
Entry cost of a large seller, $c_{\text{large}}(m)$ (\$ thousands)	349.8	196.9
	(1.5)	(0.3)
Entry cost of a fringe seller, $c_{\text{fringe}}(m)$ (\$ thousands)	20.0	22.0
	(0.8)	(0.6)
Scale of fringe sellers' payoff shocks, σ_f (\$ thousands)	15.4	
	(0.9)	
Penalty rate, χ (\$ per \$ of fraud revenue)	1.188	
	(0.091)	
Panel B. Implications for a large seller		
Detection hazard at a one-percent fraud rate, 2024 relative to before the crackdown (ratio)	30	
Planned months of operation, 2024 (twelve before the crackdown)	2.6	4.7
Monthly profit, 2024 relative to before the crackdown (ratio)	10.9	6.5

Note: Estimates of the supply model (Section 6.3). Panel A reports the main specification, with standard errors in parentheses: the penalty rate's are clustered by station, the fringe parameters' come from the likelihood, and a large seller's entry cost carries the sampling error in χ. All condition on the demand and hazard estimates. σ_f is the scale of a logistic shock, whose standard deviation is $1.81\sigma_f$. Panel B reports what the estimates imply for a large seller in the model's equilibrium. The first row is the ratio of the estimated detection hazard facing a large seller with a one-percent fraud rate in 2024 to its level before the crackdown, from the hazard estimates of Appendix I (Appendix Table I.1), which the supply model takes as given; Section 6.3 defines the other rows.

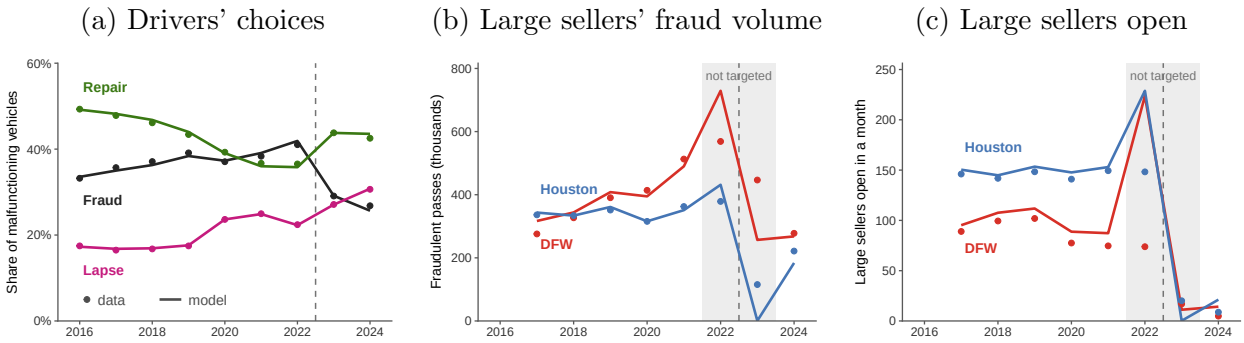
of one percent rose 30-fold with the crackdown (Panel B; Appendix I).[36] At the estimated penalty rate, a large seller's planned spell falls from a full year to 2.6 months in DFW and 4.7 in Houston, worth 8% and 15% of a year of undetected selling at the pre-crackdown number of rivals, a small fraction of the entry cost. Entry costs did not change, so free entry supports far fewer sellers, each earning far more: in the model's 2024 equilibrium a large seller earns 10.9 times its pre-crackdown monthly profit in DFW and 6.5 times in Houston. These short spells and this concentration are the model's counterpart to the short-lived, concentrated fraud stations of Section 4.2, although the model produces less concentration than the data show (Section 7.3).

7.3 Fit

Figure 3 compares the estimated model with the data on the outcomes that matter for the counterfactuals: drivers' choices in panel (a), and large sellers' volume and number in panels (b)–(c). Panel (a) fits closely by construction, because demand's metro-by-period effects pin down the shares; the tests of demand are the alternative specifications in Appendix H.5. Panels (b)–(c) are predictions rather than targets, because the supply estimator uses the

[36]The 95-percent confidence interval is 13 to 68, and the hazard also became steeper in the fraud rate (Appendix Table I.1).

Figure 3: Model Fit



Note: Dots are data and lines are the estimated model, by the year of the compliance decision. (a) Shares of repair, fraud, and lapse (scrap, idle, or unregistered) among malfunctioning vehicles, pooled over DFW and Houston; the model values are the demand model's fitted shares. (b)–(c) By metro: large sellers' fraudulent passes (thousands) and the number of large sellers open in the average month. In Houston in 2023 the model's post-crackdown equilibrium has no large sellers (panels (b) and (c)), a corner that Section 7.4 discusses. The market in a month is the drivers of malfunctioning vehicles who make their compliance choice that month (Section 6.1). The supply values are the model's equilibrium in each year at that year's estimated demand, averaged over 48 simulated draws of entrants' qualities, start dates, and detection times, enough that simulation error is below one percentage point of the counterfactual effects. The supply model is estimated on 2017–2021 and 2024; the shaded band marks 2022 and 2023, the transition between the two regimes, which the model does not target (Section 6.3). The dashed line marks the start of the crackdown.

lengths of selling spells and the numbers of sellers, not volumes; Appendix J.4 describes how the supply series are built, compares the model's price with the price the observed shares imply, and reports every outcome by year, including 2022 and 2023, which the model does not target (Appendix Table J.3).

Panel (a) shows how drivers of malfunctioning vehicles reallocated through the crackdown. Pooled over the two metros, the share of malfunctioning vehicles passed by fraud fell from 41% in 2022 to 27% in 2024. Repair recovered, from 37% to 43%, but lapse rose more, from 22% to 31%: of the drivers who left fraud, more moved to the other form of noncompliance, a lapsed registration, than back into compliance. The model carries both movements into the counterfactuals. Panel (b) shows that the supply model reproduces large sellers' volume in the two steady states. The model's 2024 large-seller volume is within 4% of the data in DFW and within 17% in Houston, and the fall in total fraud between the two steady states differs from the data's by at most 0.08 log points in either market. Houston is the weaker fit, in large sellers' share of the market and in their price (Appendix J.3).

Panel (c) shows the margin the model fits least well, the number of large sellers after the crackdown. In 2024 the model has 14.2 large sellers open in the average month in DFW, against 4.8 in the data,[37] because in each metro a single seller operated for most of the year and sold 38% of the fraudulent passes in DFW and 32% of those in Houston. A model in which large sellers plan spells of a few months spreads that seller's volume across more entrants. The counterfactuals depend on volume and price rather than which stations sell.

[37] In Houston, the model has 21.2 large sellers open in the average month, against 8.8 in the data.

7.4 Counterfactuals

The counterfactual market. We measure the crackdown's effect by simulating the market with the detection hazard at its pre-crackdown form and comparing it with the model's own market with the crackdown. Everything else stays as estimated: demand, including its estimated shifts in 2023 and 2024; each station's quality; the entry costs; and the sets of potential sellers. Each of these is plausibly invariant to enforcement, since a station's quality is its visibility to drivers and the entry costs are the costs of setting up a station. The penalty rate also stays at its estimate, because the crackdown raised the intensity of enforcement rather than what a detection costs a station (Section 5.4). What the crackdown moves is the hazard, and through it entry, the plans and posted prices of the stations that enter, and the choices drivers make among them. We measure the effect against the model rather than the data so that it does not mix the crackdown with the model's misfit, and we report it for 2024, the post-crackdown steady state on which the supply model is estimated.

The crackdown's effect. Figure 4 shows fraud volume by year in the data, in the model with the crackdown, and in the model without it; the first two columns of Table 5 describe the 2024 market in the two cases. Fraud volume counts every fraudulent pass a metro's drivers buy, at large sellers, at the fringe, or outside the metro, so fraud that the crackdown moves out of the metro does not count as fraud removed. Without the crackdown, the model has 228 large sellers open in the average month of 2024 in DFW and 250 in Houston, more than before the crackdown.[38] The crackdown cut fraud volume in 2024 by 48% across the two metros, 48% in DFW and 48% in Houston, far less than the number of large sellers, because the sellers that remained sold more at a higher price and part of the demand moved outside the metro.[39] Purchases at stations outside the metro rose from 9 to 109 thousand in DFW and from 10 to 66 thousand in Houston, so that, counted at the metro's own stations alone, the fall is 62% and 59%. Of the drivers who did leave fraud, 70% repaired their vehicles and the rest lapsed, adding 108 thousand unregistered vehicles across the two metros.

Equilibrium margins. Appendix Table J.5 separates the effect into its equilibrium margins, adding one at a time from the no-crackdown market to the model with the crackdown. The accounting a reader would do from the enforcement record alone, removing the large sellers the crackdown drove out and sending every displaced buyer to repair or lapse, attributes a fall of 73% to the crackdown in DFW and 49% in Houston. Letting the displaced buyers choose among the remaining options, at unchanged plans and prices, lowers it to 48% and 34%, since in DFW 6% of them go to the fringe and 27% to stations outside the metro. Letting the surviving sellers re-plan their capacity, months, and prices under the crackdown's hazard lowers it further, to 30% and 27%, as survivors expand into the demand left to them. Free entry returns it to 48% and 48%, because a spell under the crackdown's hazard is worth a fraction of the entry cost, so far fewer sellers enter than survive the removal. The crack-

[38] At the old hazard, a spell in DFW in 2024 would have been worth 2.34 times its pre-crackdown value at the same number of rivals, and more stations enter to share that demand.

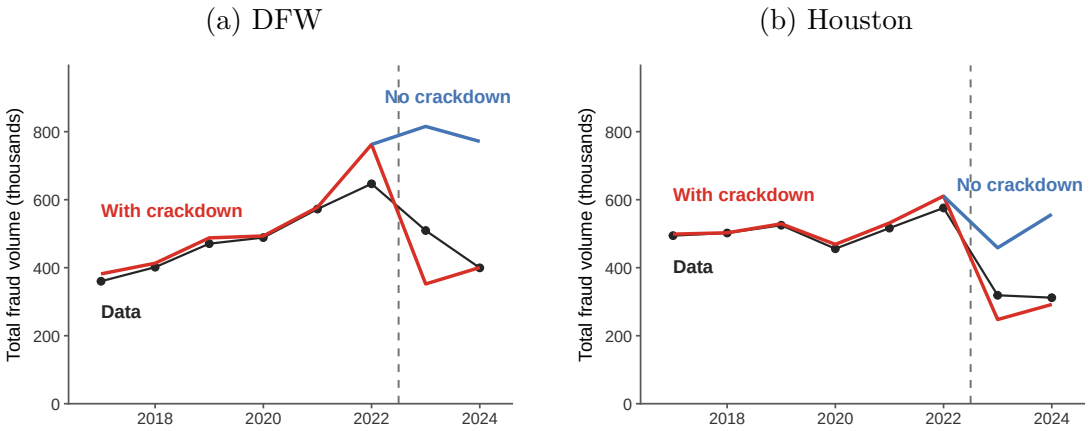
[39] In DFW, the 2024 price at large sellers is \$136 with the crackdown and \$110 without it, both above the \$87 at which the fringe sells. Large sellers' own volume fell by 65% and 66%, more than the in-metro total, as some of their buyers moved to the fringe.

Table 5: The 2024 Market Under Three Enforcement Policies

	No crackdown	Supply crackdown	Demand penalty at t^*	Both
Penalty per pass, DFW / Houston ($)	–	–	264 / 180	264 / 180
Panel A. The market, 2024				
DFW				
Fraud volume (000)	771	401 (−48%)	403 (−48%)	86 (−89%)
of which outside the metro (000)	9	109	7	69
Large sellers open, average month	227.6	14.2 (−94%)	132.9 (−42%)	0.1 (−100%)
Price at large sellers ($)	110	136 (+24%)	122 (+11%)	167 (+52%)
Houston				
Fraud volume (000)	557	292 (−48%)	295 (−47%)	89 (−84%)
of which outside the metro (000)	10	66	9	53
Large sellers open, average month	249.6	21.2 (−92%)	147.4 (−41%)	0.2 (−100%)
Price at large sellers ($)	94	110 (+17%)	105 (+13%)	133 (+43%)
Panel B. Where the fraud goes, both metros				
Share of fraud purchases at the fringe (%)	0.7	9.5	1.0	28.7
Share of fraud purchases outside the metro (%)	1.4	25.3	2.2	69.5
Unregistered vehicles, change (000)	–	+108	+115	+211
Leaving fraud: share to repair (%)	–	70	68	71
Leaving fraud: share to lapse (%)	–	30	32	29
Panel C. Welfare against no crackdown, both metros ($ millions)				
Driver surplus	–	−407	−375	−714
Repair-sector profit	–	131	126	266
Fraud-station rents	–	3	0	5
Penalty revenue (a transfer)	–	–	160	39
Avoided damages (100% VMT)	–	278	273	585
Net (100% VMT)	–	4	185	180
Break-even damage per vehicle-year ($)	–	239	94	182

Note: Each column is a 2024 market. The first is the market without the crackdown, which keeps the pre-crackdown detection hazard; the other three are compared with it, with the change from no crackdown in parentheses. The demand penalty keeps the pre-crackdown hazard and charges each driver who buys a fraudulent pass an expected penalty t, the probability of being caught times the fine, modeled as a wedge between what drivers pay and what sellers receive at every seller; t^* is set in each metro so that fraud equals its volume under the supply crackdown. Both adds the same penalties to the supply crackdown; if the two policies acted independently, the fraud remaining would be 27% of the no-crackdown volume (27% in DFW, 28% in Houston). Fraud volume in Panel A and the shares in Panel B count the same passes, every fraudulent pass a metro's drivers buy, at large sellers, at the fringe, or at stations outside the metro; the drivers who leave fraud, the unregistered vehicles, and the welfare rows follow the choices the demand model predicts, with the population of drivers held fixed at its size in each policy's market. VMT is vehicle miles traveled: at 100% an unregistered vehicle is assumed to be driven as much as a registered one; the text reports the net at half that mileage and with the avoided damages priced at the supply model's sales. Fraud-station rents are revenue less station penalties and entry costs, with entry costs treated as real resources and station penalties as a loss credited to no one; under free entry they are near zero in every market. Penalty revenue is the drivers' penalties, a transfer to the regulator counted on both sides. All statistics average 48 simulated paths, paired across columns by path.

Figure 4: Fraud Volume With and Without the Crackdown



Note: Fraudulent passes bought by each market's drivers by year, at large sellers, at the fringe, and at stations outside the metro: the data, the estimated model with the crackdown, and the model with the pre-crackdown detection hazard in 2023 and 2024 (no crackdown). The counterfactual replaces the hazard from January 2023, so before 2023 the two model series coincide. The crackdown's effect, reported in the text and in Table 5, is the 2024 difference between the two model series as a share of the no-crackdown volume. Model series average 48 simulated paths. The dashed line marks the start of the crackdown.

down's effect therefore comes from the sellers it deterred from entering, not from the ones it shut down: the naive accounting overstates the effect in DFW by half, and lands near the model's answer in Houston only because the margins it ignores offset each other.[40]

Welfare. Panel C of Table 5 puts the crackdown in dollars, valuing every driver in the demand model's population at the choice the model predicts for that driver in the two markets. Drivers lost 407.4 million dollars of surplus in 2024, the value they placed on a fraudulent pass over their next-best option, for most of them a repair they would rather not have paid for.[41] Repair shops gained 130.8 million at a margin of one half on the added repairs, and fraud sellers' rents, their revenue less the penalties they pay and the entry costs they sink, changed by only 2.9 million, because under free entry a spell is worth its entry cost in expectation in every market we simulate.[42]

Against the drivers' loss stand the damages avoided when vehicles that would have passed by fraud were repaired, scrapped, or idled: 278.0 million at the damage estimates of Appendix D, counting an unregistered vehicle as driven as much as a registered one. On these numbers the crackdown roughly broke even: the ledger nets to 4.4 million dollars, a loss

[40] Varying the regulator's two levers separately shows how much of the effect each carries. Halving the post-crackdown hazard leaves an effect of 32% in DFW and 28% in Houston, and halving the penalty rate 37% and 34%. Doubling either removes about two thirds of the fraud (66% and 66% for the hazard, 63% and 63% for the penalty rate) rather than nearly all of it, because purchases move to stations outside the metro, which the hazard does not reach; at the metro's own stations the fall reaches 97% and 91% (Appendix J.5).

[41] We convert utility to dollars driver by driver at each driver's own cost sensitivity, so a dollar of lost surplus is a dollar the driver would have paid to keep the option. The ledger leaves out the regulator's enforcement costs and any effect beyond the two metros. Appendix J.6 reports the full ledger.

[42] The fraud revenue the crackdown removes is matched, inside this row, by entry costs no longer sunk and by penalties paid. The ledger counts the entry costs as real resources, the costs of setting up a station, and the penalties as a loss to the stations that it credits to no one.

of 6.4 million in DFW and a gain of 10.8 million in Houston. Since rents are small, the question is the drivers' lost surplus against the repair profit and the pollution avoided, and its answer turns on the damage estimate: the ledger balances at 239 dollars per fraud-passed vehicle-year, against the 249 dollars we estimate for the vehicles the crackdown displaced. Two conventions matter. Counting unregistered vehicles at half the mileage raises the net to 19 million, and pricing the avoided damages at the fraud the supply model sells rather than at the fraud drivers would buy at its prices turns the net into a loss of 144 million.

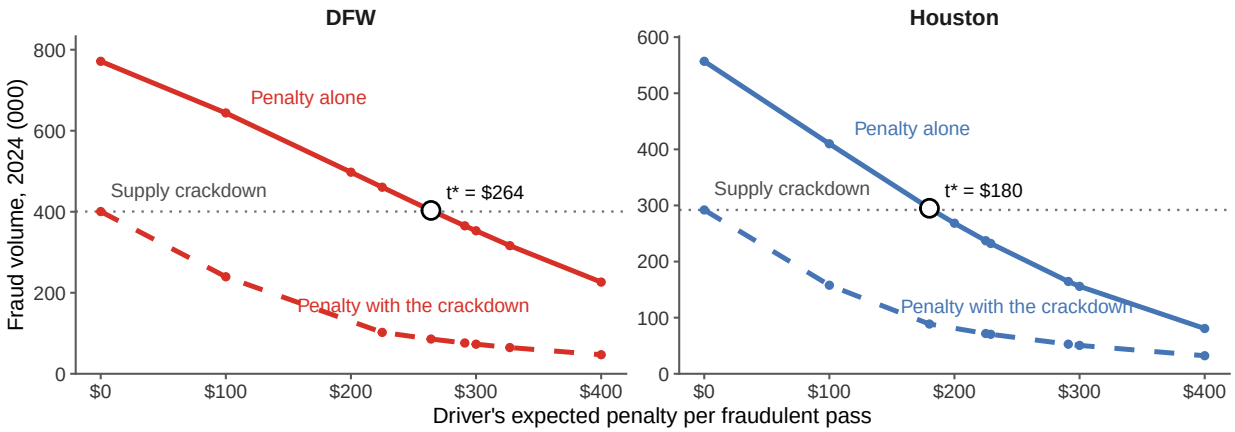
Alternative policies. The third column of Table 5 compares the crackdown with a policy aimed at drivers rather than sellers: an expected penalty on the driver who buys a fraudulent pass, the probability of being caught times the fine, which enters the model as a wedge between what drivers pay and what sellers receive at every seller. We set the penalty in each market so that fraud in 2024 falls by as much as under the crackdown, to within one percent, with the detection hazard at its pre-crackdown level: $264 per pass in DFW and $180 in Houston. The two policies remove the same fraud from different markets. Under the crackdown, 35 large sellers remain open across the two metros, the price of a pass at large sellers rises to $125, and a third of the remaining purchases move to the fringe and out of the metro. Under the penalty, 280 large sellers remain open, the price rises less, to $115, and the fringe and the stations outside the metro keep their small share of the remaining purchases, 3.2% against 34.8% under the crackdown.

The two policies avoid nearly the same damages but differ in what drivers give up. The drivers who leave fraud split between repair and lapse in nearly the same proportions under either policy (Panel B), so the damages avoided are 273 million dollars under the penalty against 278 million under the crackdown. Drivers lose 375 million dollars of surplus under the penalty against 407 million under the crackdown, even though the penalty's figure includes the 160 million they pay in penalties. The regulator collects the wedge that drivers pay under the penalty. Under the crackdown, drivers lose the sellers themselves, and the scarcity that raises the survivors' price earns no one a rent, because entrants spend it on entry costs and on spells that end in detection. The penalties are a transfer to the regulator, so the ledger's net gain is 185 million dollars under the penalty against 4 million under the crackdown.

The two instruments are complements. Adding the same penalty to the crackdown leaves 13% of the no-crackdown fraud, against the 27% the two single-policy effects would leave if they acted independently. With few sellers left and the demand for their passes cut, almost no spell covers its entry cost, and the large sellers disappear. The combination's net gain of 180 million dollars is far above the crackdown's, so a regulator that has closed the large sellers would gain from adding a driver-side penalty. It is no better than the penalty alone in this ledger, because the last fraud the combination removes costs drivers about as much surplus as the damages it avoids. Figure 5 traces fraud against the penalty with and without the crackdown; the point where the penalty alone crosses the crackdown's level is the penalty it would take to match the crackdown without the regulator shutting a single station.

Limitations. A limitation of the model is that it holds each market's primitives fixed within the pre-crackdown window, while DFW's market consolidated within it, with fewer large sellers each earning more (Appendix J.3). If that decline would have continued without

Figure 5: Fraud Volume Against the Driver Penalty



Note: Total fraud volume in 2024 (every fraudulent pass the metro's drivers buy, including those bought outside the metro), by metro, against the expected penalty per fraudulent pass charged to the driver, with the detection hazard at its pre-crackdown level (penalty alone) and at its level with the crackdown (penalty with the crackdown). The dotted line is the volume under the supply crackdown alone, and the open circle marks the penalty t^* at which the penalty alone reproduces it. Each point is a re-solved equilibrium, averaged over 48 simulated paths; a point is drawn only where the equilibrium converged.

the crackdown, the model overstates the number of large sellers that the crackdown removed in DFW. Houston's 2023 shows the limit of a model with no adjustment path: its post-crackdown equilibrium at 2023 demand has no large sellers at all, while the data show a market still adjusting (Appendix Table J.3).

The no-crackdown market also carries the demand model's 2024 metro-year effects, which we estimate from a year in which few large sellers operated; with those effects and the pre-crackdown hazard, the model fills DFW with large sellers and leaves the fringe under one percent of purchases, well below its share in any pre-crackdown year, so the shares in Panel B of Table 5 describe a no-crackdown market more concentrated than the one the data show before 2022. Finally, the demand-side crackdown assumes that drivers respond to an expected fine as they respond to a price; the regulator almost never penalized drivers in our sample, so nothing in the data separates the two responses.

8 Conclusion

Regulators rely on inspectors to certify compliance. When compliance is costly, a market for fraudulent certification can form. We study this phenomenon in the context of vehicle emissions inspections in Texas and ask how a supply-side crackdown affects this market in equilibrium. We first develop a method that recovers fraudulent inspections from a set of fraud indicators in the program's own administrative records. We find that 9.5% of passing inspections in our sample are fraudulent, with each fraudulently passed vehicle causing about $251 of health damages a year. We document how the market responded to a 2022 enforcement crackdown, which shut down a subset of inspection stations that sold much of the fraudulent passes. While fraud fell in response to this crackdown, both sides of the market

adjusted: many of the displaced buyers turned to other nearby sellers, to stations outside the metropolitan area, or to driving unregistered, and new sellers entered the market.

We develop and estimate an equilibrium model of supply and demand that captures these adjustment margins, and we use it to measure the crackdown's effect in equilibrium and to compare it with alternative enforcement policies. We find that the crackdown reduced fraudulent inspections in 2024 by 48%, even though the number of large sellers open fell by 94% in Dallas-Fort Worth and 92% in Houston. The higher likelihood of the regulator detecting and sanctioning fraudulent inspections made a station's spell worth a fraction of the cost of entering, so much of the crackdown's effect comes through the sellers who did not enter rather than through the ones it shut down. Of the drivers who stopped buying fraudulent passes, 70% repaired their vehicles and the rest let their registrations lapse. In our welfare accounting the crackdown roughly broke even: the surplus that drivers lost is about the size of the health damages it avoided and the repair profits it created together, so its sign turns on the damage estimate. A penalty on drivers who buy a fraudulent pass, set at $264 in Dallas-Fort Worth and $180 in Houston, removes the same amount of fraud at a net gain of $185 million, because what drivers pay in penalties is collected by the regulator, while what they lose under the crackdown is dissipated in sellers' entry costs and detections. The two policies are complements: together they leave 13% of the fraud that would occur without the crackdown, against 27% if their effects were independent.

Our findings speak to the design of enforcement wherever regulators rely on inspectors. The conditions that blunted this crackdown are not special to vehicle inspections: buyers who are sensitive to price and willing to substitute among sellers, sellers who can enter freely, and a jurisdiction whose boundary the buyers can cross. Under these conditions, shutting down the inspectors who sell fraudulent certifications removes less fraud than the regulator's own records suggest, because the market replaces part of what was removed. Two lessons follow. Enforcement against the demand side, which makes the regulated party liable for the false certificate, complements enforcement against inspectors rather than substituting for it, and in our setting it removes the same fraud at a lower cost. The records that the regulator collects carry the information needed to identify fraud, so informational constraints are less binding than the delegated structure suggests. What enforcement achieves in a market for fraudulent certification is decided not by the inspectors it removes but by how buyers and sellers respond to their removal, and enforcement in delegated regimes should be designed against that response.

References

Abaluck, Jason and Jonathan Gruber (2016) "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance," *American Economic Review*, 106 (8), 2145–2184.

Ale-Chilet, Jorge, Cuicui Chen, Jing Li, and Mathias Reynaert (2025) "Colluding Against Environmental Regulation," *The Review of Economic Studies*, rdaf024.

Alexander, Diane and Hannes Schwandt (2022) "The Impact of Car Pollution on Infant and Child Health: Evidence from Emissions Cheating," *Review of Economic Studies*, 89 (6), 2872–2910.

Allman, Elizabeth S., Catherine Matias, and John A. Rhodes (2009) "Identifiability of parameters in latent structure models with many observed variables," *The Annals of Statistics*, 37 (6A), 3099–3132.

Arcidiacono, Peter and Robert A. Miller (2011) "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79 (6), 1823–1867.

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov (2010) "Estimating Static Models of Strategic Interactions," *Journal of Business & Economic Statistics*, 28 (4), 469–482.

Beber, Bernd and Alexandra Scacco (2012) "What the Numbers Say: A Digit-Based Test for Election Fraud," *Political Analysis*, 20 (2), 211–234.

Becker, Gary S. (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76 (2), 169–217.

Becker, Gary S., Kevin M. Murphy, and Michael Grossman (2006) "The Market for Illegal Goods: The Case of Drugs," *Journal of Political Economy*, 114 (1), 38–60.

Bennett, Victor Manuel, Lamar Pierce, Jason A. Snyder, and Michael W. Toffel (2013) "Customer-Driven Misconduct: How Competition Corrupts Business Practices," *Management Science*, 59 (8), 1725–1742.

Berry, Steven T. (1992) "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60 (4), 889–917.

——— (1994) "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25 (2), 242–262.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan (2007) "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption," *Quarterly Journal of Economics*, 122 (4), 1639–1676.

Bishop, Gary A. (2019) "Three Decades of On-Road Mobile Source Emissions Reductions in South Los Angeles," Technical Report CRC Project E-124, Coordinating Research Council.

Bishop, Gary A. and Donald H. Stedman (2008) "Fuel-Based Motor Vehicle Emission Inventory," *Journal of the Air & Waste Management Association*, 58 (10), 1351–1359.

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer (2020) "Escalation of Scrutiny: The Gains from Dynamic Enforcement of Environmental Regulations," *American Economic Review*, 110 (8), 2558–2585.

Bresnahan, Timothy F. and Peter C. Reiss (1991) "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99 (5), 977–1009.

California Bureau of Automotive Repair (2020) "Nine Smog Check Stations Engaged in Fraud," News Release, February.

Celhay, Pablo, Nicolás Figueroa, Carla Guadalupi, and Matías Oteíza (2025) "Cracking Down on Fraud in Social Insurance: Evidence from Chile's Worker Compensation Scheme," Working paper, Pontificia Universidad Católica de Chile.

Dell, Melissa (2015) "Trafficking Networks and the Mexican Drug War," *American Economic Review*, 105 (6), 1738–1779.

DellaVigna, Stefano and Eliana La Ferrara (2010) "Detecting Illegal Arms Trade," *American Economic Journal: Economic Policy*, 2 (4), 26–57.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi (2010) "State Dependence and Alternative Explanations for Consumer Inertia," *RAND Journal of Economics*, 41 (3), 417–445.

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," *The Quarterly Journal of Economics*, 128 (4), 1499–1545.

——— (2018) "The Value of Regulatory Discretion: Estimates From Environmental Inspections in India," *Econometrica*, 86 (6), 2123–2160.

Dyck, Alexander, Adair Morse, and Luigi Zingales (2010) "Who Blows the Whistle on Corporate Fraud?" *Journal of Finance*, 65 (6), 2213–2253.

Eastern Research Group (2020) "Evaluation of the Texas Vehicle Emissions Inspection and Maintenance Program in the Dallas-Fort Worth and Houston-Galveston-Brazoria Nonattainment Areas," Technical Report ERG No. 0433.00.001.002, Prepared for Texas Commission on Environmental Quality.

——— (2022) "Evaluation of the Texas Vehicle Emissions Inspection and Maintenance Program in the Dallas-Fort Worth and Houston-Galveston-Brazoria Nonattainment Areas," Technical Report ERG No. 0433.00.007.002, Prepared for Texas Commission on Environmental Quality.

——— (2024a) "Evaluation of the Texas Vehicle Emissions Inspection and Maintenance Program in the Dallas-Fort Worth and Houston-Galveston-Brazoria Nonattainment Areas," Technical Report ERG No. 0488.00.002.002, Prepared for Texas Commission on Environmental Quality.

——— (2024b) "Vehicle Emissions Inspection Program Test Fee Analysis for AirCheckTexas Program," technical report, Prepared for Texas Commission on Environmental Quality.

Fisman, Raymond and Shang-Jin Wei (2004) "Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China," *Journal of Political Economy*, 112 (2), 471–496.

Friedman, Scott and Eva Parks (2023) "Fake Inspections, Real License Plates: Investigators Believe Millions of Texas Cars Were Never Checked for Safety & Emissions," NBC DFW (NBC 5 Investigates), March, https://www.nbcdfw.com/investigations/fake-inspections-real-license-plates-investigators-believe-millions-of-texas-cars-were-never-checked-for-safety-emissions/3220698/, March 22, 2023.

Fullerton, Don and SunLee Stechuk (2026) "The Good, the Bad and the Ugly," Working paper, University of Illinois at Urbana-Champaign.

Galenianos, Manolis and Alessandro Gavazza (2017) "A Structural Model of the Retail Market for Illicit Drugs," *American Economic Review*, 107 (3), 858–896.

Georgia Office of the Attorney General (2015) "Emissions Inspection Repeat Fraudsters Headed to Prison," Press Release, August, https://law.georgia.gov/press-releases/2015-08-10/emissions-inspection-repeat-fraudsters-headed-prison.

Greenstone, Michael, Guojun He, Ruixue Jia, and Tong Liu (2022) "Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution," *American Economic Review: Insights*, 4 (1), 54–70.

Grieco, Paul L. E., Charles Murry, Joris Pinkse, and Stephan Sagl (2025) "Optimal Estimation of Discrete Choice Demand Models with Consumer and Product Data," Working Paper 33397, National Bureau of Economic Research.

Handel, Benjamin R. (2013) "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103 (7), 2643–2682.

Hopenhayn, Hugo A. (1992) "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60 (5), 1127–1150.

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

Hui, Siu L. and Steven D. Walter (1980) "Estimating the error rates of diagnostic tests," *Biometrics*, 36 (1), 167–171.

Jacob, Brian A. and Steven D. Levitt (2003) "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *The Quarterly Journal of Economics*, 118 (3), 843–877.

Jia, Panle (2008) "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76 (6), 1263–1316.

Kobak, Dmitry, Sergey Shpilkin, and Maxim S. Pshenichnikov (2016) "Integer Percentages as Electoral Falsification Fingerprints," *Annals of Applied Statistics*, 10 (1), 54–73.

Kreps, David M. and Jose A. Scheinkman (1983) "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *The Bell Journal of Economics*, 14 (2), 326–337.

Kruskal, Joseph B. (1977) "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, 18 (2), 95–138.

KXAN Austin (2023) "Undercover Sting, Rejected Car Inspections: How Emissions Testing Fraud Impacts Austin," November, https://www.kxan.com/investigations/undercover-sting-rejected-car-inspections-how-emissions-testing-fraud-impacts-austin/.

Lugthart, Peter (2026) "The Environmental Costs and Consumer Benefits of Competitive Smog Check Markets," Job market paper, The Wharton School, University of Pennsylvania.

Markussen, Simen and Knut Røed (2017) "The Market for Paid Sick Leave," *Journal of Health Economics*, 55, 244–261.

McFadden, Daniel (1978) "Modelling the Choice of Residential Location," in Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jörgen W. Weibull eds. *Spatial Interaction Theory and Planning Models*, 75–96, Amsterdam: North-Holland.

Muller, Nicholas Z. and Robert Mendelsohn (2009) "Measuring the Damages of Air Pollution in the United States," *Journal of Environmental Economics and Management*, 58 (1), 54–75.

Muller, Nicholas Z., Robert Mendelsohn, and William Nordhaus (2011) "Environmental Accounting for Pollution in the United States Economy," *American Economic Review*, 101 (5), 1649–1675.

New Jersey Office of the Attorney General (2016) "Former Motor Vehicle Inspector and Two Current Inspectors Indicted in Alleged Scheme to Use Data Simulators to Falsify Emissions Tests," Press Release, December, https://www.njoag.gov/former-motor-vehicle-inspector-and-two-current-inspectors-indicted-in-alleged-scheme-to-use-data-simulators-to-falsify-emissions-tests/.

Oliva, Paulina (2015) "Environmental Regulations and Corruption: Automobile Emissions Inspections in Mexico City," *Journal of Political Economy*, 123 (3), 686–723.

Polinsky, A. Mitchell and Steven Shavell (2000) "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature*, 38 (1), 45–76.

Reddit, r/plano (2023) "Anybody Know a Place That Will Pass My Inspection with a Check Engine Light On?" Online forum thread, January, https://web.archive.org/web/20260715044514/https://www.reddit.com/r/plano/comments/10jero7/.

Reynaert, Mathias and James M. Sallee (2021) "Who Benefits When Firms Game Corrective Policies?" *American Economic Journal: Economic Policy*, 13 (1), 372–412.

Sanders, Nicholas J. and Ryan Sandler (2020) "Technology and the Effectiveness of Regulatory Programs over Time: Vehicle Emissions and Smog Checks with a Changing Fleet," *Journal of the Association of Environmental and Resource Economists*, 7 (3), 587–618.

Seim, Katja (2006) "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *RAND Journal of Economics*, 37 (3), 619–640.

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

Singer, Brett C. and Robert A. Harley (1998) "A Fuel-Based Approach to Estimating Motor Vehicle On-Road Emissions," *Journal of the Air & Waste Management Association*, 48 (12), 1202–1215.

Soliman, Adam (2025) "Disrupting Drug Markets: The Effects of Crackdowns on Rogue Opioid Suppliers," *American Economic Journal: Economic Policy*, 17 (4), 165–191.

Texas Department of Public Safety (2023) "DPS' Operation Cinderblock Cracks Down on Fraudulent Vehicle Inspections," Press release, March, https://www.dps.texas.gov/news/dps-operation-cinderblock-cracks-down-fraudulent-vehicle-inspections, March 30, 2023.

U.S. Department of Justice, Eastern District of California (2024) "12 Members of Smog Inspection Cheating Ring Indicted," Press Release, https://www.justice.gov/usao-edca/pr/12-members-smog-inspection-cheating-ring-indicted.

U.S. Environmental Protection Agency (2026a) "How Are Oxides of Nitrogen (NOx) Defined in the NEI?" Air Emissions Inventories webpage, https://www.epa.gov/air-emissions-inventories/how-are-oxides-nitrogen-nox-defined-nei, Accessed October 2026.

——— (2026b) "Vehicle Emissions Inspection & Maintenance (I/M): General Information for Motorists," https://www.epa.gov/state-and-local-transportation/vehicle-emissions-inspection-maintenance-im-general-information, Last updated September 18, 2026. Accessed October 1, 2026.

Zou, Eric Yongchen (2021) "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality," *American Economic Review*, 111 (7), 2101–2126.

Appendix Contents

A Additional Tables and Figures **A.1**
A.1 Setting and Data (Section 2) . . . A.1
A.2 Measuring Fraud (Section 3) . . . A.3
A.3 The Enforcement Crackdown (Section 4) . . . A.4
A.4 Estimation and Identification (Section 6) . . . A.11

B Data **A.12**
B.1 Sources . . . A.12
B.2 Cleaning . . . A.12
B.3 Inspection Cycles . . . A.13
B.4 Stations . . . A.14
B.5 Roadside Remote Sensing . . . A.14

C Fraud Detection Model Details **A.14**
C.1 Fraud Indicators . . . A.14
C.2 Estimation . . . A.20
C.3 Identification . . . A.23
C.4 Testing Conditional Independence . . . A.25
C.5 Robustness . . . A.32
C.6 Fraudulent Failed Inspections . . . A.33

D On-Road Emissions and Pollution Damages **A.37**
D.1 Sample and Specification . . . A.37
D.2 Emission Rates . . . A.37
D.3 Selection on the Severity of the Defect . . . A.38
D.4 Excess Emissions to Monetary Damages . . . A.39
D.5 Robustness . . . A.40

E Station Licensing and Turnover **A.41**
E.1 How an Inspection Station Gets a License . . . A.41
E.2 Links Between New and Closing Fraud Stations . . . A.43

F Repair Costs **A.45**
F.1 The Predicted Repair Cost . . . A.45
F.2 Descriptive Evidence . . . A.46
F.3 External Validation . . . A.48
F.4 Repair Costs of Vehicles That Are Never Tested . . . A.48

G Scrap, Idle, or Driving Unregistered: The States of Lapsed Vehicles **A.52**
G.1 Motivation and Setup . . . A.52
G.2 Identifying On-Road Share from Roadside Remote Sensing . . . A.53
G.3 Separating Scrap from Idle: The Return-Survival Model . . . A.56
G.4 Results . . . A.59
G.5 Estimation and Implementation Details . . . A.63

H Demand: Estimation, Identification and Fit **A.67**
H.1 Estimation Sample . . . A.67
H.2 Estimation . . . A.68
H.3 The Fraud Inclusive Value . . . A.71
H.4 λ_F Across Time-Effect Specifications . . . A.72
H.5 Full Estimates and Fit . . . A.72
H.6 Measuring the Set of Active Stations . . . A.73
H.7 Re-timing, or Substitution? . . . A.74

I The Detection Hazard **A.75**
I.1 Sample and Estimation . . . A.75
I.2 Estimates . . . A.78
I.3 Fit . . . A.78
I.4 Robustness . . . A.78

J Details of the Supply Model **A.81**
J.1 Inputs . . . A.81
J.2 Solving the Model . . . A.82
J.3 Estimation . . . A.84
J.4 The Model and the Data by Year . . . A.87
J.5 Robustness . . . A.89
J.6 The Welfare Ledger . . . A.90

Appendix A. Additional Tables and Figures

A.1 Setting and Data (Section 2)

Table A.1: Inspection Cycle Outcomes

		Share of cycles (%)		
	N cycles	Pass first test	Fail, then pass	Fail, never pass
Panel A: All metros, by period				
2016–2018	29,494,569	94.4	5.1	0.4
2019–2021	31,898,366	95.3	4.3	0.4
2022–2024	32,975,174	95.1	4.4	0.5
Panel B: By metro, 2016–2024				
Dallas–Fort Worth	42,458,415	95.4	4.2	0.4
Houston	36,839,971	94.9	4.7	0.4
Austin	10,173,560	93.4	5.8	0.8
El Paso	4,896,163	95.2	4.5	0.3
Panel C: By vehicle age, 2016–2024				
Age 2–15	81,349,990	95.5	4.1	0.3
Age 16–19	9,155,626	91.4	7.6	1.0
Age 20–24	3,862,354	91.1	7.6	1.3
Panel D: By odometer, 2016–2024				
Below median[a]	66,743,543	95.9	3.8	0.3
Above median[a]	27,306,093	92.6	6.6	0.8
All metros, 2016–2024	94,368,109	95.0	4.6	0.4

Note: Inspection cycles in the 17 program counties, 2016–2024. We assign each cycle to the calendar year of its decision: the date of the pass or, for a cycle that ends without a pass, the date of its first failed test. The three outcome columns sum to 100 percent up to rounding. Passes include fraudulent passes. Panel B assigns each cycle to the metro area of its final test. Panels C and D use the vehicle age bins and odometer split of the demand model (Section 5). [a]Below or above the median odometer reading for vehicles in the same age bin; the medians are those of the demand model's sample, so they need not split inspection cycles evenly. Panels C and D exclude the few cycles that fall outside the model's age bins, and Panel D also excludes the 318,334 cycles (0.34%) without a usable odometer reading.

About 5,700 stations perform emissions inspections in a typical year, a number that stays roughly flat over the sample, with a slight rise in the middle years (Appendix Figure A.1). In 95.0% of inspection cycles, the vehicle passes its first test (Appendix Table A.1).

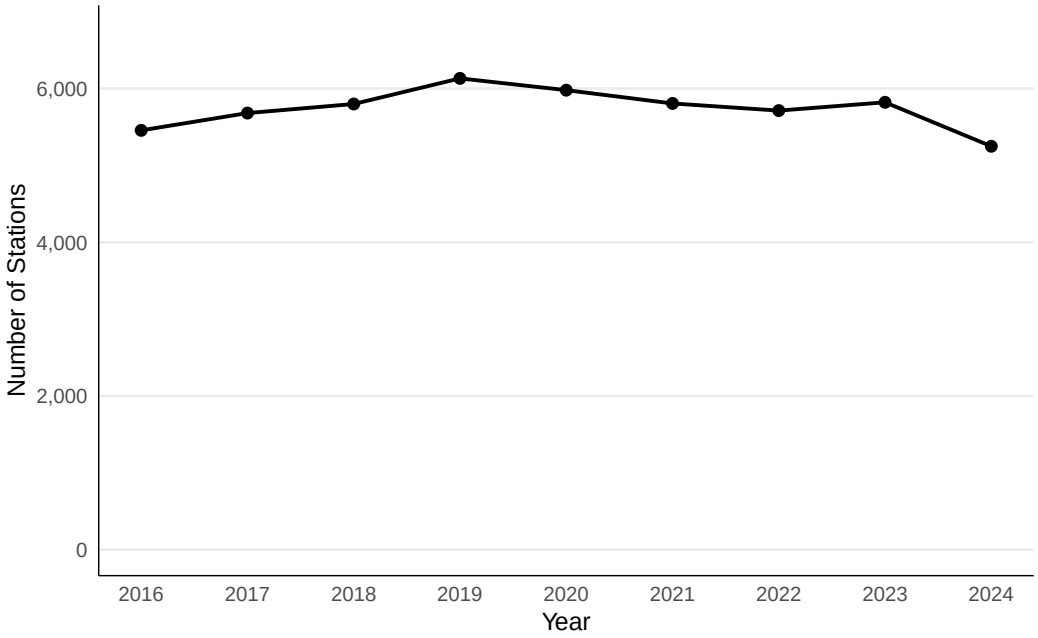


Figure A.1: Number of Stations Performing Inspections by Year

Note: Number of private licensed stations that performed at least one emissions inspection in each year, 2016–2024, in the 17 program counties.

A.2 Measuring Fraud (Section 3)

Table A.2: Sampling Uncertainty in the Fraud Rate

Year	Fraud rate (%)	Bootstrap SE (pp)	95% interval	Plug-in SE (pp)
2016	9.64	0.45	[8.82, 10.60]	0.46
2017	9.40	0.45	[8.67, 10.37]	0.47
2018	9.73	0.55	[8.83, 10.82]	0.57
2019	10.11	0.66	[8.98, 11.59]	0.62
2020	9.87	0.82	[8.47, 11.56]	0.77
2021	10.43	0.99	[8.97, 12.94]	0.94
2022	11.62	0.78	[10.30, 13.29]	0.79
2023	7.84	0.92	[6.22, 10.09]	0.94
2024	7.00	1.83	[4.27, 11.47]	1.73
2016–2024	9.52	0.46	[8.68, 10.51]	0.44

Note: Fraud rate: share of passing inspections with fraud posterior $\hat{\tau}_i > 0.5$ (%). Bootstrap: 200 draws resampling stations with replacement within metropolitan area, one draw shared across years, re-estimating the detection model in each draw; the interval is the 95% percentile interval. Plug-in: station-clustered standard error holding the model parameters fixed.

Table A.3: Sensitivity of the Fraud Rate to the Classification Cutoff

	Share with $\hat{\tau}_i >$ cutoff (%)			Mean	Share in
Year	0.3	0.5	0.7	$\hat{\tau}_i$ (%)	(0.2, 0.8) (%)
2016	9.9	9.6	9.5	9.8	0.9
2017	9.6	9.4	9.3	9.5	0.6
2018	9.9	9.7	9.6	9.8	0.5
2019	10.2	10.1	10.0	10.2	0.4
2020	10.0	9.9	9.8	9.9	0.3
2021	10.5	10.4	10.3	10.5	0.3
2022	11.7	11.6	11.5	11.6	0.3
2023	7.9	7.8	7.8	7.9	0.2
2024	7.1	7.0	6.9	7.0	0.3
2016–2024	9.6	9.5	9.4	9.6	0.4
DFW + Houston	10.9	10.7	10.6	10.8	0.5

Note: Passing inspections in the 17 program counties, 2016–2024 (95,470,830 inspections). Columns 1–3 give the share with fraud posterior $\hat{\tau}_i$ above the cutoff; 0.5 is the headline convention. Mean $\hat{\tau}_i$ is the expected fraud share. The last column is the share with $\hat{\tau}_i$ in (0.2, 0.8). DFW + Houston pools the two metropolitan areas of the market analysis.

Stations selling most of the fraudulent inspections share several common characteristics (Appendix Figure A.3). In particular, many are newly licensed: stations in their first two years of operation perform 21.8% of passing inspections but 55.4% of fraudulent ones, consistent with short-lived operations. They tend to be independent garages and used-car dealers rather than chains or new-car dealers. Stations that sell fraud also sit in lower-income, central parts of each metropolitan area: stations in the poorest quarter of ZIP codes perform a quarter of inspections but 63.1% of fraud (Appendix Figures A.3c and A.4). Finally, many

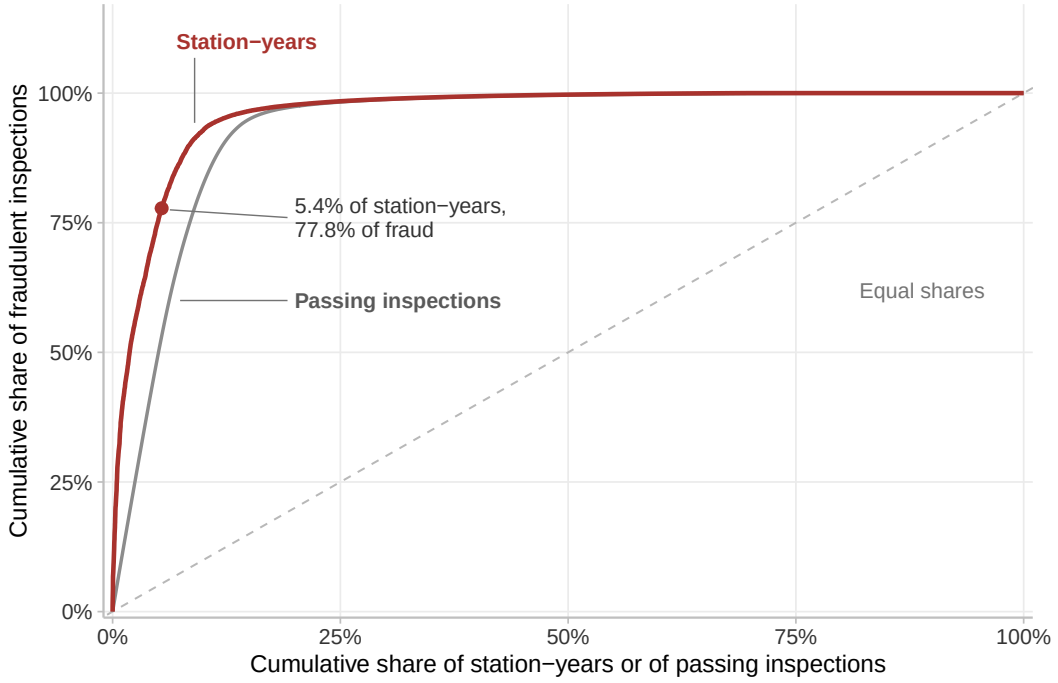


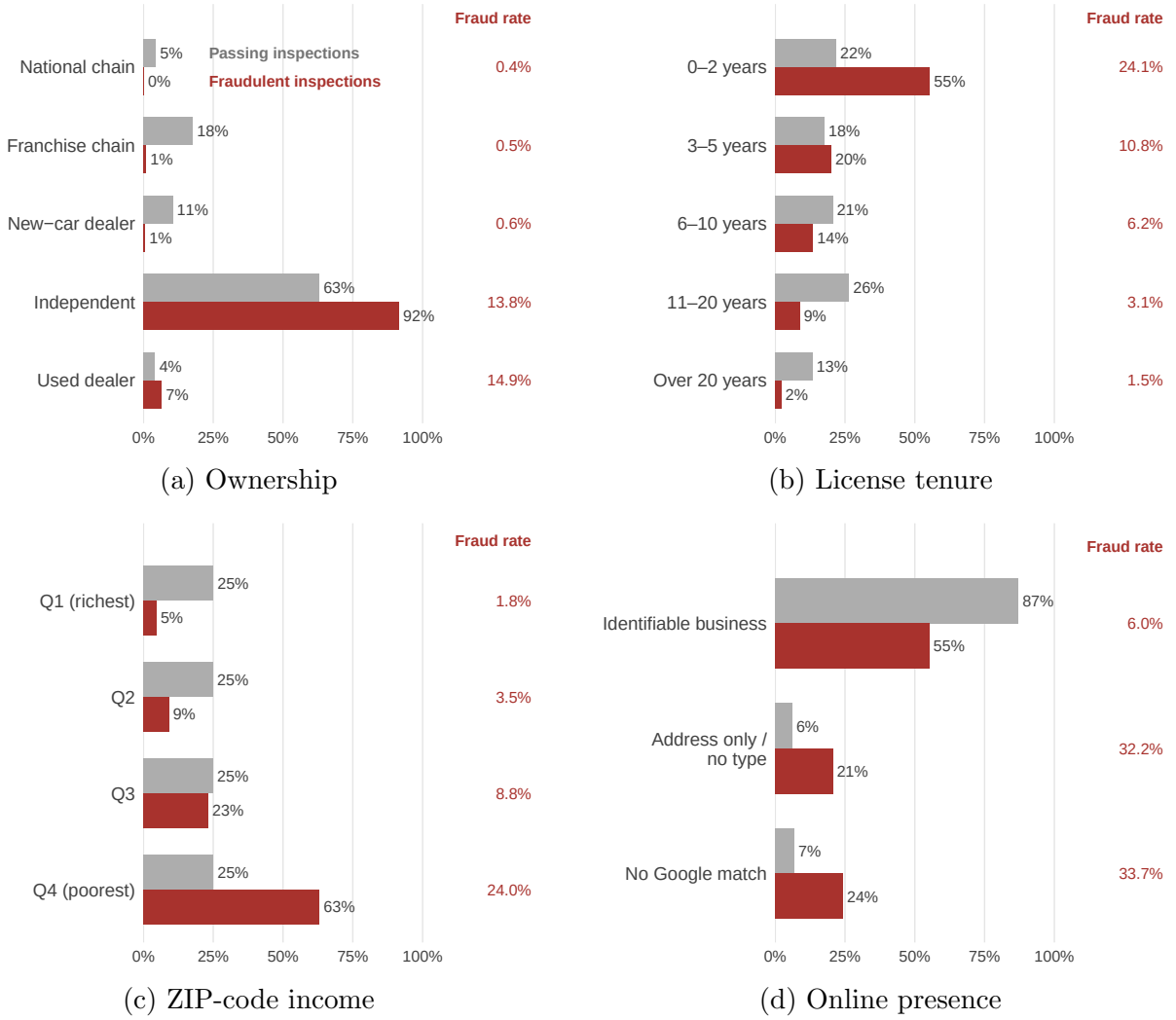
Figure A.2: Cumulative Share of Fraudulent Inspections across Station-Years

Note: Station-years are ranked by fraud rate (the share of passing inspections with posterior fraud probability above 0.5), highest first, ties broken by fraudulent inspections. Both curves plot the cumulative share of fraudulent inspections: the red curve against the cumulative share of station-years, each counting once, the grey curve against the cumulative share of passing inspections. The dot marks station-years with a fraud rate of 50% or more: 5.4% of station-years and 9.0% of passing inspections, holding 77.8% of fraudulent inspections. Sample: passing inspections in the 17 program counties, 2016–2024, 51,844 station-years. Of these, 29.9% have no fraudulent inspection; the median, 75th, 95th and 99th percentiles of the station-year fraud rate are 0.181%, 1.04%, 53.4% and 92.6%.

have no identifiable business presence online: stations whose address matches no listing on Google perform 6.8% of inspections but 24.2% of fraud.

A.3 The Enforcement Crackdown (Section 4)

Turnover in the market changed dramatically during the crackdown. Appendix Figure A.9 decomposes each quarter's change in major fraud stations into stations that start and stop selling fraud. Before the crackdown, the market was close to a steady state: in a typical quarter from January 2019–June 2022, about 12 stations started and 14 stopped in DFW, out of roughly 93 major fraud stations, and about 27 started and 24 stopped in Houston, out of roughly 168. Turnover then surged starting in October 2022. Over the following year, about 45 stations started and 60 stopped each quarter in DFW, and about 37 started and 76 stopped in Houston. In DFW, new fraud stations nearly kept pace with closures through the crackdown's first two quarters, so the number of major fraud stations barely fell until the spring of 2023; in Houston, closures outpaced new fraud stations from the start. Many of these exits were a direct result of DPS disciplinary actions, but the crackdown also coincided with many closures that have no documented sanction attached to them. These stations may have shut down for fear of a sanction, or moved their business to a new license before



(a) Ownership

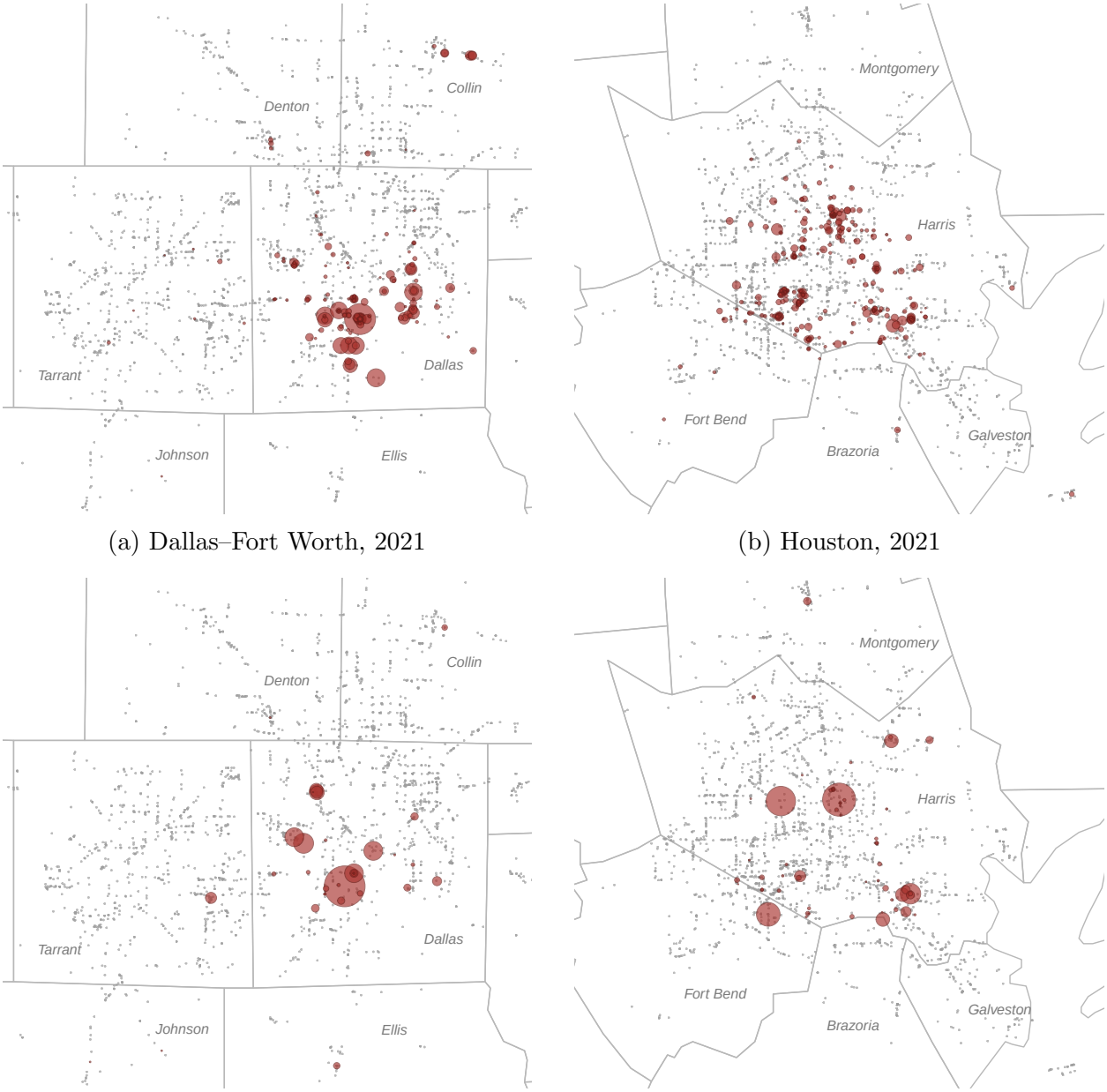
(b) License tenure

(c) ZIP-code income

(d) Online presence

Figure A.3: Shares of Passing and Fraudulent Inspections by Station Characteristic

Note: Bars show each group's share of passing (grey) and fraudulent (red) inspections; the right column gives its fraud rate. Private listed stations in the 17 program counties, 2016–2024 (236 unlisted station-years dropped); an inspection is fraudulent if its posterior fraud probability exceeds 0.5. Ownership follows Table B.4; tenure is years since first licensing. ZIP quartiles, by the ZIP's share of households earning under $25,000 (ACS), each hold a quarter of passing inspections. Online presence codes the Google Places listing (February–March 2026): named business, street address or no type only, or no match. Panels (b) and (c) omit 1 and 36 station-years lacking a license date and ZIP income, respectively.



(a) Dallas–Fort Worth, 2021

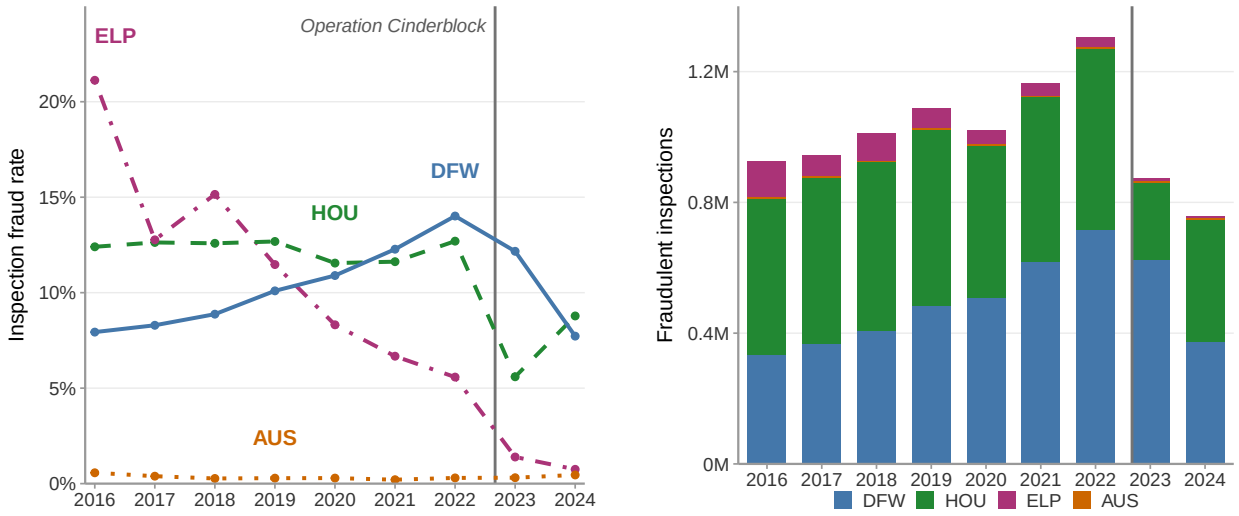
(b) Houston, 2021

(c) Dallas–Fort Worth, 2024

(d) Houston, 2024

Figure A.4: Inspection Stations and Fraud Volume, 2021 and 2024

Note: Each dot is a station with at least one passing inspection in the year. Red dots mark high-fraud stations, with at least 100 passing inspections of which over 25% have a posterior fraud probability above 0.5, and grey dots mark the rest. Red dot area is proportional to the station's fraudulent inspections, on a scale common to all panels, with the largest dot at 137,657. Coordinates are Google Places matches (February–March 2026 snapshot). Stations matched more than 250 metres from their address sit at the median of well-matched stations in their ZIP code, and these points carry 34.3% of fraudulent inspections. Stations without a usable coordinate (32) are omitted. The crop omits at most 1.0% of a panel's fraudulent inspections. The protocol indicator loses signal in 2024, so 2024 fraud may be understated.



(a) Inspection fraud rate

(b) Fraudulent inspections

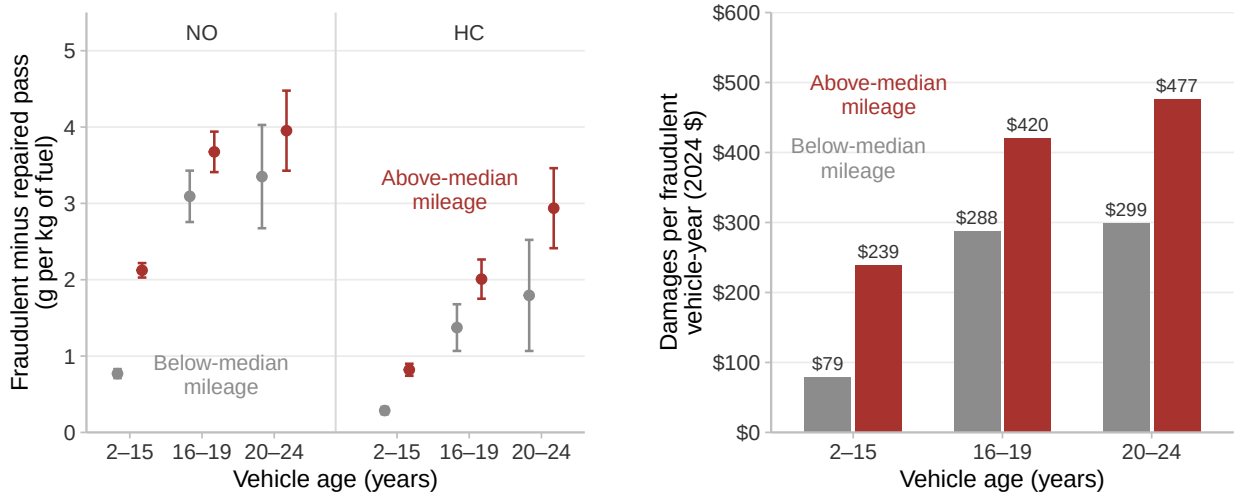
Figure A.5: Fraud Rate and Fraud Volume by Metropolitan Area, 2016–2024

Note: Estimated fraud in the four metropolitan areas of the Texas inspection and maintenance program—Dallas–Fort Worth, Houston, El Paso and Austin—by inspection-cycle year, 2016–2024. An inspection counts as fraudulent when its posterior fraud probability from the mixture model exceeds 0.5. Panel (a) plots the share of *passing* inspections that are fraudulent, so an honest market sits at zero and the axis is not a fail rate; Panel (b) plots the corresponding count, stacking the areas DFW, HOU, AUS, ELP from the bottom of each bar. Rates aggregate inspection-level posteriors within an area and year rather than averaging cell-level rates. Pooled across the four areas and the window the rate is 9.5%, peaking at 11.6% in 2022, and the window total is 9.1 million fraudulent inspections. The vertical rule marks Operation Cinderblock (August 2022). Estimation sample and screens are in Appendix C.2.

Figure A.6: Fraudulent Passes by Vehicle Age and Mileage

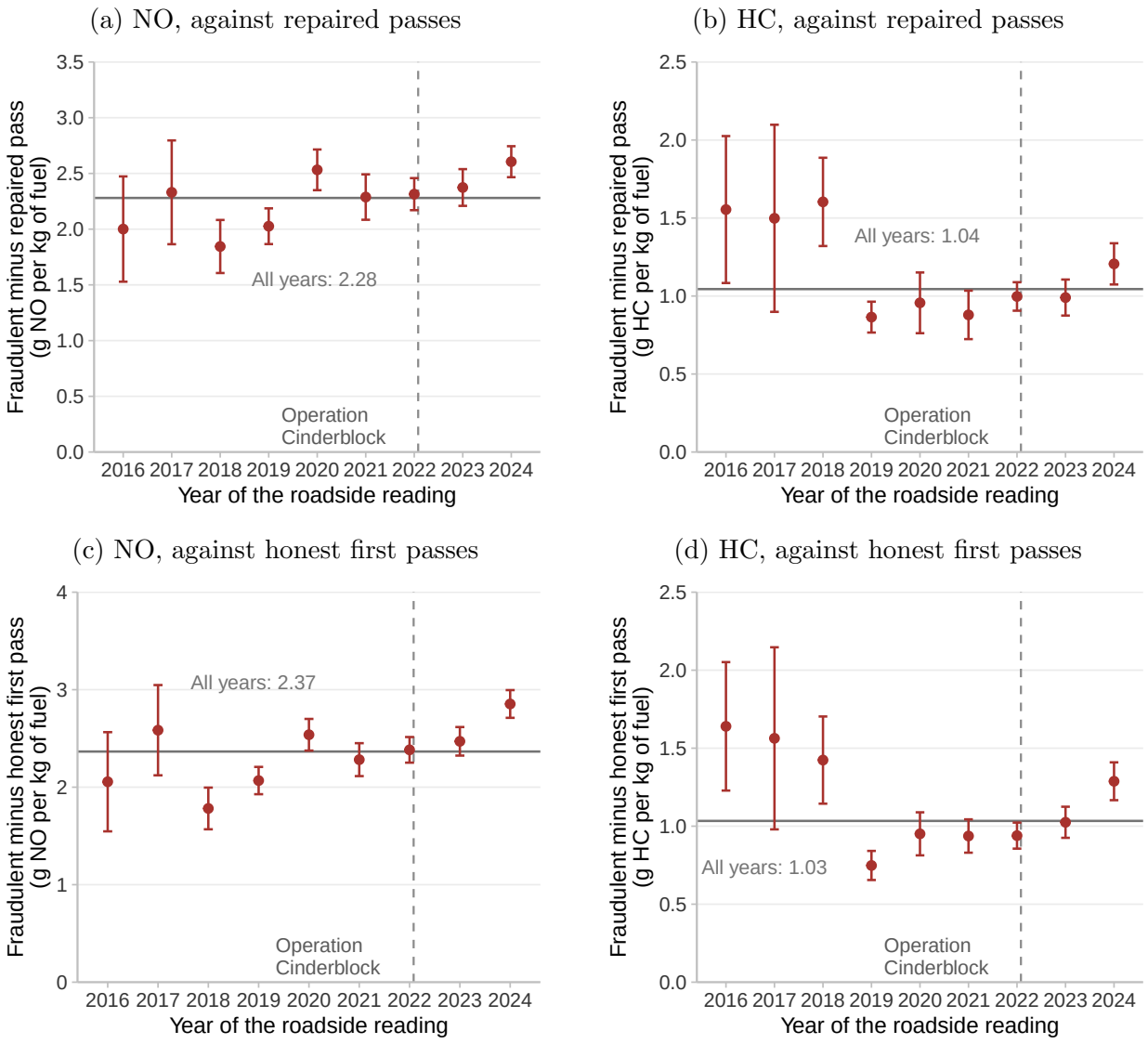
(a) Excess emissions, fraudulent minus repaired

(b) Damages per fraudulent vehicle-year



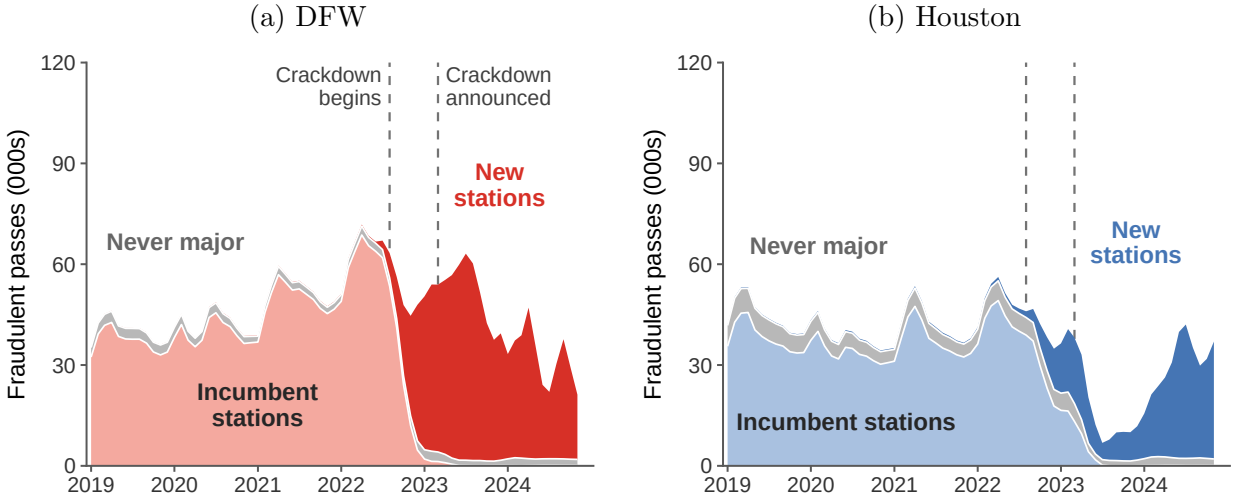
Note: Panel (a): the excess roadside emissions of fraudulent over repaired passes, in grams per kilogram of fuel, for each age group split at its median odometer, from one regression per pollutant of the form of Table 2, columns 1 and 3; bars are 95% confidence intervals. Panel (b): damages per fraudulent vehicle-year in 2024 dollars, after lowering excess emissions by 15% for NO and 18% for HC (Appendix D.3).

Figure A.7: Excess Emissions of Fraudulent Passes by Year



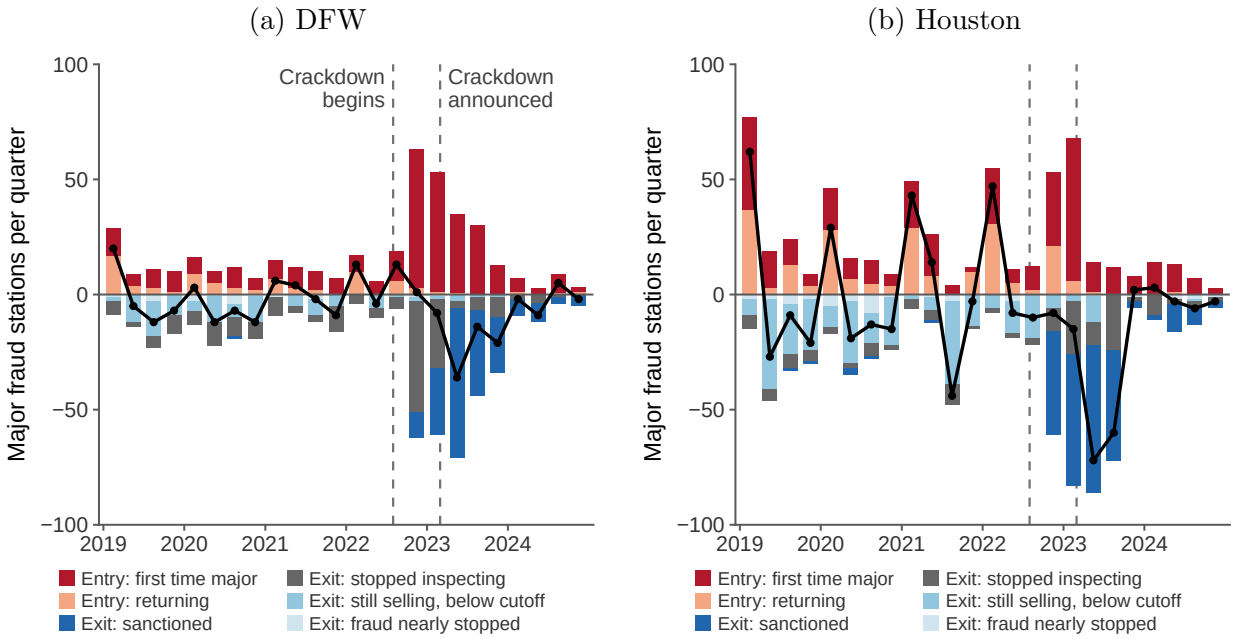
Note: Panels (a) and (b): each point is the excess roadside emissions of fraudulent over repaired passes among readings taken in one calendar year, 2016–2024, in grams per kilogram of fuel; panels (c) and (d) compare fraudulent passes with honest passes on the first test instead. The bars are 95% confidence intervals, with standard errors clustered by vehicle and by day. All years come from one regression of the same form as columns 1 and 3 of Table 2, pooled over vehicle groups, with the fraud and repair indicators interacted with the year of the reading. The solid horizontal line is the gap pooled over all years; the dashed vertical line marks the start of Operation Cinderblock in August 2022.

Figure A.8: Fraud Volume by Station Cohort



Note: Monthly fraudulent passes, January 2019–December 2024, by the station's metro, stacked by station group. Fraudulent pass: a vehicle's final passing inspection with posterior fraud probability above 0.5. New stations: first month as a major fraud station (more than 100 fraudulent passes) in or after August 2022; incumbent stations: first such month earlier (data from 2016). Grey: stations never major. Bands are centered 3-month moving averages of each group's monthly volume. Dashed lines: crackdown begins (August 2022); crackdown publicly announced (March 2023).

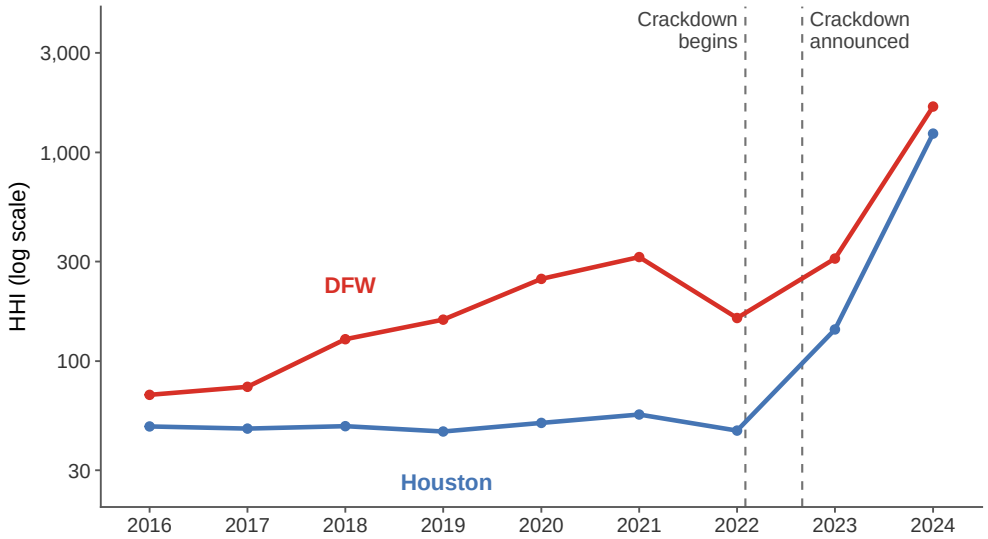
Figure A.9: Entry and Exit of Major Fraud Stations



Note: Quarterly, by the station's metro. A major fraud station sells more than 100 fraudulent passes in some month of the quarter. Above zero, entrants: major this quarter but not last, first time ever (data from 2016) or returning. Below zero, exits: major last quarter but not this one, classed in order as sanctioned (DPS action last or this quarter), stopped inspecting, still selling (30 or more fraudulent passes), or fraud nearly stopped. Black line: entries minus exits. Dashed lines: crackdown begins (August 2022) and is publicly announced (March 2023).

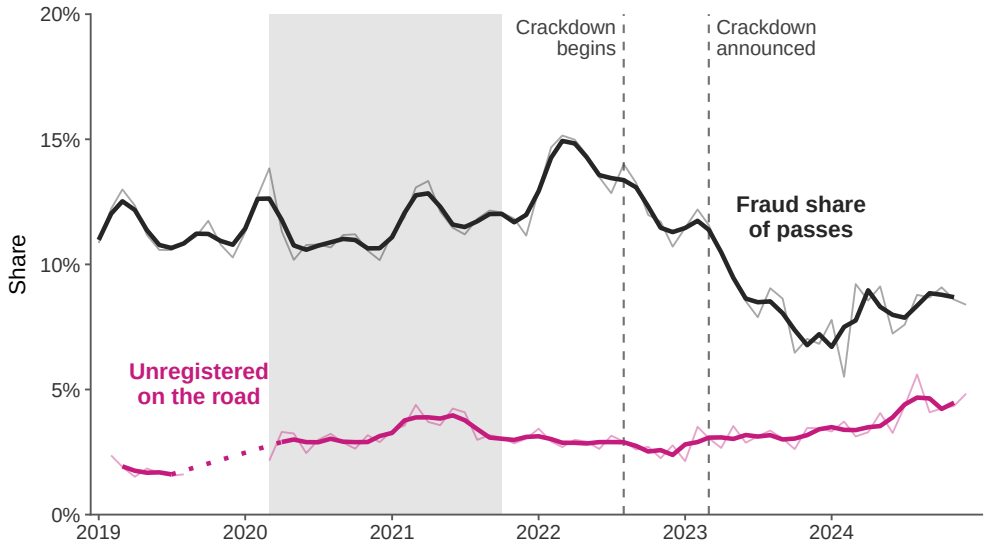
DPS could act.

Figure A.10: Concentration of the Fraud Market



Note: Herfindahl-Hirschman index of the fraud market by metro and calendar year: 10,000 times the sum of squared station shares of the metro's fraudulent passes. Fraudulent pass: a vehicle's final passing inspection with posterior fraud probability above 0.5. Stations are assigned to their own metro; log scale. Dashed lines: crackdown begins (August 2022); crackdown publicly announced (March 2023).

Figure A.11: Fraud and Unregistered Driving, Registration Expired a Year or More



Note: As panel (f) of the main figure, with the unregistered share defined as registration expired 12 or more months (the definition in earlier drafts). Metros pooled; monthly. Dotted where registration records were not updated (September 2019–February 2020). Grey band: COVID registration waiver and renewal backlog (March 2020–September 2021). Thick: 3-month centered average; thin: monthly. Dashed lines: crackdown begins (August 2022); crackdown publicly announced (March 2023).

On a scale from near zero (many small sellers) to 10,000 (a single seller), the Herfindahl–Hirschman index of station shares of fraud rose from 315 to 1,659 in DFW and from 55 to

1,231 in Houston between 2021 and 2024 (Appendix Figure A.10).

A.4 Estimation and Identification (Section 6)

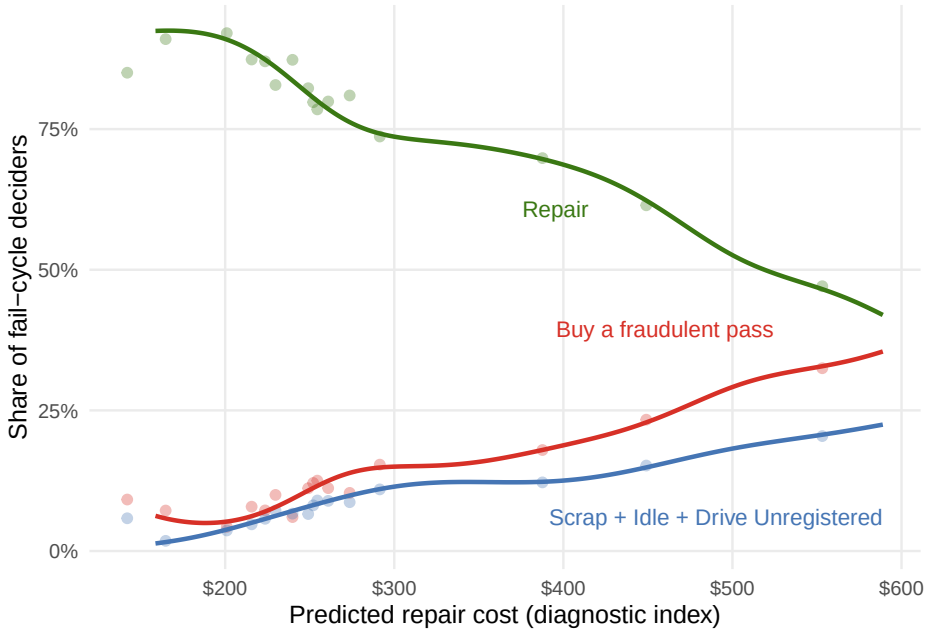


Figure A.12: Compliance Choices by Predicted Repair Cost

Note: Shares of tested malfunctioning vehicles choosing each outcome, by predicted repair cost $\hat{r}_i$ (2024 dollars; the diagnostic-only cost of Appendix F.1). Lines are weighted binomial GAM fits ($k = 12$) evaluated between the 1st and 99th percentiles of $\hat{r}_i$; faint points are means within equal-frequency bins of $\hat{r}_i$, plotted at the weighted mean cost within each bin, and are a visual reference only. 2016–2024. Drivers who never take an honest test have no diagnostics and do not appear. The third series pools scrapping, idling and driving unregistered.

Appendix B. Data

B.1 Sources

Table B.1 lists the data sources. The inspection records are the core of the analysis; the other sources measure what happens to vehicles after they leave the inspection system and how the regulator enforces the program.

Table B.1: Data Sources

Source	Provider	Coverage	Use
Emissions inspection records	TCEQ, purchased	17 program counties, January 2007–December 2025; 224,038,908 raw test records	Inspection cycles, fraud indicators, fraud detection, repair diagnostics and costs
TORVET roadside remote sensing	DPS, public information request	14 Dallas–Fort Worth and Houston counties, May 2015–October 2025; 3.9 million readings at 13–63 sites a year	On-road use of lapsed vehicles, on-road emissions
Disciplinary actions	DPS, monthly public reports	Statewide, January 2016–May 2026; 2,381 actions	Timing and targets of sanctions
Overt and covert audits	DPS, public information request	All counties, county by year, 2015–2025	Audit intensity
Master Station List	DPS, public information request	All 15,269 licensed stations, with location and license dates	Station locations, entry, and exit
Google Maps business listings	Google Places API, collected February–March 2026	Stations on the Master Station List matched to a listing	Business category, rating, reviews, hours

Note: We count raw test records before the cleaning steps of Appendix B.2.

Each inspection record contains the OBD download described in Section 2.3, along with the station and analyzer that ran the test, its date and time, the result, the odometer reading, and the vehicle's make, model, and model year. When a vehicle that failed returns for a retest, the inspector may also record the category of repair performed and its total cost.[43] DPS publishes its disciplinary actions against stations and inspectors in monthly reports; each gives the date, the type of action (such as a suspension or revocation), and a description of the case.

B.2 Cleaning

Modern emissions inspections use On-Board Diagnostics (OBD-II), a system standardized on all 1996 and later model-year vehicles.[44] During an honest test, the inspector connects an

[43] Reporting is voluntary: over 2016–2024, 4.8% of retests report a positive repair cost and 12.4% report a repair category. Appendix F describes these data.

[44] In the program counties, gasoline vehicles are tested every year from their second through their 24th model year: new vehicles receive a two-year initial inspection, and vehicles 25 or more years old are exempt. Diesel vehicles, motorcycles, dual-fuel vehicles that cannot run on gasoline, and vehicles with antique registration are also exempt (30 TAC §114.50).

analyzer to the vehicle's diagnostic port and the analyzer downloads the vehicle's *electronic identity*—the electronic VIN ("eVIN") stored in the engine control module, calibration identifiers, and the count of supported parameters—together with the status of up to 13 *readiness monitors* and any stored Diagnostic Trouble Codes and the malfunction indicator light status (i.e., the check-engine light). A vehicle passes if the light is off, no critical codes are present, and enough monitors are "ready." The data from each inspection is transmitted and stored in a central database.

The raw inspection files run from January 2007 to December 2025 and contain 224,038,908 test records in the 17 program counties. We clean every year the same way. We keep only on-board diagnostic emissions tests (186,134,646 records), and drop records with a missing, malformed, or flagged VIN (463,412), an implausible model year (72), or an invalid test sequence number (64,483); tests at fleet and government stations, which are not open to the public (1,092,929); and duplicate records created by transmission errors between the analyzers and the central database (821,430). We build inspection cycles from the remaining 183,692,320 tests (Appendix B.3), then drop cycles that end in the same month as a later cycle for the same vehicle (912,751) and cycles on vehicles younger than two or older than 24 model years, for which the program does not require testing (3,576,269). Waivers are essentially absent (13 cycles), and we drop them. This leaves 169,180,792 cycles.

Our main analysis uses the 94,368,109 cycles with decisions in 2016–2024, but we use the other years in three ways. First, we run the fraud detection model in every year, so that we can classify passes before 2016. Second, we use each vehicle's cycles before 2016 to tell whether it enters the sample period with an unresolved malfunction: whether its last earlier cycle ended without a pass or with a fraudulent pass. Third, we use tests in 2025 to complete the cycles that began late in 2024. Before 2014, changes in analyzer technology and in how the data were reported shift the rates of two of the fraud indicators that feed the detection model, the eVIN anomaly and the communication protocol (Section 2.3). In these early years, the model therefore separates fraudulent from legitimate passes less precisely. This matters little for the main sample, because early passes affect it only through the state in which a vehicle begins 2016.

B.3 Inspection Cycles

We sort each vehicle's tests by date and time. A pass ends a cycle, and the next test starts a new one. A fail does not end a cycle, unless the vehicle has not passed within 300 days of the cycle's first test: the cycle then ends without a pass, and any later test starts a new cycle. For example, a vehicle that fails in March 2019 and next tests in May 2020 has two cycles: one ending without a pass in 2019 and one beginning in 2020. The 300-day rule is rarely needed: only 205,744 cycles (0.22% of the total) begin because the vehicle's previous cycle reached 300 days without a pass.

We assign each cycle to the calendar year of its decision: the date of the pass or, for a cycle that ends without a pass, the date of its first failed test. Most cycles are short: 95.3% consist of a single test, and the mean cycle has 1.05 tests. Of the 4,751,479 cycles that begin with a fail, 92.4% include at least one retest. Because renewal dates drift, a vehicle can have two cycles in one calendar year: 5.3% of vehicle-years in Dallas–Fort Worth and Houston do.

B.4 Stations

Table B.2 describes the inspection stations. In a typical year, about 5,700 stations perform emissions inspections. Most are independent repair shops; franchise chains, new-car dealers and national chains make up most of the rest. Station turnover is high: of the 10,865 stations active at some point between 2016 and 2024, 5,617 stopped inspecting before 2024. 15.2% of station-years are at a station that DPS sanctioned at least once over the period.

B.5 Roadside Remote Sensing

The TORVET data contain 3.9 million readings taken between May 2015 and October 2025 at 13 to 63 sites a year in the 14 Dallas–Fort Worth and Houston counties. We use the 3,180,600 readings from 2016 to 2024, the main sample period. Each reading records the date, time, and location, the measured exhaust concentrations, and the vehicle's VIN, make, model year, and registration expiration date, which DPS obtains from the registration record of its license plate. We link readings to the inspection records through the VIN.

Table B.3 summarizes the readings. Panel A shows that coverage grows over the sample, from 102,579 readings in 2016 to 600,483 in 2024, and that we match 92.7% of readings to a vehicle in the inspection records. Panel B splits matched readings by registration status: 86.1% are on vehicles with a current registration, 7.3% on vehicles whose registration expired one to six months earlier, and 6.6% on vehicles whose registration expired more than six months earlier. The longer the registration has been expired, the older the vehicle and the more it emits of all three pollutants.

The state uses the TORVET readings mainly to evaluate the program: the program's independent evaluations use them to compare the on-road emissions of vehicles inside and outside the inspection areas (Eastern Research Group, 2022). DPS can also use the readings to enforce against drivers. When a vehicle is measured as a high emitter, DPS can send its owner a notice requiring the vehicle to pass an inspection within 30 days, and the TORVET data record each notice. Table B.4 shows that these notices are rare. Over 2016–2024, DPS sent 5,291 notices, or 0.65 per 10,000 vehicles inspected in the two areas each year, about one in 15,000. The rate ranges from 0.11 per 10,000 in 2018 to 1.15 in 2024. It rises mainly because DPS took more readings: per reading, the notice rate fluctuates between 5 and 23 per 10,000 without a trend. The vehicles notified are old: their median model year is 2004. The notices target gross emitters, not necessarily vehicles with a fraudulent pass.

Appendix C. Fraud Detection Model Details

C.1 Fraud Indicators

Section 2.3 describes our four indicators and the kind of fraud each detects. Choosing them took care, because the model's identifying assumption depends on which indicators enter it. This appendix explains where our candidate indicators came from, how we build the four we use, how we narrowed a longer list down to them, and how we tested the choice.

Table B.2: Inspection Stations, 2016–2024

Panel A. Continuous characteristics

	Mean	25th pct.	Median	75th pct.	Std. Dev.
Station-years (N = 51,615)					
Inspections per year	1,868	519	1,190	2,233	2,480
Analyzers	1.2	1.0	1.0	1.0	0.5
Years in operation	8.7	2.0	6.0	13.0	8.6
Matched stations (N = 9,856)					
Rating (1–5)	4.42	4.30	4.50	4.70	0.48
Number of reviews	687	63	225	691	1,475

Panel B. Shares (%)

	Share (%)
Station-years	
Ownership type (sums to 100)	
Independent	63.9
National chain	7.5
Franchise chain	12.3
New-car dealer	10.9
Used dealer	5.4
Ever sanctioned 2016–2024	15.2
Stations active 2016–2024 (N = 10,865)	
Matched to a Google listing	90.7
Matched stations	
Google business category (sums to 100)	
Car repair	59.5
Dealer	13.5
Tire shop	8.7
Car wash / gas station	3.1
Other / unclassified	15.1
Listing features (each yes/no)	
Has website	62.6
Has phone listing	89.3
Open Saturday	71.3
Open Sunday	21.0

Note: Licensed private stations in the 17 program counties. In Panel A the unit is a station-year: each station counts once for every year from 2016 to 2024 in which it performed at least one emissions inspection (51,615 station-years), so a station active in all nine years counts nine times. We count inspections as inspection cycles, years in operation from the station's original license date. We classify ownership type from the station's name, the chain registry of Appendix C.4.1, and its Google listing: national chains are the company-operated brands in the registry, the brands used in Appendix C.4.1; franchise chains are its franchised brands, such as Meineke, together with the few brands whose stores are partly company-operated and partly franchised; new-car dealers carry a vehicle make in their name, using only makes that mostly appear in dealer names; used dealers are listed on Google as car dealers but carry no make in their name; and all other stations are independent. Ever sanctioned means the station received at least one suspension, revocation, or closure from DPS between 2016 and 2024. In Panel B the unit is a station: each of the 10,865 stations active in 2016–2024 counts once. The rows below the match rate use only the 9,856 matched stations, and they describe each listing as it stood when we collected it, in February and March 2026, not during the sample period. Because the listings postdate the sample, we match fewer of the stations that closed earlier: 95.4% of stations still inspecting in 2024, against 86.3% of those that stopped earlier.

Table B.3: TORVET Roadside Readings, 2016–2024

Panel A. Coverage by year

Year	Readings	Sites	Distinct vehicles	Matched to an inspected vehicle (%)
2016	102,579	24	93,068	92.6
2017	59,557	13	54,678	93.5
2018	176,921	28	148,638	93.9
2019	544,853	63	427,494	93.5
2020	352,674	48	268,827	93.4
2021	346,844	44	279,732	93.6
2022	486,816	42	368,714	92.9
2023	509,873	41	389,580	92.8
2024	600,483	49	451,945	90.2
2016–2024	3,180,600	92	1,981,352	92.7

Panel B. Matched readings by registration status, 2016–2024

	Registration current	Expired 1–6 months	Expired more than 6 months
Share of matched readings (%)	86.1	7.3	6.6
Mean vehicle age (years)	8.3	8.6	10.0
Mean CO (% by volume)	0.155	0.183	0.220
Mean HC (ppm, propane)	13.3	17.1	29.9
Mean NO (ppm)	105.1	123.5	165.3
Readings	2,307,139	195,868	176,030

Note: Readings at sites in the Dallas–Fort Worth and Houston counties. A site is a distinct measurement location; distinct vehicles are distinct VINs. A reading is matched if its VIN appears in the inspection records. Panel B pools all matched readings from 2016 to 2024 on vehicles aged 2 to 24 and weights each reading equally, so later years, with more readings, carry more weight. We take registration status from the expiration date in the TORVET record, which gives the first day of the month in which the registration expires. Because a Texas registration is valid through the end of that month, and the law allows five working days after expiration before driving becomes an offense, we count a registration as current until 31 days after the recorded date. We exclude the few readings with no expiration date, and, in both panels, the 2,867 readings with no recorded site county. CO, HC, and NO are exhaust concentrations. We drop readings with CO_2 below 5% and subtract from each HC reading a site-day zero offset, the median HC reading of vehicles aged 2 years or less at the same site and day (Bishop and Stedman, 2008); Appendix D.2 explains why the damage estimates need no such offset.

C.1.1 Where the Candidates Come From

We assembled candidate indicators from four sources. The program's independent evaluator has documented anomalies in the inspection records for years, including mismatched electronic VINs and unexpected communication protocols, and it tracks several others, such as changes in a vehicle's electronic profile or readiness status between a failed test and the next pass, quick retests and changes in weight class (Eastern Research Group, 2020, 2022, 2024a). Our conversations with the Texas Department of Public Safety and the Clean Air Task Force in the Travis County Constable's office pointed to more. They made clear that some fraudulent inspections are sold online and performed without the car present, which led us to check whether the inspector scanned the VIN barcode or typed the VIN in. Prior

Table B.4: TORVET High-Emitter Notices, 2016–2024

	Roadside sensing			Inspected fleet (DFW + Houston)		
Year	Readings	Notices	Per 10,000 readings	Vehicles inspected	Vehicles notified	Per 10,000 vehicles
2016	102,579	192	18.72	7,669,540	184	0.24
2017	59,557	99	16.62	8,101,252	91	0.11
2018	176,921	92	5.20	8,320,560	89	0.11
2019	544,853	563	10.33	8,639,706	454	0.53
2020	352,674	796	22.57	8,331,177	762	0.91
2021	346,844	789	22.75	8,881,891	756	0.85
2022	486,816	897	18.43	9,049,150	865	0.96
2023	509,873	813	15.95	8,972,912	789	0.88
2024	600,483	1,050	17.49	8,720,627	1,004	1.15
2016–2024	3,180,600	5,291	16.64	76,686,815	4,994	0.65

Note: Readings and notices from the TORVET data, for the same readings as Table B.3. Vehicles inspected are the distinct vehicles with an inspection in the Dallas–Fort Worth and Houston counties that year; vehicles notified are those among them that received at least one notice.

work motivates two of our indicators. The profile double match is similar in spirit to Oliva (2015), who detects fraud from tailpipe readings that are too similar from one car to the next; we look instead for electronic profiles that repeat across consecutive inspections of different vehicles. Typed or fabricated figures cluster on round numbers (Beber and Scacco, 2012; Kobak et al., 2016), the idea behind the round odometer. Finally, we built a few candidates from patterns in the data, such as high testing volume and inspections outside business hours. Table C.1 lists every candidate.

C.1.2 Construction

eVIN anomaly. The indicator flags a passing inspection of a vehicle of model year 2005 or later whose computer reports no VIN, or a VIN different from the one the inspector entered. Federal rules require these vehicles to report their VIN to the analyzer; older vehicles need not, so the indicator does not apply to them. Before comparing the two VINs we convert both to upper case, remove surrounding spaces and correct common transcription errors, such as the letter O for the digit 0.

Communication protocol. The vehicle's computer also reports the protocol it uses to communicate with the analyzer, and the protocol is a property of the vehicle's make, model and year. We learn the expected protocols from the data. We group vehicles by the characters of the VIN that identify the manufacturer, the platform and the model year, and call a protocol expected for a group when it appears on at least a quarter of the group's tests, among groups with at least ten tests. The indicator flags a passing inspection whose protocol is not expected for its group. Because the expectation is set by the kind of vehicle, the indicator fires when the analyzer is connected to a different kind of vehicle from the one entered, as with a surrogate, but not to a different vehicle of the same kind. Analyzers changed how they record the protocol in 2024, which weakens the indicator in that year (Appendix C.5.1).

Profile double match. Each inspection records an electronic profile of the vehicle's computer: its module identifier, its calibration identifier, a calibration check number and the number of data items it supports. The indicator flags an inspection whose profile exactly matches that of the previous inspection on the same analyzer of a different VIN. Two consecutive inspections of nominally different vehicles with identical profiles indicate that the same device, a surrogate vehicle or a simulator, was connected for both.

Round odometer. The indicator flags a passing inspection whose recorded odometer reading is positive and an exact multiple of 1,000. It is the weakest of the four: it flags only 0.62% of passing inspections, and fraud raises its firing rate far less than it raises the others. It earns its place in two ways. The inspector types the reading by hand, so it shares no equipment or data with the other three indicators, which gives the restrictions of Appendix C.4.2 power against dependence between it and each of the others. And it keeps four indicators in the model after the protocol indicator weakens in 2024.

C.1.3 Choosing the Set

Many candidates are poor choices for the model, so we screened them before estimating anything, on three criteria. First, an indicator must fire often enough to be informative. A candidate that flags fewer than 0.5% of passing inspections gives the model few flagged inspections at each station and vehicle type, and gives the restrictions of Appendix C.4.2 little power. Second, no two indicators should read the same part of the inspection record. Many candidates are rough transformations of one another, and two flags computed from the same record move together for reasons other than fraud, which would violate Assumption 1A by construction. The eVIN anomaly and the protocol indicator come closest: both use what the vehicle's computer reports when the analyzer connects to it. We keep both because they flag different failures, a missing or wrong VIN against a protocol that does not fit the vehicle, and Appendix C.4.1 measures the dependence this creates. Third, an indicator must be set by the inspection itself, by what the analyzer was connected to and what the inspector entered, rather than by the analyzer's workload or the station's practices. An indicator set by the workload, such as the number of tests an analyzer runs that day, is shared by every inspection on that analyzer that day rather than describing the inspection.

Table C.1 shows how the candidates fare. Two of them, high volume and a VIN typed rather than scanned, pass the first two screens and fail only the third; we test them alongside our indicators in Appendix C.1.4. The typed VIN fires on 39.36% of passing inspections, which mostly reflects older equipment rather than fraud. The remaining candidates fail an earlier screen. Most fire too rarely, often because they apply only to retests, a minority of inspections. Others are close variants of an indicator we keep or test: a match on the number of data items alone is a weaker version of the profile double match; rapid succession, like high volume, measures the analyzer's workload; and 99% of inspections outside business hours also have a typed VIN.

Table C.1: Candidate Fraud Indicators

Indicator	Record it reads	Passes flagged (%)	Status
A. Indicators used			
eVIN Anomaly	The VIN the car's computer reports to the analyzer: missing, or different from the VIN the inspector entered	7.46	Used
Communication Protocol	The data-link standard the car's computer uses with the analyzer, against the standard usual for that make, model and year	8.92*	Used
Profile Double Match	The car's computer identifiers, against the analyzer's previous test of a different VIN	8.29	Used
Round Odometer	The odometer reading the inspector typed in: an exact multiple of 1,000 miles	0.62	Used
B. Candidates tested but not used			
High Volume	Number of tests the analyzer ran that day (over 100)	1.84	Set by the analyzer's workload that day, not by the inspection; tested in Table C.2
VIN Typed, Not Scanned	Whether the inspector scanned the VIN barcode or typed the VIN	39.36	Set by the station's equipment and filing practice, not by the inspection; tested in Table C.2
C. Candidates not tested			
Data-Count Triple Match	The number of data items the car's computer supports, against the analyzer's previous two tests of different VINs	7.98	Close variant of profile double match, which requires the whole profile to agree
Electronic Profile Change	The car's computer identifiers on a retest, against its first test in the inspection cycle	0.40	Close variant of profile double match, on retests only; too rare (under 0.5% of passes)
Readiness Profile Change	The status of the car's emissions self-checks on a retest, against its first test in the inspection cycle	0.12	Close variant of profile double match, on retests only; too rare (under 0.5% of passes)
Rapid Succession	Time since the analyzer's previous test of a different VIN (under 5 minutes)	2.84	Close variant of high volume: both measure the analyzer's workload
Outside Hours	Time of day the test ended (before 6 am or after 9 pm)	0.70	Close variant of the typed VIN: 99% of passes outside business hours also have a typed VIN, against 39% of all passes (2022)
Warning-Light Mismatch	The dashboard warning light as the inspector recorded it, against the car's computer, with fault codes stored	0.09	Too rare (under 0.5% of passes)
Quick Retest	Time since the same car failed at the same station (under 15 minutes)	0.43	Retests only; too rare (under 0.5% of passes)
Weight-Class Switch	Weight class entered on a failed test and on the later pass	0.02	Retests only; too rare (under 0.5% of passes)
Fourth or Later Attempt	Number of attempts in the car's inspection cycle	0.07	Retests only; too rare (under 0.5% of passes)

Note: The car's computer is the onboard diagnostic system the analyzer plugs into. *Passes flagged* is the share of all passing inspections in the program counties, 2016–2024 pooled, that the indicator flags. eVIN Anomaly flags a model-year 2005 or later vehicle whose computer reports no VIN or a VIN different from the one entered. A candidate that flags under 0.5% of passes leaves the test of the restrictions little power. * Computed over 2016–2022, before the change in how the protocol is reported; the flag's rate falls in 2023 and collapses in 2024.

C.1.4 Testing the Choice

The screens rest on reasoning about how each indicator is generated. The over-identifying restrictions of Appendix C.4.2 let us check that reasoning against the data: with four indicators, the model implies two restrictions per year on how the indicators fire together, which fail when two of them share a cause other than fraud. Table C.2 re-estimates the model on every set of four drawn from six candidates: our four indicators plus high volume and the typed VIN, the two candidates that fail only the third screen.

The results support the screens. Every set without the typed VIN gives nearly the same answer as ours, with fraud rates between 8.79% and 9.52% and station estimates that track ours closely. Adding the typed VIN raises the fraud rate, up to 15.09%, and moves the station estimates away from ours, as expected of an indicator that mostly reflects older equipment. The year-by-year test rejects every set that contains the typed VIN without high volume, and decisively (pooled $|t| \geq 7.7$); 11 of the 15 sets pass it.

Our main set passes the year-by-year test but fails the pooled test, which Appendix C.4.2 discusses. Replacing the round odometer with high volume passes both, gives a fraud rate of 9.48% against 9.52%, and classifies 99.5% of stations the same way. The comparison is not quite like for like: high volume almost never fires in the first years of the sample, so the swap has fewer testable restrictions (14 against 16) and less power to reject. The swap would be a reasonable main specification too. Because the two sets give nearly identical estimates, we keep our main set, whose indicators are all set by the inspection itself. Passing the restrictions is not enough on its own, however: some sets that pair the typed VIN with high volume pass the year-by-year test while giving fraud rates well above ours. The screens choose the indicators, and the test confirms that the set they choose is sound.

C.2 Estimation

This appendix describes the sample, likelihood and algorithm behind the estimates of Section 3.2, and how we compute their standard errors.

C.2.1 Sample

We estimate the model on passing OBD inspections at stations in the seventeen program counties. Each inspection cycle contributes one test, its final one, assigned to the year in which it falls. We leave out failed tests because a station that sells a fraudulent pass records a pass, not a failure. In Appendix C.6, we consider whether fraud also occurs on failed inspections.

A vehicle type is a make crossed with a 3-year model-year bin starting in 1996, the first model year subject to OBD-II. The 10 makes with the most inspections are kept separately and the rest are pooled. A station and vehicle type enters a year's fit only with at least 5 inspections, and a station only with at least 30. Over 2016–2024 the sample holds 95.5 million passing inspections at 10,334 stations.

We fit the model separately in each year, pooling the metropolitan areas within each fit. Pooling lets the smaller markets borrow detection and false-positive rates from the larger ones, and we recover metropolitan fraud rates afterwards from the inspection-level posteriors. All 9 yearly fits converge.

Table C.2: Fraud Estimates Across Indicator Sets

	Indicators in the set											
	eVIN	Profile	Protocol	Odometer	High volume	Typed VIN	Fraud rate (%)	Rejected, year by year?	Rejected, pooled?	Testable restrictions	Correlation with main set	Stations alike (%)
Main set	✓	✓	✓	✓			9.52	No	Yes	16 of 18	1.000	100.0
Swap	✓	✓	✓		✓		9.48	No	No	14 of 18	0.997	99.5
		✓	✓	✓	✓		9.28	No	No	16 of 18	0.989	98.3
	✓		✓	✓	✓		8.79	No	No	14 of 18	0.985	98.8
	✓	✓		✓	✓		9.26	No	No	16 of 18	0.982	96.8
		✓	✓		✓	✓	10.77	No	No	14 of 18	0.965	91.2
	✓	✓			✓	✓	11.20	No	Yes	15 of 18	0.938	86.1
	✓		✓		✓	✓	11.97	No	No	14 of 18	0.934	83.6
			✓	✓	✓	✓	12.23	No	No	14 of 18	0.876	83.1
	✓			✓	✓	✓	14.99	No	No	14 of 18	0.799	75.6
		✓		✓	✓	✓	15.09	No	No	16 of 18	0.697	76.0
	✓	✓	✓			✓	10.64	Yes	Yes	16 of 18	0.988	94.2
		✓	✓	✓		✓	11.15	Yes	Yes	16 of 18	0.964	89.4
	✓	✓		✓		✓	12.23	Yes	Yes	18 of 18	0.923	82.1
	✓		✓	✓		✓	12.96	Yes	Yes	16 of 18	0.877	81.2

Note: Each row re-estimates the fraud model with a different set of four indicators, drawn from the six listed and marked with a check: the eVIN anomaly, the profile double match, the communication protocol, the round odometer, high volume and a VIN typed rather than scanned. The swap replaces the round odometer with high volume. Estimation uses every passing inspection in the program counties, one estimation per year. Every number is pooled over 2016–2024, the main sample. The fraud rate is the share of passing inspections whose estimated probability of fraud exceeds 0.5. The model implies two testable restrictions per year on how the four indicators occur together, 18 in all. A set's restrictions are rejected if any testable restriction is rejected at the 5% level after a Bonferroni correction for all 18 (a p-value below 0.05/18), with standard errors clustered by station. The pooled test stacks all testable years and rejects if either of its two restrictions has a p-value below 0.05/2. Sets that include the protocol indicator have at most 16 testable restrictions because the protocol is not recorded consistently in 2024; a restriction also cannot be tested in a year in which one of the set's indicators almost never fires, which happens to the high volume indicator in the first years of the sample. The last two columns compare each set's estimated station fraud shares with those of the main set, across station-years: their correlation, and the percentage of station-years that both sets place on the same side of a 0.05 fraud share.

C.2.2 Likelihood

Write $Y_i = (Y_{i1}, \ldots, Y_{iK})$ for the firing pattern of inspection i. Under Assumption 1A, the probability of each pattern at station s for vehicle type v is a mixture of two products, one for fraudulent and one for honest inspections:

$$\Pr(Y_i = y \mid s, v) \;=\; \rho_{sv} \prod_{j=1}^{K} \beta_{jv}^{y_j} (1-\beta_{jv})^{1-y_j} \;+\; (1-\rho_{sv}) \prod_{j=1}^{K} \Lambda_{jv}^{y_j} (1-\Lambda_{jv})^{1-y_j}. \tag{16}$$

Conditional independence is what makes each branch a product, and the product form is what Appendix C.4 tests.

The fraud share ρ_{sv} is free for every station and vehicle type, so nothing in the specification pushes a station toward fraud or honesty. The detection and false-positive rates vary by vehicle type but are shared across stations, with $\Lambda_{jv} = \text{logit}^{-1}(\zeta_{jv})$. In principle we could identify the detection and false-positive rates separately at each station (Appendix C.3), but many stations inspect few vehicles of a given type, so those estimates would be noisy.

Because our interest is in the fraud shares, we pool the rates across stations, estimating each from every station that inspects that vehicle type. The log-likelihood sums over stations and vehicle types,

$$\log \mathcal{L} \;=\; \sum_{s,v} \sum_{y \in \{0,1\}^K} c_{sv}(y)\, \log \Pr(Y = y \mid s, v), \tag{17}$$

where $c_{sv}(y)$ counts the inspections at station s of vehicle type v with firing pattern y.

C.2.3 EM Algorithm

Because fraud status is unobserved, we maximize the likelihood with the EM algorithm, which alternates between two simple steps.

E-step. Given the current parameters, Bayes' rule gives the probability that an inspection with pattern y is fraudulent:

$$\tau_{sv}(y) \;=\; \frac{\rho_{sv} \prod_j \beta_{jv}^{y_j} (1-\beta_{jv})^{1-y_j}}{\rho_{sv} \prod_j \beta_{jv}^{y_j} (1-\beta_{jv})^{1-y_j} + (1-\rho_{sv}) \prod_j \Lambda_{jv}^{y_j} (1-\Lambda_{jv})^{1-y_j}}. \tag{18}$$

Inspections with the same pattern at the same station and vehicle type share a posterior, so with four indicators there are at most sixteen distinct posteriors for each. The posterior $\hat{\tau}_i$ of Section 3.1 is $\tau_{sv}(y_i)$ at convergence.

M-step. Each parameter updates to a posterior-weighted share. The fraud share is the expected share of fraudulent inspections at the station,

$$\hat{\rho}_{sv} \;=\; \frac{1}{n_{sv}} \sum_{y} c_{sv}(y)\, \tau_{sv}(y), \tag{19}$$

and the detection rate is the share of fraudulent inspections of vehicle type v, across all stations, on which indicator j fires,

$$\hat{\beta}_{jv} \;=\; \frac{\sum_s \sum_{y:\, y_j = 1} c_{sv}(y)\, \tau_{sv}(y)}{\sum_s \sum_y c_{sv}(y)\, \tau_{sv}(y)}. \tag{20}$$

The false-positive rate is the same share computed with the weights $1 - \tau_{sv}(y)$. The honest weights come from the posteriors rather than from stations assumed to be honest, so we estimate the false-positive rates on the same sample as the fraud rate.

Convergence and labels. We stop when the log-likelihood rises by less than 10^{-8} in relative terms, or varies by less than 10^{-3} over ten iterations. Swapping the two groups leaves the likelihood unchanged, so the data alone cannot say which group is fraudulent. Assumption 1B(ii) fixes the labels: fraud raises every indicator's firing rate, $\beta_{jv} > \Lambda_{jv}$. We start the algorithm from each indicator's excess firing rate, which orients the two groups this way from the first iteration.

C.2.4 Outputs

The estimation returns two objects: the fraud share $\hat{\rho}_{sv}$ of each station and vehicle type, and the posterior $\hat{\tau}_i$ of each inspection. Every fraud rate in the paper is built from the posteriors. We classify an inspection as fraudulent when $\hat{\tau}_i > 0.5$, and a fraud rate, whether statewide, by metropolitan area or by station, is the share of inspections so classified. This classification rate would differ from the inspection-weighted mean of $\hat{\rho}_{sv}$ only if many posteriors were far from zero and one, and few are: only 0.4% of inspections have a posterior between 0.2 and 0.8. A station's fraud volume is the sum of $\hat{\tau}_i$ over its inspections. The model has no station-level parameter, so these station measures summarize the two outputs rather than estimate anything new.

C.2.5 Standard Errors

We compute standard errors by bootstrapping stations, the level at which fraud is sold. Each of 200 draws resamples stations with replacement within each metropolitan area and keeps every year of a drawn station's inspections, so that the standard error of the pooled rate reflects how fraud persists within a station across years. We re-estimate the model on each draw, starting from the main estimates, and recompute the fraud rate. The pooled 2016–2024 rate has a standard error of 0.46 percentage points, and yearly standard errors range from 0.45 in 2016 to 1.83 in 2024 (Appendix Table A.2). Re-estimating a subset of draws from the model's own starting values instead changes the fraud rate by at most 0.34 percentage points, so the starting point does not drive the results.

C.3 Identification

The covariances between indicators. Under Assumption 1A, the joint firing rate of two indicators is $\Pr(Y_{ij} = 1, Y_{ik} = 1 \mid s, v) = \rho_{sv}\beta_{jv}\beta_{kv} + (1 - \rho_{sv})\Lambda_{jv}\Lambda_{kv}$. Subtracting the product of their firing rates in Equation (1) gives the covariance in Equation (2). Its first term, $\rho_{sv}(1 - \rho_{sv})$, is the variance of fraud status among inspections of vehicles of type v at station s. It is largest at stations that mix fraudulent and honest inspections and zero at stations that do only one or the other. The second and third terms are the two indicators' informativeness, $\beta_{jv} - \Lambda_{jv}$, which measures how much more often an indicator fires on fraudulent than on honest inspections. Their product is large when both indicators fire much more often on fraud, and small when either fires at nearly the same rate on both kinds of inspection.[45] In scenario B of Section 3.1, the covariance of the two indicators, $4.8\% - 0.5\% \approx 4.3$ percentage points, equals $0.1 \times 0.9 \times 0.8 \times 0.6$, as Equation (2) implies.

A third scenario. Consider a third scenario, C, in which only 5% of inspections are fraudulent but the indicators are sharper: indicators 1 and 2 detect 99% and 97% of fraudulent inspections and fire on about 3% and 1% of honest ones. Scenario C produces the same firing rates as scenario B, 8% and 6%, and the same co-firing rate, 4.8%, and therefore the

[45] An indicator whose detection rate is close to its false-positive rate carries little information about fraud, which is why Assumption 1B(ii) requires every indicator to fire more often on fraudulent than on honest inspections.

same covariance. With only two indicators, scenarios B and C cannot be distinguished: the pairwise covariances reveal how informative each indicator is only up to a common scale that mixes the amount of fraud with the indicators' informativeness. The three-way co-firing of a third indicator separates the two (Equation (22) below).

Section 3.1 counts moments and unknowns. Equal counts are necessary for identification but not sufficient, so this appendix shows that the mapping from moments to parameters can be inverted. We work with three indicators, the case in which the model is exactly identified, and fix a station s and vehicle type v, dropping those subscripts.

The general result. Kruskal (1977) gives a rank condition for this kind of three-way decomposition, and Allman et al. (2009) apply it to latent class models: with three conditionally independent indicators and two latent groups, the parameters are identified up to relabelling the groups. Assumption 1B(ii) fixes the labels. The same structure, an unobserved binary state measured by several imperfect tests whose error rates are estimated rather than assumed, is the design of Hui and Walter (1980) in the diagnostic-testing literature, where our detection and false-positive rates are a test's sensitivity and its false-positive rate.

A constructive proof. The argument is also direct. Write $\kappa = \rho(1-\rho)$ and $d_j = \beta_j - \Lambda_j$, so that each covariance in Equation (2) is $\Sigma_{jk} = \kappa\, d_j d_k$, and let $M = \mathbb{E}\big[\prod_{j=1}^{3}(Y_j - \bar{Y}_j)\big] = \kappa(1-2\rho)\, d_1 d_2 d_3$ denote the three-way co-movement of the indicators around their firing rates $\bar{Y}_j$. Four steps recover every parameter.

1. *Each indicator's strength, up to scale.* Ratios of covariances isolate one indicator at a time:
$$\frac{\Sigma_{12}\,\Sigma_{13}}{\Sigma_{23}} \;=\; \kappa\, d_1^2, \tag{21}$$
and likewise $\kappa\, d_2^2$ and $\kappa\, d_3^2$.

2. *The fraud share.* The three-way moment separates the amount of fraud from the strength of the indicators:
$$\frac{M^2}{\Sigma_{12}\,\Sigma_{13}\,\Sigma_{23}} \;=\; \frac{(1-2\rho)^2}{\rho(1-\rho)}. \tag{22}$$
The right-hand side falls steadily as ρ rises from zero to one half, so it pins down ρ up to the swap $\rho \leftrightarrow 1-\rho$. Because every d_j is positive, the sign of M is the sign of $1-2\rho$, which settles the swap.

3. *Each indicator's strength.* With κ known, step 1 gives each d_j, taking the positive root.

4. *The rates.* The firing rate of each indicator is $\bar{Y}_j = \Lambda_j + \rho d_j$, so $\Lambda_j = \bar{Y}_j - \rho d_j$ and $\beta_j = \Lambda_j + d_j$.

The steps require that the station has both fraudulent and honest inspections of the vehicle type, $0 < \rho < 1$, and that every indicator is informative, $d_j > 0$, which is Assumption 1B.

More indicators. With K indicators the data have $2^K - 1$ moments and the model has $1 + 2K$ unknowns.

Indicators	Moments	Unknowns	
2	3	5	Not identified
3	7	7	Exactly identified
4	15	9	Six restrictions to spare

Every indicator beyond the third therefore adds restrictions the data must satisfy rather than identifying power. With four indicators we use two of them, the restrictions on the pairwise covariances, to test Assumption 1A in Appendix C.4.2.

The remaining cells. The argument above applies to any cell that contains both fraudulent and honest inspections. Because the rates are shared within a vehicle type, the fraud shares of the remaining cells then follow from Equation (1). Most cells are too small to estimate the rates on their own, which is why we pool them.

C.4 Testing Conditional Independence

The detection model rests on Assumption 1A: once we know whether an inspection is fraudulent, the fraud indicators fire independently of one another. The assumption fails when two indicators share a cause other than fraud. They then fire together on honest inspections, and the model mistakes the excess co-firing for fraud. This appendix reports the evidence summarized in Section 3.4. We start at stations that rarely commit fraud, where any excess co-firing must have an honest cause. We then test restrictions that the assumption imposes on the data, and bound how far the dependence we find could move the fraud rate.

C.4.1 Stations That Rarely Commit Fraud

The most direct evidence comes from stations that commit almost no fraud. There, the model should find little fraud, and if the indicators are independent given fraud status they should fire together about as often as their individual firing rates predict. Any excess must come from a shared honest cause.

We compare three groups of stations. The first, which we call the national chains, is the outlets of repair chains whose Texas stores are all operated by the company itself, such as Firestone, NTB, Pep Boys and Goodyear, excluding every station ever sanctioned for clean scanning.[46] These 602 stations performed 3,638,701 passing inspections over 2016–2024. The second group is all other stations never sanctioned for clean scanning, including franchised chains and dealerships, and the third is stations that were. We define the groups from ownership, station name and the sanction record, not from any output of the model.

[46] A station counts as a chain outlet in a year only if both the name recorded on its inspections that year and the name on its license record match a listed chain. The second requirement keeps out short-lived license numbers that record a chain's name without belonging to the chain. The list also includes several regional chains whose stores are all company-operated.

The model finds fraud where it should. Figure C.1 reports the share of passing inspections the model classifies as fraudulent in each group, for all vehicles and for vehicles so young that they have little to fail and their owners little reason to buy a fraudulent pass. The model classifies 22.6% of passing inspections at sanctioned stations as fraudulent, 7.0% at other never-sanctioned stations and 0.36% at the national chains. The ordering survives among nearly new vehicles, where the rates are lower: 4.0%, 1.6% and 0.64%. A false-positive rate common to all stations would not vary across the groups in this way. The model finds fraud where the regulator found it and almost none at the chains, which makes the chains a natural benchmark: nearly all of the co-firing there comes from honest inspections.[47]

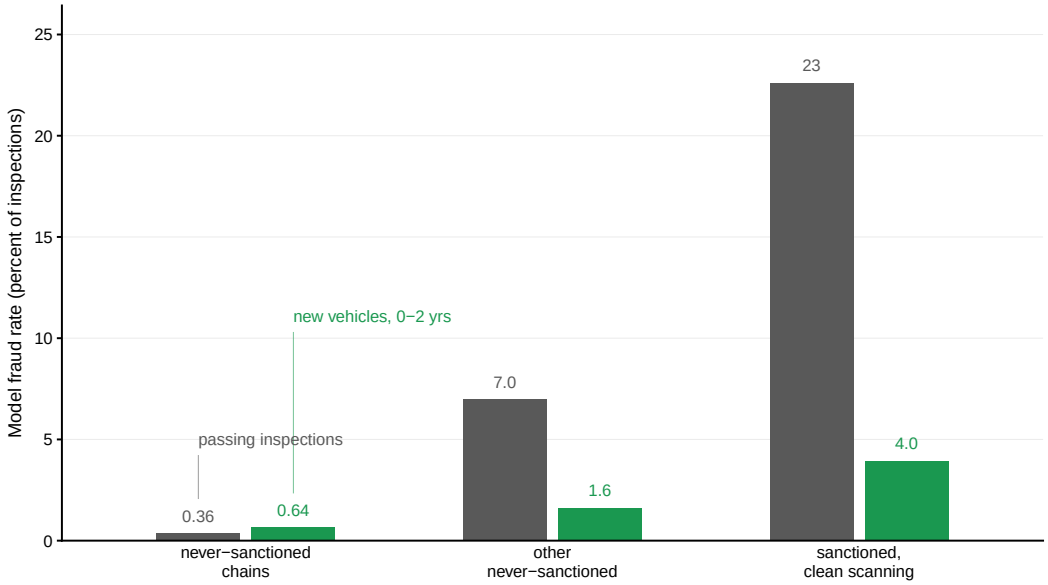


Figure C.1: Estimated Fraud by Population and Station Group, 2016–2024

Note: Share of inspections with posterior probability of fraud above one half, for passing inspections and for passing inspections of nearly new vehicles, at three groups of stations, pooled over 2016–2024. The national chains comprise 602 company-operated chain outlets never sanctioned for clean scanning. The groups are defined from ownership, station name and the sanction record, not from any output of the model.

Co-firing at the chains. For each pair of indicators, Table C.3 reports how often the two fire together at the chains, how often they would fire together if they were independent, and their correlation among inspections of the same vehicle type at the same station. The three pairs that involve the round odometer behave just as the assumption predicts: they fire together about as often as independence implies, with correlations of at most 0.006. One pair stands out. The eVIN anomaly and the protocol indicator fire together 12 times per 10,000 inspections, against 0.75 under independence, a correlation of 0.304. The profile double match fires together with each of these two somewhat more often than independence predicts, but by far less.

[47] Even a nearly new vehicle can be sold a fraudulent pass, so neither population is fraud-free by construction. Appendix C.6 treats failed inspections.

Table C.3: Co-Firing of the Indicators at the National Chains

	Fire together per 10,000				Below
Indicator pair	Observed	Independence	$\hat{\phi}$	95% interval	0.05
Pairs not involving the round odometer					
eVIN × protocol	12	0.75	0.304	$[0.279, 0.328]$	No
eVIN × profile	7.0	1.3	0.106	$[0.067, 0.146]$	No
profile × protocol	2.3	0.38	0.072	$[0.058, 0.086]$	No
Pairs involving the round odometer					
eVIN × odometer	0.42	0.20	0.005	$[0.003, 0.008]$	Yes
profile × odometer	0.15	0.11	0.001	$[-0.001, 0.004]$	Yes
protocol × odometer	0.21	0.078	0.006	$[0.003, 0.010]$	Yes
Model fraud rate, these stations vs. statewide				0.22–0.47% vs. 7.0–11.6%	

Note: The national chains are the 602 outlets of company-operated repair chains in the program counties, 2016–2024, excluding every station ever sanctioned for clean scanning. Observed is how often the two indicators fire on the same inspection; independence is how often they would if they were independent among inspections of the same vehicle type at the same station. $\hat{\phi}$ is their correlation among such inspections, pooled over stations and years, with 95% intervals from resampling stations; the last column marks pairs whose interval lies below 0.05 in absolute value. Inspections: 3,638,701, or 3,412,004 for pairs with the eVIN anomaly, which is defined from model year 2005. The bottom row gives the fraud rate at these stations and statewide, each as a share of all passing inspections in the year.

Where the excess comes from. The eVIN anomaly and the protocol indicator fire together because both depend on the vehicle identification number that the analyzer reads from the vehicle's computer. When that VIN matches the one the inspector enters, the protocol indicator fires on only 0.04% of inspections. When the computer reports no VIN, it fires on 5.4%, and when the computer reports a different VIN, on 22.9%.[48] Inspections whose computer reports no VIN, mostly of older vehicles, account for 26.4% of the times the two indicators fire together at the chains, and inspections whose computer reports a different VIN account for the remaining 73.6%. An honest inspection can turn up a different VIN when, for example, the vehicle's computer has been replaced with one taken from a different kind of vehicle.

Some of this co-firing may instead reflect a little fraud at the chains themselves, and no choice of benchmark can fully separate the two: any group of stations chosen because it rarely commits fraud also has few honest inspections in which both indicators fire. Failed inspections offer a cleaner look, because a vehicle that fails has not bought a pass. Among initial failed inspections at the same stations, the correlation for this pair is similar (0.2622 [0.2343, 0.2901]), which suggests that most of the co-firing is honest. To be safe, the bounds in Appendix C.4.3 treat all of it as honest, the case least favorable to the model.

C.4.2 Over-Identifying Restrictions

The restrictions. With four indicators, the model has more moments than unknowns, and the extra moments make Assumption 1A testable. Number the indicators 1 (eVIN anomaly), 2 (profile double match), 3 (communication protocol) and 4 (round odometer).

[48] These rates are for vehicles of model year 2005 or later, which report an electronic VIN.

Fix a station s and a vehicle type v, and let Σ_{jk} denote the covariance of indicators j and k among its passing inspections. Write $\kappa = \rho_{sv}(1-\rho_{sv})$ for the variance of fraud status and $d_j = \beta_{jv} - \Lambda_{jv}$ for the gap between indicator j's detection and false-positive rates. Under Assumption 1A, Equation (2) says that each covariance factors into a term common to every pair and one term for each indicator:

$$\Sigma_{jk} \;=\; \kappa\, d_j\, d_k.$$

The intuition is that fraud is the only reason the indicators move together. Each indicator then has a single strength, d_j, and the covariance of any pair is the product of the two strengths, scaled by how much fraud status varies at the station. One implication follows directly. Compare indicators 2 and 3 through their covariances with indicator 1, and again through their covariances with indicator 4. Both comparisons recover the same relative strength, and the same holds for indicators 2 and 4:

$$\frac{\Sigma_{12}}{\Sigma_{13}} \;=\; \frac{d_2}{d_3} \;=\; \frac{\Sigma_{42}}{\Sigma_{43}}, \qquad \frac{\Sigma_{12}}{\Sigma_{14}} \;=\; \frac{d_2}{d_4} \;=\; \frac{\Sigma_{32}}{\Sigma_{34}}.$$

Whichever indicator we compare against, indicators 2 and 3 (or 2 and 4) must come out with the same relative strength. Cross-multiplying each pair of ratios gives the two restrictions we test, a form that also avoids dividing by covariances close to zero:

$$T_1 \;=\; \Sigma_{12}\Sigma_{34} - \Sigma_{13}\Sigma_{24} \;=\; 0, \qquad T_2 \;=\; \Sigma_{12}\Sigma_{34} - \Sigma_{14}\Sigma_{23} \;=\; 0. \tag{23}$$

A shared cause other than fraud breaks this logic. It raises the covariance of the two indicators it affects, so the comparison gives different answers depending on which indicator we compare against, and T_1 or T_2 moves away from zero.

Implementation. We test the restrictions separately in each year. For every cell, a station and vehicle type, with at least 50 passing inspections in the year, we compute the six sample covariances and the two restrictions $\hat{T}_g(s,v)$. Under Assumption 1A each has mean zero, and so does their average across cells. Our test statistic is that average divided by its standard error:

$$\bar{T}_g \;=\; \frac{1}{N}\sum_{(s,v)} \hat{T}_g(s,v), \qquad t_g \;=\; \frac{\bar{T}_g}{\widehat{\mathrm{se}}(\bar{T}_g)}, \qquad g = 1, 2,$$

where N is the number of cells and we cluster the standard error by station. An attractive feature of these restrictions is that they use sample covariances only, so our test does not depend on any estimated parameter. Two restrictions in each year of 2016–2024 give 18 tests in all. Because we run so many, we apply a Bonferroni correction: a restriction is rejected at the 5% level only when its p-value falls below $0.05/18 = 0.00278$. In 2024 a change of inspection analyzers stopped them recording the protocol indicator consistently, so we cannot test that year. We still count its two restrictions in the correction, which only makes the test more conservative.

Results. Table C.4 reports the test year by year. Each row gives the number of cells and stations, followed by the t-statistic and p-value of each restriction. No p-value falls below the threshold of 0.00278: in every testable year we fail to reject both restrictions. The closest call comes in 2019 ($p = 0.0030$). Controlling the false discovery rate instead also rejects none.

Table C.4: Tests of the Over-Identifying Restrictions

			$T_1 = 0$		$T_2 = 0$	
Year	Cells	Stations	t	p	t	p
2016	47,987	4,116	−1.94	0.0521	−0.35	0.7267
2017	50,887	4,003	−0.46	0.6447	−1.30	0.1931
2018	50,696	4,138	−0.47	0.6376	−1.98	0.0475
2019	51,830	4,429	−0.68	0.4993	−2.97	0.0030
2020	49,983	4,246	−0.39	0.6949	−1.55	0.1212
2021	54,091	4,008	−2.22	0.0262	−1.76	0.0784
2022	55,393	3,927	−1.98	0.0473	−1.34	0.1786
2023	54,072	3,766	−0.32	0.7505	2.46	0.0139
2024	51,914	3,574	*not testable*			
Pooled 2016–2023	414,939	7,618	−2.99	0.0028	−3.70	0.0002

Note: Each row tests the restrictions of Equation (23), with indicators numbered 1 eVIN anomaly, 2 profile double match, 3 communication protocol and 4 round odometer. A cell is a station and vehicle type with at least 50 passing inspections in the year. The statistic t is the average of the restriction across cells divided by its standard error, clustered by station; p is two-sided. The pooled row averages over all cells in the testable years. With a Bonferroni correction for all 18 restrictions, a restriction is rejected at the 5% level when $p < 0.0028$. The protocol indicator is not recorded consistently in 2024, so that year is not tested.

The bottom row of the table pools all cells across the testable years, and here the restrictions do reject ($t = -2.99$ and $t = -3.70$). Several yearly p-values fall below 0.05 before the correction, and nearly all the yearly statistics share the same sign, so a departure too small to detect in any single year adds up when we pool 95,470,830 passing inspections. A direct check on the fitted model points the same way: the covariance that the model leaves unexplained between pairs of indicators departs clearly from zero in every year, with a maximum $|t|$ between 50.88 and 69.24. The assumption therefore does not hold exactly. Appendix C.4.1 locates the departure, and Appendix C.4.3 shows that it moves the fraud rate very little.

How large could the violation be? The pooled test suggests that the assumption does not hold exactly. With a sample this large, however, the test can detect violations far too small to matter for our estimates, so the relevant question is how large the violation could be. We first bound the dependence between indicators that the data allow, and Appendix C.4.3 then bounds its effect on the fraud rate. Suppose indicators j and k have correlation ϕ_F among fraudulent inspections. This adds $\rho_{sv}\,\phi_F\,\sigma_{jv}\sigma_{kv}$ to their covariance, where $\sigma^2_{jv} = \beta_{jv}(1-\beta_{jv})$ is the variance of indicator j among fraudulent inspections. Because each restriction multiplies the covariance of one pair by the covariance of the other, the restriction shifts by ϕ_F times an amount we can compute from the fitted model. For indicators 1 and

2, for example,

$$\Delta T_1 \;=\; \phi_F \cdot \underbrace{\rho_{sv}\,\sigma_{1v}\sigma_{2v}\,\Sigma_{34}}_{\delta_1(s,v)}.$$

We compute this amount in every cell and average it across cells, giving $\bar{\delta}_g$. The largest correlation consistent with the data is the one whose shift reaches the upper end of a one-sided 95% confidence interval for the restriction:

$$\phi_F^{\max} \;=\; \frac{|\bar{T}_g| + 1.645\,\widehat{\mathrm{se}}(\bar{T}_g)}{\bar{\delta}_g},$$

using whichever restriction is more sensitive to the pair. Table C.5 reports $\phi_F^{\max}$ for each pair, pooling all testable years.

Table C.5: Largest Dependence Consistent with the Restrictions

Indicator pair	95% upper bound on $\lvert\phi_F\rvert$
Pairs with the round odometer	
eVIN anomaly	0.0054
Profile double match	0.0032
Protocol	0.0024
Other pairs	
eVIN anomaly × profile double match	0.0604
eVIN anomaly × protocol	0.21
Profile double match × protocol	0.0267

Note: Each entry is the 95% upper bound on the correlation ϕ_F between two indicators among fraudulent inspections that the restrictions of Table C.4 allow, pooling 414,939 cells at 7,618 stations over 2016–2023. Section C.4.2 explains how the bounds are computed. Bounds for the other pairs are wide because their dependence enters the restrictions only through covariances with the round odometer, which are small; Section C.4.1 measures those pairs directly.

Among fraudulent inspections, most of the dependence the data allow is small. The bounds for the three pairs involving the round odometer are tiny, at most 0.0054. Even the bounds for the profile double match with the eVIN anomaly and with the protocol indicator are modest, at 0.060 and 0.027. Only the eVIN anomaly and the protocol indicator can carry a sizable dependence, with a bound of 0.21. This is the pair we were worried might violate the assumption: both depend on the VIN that the analyzer reads from the vehicle's computer.

These bounds say little about dependence among honest inspections. Honest inspections rarely set off any of the indicators, so even strong dependence among them adds almost nothing to the covariances, and the restrictions cannot detect it. We therefore study dependence among honest inspections directly in Appendix C.4.1, at national chains, where fraud is rare and nearly every inspection is honest. There the eVIN anomaly and the protocol indicator do fire together more often than independence predicts, and Appendix C.4.3 shows that this dependence moves the fraud rate very little.

Two further exercises show that this dependence does not drive our results. Appendix C.4.3 merges the eVIN anomaly and the protocol indicator into a single indicator, which removes their dependence, and the fraud rate changes by only +0.38 percentage points. Appendix C.1.4 replaces the round odometer with high station volume, a stronger indicator.

That set satisfies the restrictions even in the pooled sample ($t = -1.74$ and -0.34), and the fraud rate barely moves (9.48% against 9.52%). The assumption does not hold exactly, but the violation matters little for the headline results.

C.4.3 Sensitivity of the Fraud Rate

The model has no parameter for honest co-firing. Under Assumption 1A, the probability that two indicators fire together on an honest inspection is simply the product of their false-positive rates, so the model can explain the extra joint firings we find at the chains only as fraud. This gives a natural ceiling: the model's fraud rate at the national chains, 0.22–0.47% per year, bounds the fraud that such inspections can produce there.

A direct calculation bounds this channel. At the national chains, the eVIN anomaly and the protocol indicator fire together on 12 of every 10,000 inspections. If honest inspections elsewhere set off both indicators as often, and the model classified every such inspection as fraudulent, the statewide fraud rate would rise by only about 0.1 percentage point. There is therefore little evidence that this small correlation between the indicators changes our headline fraud rate much. To gauge the effect on the statewide fraud rate in other ways, we also run two exercises. Table C.6 reports both.

Table C.6: Sensitivity of the Fraud Rate to Dependent Indicators (2022 Only)

	Change in fraud rate (pp)	Correlation of station fraud shares	Stations classified alike (%)
Strongest dependence the firing rates allow (simulation)			
Among honest inspections			
eVIN anomaly × profile double match	+0.51	0.999	98.9
eVIN anomaly × protocol	+0.04	0.999	99.3
Profile double match × protocol	+0.04	0.999	99.4
Among fraudulent inspections			
eVIN anomaly × profile double match	−0.02	0.999	99.4
eVIN anomaly × protocol	−0.05	0.999	99.4
Profile double match × protocol	−0.34	0.999	99.3
Dependence removed (re-estimation on the 2022 data)			
eVIN anomaly and protocol merged into one indicator	+0.38	0.995	97.8
Protocol cannot fire when the computer reports no VIN	+0.05	0.992	97.0

Note: Top panel: we simulate data from the 2022 estimates with the strongest dependence that the pair's fitted firing rates allow, re-estimate the model, and compare with the simulation's true values; entries are means over 15 simulations, and the change in the fraud rate is net of a simulation with no dependence. Bottom panel: we re-estimate the model on the 2022 data and compare with the main estimates. The fraud rate is the share of passing inspections classified as fraudulent. A station is classified as committing fraud when its estimated fraud share exceeds 0.05.

Our first exercise asks how far the fraud rate could move under strong dependence. We simulate data from the 2022 estimates, building in the strongest dependence that a pair's fitted firing rates allow, re-estimate the model, and compare the estimates with the simulation's true values. We do this for each pair of the three indicators other than the round odometer, separately among honest and among fraudulent inspections. The fraud rate moves by at most +0.51 percentage points, for dependence between the eVIN anomaly

and the profile double match among honest inspections. Among honest inspections, the pairs involving the protocol indicator move it by +0.04 points at most. Station fraud shares stay correlated with the truth at 0.999 or more, and at least 98.9% of stations are classified correctly. This exercise is not a strict worst case. The strongest dependence is capped by the fitted false-positive rates, which are themselves estimated under the assumption, and because the protocol indicator rarely fires on honest inspections of most vehicle types, the cap falls short of the correlation we measure at the chains.

Our second exercise removes the dependence from the real data. We re-estimate the model for 2022 in two variants. The first merges the eVIN anomaly and the protocol indicator into a single indicator, which eliminates their dependence by construction; the fraud rate changes by +0.38 percentage points, from 11.62% to 12.00%. The second stops the protocol indicator from firing on inspections whose computer reports no VIN, the source of 26.4% of the joint firings at the chains; the fraud rate changes by +0.05 percentage points. The two variants leave 99.2% and 98.3% of inspections with the same classification. Both raise the fraud rate slightly. Removing dependence among honest inspections alone would lower it, so neither refit suggests that the main specification overstates fraud.

The two exercises tell the same story. Neither moves the fraud rate by as much as a percentage point, against yearly fraud rates of 7.0–11.6%, and both leave station fraud shares nearly unchanged. The dependence is real, but it causes little bias in our fraud estimates.

C.5 Robustness

C.5.1 The 2024 Change in Protocol Reporting

In late 2023, the communication protocols that analyzers record for vehicles began to change. From 2024 on, the protocol indicator fires far less often, and its detection rate falls from 82% in 2023 to 38% in 2024: the indicator loses about 53% of its power in a single year (Table C.7). We asked the regulator, the contractor that maintains the inspection database and the analyzer manufacturers about the change, but none could tell us its cause. The same change is why we cannot test the restrictions of Appendix C.4.2 in that year.

The timing makes this a serious concern. The crackdown we study in Section 4 began in 2022, and the years after it are exactly the years in which we measure how fraud responded. If measured fraud fell partly because one of our four indicators lost its power, part of the decline we attribute to the crackdown would be an artifact of the data rather than a change in behavior.

The other indicators give no sign of a similar problem. The eVIN anomaly and the profile double match strengthen over the period as the fleet turns over and more vehicles report the information they read, the round odometer fires rarely throughout, and all four false-positive rates stay small. Before the change, the protocol indicator was stable too, with a detection rate between 78% and 86% in every year through 2023.

To check whether the change drives the decline, Figure C.2 re-estimates the model in every year without the protocol indicator; with three indicators the model is still exactly identified (Appendix C.3). The two series are nearly identical. The fraud rate under the main specification peaks at 11.6% in 2022 and falls to 7.0% by 2024; without the protocol

Table C.7: Indicator Rates and the Fraud Rate by Year

	Detection rate $\hat{\beta}_{jv}$ (%)				False-positive rate $\hat{\Lambda}_{jv}$ (%)				
Year	eVIN	Profile	Protocol	Odometer	eVIN	Profile	Protocol	Odometer	Fraud rate (%)
2016	48.1	63.0	78.5	1.5	0.91	1.40	0.10	0.32	9.6
2017	52.6	64.6	80.6	2.0	1.02	1.31	0.09	0.32	9.4
2018	59.6	66.4	84.7	2.3	1.14	1.42	0.08	0.32	9.7
2019	63.5	67.9	84.9	2.9	1.15	1.27	0.09	0.32	10.1
2020	66.6	73.3	86.3	3.7	1.05	1.17	0.09	0.34	9.9
2021	72.4	78.4	86.2	3.9	1.16	1.00	0.14	0.37	10.4
2022	75.0	80.1	85.4	3.8	1.22	1.18	0.12	0.35	11.6
2023	70.7	89.9	82.0	4.4	1.30	1.13	0.11	0.33	7.8
2024	78.7	81.7	38.4	4.5	1.01	0.19	0.06	0.28	7.0

Note: Estimates from the main specification, fitted separately in each year. The detection rate $\hat{\beta}_{jv}$ is the probability that an indicator fires on a fraudulent inspection, and the false-positive rate $\hat{\Lambda}_{jv}$ the probability that it fires on an honest one. Each is averaged over vehicle types, weighting detection rates by fraudulent inspections and false-positive rates by honest ones. The fraud rate is the share of passing inspections classified as fraudulent. The protocol indicator's detection rate falls sharply in the final year, when analyzers change how they record the communication protocol.

indicator it peaks at 11.5% and falls to 6.9%, so 98.8% of the decline survives. The two series never differ by more than 0.46 percentage points in any year. If the weakening indicator had produced the decline, removing it would shrink the last step, which coincides with the change in the analyzers. Instead, that step is slightly larger without the protocol indicator (1.00 against 0.84 percentage points). We therefore find no evidence that the change in protocol reporting drives the measured fall in fraud.

C.5.2 Vehicle-Type Grouping

The detection and false-positive rates vary by vehicle type, so our estimates could depend on how we group vehicles. Table C.8 re-estimates the model for 2022 under 19 groupings, from model year alone to the fifteen largest makes crossed with single model years. The fraud rate ranges from 10.41% to 12.70%, against 11.62% under our grouping, and every grouping classifies at least 97.5% of stations the same way as ours. The variation follows a simple pattern. Groupings that ignore make or model year give somewhat higher rates, likely because they cannot absorb differences across vehicles in how often the indicators fire on honest inspections. The finest groupings give lower rates mostly for a mechanical reason: they split the sample so finely that more inspections fall below the size rules and are counted as not fraudulent. Our grouping sits between the two.

C.6 Fraudulent Failed Inspections

In the main analysis, we estimate the detection model on the subsample of all *passing* inspections because we aim to identify clean scans, which are by definition passed inspections on malfunctioning vehicles. But a dishonest station could also potentially fake a *failed* inspection, for example to try to sell the owner a repair the car does not need, or to sell a clean scan to an owner whose car would have passed anyway. Most drivers know the worry:

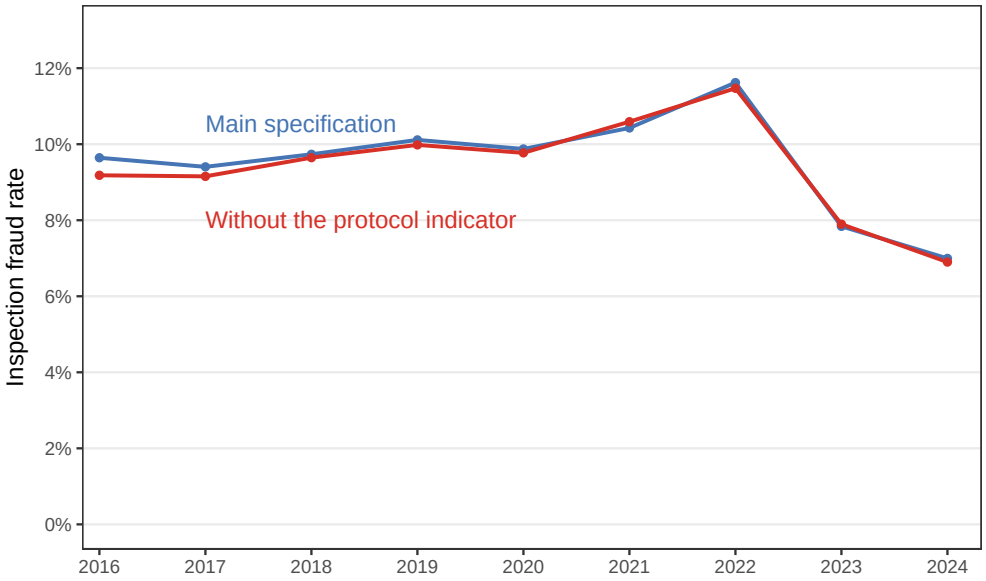


Figure C.2: Fraud Rate With and Without the Protocol Indicator

Note: Share of passing inspections classified as fraudulent in each year, 2016–2024, under the main specification and with the communication protocol indicator removed. Both series are estimated on the same inspections with the same estimator. They share three of the four indicators, so the gap between them measures the protocol indicator's contribution, not sampling variation. With three indicators the model is still exactly identified (Appendix C.3).

the mechanic says the car needs an expensive repair, and there is no easy way to tell whether it really does. This appendix provides evidence that fraudulent failures are rare.

C.6.1 Sample and Estimation

We start from every initial failed inspection between 2016 and 2024 and keep those in which the analyzer actually connected to the car's computer: the connector check passed, the analyzer and the car agreed on a communication protocol, and the analyzer read back the status of the check-engine light. When there is no connection the electronic VIN is blank for purely mechanical reasons, and the model would mistake that for an eVIN anomaly. That leaves 4,830,343 failed inspections.

We then re-estimate the detection model of Section 3 on passing and failed inspections pooled together. We specify that the parameters of the fraud detection model (i.e., how often each indicator fires under fraud and under honest testing) are common across passing and failed inspections. However, we allow each station to have two different fraud shares for passing and failed inspections, respectively. Thus, a station that sells many clean scans is not presumed to fake its failures as well.

Recall that the model estimates these fraud shares cell by cell, where a cell is one station and one vehicle type (a make and a range of model years) in one year. When a cell holds only a handful of inspections, its fraud share would be estimated from the very inspections it is then used to score. In the main analysis, we require at least 5 inspections per cell and 30 per station. For passes this barely binds: 97.1% of passing inspections sit in cells large enough to estimate.

Table C.8: Robustness to the Vehicle-Type Grouping (2022 Only)

How vehicles are grouped	Vehicle types	Inspections estimated (%)	Fraud rate (%)	Change vs. main (pp)	Stations alike (%)
Model year only, single years	28	99.6	12.13	+0.51	99.2
Model year only, 2-year groups	14	99.8	12.18	+0.56	99.0
Model year only, 3-year groups	10	99.9	12.21	+0.59	99.0
Model year only, 5-year groups	6	99.9	12.15	+0.53	98.9
Top 5 makes, no model year	6	100.0	12.70	+1.08	98.4
Top 5 makes × 5-year model-year groups	36	99.2	11.89	+0.27	99.2
Top 5 makes × 3-year model-year groups	56	98.6	11.81	+0.19	99.4
Top 5 makes × 2-year model-year groups	84	97.9	11.44	−0.18	99.5
Top 5 makes × single model years	163	95.0	11.18	−0.44	98.6
Top 10 makes, no model year	11	99.9	12.70	+1.08	98.4
Top 10 makes × 5-year model-year groups	65	98.3	11.80	+0.18	99.4
Top 10 makes × 3-year model-year groups (main)	101	97.0	11.62	—	100.0
Top 10 makes × 2-year model-year groups	150	95.5	11.22	−0.40	99.1
Top 10 makes × single model years	289	90.5	10.68	−0.94	97.9
Top 15 makes, no model year	16	99.8	12.69	+1.07	98.5
Top 15 makes × 5-year model-year groups	94	97.4	11.71	+0.09	99.5
Top 15 makes × 3-year model-year groups	145	95.4	11.47	−0.15	99.5
Top 15 makes × 2-year model-year groups	213	93.3	10.99	−0.63	98.6
Top 15 makes × single model years	411	86.9	10.41	−1.21	97.5

Note: Each row re-estimates the fraud detection model on all 11,239,617 passing inspections of 2022, changing only how vehicles are grouped into vehicle types. The inspections, the four fraud indicators and the estimation rules are those of the main estimate, whose grouping is marked (main). Makes are ranked by their number of 2022 inspections; makes outside the top group are pooled as one other make. Model-year groups start at 1996. Vehicle types: the number of vehicle types left after dropping station and vehicle-type pairs with fewer than 5 inspections and stations with fewer than 30. Inspections estimated: the share of passing inspections left after those size rules. Fraud rate: the share of all passing inspections that are more likely fraudulent than not; inspections dropped by those size rules count as not fraudulent, as in the main estimate. Change vs. main: the difference from the main fraud rate, in percentage points. Stations alike: the share of the 5,414 stations in the main estimate that are placed on the same side of a 0.05 station fraud share as in the main estimate; a station a grouping drops counts as below it.

On the other hand, failures are far less common, and most stations see only a few per vehicle type, so just 55.2% of failed inspections sit in cells with enough inspections to meet our minimum threshold for estimation. We maintain the same minimum threshold when estimating fraud shares for failed inspections. For the remaining 44.8% of failed inspections that do not meet the minimum threshold, we impute a fraud share from the station's other failed inspections, pooled across its estimated cells, or, for a station with none, the average across all stations that year.

C.6.2 Results

Fraudulent failures turn out to be rare. Over 2016–2024 the model flags 70,434 of the 4,830,343 failed inspections (1.5%), against 9,094,378 of 95,470,830 passing inspections (Table C.9, Panel A). For every hundred fraudulent passes there are fewer than one fraudulent failure: the ratio is 0.77%. Most of the flagged failures occur at stations that were later sanctioned for clean scanning, where 4.0% of failed inspections are flagged, compared with 1.0% at other stations and 0.2% at chain outlets that were never sanctioned. Like fraudulent passes, they fall off after Operation Cinderblock (Panel B).

To better understand these apparently fraudulent failed inspections, we examine what

Table C.9: Fraudulent Failed Inspections: Counts and What Follows, 2016–2024

	Passing inspections			Connected failed tests		
	N	Fraud class	%	*N*	Fraud class	%
A. All years, 2016–2024						
All stations	95,470,830	9,094,378	9.5	4,830,343	70,434	1.5
Sanctioned for clean scanning	18,750,516	4,134,885	22.1	831,775	33,478	4.0
Other unsanctioned	72,877,966	4,946,352	6.8	3,700,924	36,487	1.0
Unsanctioned chain	3,842,348	13,141	0.3	297,644	469	0.2
B. All stations, by period						
2016–2021 (pre-crackdown)	62,281,542	6,163,466	9.9	3,138,318	54,618	1.7
2022 (crackdown begins)	11,239,617	1,307,392	11.6	508,065	8,115	1.6
2023–2024 (post-crackdown)	21,949,671	1,623,520	7.4	1,183,960	7,701	0.7

C. What follows a connected failed test, all stations, 2016–2024

	Failed tests (*N*)	Next test at same station, same day (%)	First later pass in fraud class (%)	Repair recorded before that pass (%)
Fraud signature	70,434	68.8	78.3	8.3
No fraud signature	4,759,909	25.1	8.3	11.9

Note: Inspections in the fraud class (posterior probability of fraud above one half) under the detection model re-estimated on passing inspections and connected initial failed inspections together, with a separate fraud share for each station's failed inspections. Failed inspections in sparse cells are scored with the station's pooled fraud share for failed inspections (Appendix C.6.1). Station groups are defined on ownership, recorded name and the regulator's sanction record (Appendix C.4.1). Panel C follows each connected initial failed inspection to the car's next tests in the same inspection cycle (including January to March of the following year); a repair is recorded if any test between the failure and the next pass reports a repair category or a positive repair cost.

happens to the vehicle after the failure. Panel C shows that a flagged failure is usually followed by another inspection at the same station the same day (68.8% of cases, against 25.1% after other failures), and its next pass is fraudulent 78.3% of the time, against 8.3%. Repairs, if anything, are less common after a flagged failure (8.3% against 11.9%), and only 0.8% of flagged failures end in a repair and a legitimate pass at the same station. Stations do not appear to fake failures to sell repairs.

We read most of these failures as clean scans that went wrong. The station hooks up a surrogate car or a simulator, the scan fails (perhaps because the surrogate's readiness monitors came back as invalid), the analyzer logs the failure, and the station tries again. On the other hand, it is possible that some of these failures are intentional, so that the station can offer the owner a pass for a fee. Unfortunately, we cannot confidently distinguish these two explanations with the available data. Our measure also cannot see failures that leave no electronic trace, but there are few ways for an inspector to fail a car without leaving one. The main ones are reporting a missing or faulty gas cap and reporting a lit check-engine light, and both are easy for the driver to verify: the check-engine light is visible on the

dashboard, and the gas cap can be checked or replaced cheaply. It is therefore unlikely that this type of fraud is common. Given how rare fraudulent failures are, we abstract away from modeling fraudulent failures in our main analysis, and we believe they are unlikely to have any meaningful impact on our results.

Appendix D. On-Road Emissions and Pollution Damages

This appendix describes the comparison of roadside emissions in Section 3.3, how we measure emission rates, how we turn the comparison into pollution damages, and how robust it is, including whether fraudulent passes became harder to detect after the crackdown. The remote sensing data are described in Appendix B.5.

D.1 Sample and Specification

Each observation is one roadside reading of a gasoline vehicle registered in Dallas–Fort Worth or Houston and read by a sensor in the same metropolitan area, over 2016–2024. We link each reading to the vehicle's most recent inspection that ended in a pass or a fail in the 365 days before the reading, and classify the reading by that inspection: a fraudulent pass (posterior probability of fraud above one half), a repaired pass (an honest pass in a cycle whose initial test failed), an honest first-time pass, or a fail with no pass since. Readings with no inspection in the window do not enter. The sample has 2,199,150 readings, of which 235,032 follow a fraudulent pass and 75,478 a repaired pass.

For reading r of vehicle i we estimate

$$y_r = \gamma^{\text{fraud}}\,\text{Fraud}_i + \gamma^{\text{repair}}\,\text{Repair}_i + \gamma^{\text{fail}}\,\text{Fail}_i + w_r'\theta + \alpha_{\text{make}(i)} + \alpha_{\text{MY}(i)} + \alpha_{\text{date}(r)} + \varepsilon_r, \quad (24)$$

where y_r is NO or HC in grams per kilogram of fuel; Fraud_i, Repair_i and Fail_i indicate that the linked inspection was a fraudulent pass, a repaired pass or a fail not yet followed by a pass, so that honest first-time passes are the omitted group; w_r holds the vehicle's speed, acceleration and engine load at the reading; and the α terms are make, model-year and reading-date fixed effects. The excess emissions of a fraudulent pass, relative to a repaired one, are $\gamma^{\text{fraud}} - \gamma^{\text{repair}}$. We weight each reading by the inverse of the vehicle's number of readings in that calendar year, so that each vehicle-year counts once, and cluster standard errors by vehicle and by reading date. The group estimates of Appendix Figure A.6 interact Fraud_i and Repair_i with the six vehicle age and mileage groups of the demand model: its three age bins, each split at the median odometer for that age.

D.2 Emission Rates

A remote sensor measures, in the exhaust plume of a passing vehicle, the concentrations of carbon monoxide, hydrocarbons and nitric oxide relative to carbon dioxide. We convert these ratios into grams of each pollutant per kilogram of fuel burned with the carbon balance

used in the remote sensing literature (Bishop, 2019; Eastern Research Group, 2024a). Let Q, Q' and Q'' be the measured ratios of CO, HC and NO to CO_2. Then

$$\mathrm{CO} = \frac{28\,Q}{0.014\,D}, \qquad \mathrm{HC} = \frac{2 \cdot 44\,Q'}{0.014\,D}, \qquad \mathrm{NO} = \frac{30\,Q''}{0.014\,D}, \qquad D = 1 + Q + 6\,Q', \tag{25}$$

in grams per kilogram of fuel. The numerators convert moles of each pollutant into grams with its molecular weight; hydrocarbons are measured as propane, with a molecular weight of 44. The denominator counts the moles of carbon in the exhaust, which leaves the engine as CO_2, CO or hydrocarbons, and 0.014 kilograms of fuel per mole of carbon turns it into kilograms of fuel, treating gasoline as CH_2. Without the CO and HC terms, dividing by CO_2 alone would understate the fuel burned and overstate emission rates, most of all for the high emitters this comparison is about.

The two factors of two account for hydrocarbons the sensor does not see. A sensor calibrated with propane detects about half of the hydrocarbons in exhaust (Singer and Harley, 1998), so we double the measured HC ratio, both in the numerator for HC and in the carbon count of the denominator, where it appears as $6 = 2 \times 3$ carbon atoms per propane molecule. NO is measured, and kept here, as the mass of NO; Appendix D.4 values it as NO_x.

Remote sensing readings are noisy around zero, and some readings of clean vehicles are negative. We keep negative readings rather than truncate them, because truncation would raise the mean of every group. Earlier work subtracts a site-day zero offset from HC readings, the median reading of new vehicles at the same site and day (Bishop and Stedman, 2008). On most days all readings come from sensors in one county, so the reading-date fixed effects in Equation (24) absorb most of such an offset. It would mainly shift the mean of the comparison group rather than excess emissions, and we do not apply one.

D.3 Selection on the Severity of the Defect

Owners choose fraud when a repair is expensive, so fraudulently passed vehicles may have worse defects than repaired ones and would emit more even after a repair. If so, the excess emissions in Table 2 overstate what a repair would achieve. The data point to this selection: among vehicles that failed a genuine test, those that went on to pass fraudulently had a mean predicted repair cost, which we estimate from the failed test's diagnostics (Appendix F), of \$320, against \$250 for those that were repaired.

To hold the defect fixed, we use the vehicles that failed a genuine test before passing, some of which were then repaired and some of which then passed fraudulently. For these vehicles we observe the diagnostic profile of the failed test: whether the check-engine light was on, the number of stored trouble codes and which emission-control monitors were not ready. The profile is recorded before the owner chooses between repair and fraud, so it measures the defect. Adding fixed effects for the profile lowers excess emissions by 15% for NO, from 2.05 to 1.74 grams per kilogram of fuel, and by 18% for HC, from 0.93 to 0.77 (Table D.1, columns 7 and 8). Controlling instead for the predicted repair cost at the vehicle's previous genuine inspection shrinks it by a similar amount.

We do not use these estimates directly. The failed-test sample is small, and once we control for the diagnostic profile, the older vehicle age and mileage groups contain too few

fraudulently passed vehicles to estimate their excess emissions precisely. Instead, we treat the result as a measure of how large the bias from worse defects is, and make a conservative adjustment: we lower the excess emissions of every group from the main regression (Appendix Figure A.6) by 15% for NO and 18% for HC.

The adjustment also applies to the many fraudulently passed vehicles that never failed a genuine test, whose owners bought a fraudulent pass directly and whose defect we therefore do not observe. Two checks support this. First, controlling for the predicted repair cost at an earlier genuine inspection, which we observe for some of these vehicles, lowers excess emissions by a similar amount. Second, on the road these vehicles emit about as much as vehicles that failed first and then passed fraudulently, comparing vehicles of the same make, model year and age and mileage group: the difference is 0.08 grams per kilogram of fuel for NO (standard error 0.08) and 0.02 for HC (0.09). Without the adjustment, damages would be \$299 per fraudulently passed vehicle-year instead of \$251; we regard the unadjusted figure as an upper bound.

D.4 Excess Emissions to Monetary Damages

For a fraudulently passed vehicle in vehicle group g the damage per year is

$$d_g = \kappa \sum_{p \in \{\mathrm{NO},\,\mathrm{HC}\}} \rho_p \left(\gamma_{pg}^{\mathrm{fraud}} - \gamma_{pg}^{\mathrm{repair}}\right) m_p \, \frac{2.79}{\mathrm{MPG}} \, \mathrm{VMT}_g \, 10^{-6} \, \mathrm{MD}_p, \tag{26}$$

where ρ_p lowers excess emissions by the percentage in Appendix D.3 ($\rho_{\mathrm{NO}} = 0.85$, $\rho_{\mathrm{HC}} = 0.82$); 2.79 kilograms of fuel per gallon and the vehicle's EPA combined fuel economy turn grams per kilogram of fuel into grams per mile; VMT_g is annual miles; 10^{-6} converts grams to metric tons; MD_p is the AP4 marginal damage per ton of ground-level emissions in the vehicle's registration county, or the inspection station's county where the registration county is missing (Muller and Mendelsohn, 2009; Muller et al., 2011), NOx for NO and VOC for HC; $m_{\mathrm{NO}} = 46/30$ values NO as NOx by the mass of NO_2, as in the National Emissions Inventory (U.S. Environmental Protection Agency, 2026a), and $m_{\mathrm{HC}} = 1$; and $\kappa = 1.212$ converts AP4's 2020 dollars to the 2024 dollars of the repair costs with the consumer price index. We average over vehicles within each group, weighting counties by their fraudulent vehicle-years. Fraudulent vehicle-years outside the six groups, 11.6% of the total, take the excess emissions of their age bin.

Annual miles come from the odometer readings of fraudulently passed vehicles at their own inspections. Inspectors type the reading at every inspection, so a fraudulent inspection can carry the vehicle's true mileage or an invented one. We drop round readings and readings that fall outside the range set by the vehicle's readings at its honest inspections and fails before and after. Annual miles are the change between consecutive remaining readings divided by the time between them, averaged by metropolitan area, age bin and mileage group. The average fraudulently passed vehicle drives 12,128 miles a year. These miles come only from vehicles that return for another inspection, which may favor vehicles that stay on the road.

Excess emissions, and therefore damages, depend on the characteristics of the vehicle. Excess emissions rise steeply with both age and mileage (Appendix Figure A.6). Among

vehicles aged 2–15 with below-median mileage, fraudulent passes emit 0.77 grams more NO per kilogram of fuel than repaired ones; among vehicles aged 20–24 with above-median mileage, they emit 3.95 grams more. These are the vehicles that are passed fraudulently most often: older vehicles have the highest fraud rates (Table 1), and within each age group 60–74% of fraudulent passes are of vehicles with above-median mileage. Damages average $271 in Dallas–Fort Worth and $231 in Houston.

The counterfactuals of Section 7 assign unregistered driving the damage per mile of a fraudulent pass, scaled by its relative miles driven, which is conservative (Table D.1, column 6). Damages omit carbon monoxide, which AP4 does not value.

Damages range from $79 for vehicles of age 2–15, below-median mileage to $477 for vehicles of age 20–24, above-median mileage (Appendix Figure A.6).

D.5 Robustness

Appendix Table D.1 varies the specification one choice at a time; column 1 repeats the main estimate of Table 2 (2.28 grams per kilogram of fuel for NO and 1.04 for HC).

Comparison group (column 2). Repaired vehicles are a small and selected group: owners who choose to repair may maintain their vehicles differently, so the main estimate could partly reflect who repairs rather than what fraud does. We instead compare fraudulently passed vehicles with the much larger group of vehicles that passed honestly at their first test. Excess emissions are 2.37 for NO and 1.03 for HC, close to the main estimate, because repaired vehicles emit about as much as honestly passed ones.

Fraud measure (column 3). The main specification classifies an inspection as fraudulent when its posterior probability of fraud exceeds one half, so some honest inspections are counted as fraud and some fraud as honest, which would bias the estimate toward zero. Column 3 instead enters the posterior probability itself, so the coefficient is the excess emissions of an inspection that is fraudulent with certainty. It is 2.32 for NO and 1.06 for HC, so misclassification near the threshold does little to the estimate.

Timing (column 4). A vehicle read soon after its inspection may still be in the condition it was tested in, for instance just after a repair, so early readings could understate how much a repaired vehicle pollutes over the following year. We drop readings taken within 60 days of the inspection. Excess emissions are 2.26 for NO and 1.05 for HC.

Failed vehicles (column 5). Vehicles that failed and have not yet passed are on the road with a known malfunction, and how we treat them could shift the comparison. The main specification gives them their own indicator; column 5 drops them instead. Excess emissions are unchanged (2.28 for NO and 1.04 for HC).

Unregistered driving (column 6). The counterfactuals of Section 7 need the damage done by malfunctioning vehicles that stop inspecting and drive unregistered. These vehicles never pass an inspection, so the main estimate does not measure them. We observe them on the road as the demand model's lapsed pool, malfunctioning vehicles that failed or previously passed through fraud and then let their registration lapse, and compare them with repaired vehicles. They emit 2.72 grams per kilogram of fuel more NO and 2.20 more HC, more than fraudulently passed vehicles do, and the difference from the fraud estimate is statistically significant. Assigning their unregistered driving the damage of a fraudulent pass (Appendix D.4) is therefore conservative.

Columns 7 and 8 are the severity comparison of Appendix D.3.

Concealment over time (Appendix Figure A.7). A limitation of the detection model is that it can detect fraud only if fraud leaves a signal in the indicators. A fraudulent pass that sets off none of them, for example one produced with a simulator that reports the vehicle's own VIN and a profile no other test has used, looks honest to the model. This matters most for the crackdown we study in Section 4. Measured fraud fell sharply after the crackdown, but the crackdown also gave corrupt stations a reason to cover their tracks. If they changed their methods so that fewer fraudulent passes set off the indicators, the model would count those passes as honest, and part of the measured decline would reflect better concealment rather than less fraud. Fortunately, the roadside remote-sensing data allow a direct test. If stations got better at hiding fraud, a growing share of the passes we classify as honest would in fact be fraudulent, and those vehicles would still pollute. Vehicles with honest passes would then grow dirtier on the road relative to those whose passes we classify as fraudulent, and the excess emissions of a fraudulent pass would shrink over time. We find no evidence of this. Measured against honest passes that followed a failed test, excess NO emissions range from 1.84 to 2.61 grams per kilogram of fuel across years; measured against honest passes on the first test, they range from 1.78 to 2.85. Neither gap declines after the crackdown, even though the fraud market changed substantially over the period.

Appendix E. Station Licensing and Turnover

Section 4.2 shows that most of the fraud sold after the crackdown came from new fraud stations, inspection licenses that had not sold fraud before. In this appendix, we investigate whether stations operating under new licenses had links to incumbent fraud stations that were exiting the market. The answer matters for enforcement: closing a license removes a fraudulent business from the market only if the people, the location and the equipment behind it do not reappear under another license. We first describe how a station obtains a license, then trace the links between new and closing fraud stations.

E.1 How an Inspection Station Gets a License

To inspect vehicles, a station needs a certificate of appointment from DPS, which we call a license. According to DPS, the owner applies online and pays a nonrefundable fee of $100, and DPS checks the background of each owner: some criminal convictions disqualify an applicant for up to ten years, or permanently if the offense was committed as an inspector or station owner. Once DPS approves the application on this initial review, it tells the applicant to install the testing equipment. In emissions counties, including those in DFW and Houston, that means buying an approved emissions analyzer and signing a data contract with the state's inspection-data contractor. DPS then inspects the premises and the equipment before issuing the license, which must be renewed every two years.[49]

[49] Texas Department of Public Safety, *Vehicle Inspection Operations and Training Manual*, Chapter 1 (current edition, accessed October 2026); Texas Department of Public Safety, "Inspection Station Qualifications."

Table D.1: Excess Emissions: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. NO (g per kg of fuel)								
Measured minus comparison group	2.28 (0.03)	2.37 (0.03)	2.32 (0.03)	2.26 (0.04)	2.28 (0.03)	2.72 (0.06)	2.05 (0.09)	1.74 (0.09)
B. HC (g per kg of fuel)								
Measured minus comparison group	1.04 (0.03)	1.03 (0.02)	1.06 (0.03)	1.05 (0.03)	1.04 (0.03)	2.20 (0.10)	0.93 (0.08)	0.77 (0.07)
Measured group	Fraud passes	Fraud passes	Fraud posterior	Fraud passes	Fraud passes	Lapsed pool	Fraud passes	Fraud passes
Comparison group	Repaired	Honest first-time	Repaired	Repaired	Repaired	Repaired	Repaired	Repaired
Readings within 60 days dropped				✓				
Failed-not-yet-passed vehicles dropped					✓			
Sample	All	All	All	All	All	All	Failed a genuine test	Failed a genuine test
Failed-test diagnostic fixed effects								✓
Driving conditions; make, model-year and date fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Readings	2,199,150	2,199,150	2,199,150	1,802,652	2,190,592	2,230,544	87,136	86,815

Note: Each column reports the excess roadside emissions of the measured group over the comparison group, from a regression of the same form as Table 2; column 1 is that table's main estimate. Column 3 replaces the fraud indicator with the posterior probability of fraud. Column 6 measures unregistered vehicles in the demand model's lapsed pool. Columns 7 and 8 keep vehicles that failed a genuine test and compare their later fraudulent and repaired passes; the ratio of column 8 to column 7 (0.85 NO, 0.82 HC) scales the damages. Standard errors, clustered by vehicle and by day, are in parentheses.

DPS does not publish application timelines, but our data give a sense of them. Once DPS issues a license, the station usually starts inspecting within days: the median gap is 6 days, and 29 days at the 90th percentile. Obtaining a new license takes longer, as we show in Section E.2.

A license cannot be transferred. It covers the people named on it and the location it lists, and a change of owner or location ends the right to inspect until DPS licenses the new holder.[50] A new license at an existing station is therefore often the ordinary record of a sale: the owner sells the business, and the buyer must apply for a license of their own. Honest stations change licenses this way, and nothing in the rules prevents a buyer from continuing a station's practices. For example, an owner may sell a popular shop that performs many honest inspections, or a new owner may take over a business that earns most of its revenue from fraudulent inspections.

[50] Texas Transportation Code §548.403; 37 Texas Administrative Code §23.1(j).

Apart from ordinary changes in ownership, a DPS sanction can force a station to exit or to change hands. When DPS suspends or revokes a license, its rules bar the owner from taking any part in a station that takes over the site.[51] A sanctioned business that reopens there must therefore carry someone else's name, whatever the owner's actual role. Whether the person on the new license actually runs the station is hard to verify. A 2023 NBC DFW investigation reported that the person listed as the owner of one inspection station said he had never owned it and that his name had been used without his knowledge (Friedman and Parks, 2023). Our data cannot settle this either. Neither the DPS station list nor the inspection records name the people who own or run a station; owner and inspector names appear only in DPS's records of disciplinary actions, so they cover sanctioned stations alone.

E.2 Links Between New and Closing Fraud Stations

In this subsection, we show that many stations that entered the fraud market during the crackdown were linked to fraud stations that had just closed, either on their own or because DPS shut them down. Because we cannot follow owners, we establish these links through two things we observe for every station and that a business can carry to a new license: its location and its emissions analyzers. Every inspection record includes the serial number of the analyzer that performed it, so we can see when an analyzer used at one station later appears at another. We say that a new fraud station, as defined in Section 4.2, took over a closing fraud station if, within 6 months of that station's last pass, it began selling fraud at scale at the same site or with one of its analyzers.[52]

Figure E.1 shows that, of the new fraud stations that appeared between August 2022 and July 2023, 66% took over a closing fraud station, up from 24% of the stations that began selling fraud at scale in 2018–2021. Almost half of these takeovers shared both the site and an analyzer with the closing station (46%), about two in five shared only an analyzer (41%), and about one in eight shared only the site (13%).

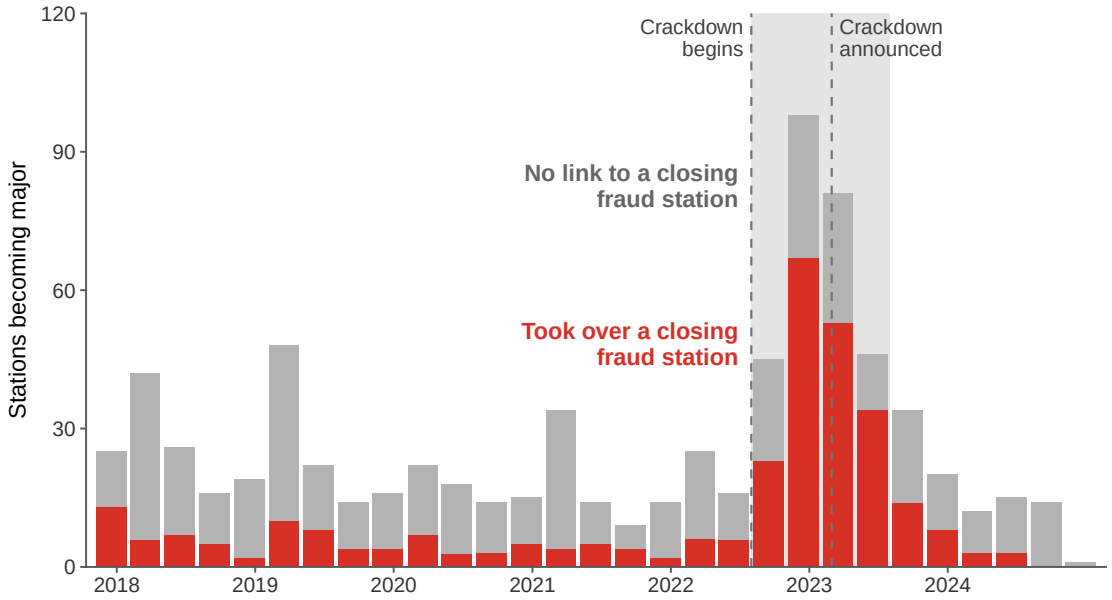
While the number of links is striking, it does not show how many of these businesses were truly connected. Links also arise by chance: any station that closes is likely to sell or release its storefront and its testing equipment. Suppose that DPS shuts down a station and a new owner with no ties to the old business buys the property and its inspection equipment. Our measure would count this as a link, even though the new and old owners have no relation whatsoever. We therefore ask how often a new fraud station would be linked to a closing fraud station by chance. We calculate that random chance alone would link about 47% of new fraud stations to a recently closed fraud station in the crackdown year.[53] Therefore,

[51] *Vehicle Inspection Operations and Training Manual*, Chapter 1, §01.50.

[52] We treat two stations as being at the same site if any of three conditions holds: they share a street address and ZIP code, after standardizing the address and ignoring suite or unit numbers; they share a Google Maps place, which catches the same storefront recorded under different spellings of its address; or they lie within 30 meters of each other, using Google Maps coordinates.

[53] We treat each closing station that a new fraud station is linked to as a random draw from the stations that closed in its metropolitan area around the time it began selling fraud at scale. A station linked to one closing station is linked to a fraud station by chance with probability equal to the fraud share of those closing stations; a station linked to several has a higher chance, because any of its links could land on a fraud station. 55% of the new fraud stations with any link were linked to two or more closing stations. Averaging over all new fraud stations, and counting those with no link as zero, gives 47%. The benchmark is, if anything,

Figure E.1: New Major Fraud Stations and Takeovers of Closing Fraud Stations



Note: Stations in Dallas-Fort Worth and Houston (chains excluded), by three-month period (August–October, November–January, etc.) of their first month as a major fraud station, with more than 100 fraudulent passes. Bars from August 2022 on are new fraud stations. Red: took over a closing fraud station (same site, defined in the text, or one of its analyzers), with that first month within 6 months of its last pass. Shading: the crackdown year (August 2022–July 2023). Dashed lines: crackdown begins (August 2022), announced (March 2023). The last bar covers November–December 2024 only.

the fact that we can link 66% of the new stations to closing fraud stations suggests that a substantial share of the stations that entered during the crackdown were connected to former fraud businesses: run by the same people, or taken over by arrangement with them, rather than bought by an unrelated owner. However, we cannot say with confidence which ones were, or how.[54]

Many of these takeovers appear to have been planned. Among takeovers by new fraud stations that opened under a new license, the license was issued before the old station stopped testing in 31% of cases. One reason may be that the old license holder expected DPS to sanction it and moved the business to a new license, perhaps with a different name on the paperwork, before the sanction arrived. By contrast, when the new license never sold fraud at scale, it was rarely issued before the old station closed (9% of cases).

In our model, we formulate sellers as playing a static game each year (Section 5), in part because we cannot confidently link who operates each license over time or how the operators

generous to chance, because several links to fraud stations are themselves a sign of a network. Before the crackdown, when few closing stations were fraud stations, chance would give 6%, against the observed 24%.

[54]The link weakened after the first year: new fraud stations that appeared after July 2023 took over a closing fraud station 29% of the time, partly because few fraud stations were still closing by then. From the side of the closing stations, 38% of the fraud stations that closed in the crackdown year were followed by a new fraud station at their site or with their equipment, against 9% of other closing stations. Widening the window to 12 months moves no share in Figure E.1 by more than 8.1 percentage points, and adding links through shared business names, phone numbers or people named in DPS disciplinary records moves none by more than 4.8. The links are stronger in DFW (81% in the crackdown year) than in Houston (44%).

are related to one another. Nonetheless, whether the same people moved to a new license or a new operator took over a site set up for fraud (Section E.1), the lesson for enforcement is similar: closing a license often left the location or the equipment that continued to sell fraudulent inspections. Importantly, the model does capture the incentive for further entry when licenses are removed. Every licensed station in a metropolitan area is a potential seller, so when DPS shuts some down, the stations that remain face less competition, and some that did not sell fraud before find it worthwhile to start.

Appendix F. Repair Costs

This appendix describes the predicted repair bill $\hat{r}$ that the demand model uses as the price of repair. Appendix F.1 defines it, Appendix F.2 describes its distribution, Appendix F.3 checks its level against outside benchmarks, and Appendix F.4 gives the evidence for applying its distribution to drivers who never take an honest test.

F.1 The Predicted Repair Cost

Measuring and predicting repair costs. We do not directly observe each driver's repair cost. Inspectors can record repair costs at retest, but reporting is effectively voluntary and the large majority of entries are recorded as zero; Table F.2 summarizes the tests that do carry a positive recorded cost. More importantly, we observe OBD diagnostic data for *all* vehicles that fail their initial test—including those whose drivers subsequently obtain a fraudulent pass or let the registration lapse.

We therefore estimate a model of log repair costs on diagnostic variables, using the cycles with a positive recorded cost, and predict a cost for every cycle that begins with a failure.[55] For the regression we collapse multiple tests within the same inspection cycle into a single observation and restrict to cycles ending in repair, since only those cycles can have an incurred cost. The specification is

$$\begin{aligned}\log(c_i) = \beta_0 + \beta_1 \cdot \mathrm{MIL}_i + \beta_2 \cdot \mathrm{MIL_Run}_i + \sum_{k=1}^{6} \beta_{3k} \cdot \mathbb{1}\{\mathrm{DTC}_i = k\} \\ + \beta_4 \cdot \mathrm{GasCap}_i + \sum_j \beta_{5j} \cdot \mathrm{Monitor}_{ij} + \sum_a \gamma_a \cdot \mathbb{1}\{\mathrm{Age}_i = a\} + \gamma_{\mathrm{odo}} \cdot \mathrm{HiOdo}_i \\ + \delta_{m(i)t(i)} + \delta_{\mathrm{make}} + \delta_{\mathrm{station}} + \varepsilon_i, \quad (27)\end{aligned}$$

where MIL_i is the check engine light indicator, $\mathrm{MIL_Run}_i$ the inspector's recorded MIL-on-while-running check, DTC_i the count of stored diagnostic trouble codes (entered as indicators, top-coded at six or more), GasCap_i the gas cap test result, and $\mathrm{Monitor}_{ij}$ the set of OBD readiness monitors that were not ready at the failed test.[56] The vehicle level block is

[55] Predicted repair costs are defined only for vehicles with an observed initial failure. Some drivers buy a fraudulent pass without first failing an honest test. Section 6.1 describes how the model prices repair for these drivers (Assumption 4), and Appendix F.4 gives the evidence for that assumption.

[56] A/C refrigerant and PCV readiness monitors are excluded: A/C does not count toward the Texas I/M not-ready threshold, and PCV readiness data were not reported in our data.

entered in bins: $\mathbb{1}\{\text{Age}_i = a\}$ are indicators for the vehicle-age terciles (2–15, 16–19, and 20–24 years), and HiOdo_i indicates an odometer reading above the age-conditional median (a missing or out-of-range reading is imputed to the fit-sample median log odometer before binning). The remaining terms are fixed effects: $\delta_{m(i)t(i)}$ for metro × year, where metro is Dallas–Fort Worth, Houston, or the remaining program counties and the early sensing-era years 2016–18 are pooled into a single year level, and δ_{make} and δ_{station} for vehicle make and inspection station, both absorbed in estimation. Standard errors are clustered by station. The *continuous* odometer reading is excluded in favor of the binned term: entered continuously its coefficient is positive and significant, consistent with higher-mileage vehicles facing more expensive repairs, but it leaves the diagnostic coefficients and the fit essentially unchanged once the make fixed effects and the age bins are in, and self-reported odometer readings carry measurement error that would attenuate the estimate.

The estimated coefficients confirm that diagnostic severity predicts repair cost: an illuminated check engine light and a higher count of stored trouble codes both raise the predicted cost substantially, with a monotone gradient in the trouble-code count; a gas cap failure — the cheapest repair in the menu — lowers it. Among the readiness monitors, catalyst, evaporative system, oxygen-sensor-heater, and secondary-air failures raise predicted cost, while the rarely triggered monitors are imprecisely estimated. Because $\hat{\boldsymbol{\beta}}$ is estimated conditional on the fixed effects, predicted repair cost varies only with the diagnostic profile—determined by the vehicle's mechanical condition rather than by the driver's compliance decision. We convert the index to dollars using $\hat{r}_i = \bar{c} \cdot \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}}) / \mathbb{E}[\exp(\mathbf{x}'\hat{\boldsymbol{\beta}})]$, with the renormalization anchor $\bar{c} = \$266$ — the mean self-reported positive repair cost among the cycles the regression is estimated on — in 2024 dollars.[57]

Two features of this construction should be read literally. First, there is one predicted-cost object, and it is the diagnostic one: the index $\mathbf{x}_i'\hat{\boldsymbol{\beta}}$ is evaluated with the diagnostic terms only: the vehicle-age and odometer bins and the metro × year effects are set to zero and the absorbed make and station fixed effects are omitted. So $\hat{r}_i$ is not a vehicle-specific expected bill but the cost a typical car would face for *this* diagnostic problem, in a reference metro-year, at a typical station, with the level pinned to $\bar{c}$. Age and odometer enter the demand utilities of Section 5 directly, so including them in $\hat{r}_i$ as well would double-count the same variation; the omission is a deliberate alignment choice, not an approximation. Second, the scale of $\hat{r}_i$ is absorbed by the price coefficient α_i in estimation: rescaling $\bar{c}$ leaves the fit and the choice probabilities unchanged and only changes the dollar denomination in which the price side and the welfare figures are reported.

F.2 Descriptive Evidence

Table F.1 reports the distribution of the predicted repair cost by vehicle age. The median predicted cost is $248, and the median rises with vehicle age. Table F.2 summarizes the tests that carry a positive recorded cost.

[57] Repair costs are deflated to 2024 dollars using the BLS motor vehicle maintenance and repair price index (FRED series CUUR0000SETD).

Table F.1: Predicted Repair Costs by Vehicle Age

Age bin	N (assess)	p10	Median	Mean	p90
0–4	519,701	159	222	229	270
5–9	1,107,653	201	226	248	391
10–14	1,171,991	201	252	271	432
15–19	843,990	201	254	289	451
20–30	277,265	201	257	306	501

Note: Predicted repair cost distribution (dollars, 2024 real) among inspection cycles that begin with a failed honest inspection, by vehicle-age bin at the time of the failure. N (assess) is the count of these cycles in the bin.

Table F.2: Repair Cost Summary Statistics

	Mean	Std. Dev.
Total tests	101,764,403	
Tests with repair reported	670,357	
Tests with positive repair cost	250,721	
Repair cost (2024 $)	$268	$1,468
Repair Type (share)		
Fuel system	29.9%	
Ignition/electrical	12.2%	
Emissions system	25.5%	
Engine mechanical	2.8%	
Misc.	29.6%	

Note: Sample includes all emissions tests (2016–2024). A repair is identified when the repair group code (REP_GRP) is 1–5. Repair cost statistics are computed among tests with positive reported repair costs, expressed in 2024 dollars using the BLS motor vehicle maintenance and repair price index (CUUR0000SETD). Repair type shares are computed among tests with positive repair costs.

Fraud rises with predicted repair cost, and rises from a much larger base than the off-road channels. Appendix Figure A.12 plots the three compliance outcomes—repair, a fraudulent pass, and the scrap/idle/unregistered composite—against the diagnostic-only predicted repair cost, over 3,920,600 drivers of malfunctioning vehicles who failed an honest inspection. Each series is a weighted smooth (binomial GAM) fit across the 1st to 99th percentile of predicted cost; the faint points are equal-frequency bin means, shown as a visual reference for the data behind the fits rather than as reported quantities. Read at the two ends of the fitted range, repair falls from 92.5% of drivers to 42.0%. Both non-repair channels rise: the fraud share rises from 6.3% to 35.5% (a factor of 5.7, an elasticity of 1.32 with respect to predicted cost), and the scrap/idle/unregistered composite from 1.3% to 22.5% (a factor of 16.7). Fraud is thus the dominant exit at every point of the cost distribution, but its dominance is widest where bills are small: fraud accounts for 82.3% of all non-repair at the low end of the fitted range and 61.2% at the high end. Cheap failures are resolved either by repairing or by buying a pass, and it is only as the repair bill grows that taking the vehicle off the road becomes a materially used alternative — so the fraud

channel rises steeply in absolute terms while giving back ground in relative terms. This is a substitution pattern, not an orthogonality result, and we do not read it as evidence that the diagnostic profile is unrelated to fraud access. What the reduced form establishes is the first-order fact the structural model must reproduce — that the fraud margin is steeply cost-sensitive — and the model in Section 5 is what disciplines the split between fraud and the off-road alternatives, rather than an assumption imposed on it here.

F.3 External Validation

External validation of the cost scalar. Because repair cost records are self-reported and voluntary, we cross-check the renormalization anchor $\bar{c} = \$266$ — itself the mean of those self-reports over the estimation sample — against two independent benchmarks. The anchor is a level normalization, not a sample average of the fitted prior: the implied mean of the repair-cost prior over the estimation sample is \$300, 1.13× the anchor, the gap reflecting the lognormal variance term $\exp(\sigma_r^2/2)$ and sample composition.

The first is RepairPal, an independent online repair cost estimator. We scraped 114 quotes covering 7 common emissions-related repair types (oxygen sensor, catalytic converter, EGR valve, evaporative canister, MAF sensor, spark plugs, and PCV valve) across 20 make–model-year combinations. Individual repair types span a wide range, from \$290 for spark plugs to \$2,347 for a catalytic converter. RepairPal quotes sit above our scalar, for two likely reasons: the repair type codes in our data map imprecisely onto specific RepairPal repairs, and RepairPal covers only professionally diagnosed jobs at retail labor rates, while the repairs recorded in our data may include lower-cost jobs.

The second benchmark is a survey of the same Texas inspection stations conducted by Eastern Research Group (2024b). Station owners reported average repair costs following a failed emissions inspection of \$293–\$379 across metros (nominal 2023 dollars), with medians of \$150–\$250, in the same range as our scalar in 2024 dollars.

Taken together, these benchmarks put $\bar{c}$ in a defensible range for the typical Texas emissions repair. We note that the demand elasticity estimated in Section 6.1 is invariant to the scalar: because the price coefficient scales inversely with $\bar{c}$, any rescaling leaves relative rankings and elasticities unchanged. Counterfactual results that depend on dollar-valued cost levels—welfare calculations in particular—do depend on the scalar being correctly calibrated, and should be read with the uncertainty in $\bar{c}$ in mind.

F.4 Repair Costs of Vehicles That Are Never Tested

Non-testers have no diagnostics, and therefore no predicted repair cost $\hat{r}$. They matter for demand: 76% of fraud buyers never take an honest test. For these drivers, the model integrates the repair bill over $G(r \mid x)$, the distribution of the predicted repair cost among tested vehicles with the same covariates x:

$$\log r \mid x \;\sim\; N\big(\mu_r(x),\, \sigma_r(x)^2\big).$$

Here $\mu_r(x)$ is the fitted value of a regression of $\log \hat{r}$ among tested vehicles on vehicle age (terciles), odometer bin, and metro and year, and $\sigma_r(x)$ is the residual spread within groups

defined by age, odometer bin, and metro. We compute the integral by Gauss–Hermite quadrature (Appendix H.2).

Using $G(r \mid x)$ for non-testers rests on the first part of Assumption 4 in Section 6.1. We cannot test that assumption directly, because we never observe a non-tester's diagnostics. The rest of this section gives three pieces of indirect evidence from fraud buyers. First, repair cost rises with on-road emissions, so on-road emissions are informative about repair costs we do not observe (Appendix F.4.1). Second, untested and tested fraud buyers have nearly identical on-road emissions, so we find no evidence that untested fraud buyers have hidden, more expensive defects (Appendix F.4.2). Third, the two groups drive similar makes, but untested fraud buyers drive younger, lower-mileage vehicles; this difference, if anything, lowers their repair costs, and x already includes age and mileage (Appendix F.4.3).

F.4.1 Repair Cost Rises with On-Road Emissions, as a Hinge

Among fraud buyers with an observed predicted repair cost (matched by VIN to their most recent failed inspection), mean predicted repair cost is flat below on-road NO ≈ 1.5 g/kg-fuel and rises monotonically thereafter, reaching 15.1% above the flat-region mean in the top emissions bin. Figure F.1, Panel A, plots this relationship in 20 equal-count bins of on-road NO.

F.4.2 Untested and Tested Fraud Buyers Have Near-Identical Emissions

Figure F.1, Panel B, overlays the on-road NO distribution (asinh scale) for three groups: untested fraud buyers (fraud buyers with no prior failed inspection), tested fraud buyers (fraud buyers matched to a prior failed inspection), and a placebo sample of compliant vehicles. The Kolmogorov–Smirnov distance between untested and tested fraud buyers is $D = 0.035$, versus $D = 0.281$ against the compliant placebo—an order of magnitude closer. The same construction cleanly separates compliant vehicles from both groups of fraud buyers, confirming that the overlap between untested and tested fraud buyers reflects the underlying data rather than an artifact of the comparison.

A comparison that adjusts for make, model-year, and capture-date fixed effects and for driving conditions finds that untested fraud buyers are in fact slightly *dirtier* than tested fraud buyers. We estimate it on all remote sensor readings of noncompliant vehicles, fraud-flagged or unregistered, with a control for fraud status. For NO, the premium of untested over tested fraud buyers is 0.162 g/kg (SE 0.042). The raw near-equality therefore masks a small adjusted gap, and the gap goes the "wrong" way for the concern that untested fraud buyers hide more expensive defects. Translating this gap into repair costs through the relationship between diagnostics and emissions implies a repair-cost gap for untested fraud buyers of 2.2% (95% CI [1.1%, 3.3%]). This gap is small enough that treating their repair-cost distribution as equal to that of tested fraud buyers is a reasonable approximation for the demand model. The emissions premium of tested fraud buyers itself reflects repair rather than information: compliant vehicles that once failed get repaired and keep emissions low, while fraud vehicles that once failed do not, implying a repair effect of 2.2 g/kg NO per averted instance of noncompliance.

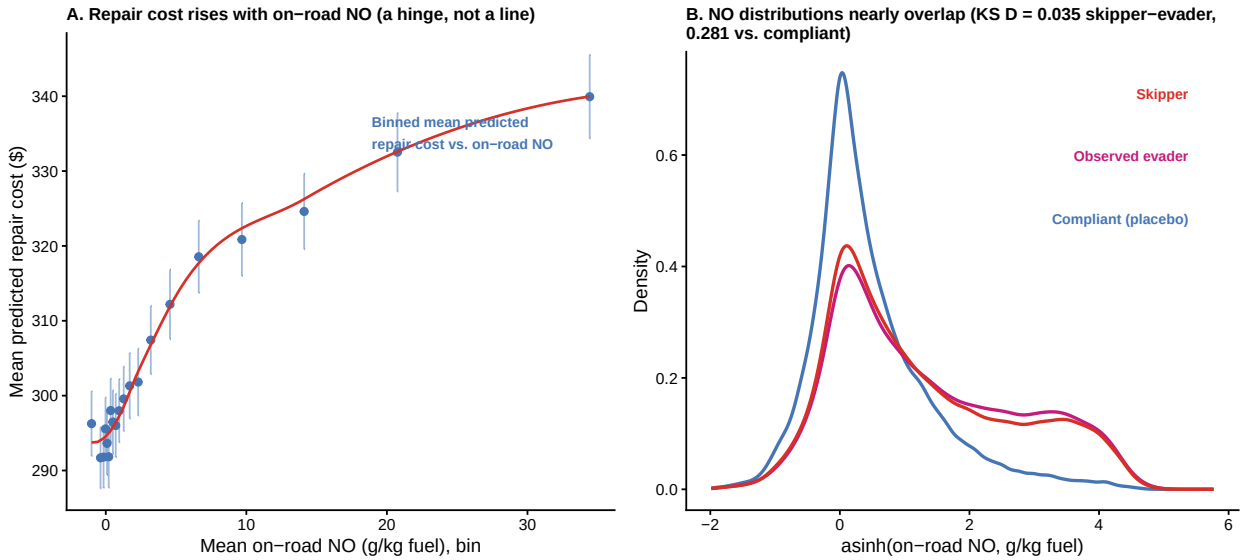


Figure F.1: Repair Cost, On-Road Emissions, and the Selection-Null Result

Note: Panel A: 20 equal-count bins of on-road NO (g/kg-fuel) among fraud buyers with an observed predicted repair cost; points are bin means with 95% CI, red line is a loess smooth. Panel B: kernel density of asinh(on-road NO) for untested fraud buyers, tested fraud buyers, and a placebo sample of compliant vehicles; D statistics are Kolmogorov–Smirnov distances. Both panels use remote sensor readings from 2017 onward matched to inspections flagged as fraudulent; predicted repair cost is the diagnostic-only cost of Appendix F.1.

F.4.3 Make and Model-Year Composition: Untested vs. Tested Fraud Buyers

The emissions-based selection test above has power against selection on emission-*visible* severity but is blind to emission-orthogonal cost components such as parts prices or labor rates, which are also absent from $\hat{r}$ by construction. We check this directly by comparing the make and vehicle-age/odometer composition of untested fraud buyers against tested fraud buyers (Table F.3). The two margins give different answers. Make composition is close: per-make standardized mean differences across the top-20 makes are all under 0.1 in absolute value (Table F.3), though with this many vehicles the χ^2 test on the make $\times$ group table still rejects exact equality ($p < 0.001$). Vehicle age and odometer, by contrast, are materially imbalanced: the untested-minus-tested mean age gap is -1.41 years and the odometer gap -12,785 miles, so untested fraud buyers drive younger, lower-mileage vehicles, and the age imbalance is the largest standardized difference anywhere in the comparison ($|\text{SMD}| = 0.292$).

That figure exceeds the conventional 0.10 covariate-balance threshold, so the composition of the two groups differs materially. We therefore do not claim compositional equivalence between the two groups. What survives is narrower than compositional balance: the check was designed to detect selection on emission-orthogonal cost components such as parts prices, and the make distribution—the observable most directly tied to parts prices and labor rates—is balanced, while the imbalance that does appear is on age and mileage, which are correlated with the diagnostic-visible severity margin the emissions-based bound in Appendix F.4.2 already tests. Applying the tested fraud buyers' repair-cost distribution to untested fraud buyers is thus not justified by compositional balance alone; it rests on the emissions-based

bound, with the age and mileage gap a remaining caveat.

The natural next question is which way that imbalance pushes repair costs, and Table F.4 answers it. Among tested fraud buyers, older and higher-mileage vehicles are modestly *more* expensive to repair: mean predicted repair cost rises from $291 in the youngest age tercile to $305 and $324 in the middle and oldest, and from $294 to $309 and $313 across odometer terciles. Conditional on make, the gradient is 0.92% per year of vehicle age and 0.11% per 10,000 miles. Untested fraud buyers drive younger, lower-mileage vehicles, so both margins point the same direction: evaluating the gradient at the observed composition gaps implies a repair-cost difference for untested fraud buyers of -1.44% relative to tested fraud buyers (delta-method SE 0.06 percentage points). The difference is small and points downward, so the age and mileage imbalance does not hide higher repair costs among untested fraud buyers. Two caveats keep this from being a clean-bill-of-health argument. The gradient is estimated on tested fraud buyers and extrapolated to untested ones, and $\hat{r}$ is the diagnostic-only cost object, in which the repair-cost model's own age, odometer, and metro-year level terms are zeroed; the age/mileage gradient measured here therefore operates through the OBD diagnostic terms, so it speaks to the same diagnostic-visible severity margin as the bound in Appendix F.4.2 and not to the emission-orthogonal parts-price channel. What it does establish is that the direction and magnitude of the observed age/mileage imbalance, taken at face value, are too small to matter for the repair-cost distribution we apply to untested fraud buyers.

Table F.3: Make Composition: Skippers vs. Observed Evaders

Make	Observed evader share (%)	Skipper share (%)	SMD
CHEV	23.0	22.7	-0.01
FORD	15.4	17.0	0.04
TOYT	11.1	9.4	-0.06
NISS	8.5	9.1	0.02
DODG	6.9	6.6	-0.01
HOND	6.6	6.2	-0.02
OTHER	6.0	5.6	-0.02
GMC	4.6	4.8	0.01
JEEP	2.0	1.9	-0.01
HYUN	1.9	1.8	-0.01
Mean vehicle age: observed evader = 13.80, skipper = 12.39 (SMD = -0.292)			
Mean odometer: observed evader = 181,632, skipper = 168,847 (SMD = -0.175)			
Chi-square test (make × group), $\chi^2 = 254.6$, df = 20, p ¡ 0.001			

Notes: Top-10 makes by combined fraud-buyer volume (top-20 used for the underlying chi-square test and SMD summary); remaining makes pooled into OTHER. Observed evaders = fraud buyers with a matched fail-cycle predicted repair cost under the Design-A VIN-history merge; skippers = fraud buyers with no such match. SMD = standardized mean difference (skipper share minus evader share, pooled-SD denominator); age/odometer SMDs use pooled-SD of the continuous variables. Sample: 2017+ fraud-flagged TORVET captures, VIN level.

Table F.4: Vehicle Age, Odometer, and Predicted Repair Cost Among Tested Fraud Buyers

	Vehicle age (per year)	Odometer (per 10,000 mi)
Gradient in $\log \hat{r}$ (%)	+0.92	+0.11
	(0.04)	(0.03)
Skipper − evader gap	-1.41 yrs	-12,785 mi
Implied $\hat{r}$ difference for skippers: -1.44% (SE 0.06)		
Mean $\hat{r}$, low tercile	\$291	\$294
Mean $\hat{r}$, middle tercile	\$305	\$309
Mean $\hat{r}$, high tercile	\$324	\$313
Observations (VINs)	35,497	35,497
Make fixed effects	Yes	Yes

Note: Estimated on observed evaders (fraud buyers with a matched fail-cycle predicted repair cost $\hat{r}$), one observation per VIN, 2017+ TORVET captures. The top panel regresses $\log \hat{r}$ on vehicle age and odometer with make fixed effects; heteroskedasticity-robust standard errors in parentheses. The implied difference evaluates that gradient at the observed skipper-minus-evader composition gaps, with a delta-method standard error. Terciles are cut on the observed-evader distribution of each variable. $\hat{r}$ is the severity-only `rhat_diag` projection, in which the repair-cost model's age and odometer level terms are set to zero, so any gradient here operates through the OBD diagnostic terms rather than through age or mileage directly.

Appendix G. Scrap, Idle, or Driving Unregistered: The States of Lapsed Vehicles

G.1 Motivation and Setup

To evaluate the emissions-inspection program comprehensively, we must account not only for the vehicles that pass inspection, honestly or fraudulently, but also for the vehicles that disappear from the inspection record. Consider a vehicle that fails an inspection and never returns with a passing result. It could have been scrapped; it could still be driving in the regulated counties without valid registration; or it could be sitting idle—garaged, parked off the road, or temporarily outside the regulated counties with the option of coming back. The three states have very different implications for pollution and welfare: a vehicle on the road pollutes now; a scrapped vehicle pollutes neither now nor in any future period; an idle vehicle does not pollute now but may return to the road. Quantifying how these vehicles divide among the three states is therefore essential for the policy counterfactuals: enforcement that shuts down fraud suppliers has very different welfare implications if it pushes drivers into unregistered driving than if it steers them toward scrap. It also matters for the level of damages—treating every such vehicle as an unregistered driver would overstate the pollution.

In this section we explain how we identify the state of these vehicles. The population is the set of vehicles that are emissions-noncompliant at their last observed inspection—a failed result or a fraudulent pass—and that have not subsequently obtained a passing inspection, whether honest or fraudulent. A vehicle's lapse begins when its compliance actually ends: at

the failed inspection for a failed result, and at sticker expiry—one year after the test—for a fraudulent pass. The first lapse year is therefore a partial calendar year, and lapse duration k counts calendar years from that onset. We call such an emissions-noncompliant, lapsed vehicle-year simply *lapsed*, and the set of them in a given year the *lapsed pool*, indexed by lapse duration k, the number of years the vehicle has been without a valid sticker. Section 4's count of vehicles driving unregistered measures something different: the vehicles whose registration has been expired for more than a month at a point in time. A lapse in the demand model of Section 6.1 is a choice in an inspection cycle.

Identifying the state of vehicles that have disappeared from the inspection record is inherently difficult, but two features of the data make it possible. First, we observe the universe of Texas emissions inspections over time, so a lapsed vehicle that skips inspection this year may still be seen returning in a future year—and only a *surviving* vehicle, one not yet scrapped and hence either idle or driving unregistered, can return, which is informative about scrappage. Second, auxiliary roadside remote-sensing data can sight lapsed vehicles in traffic, which is informative about which lapsed vehicles are driving on the road without registration. Sightings alone split the lapsed pool into driving and off the road, but cannot separate the off-road vehicles into scrapped and idle—a vehicle that is never sighted could be scrapped or could simply be sitting unused—and the return margin separates the two. Section G.2 estimates the on-road share of the lapsed pool, Section G.3 estimates its survival, and Section G.4 combines the two into the full three-way decomposition.

G.2 Identifying On-Road Share from Roadside Remote Sensing

Texas operates a network of roadside remote-sensing monitors (the TORVET program) that read the exhaust plume and license plate of vehicles passing fixed sites on public roads. The monitors are sited on ordinary arterial and highway segments and record whatever drives past, so a sighting is a draw from the traffic stream rather than from an enforcement action: unlike a traffic stop or an audit, being sighted is not a consequence of being suspected. That gives us the first assumption.

> Assumption G.1 (representative sampling). Conditional on being on the road in a monitored metro-year, a lapsed vehicle is sighted by remote sensors with the same probability as a registered vehicle with the same observed characteristics.

Sightings alone, however, cannot pin the level of the on-road share. The monitors observe a small and time-varying share of traffic—campaigns, a metro-year's deployment of monitors, run for weeks, not years, and site counts change across the sample—so a low sighting rate for lapsed vehicles is jointly consistent with "few of them drive" and "the monitors saw little of anyone." What is needed is a reference group of *registered* vehicles, whose on-road status is known, measured by the same monitors in the same campaign.

> Assumption G.2 (normalization). A vehicle with a valid registration, having passed inspection, is on the road with probability one during its registration term.

Any registered vehicle could serve as this reference group. We use registered vehicles that our detection model classifies as fraudulent passers: they are the natural comparison group for an empirical model of vehicles on the margin of non-compliance.

Neither assumption is innocuous. Unregistered drivers have an incentive to keep off the highways where monitors sit, and if they do, Assumption G.1 undersamples them and we understate unregistered driving. We can provide evidence on the validity of this assumption by comparing relative sighting rates between on- and off-highway monitor sites. Assumption G.2 fails when a fraudulently passed vehicle breaks down partway through its term—it bought a year of driving but delivered less—which depresses the registered group's sighting rate and overstates unregistered driving. We can assess its validity by comparing fraudulent passers' on-road rates with those of observably similar honest passers.

With the reference group in hand the identifying content of the sighting data can be stated in one line. Comparing lapsed vehicles to registered vehicles observed by the same monitors in the same metro-year,

$$\frac{\Pr(\text{sighted} \mid \text{lapsed})}{\Pr(\text{sighted} \mid \text{registered})} = \frac{\Pr(\text{sighted} \mid \text{lapsed})}{\Pr(\text{sighted} \mid \text{lapsed, on road})} = \Pr(\text{on road} \mid \text{lapsed}) \equiv f. \tag{28}$$

The first equality is where the assumptions work: registered vehicles are on the road (Assumption G.2) and, once on the road, are sighted at the same rate as lapsed vehicles with the same characteristics (Assumption G.1), so the observable registered sighting rate stands in for the unobservable sighting rate of on-road lapsed vehicles. The second equality is arithmetic: an off-road vehicle cannot be sighted, so the lapsed sighting rate is the on-road sighting rate times the on-road share. Everything about the sensing network that is common to the two groups—how many sites operated, for how long, how reliably they read a plate—cancels in the ratio. The ratio is the object we want, and it is a relative on-road share: f compares driving by a lapsed vehicle to driving by a vehicle known to be on the road.[58] In reality a vehicle's on-road status and how much it drives once on the road are separate decisions, and sightings cannot tell them apart: a sighting reflects both at once. We therefore read f throughout as a measure of on-road driving weighted by miles driven, relative to the reference group, not a pure share of vehicles.

Following the intuition of equation (28), we specify the following empirical model to estimate f, the miles-weighted on-road share of lapsed vehicles. Two features of the data require it to be richer than the ratio. First, the roadside monitors do not run all year—each metro-year's deployment, which we call a *campaign*, runs for weeks—and some vehicles are lapsed for only part of a year; we correct for this with an offset ω_{it}, the share of the campaign's sensing that overlaps the window in which vehicle i is lapsed in year t—or, for a registered vehicle, its registration term.[59] Second, a campaign offers many detection opportunities, so the outcome we model is being sighted *at least once*. Indexing vehicle-years by (i, t) for

[58] The largest fitted on-road share across all 36,298 vehicle groups is 0.7227 , and 0 groups exceed one. No fitted value is ever censored or clamped.

[59] We write the sensing overlap as ω_{it} to avoid collision with $\hat{\tau}_i$, the fraud posterior of Section 3.1. ω_{it} is a property of the sensor calendar, not of the vehicle: a driver whose lapse begins in December is not a driver who drives one twelfth as much, but one who is watched for one twelfth of the campaign. It is therefore an offset with a coefficient fixed at one, never a covariate, which is what makes f a full-year on-road share comparable across vehicles whose lapses begin in different months.

vehicle i in calendar year t, let $d(x_{it})$ denote the probability that an on-road vehicle with characteristics x_{it} is sighted at least once over the campaign, where x_{it} collects vehicle age bin, odometer bin, and metro-year; detection depends on x_{it} because a vehicle's characteristics affect how easily a monitor captures and reads it, and because the density of sensing varies across markets and years. The on-road share $f(x_{it}, k_{it})$ depends additionally on the lapse duration k_{it}, the number of years since the vehicle lapsed. A vehicle on the road with probability $f(x_{it}, k_{it})$, watched for a share ω_{it} of the campaign, faces the fraction $f(x_{it}, k_{it})\,\omega_{it}$ of the campaign's sensing and goes unsighted with probability $(1 - d(x_{it}))^{f(x_{it},k_{it})\,\omega_{it}}$, so

$$\Pr(\text{sighted}_{it}) \;=\; 1 - \Big(1 - d(x_{it})\Big)^{\omega_{it}\,(f(x_{it},k_{it}))^{\mathbf{1}[\text{lapsed}_{it}]}}, \tag{29}$$

where the indicator exponent encodes Assumption G.2: a registered vehicle is on the road with probability one, so it faces every opportunity its sensing window offers and its sighting rate reveals $d(x_{it})$ directly.[60] Detection is rare in our setting, and when $d(x_{it})$ is small (29) collapses to a product of three intuitive terms:

$$\Pr(\text{sighted}_{it}) \;\approx\; \underbrace{d(x_{it})}_{\substack{\text{detection rate}\\ \text{given on the road}}} \times \underbrace{f(x_{it},k_{it})^{\mathbf{1}[\text{lapsed}_{it}]}}_{\substack{\text{probability the}\\ \text{vehicle is on the road}}} \times \underbrace{\omega_{it}}_{\substack{\text{share of the campaign}\\ \text{it overlaps}}}.$$

A vehicle is sighted if the monitors pick it up, given that it is on the road, during the portion of the campaign it is around for. This three-way separation is only approximate as a product, but it holds *exactly* after a complementary log-log transform, which turns the product into a sum and gives the linear estimating equation:

$$\log\!\big(-\log\big(1-\Pr(\text{sighted}_{it})\big)\big) \;=\; \underbrace{\mathbf{1}[\text{lapsed}_{it}]\,\log f(x_{it},k_{it})}_{\text{lapsed on-road share}} + \underbrace{\log\!\big(-\log(1-d(x_{it}))\big)}_{\text{on-road detection rate}} + \underbrace{\log\omega_{it}}_{\substack{\text{known}\\ \text{offset}}}. \tag{30}$$

Equation (30) makes the identification argument explicit. The reference group pins $d(x_{it})$, the detection rate, since its on-road share term is zero. $f(x_{it}, k_{it})$ is then identified from the relative sighting rate of lapsed vehicles, after adjusting for how long the vehicle has been lapsed and how its lapse period overlapped that year's sensing campaign. We estimate both in a single complementary log-log regression on the stacked registered and lapsed vehicle-years, with ω_{it} entering as a known offset.

Both components of equation (30) are linear indices on the complementary log-log scale:

$$\log f(x_{it},k_{it}) \;=\; a_1\,k_{it} + a_2\,k_{it}^2 + \psi(x_{it}), \qquad \log\!\big(-\log(1-d(x_{it}))\big) \;=\; \zeta(x_{it}), \tag{31}$$

[60] Equation (29) evaluates the escape probability at the vehicle's expected share of the campaign's sensing, $f(x_{it}, k_{it})\,\omega_{it}$. Treating on-road status as a vehicle-level Bernoulli draw instead would give $f(x_{it}, k_{it})\big[1 - (1 - d(x_{it}))^{\omega_{it}}\big]$. The two differ only at second order in $d(x_{it})$: they coincide as $d(x_{it}) \to 0$ and diverge materially only when detection over a campaign is close to certain. At the detection rates we estimate they are numerically indistinguishable, and we use (29) because its complementary log-log transform is exactly linear.

where $\psi(x_{it})$ and $\zeta(x_{it})$ are each a linear index in age bin, odometer bin, metro, and metro-by-year terms, estimated separately for the two components.[61][62] Because age and odometer bins appear in both indices, we are comparing lapsed vehicles' sighting rates against registered vehicles with similar characteristics; the metro and metro-by-year terms absorb the level of driving and of sensing in each market and year.

Table G.1 reports the estimates. The on-road share falls steeply with lapse duration ($\hat{a}_1 = -0.8602$, se 0.0132 ; $\hat{a}_2 = 0.00445$, se 0.00329): a vehicle in its first lapse year drives a substantial fraction of what a registered vehicle drives, and one lapsed five years or more barely appears on the road at all. Older and higher-mileage lapsed vehicles drive less. Averaged over the pool, the on-road share rises across the sample period, from 0.0723 in 2016 to 0.1267 in 2024, a log change of 0.5604 or 75.1 %.

G.3 Separating Scrap from Idle: The Return-Survival Model

The on-road share answers how much the lapsed pool drives, but not how much of the pool still exists: a pool in which few vehicles survive and a pool in which the survivors rarely drive generate identical sighting data. We use data on lapsed vehicles' likelihood of returning to the inspection record after a lapsed-registration spell to separate vehicles that are scrapped versus idle.

Let Scrap_{it} equal one if vehicle i has been scrapped by the end of year t and zero otherwise; scrap is irreversible, so once it is one it stays one and a scrapped vehicle ceases to exist in every future period. Survival is its complement, and it is what the return margin speaks to: returning to inspection is something only a surviving vehicle can do. A vehicle that is idle in the sense above—garaged, parked off the road, or temporarily outside the regulated counties but able to come back—remains capable of returning, and the data show that many such vehicles do return, sometimes after several years. A scrapped vehicle never returns. Writing $\text{Return}_{i,t+1}$ for the event that a lapsed vehicle is next seen at any inspection within a fixed look-ahead window, that logic is an identity:

$$\underbrace{\Pr\big(\text{Return}_{i,t+1} \mid \text{lapsed}_{it}\big)}_{\substack{\text{observed in the} \\ \text{inspection record}}} = \underbrace{\Pr\big(\text{Scrap}_{it} = 0 \mid \text{lapsed}_{it}\big)}_{\text{survival}} \times \underbrace{\Pr\big(\text{Return}_{i,t+1} \mid \text{lapsed}_{it},\, \text{Scrap}_{it} = 0\big)}_{\substack{\text{return rate of survivors:} \\ \text{unobserved}}}.$$

[61] The year factors in the on-road share, detection, and return-survival indices pool 2016 through 2018 into a single coefficient per metro. In those years the network consisted of a small number of sites operating for a few weeks at a time, and the resulting year effects are not separately identified: their confidence intervals are several times wider than those of the later years and overlap the entire post-2018 range. Pooling is a statement about what the sensing era can distinguish, not a claim that behavior was identical across those years; the same pooled factor is used in all three indices so that the ratios formed from them compare like with like.

[62] Lapse duration is capped at $k^* = 6$: rows with $k_{it} > k^*$ are evaluated at k^*, so the index is held flat beyond k^* years of lapse, where too few lapsed vehicles are sighted to measure anything; the flat region is discussed in Section G.5.

Table G.1: Lapsed Vehicle On-Road Share: Parameter Estimates

	Estimate	Std. error
Panel A: Detection Rate (Registered Vehicles)		
Vehicle		
Age 16–19	-0.2757	(0.0082)
Age 20–24	-0.4724	(0.0123)
Odometer above median	0.2262	(0.0071)
Metro and metro × year FEs	Yes	
Panel B: Relative On-Road Share (Lapsed Vehicles)		
Vehicle		
Age 16–19	-0.1593	(0.0169)
Age 20–24	-0.2127	(0.0232)
Odometer above median	-0.1502	(0.0147)
Lapse duration		
Lapse duration k (linear)	-0.8602	(0.0132)
Lapse duration k^2 (quadratic)	0.0044	(0.0033)
Metro and metro × year FEs	Yes	
Vehicle-years	17,552,506	
SEs clustered by VIN	Yes	

Note: The remote-sensing detection hazard and the on-road share hazard are estimated jointly by complementary log-log (cloglog) regression on 17,552,506 target vehicle-years. Each row reports the coefficient with its standard error in parentheses; **standard errors are clustered by VIN**. Panel A is the detection hazard on the registered reference group – registered vehicles observed passing an inspection fraudulently, which are on the road by construction and so pin the level of detection. Panel B is the on-road share of lapsed vehicles measured *relative* to that registered group: its coefficients are read as $\log f$, shifts in the on-road share, not detection rates. Both panels additionally include a metro level and metro × year fixed effects, entered as a year main effect plus a Houston interaction. Intercepts, metro levels and metro-year effects are estimated but not reported: they are absorbed in the omitted categories and the fixed-effect block and carry no separate interpretation. In both panels the year factor pools 2016–18 into the omitted base cell because of data sparsity in the early, two-site sensing era. Other omitted categories are age bin 2–15, odometer below median, and metro DFW. Lapse duration k enters Panel B as a quadratic.

A vehicle comes back for inspection after lapsing only if it still exists and, existing, chooses to come back, so survival is the ratio of the observed return rate to the survivors' return rate. The identification problem is that the return rate conditional on survival is not directly observable: we need a group of vehicles whose survival is observed and whose return behavior can be used to estimate the population return rate. A lapsed vehicle captured by the remote sensor at time t must not be scrapped at time t, and so we can estimate the return rate amongst this set of surviving vehicles. Our key identifying assumption is as follows:

> Assumption G.3. Among lapsed vehicles of the same duration and characteristics, those observed on the road return to inspection at the same rate as surviving vehicles that were not observed: $\rho(x,k) = \rho_S(x,k)$.

Neither a return rate nor a survivors' return rate is defined for a single vehicle, which

either returns or does not, so we pool lapsed vehicles that share the same characteristics and lapse duration; the pooled objects accordingly carry arguments (x, k) rather than vehicle-year subscripts. Let x collect those characteristics (age bin, odometer bin, metro, and metro-year), the same ones on which the on-road share is estimated, and k the lapse duration. With $R(x,k)$ the share of the pooled vehicles that return within the window, $S(x,k) = \Pr(\text{Scrap}_{it} = 0)$ their survival rate, and $\rho(x,k)$ the window return rate of their survivors, the identity above becomes

$$S(x,k) \;=\; \frac{R(x,k)}{\rho(x,k)}, \tag{32}$$

survival is the observed return rate deflated by the survivors' return rate, and Assumption G.3 lets us measure the denominator by the sighted vehicles' return rate $\rho_S(x,k)$.[63] For example, if 20 percent of all lapsed vehicles return within a year but 40 percent of those seen on the road do, half of the lapsed vehicles survive.

Assumption G.3 is a strong assumption. Being on the road is precisely what makes both a sighting and a prompt re-inspection likely: a vehicle someone is still driving is a vehicle whose owner has a reason to bring it back into compliance. Sighted survivors are therefore relatively early returners. If this is true, then $\rho_S(x,k) \geq \rho(x,k)$, and since $S(x,k) = R(x,k)/\rho(x,k)$, this would imply estimated survival $\hat{S}(x,k) = R(x,k)/\rho_S(x,k)$ is a lower bound on true survival. Because the unregistered-driving share is increasing in survival—more surviving vehicles means more vehicles that could be on the road—the estimates below understate unregistered driving and understate the emissions damages attributed to it. Therefore, we interpret our estimates of unregistered on-road driving and associated damages as conservative.

Call each pool of vehicles sharing (x,k) a vehicle group. The group-level ratios of equation (32) are noisy survival measurements: their denominator rests on sighted lapsed vehicles, a population three orders of magnitude smaller than the pool itself, so groups carry very unequal information and the longest durations carry almost none. Rather than report the raw ratios, we fit a parametric surface through them, taking the same functional form as the on-road share and detection indices of equation (31)—a linear index on the complementary log-log scale:

$$\log\!\left(-\log\frac{R(x,k)}{\rho_S(x,k)}\right) \;=\; b_1\,k + b_2\,k^2 \;+\; \eta(x) \;+\; \varepsilon(x,k), \tag{33}$$

where the left-hand side is measured group by group, the right-hand side is a linear index in the same quadratic duration basis and the same age, odometer, metro and metro-year terms with $\eta(x)$ collecting the covariate part, and $\varepsilon(x,k)$ is sampling error.[64] The parameters minimize the weighted distance between the two sides with weights $1/\text{se}^2$, the inverse of each ratio's sampling variance, so groups with more observations get more weight. This is the standard discrete-time proportional-hazards model: exponentiating the index twice gives the implied survival function $S(x,k) = \exp\!\big(-\Lambda_0(k)\,e^{\eta(x)}\big)$ with cumulative baseline hazard

[63] Returns are scored off the full inspection record, independently of outcome, so a lapsed vehicle that comes back and passes cleanly is a return—the commonest way a repaired vehicle re-enters the record. The identical event definition is used in both arms of the ratio.

[64] As in the on-road share model, duration is capped at k^*: rows with $k > k^*$ are evaluated at k^*, so the index is held flat beyond k^* years of lapse, where too few lapsed vehicles are sighted to measure anything.

$\Lambda_0(k) = \exp\big(b_1 k + b_2 k^2\big)$, so the covariates rescale the entire hazard profile proportionally rather than shifting it by a constant.

The look-ahead window must be the same for every cohort. We score R and ρ_S over a window of $T = 1$ year for every calendar cohort—the only horizon observable for all of them—and survival is defined and reported on that same one-year scale. One exception to the uniform window is mechanical: the first lapse year runs from lapse onset to December 31, so $S(x, 0)$ is a pooled survival probability over partial first-year windows of heterogeneous length (mean 0.53 years). We do not rescale it to a full-year equivalent; $S(x, 0)$ is defined on the entry-year window itself.

Finally, the fitted vehicle groups reach only part way to k^*, so the level of survival at the longest durations is pinned by a terminal anchor rather than left to the functional form; the anchor and the fitted coverage are described in Section G.5.

Table G.2 reports the survival parameters ($\hat{b}_1 = 0.7293$, se 0.0207 ; $\hat{b}_2 = -0.0810$, se 0.0037). Scrappage is heavily front-loaded: a fraction 0.4393 of the vehicles surviving at the start of the first lapse year are gone by the second, and the hazard remains in a similar range through the fourth. Beyond k^* the estimated scrappage hazard is held at zero, because the return windows identify survival only that far. This is a statement about the *hazard*, not about the survival level: the aggregate survival of the lapsed pool at k^* is 0.0476 and continues to drift downward at longer durations, because the composition of the vehicles that reach those durations keeps shifting toward older and shorter-lived vehicle groups. Roughly 24.4 % of lapsed vehicle-years (4,288,807 of them) sit beyond k^* and are governed by the zero-hazard convention; it is the conservative direction, in that it keeps the largest possible number of vehicles surviving in the tail, but it is a convention and every exhibit that uses it says so.

G.4 Results

Together the two estimated components deliver every share the decomposition requires. Writing $f(x, k)$ for the fitted on-road share of Section G.2 and $S(x, k)$ for the fitted survival of Section G.3, the share of *surviving* vehicles that are on the road is

$$g(x, k) \;=\; \frac{f(x, k)}{S(x, k)}. \tag{34}$$

This is the object our structural model of drivers' compliance choices requires: the model treats scrap, idle, and unregistered driving as a choice among *surviving* vehicles, while a previously scrapped vehicle makes no active choice—the vehicle no longer exists. The construction is also where the shared design of the two layers pays off: because the on-road share and survival models are complementary log-log models in one design—the same duration columns and the same covariate set—they see the same variation, and the ratio g is not contaminated by one model conditioning on covariates the other omits. The scrappage hazard is $h(x, k) = 1 - S(x, k)/S(x, k-1)$. Among vehicles with characteristics x surviving at the start of lapse year k, a fraction $h(x, k)$ is scrapped during the year, a fraction $\big(1 - h(x, k)\big)\, g(x, k)$ drives unregistered, and the remainder, $\big(1 - h(x, k)\big)\big(1 - g(x, k)\big)$, is idle. Every exhibit below is an aggregation of these group-level shares over the *decision stock*: lapsed vehicle-years weighted by the probability that the vehicle is surviving and off

Table G.2: Lapsed Vehicle Return-Survival Model: Parameter Estimates

	Estimate	Std. error
Duration baseline $\Lambda_0(k)$		
Lapse duration k (linear)	0.7293	(0.0207)
Lapse duration k^2 (quadratic)	-0.0810	(0.0037)
Covariate shifters $\eta(x)$		
Age 2–15	-0.1070	(0.0260)
Age 20–24	0.1248	(0.0394)
Odometer below median	0.0831	(0.0214)
Metro and metro × year FEs	Yes	
Fitted cells	216	

Note: Parametric return-survival model $S(x,k) = \exp\{-\Lambda_0(k)\, e^{\eta(x)}\}$, where $\Lambda_0(k)$ is the cumulative baseline over lapse duration k (quadratic, held flat past k^*) and x collects the age, odometer, metro and metro-year covariates that shift survival across cells. Fitted by weighted minimum distance to the nonparametric survival moments, with weights fixed at the inverse moment variances. Each row reports the coefficient with its standard error in parentheses; standard errors are from the criterion-based sandwich $(G'WG)^{-1}$, the same construction used for all derived margins, which treats the fitted cells as independent and takes the nonparametric moment standard errors as known. The omitted categories are age bin 16–19 and odometer above median. The model additionally includes metro levels and metro × year fixed effects, entered as a year main effect plus a Houston interaction. Metro levels and metro-year effects are estimated but not reported: they are absorbed in the fixed-effect block and carry no separate interpretation (the $x'\beta$ design carries no intercept, so the metro levels rather than a constant hold the baseline). The year factor pools 2016–18 into the omitted base cell because of data sparsity in the early, two-site sensing era. *Rank note.* The criterion prices 18 of 19 parameter directions: because the survival level is pinned by the terminal anchor, a common shift in $\eta(x)$ is offset exactly by a shift in the level of $\Lambda_0(k)$, leaving one direction unpriced. That direction loads on the metro levels, which are not reported; every coefficient shown here is orthogonal to it and is priced normally.

any purchased registration term, which is the population the demand model treats as making a choice.

Continuing the example of Section G.3 for a vehicle's first year after lapsing, suppose the sensors record a comparable fraud-passed vehicle four times as often as a lapsed vehicle. Then $f = 1/4$: a quarter of all lapsed vehicles are driven unregistered. Since half of all lapsed vehicles survive, $g = 1/2$: half of the survivors are driven unregistered and the other half are idle. Of all lapsed vehicles, half are scrapped, a quarter are idle, and a quarter are driven unregistered.

Table G.3 reports the estimated decomposition, marginally by calendar year, metro, vehicle age, odometer and lapse duration. Each block is an independent reconstruction of the same population, so the blocks must and do reproduce the overall row exactly. Over the pool as a whole, mean survival is 0.1879 : just under one in five lapsed vehicle-years corresponds to a vehicle that is still surviving and still making a compliance decision, and the remainder have left the market. Among vehicles surviving at the start of a lapse year, 37.8 % are scrapped during it, 37.7 % are idle, and 24.4 % are driving unregistered. The composition varies sharply with duration: unregistered driving is concentrated in the first lapse year, when the vehicle was recently in use and its owner has not yet resolved the

failed inspection, and it declines monotonically thereafter as the surviving vehicles settle into long-term idleness.

Table G.3: On-Road Outcomes for Lapsed Vehicle-Years: Marginal Distributions

				conditional on not yet scrapped at entry to k			
	N (lapsed veh-yrs)	% of N	S (surv)	scrap	idle	unreg	$g > 1$ mass %
ALL	17,552,433	100.0%	0.1879	0.3784	0.3773	0.2442	0.35
year							
2016	1,081,759	6.2%	0.1704	0.4742	0.3145	0.2113	0.00
2017	1,365,973	7.8%	0.1767	0.4752	0.3132	0.2116	0.00
2018	1,507,234	8.6%	0.1453	0.4701	0.3204	0.2096	0.00
2019	1,804,976	10.3%	0.1568	0.4504	0.3583	0.1913	0.00
2020	2,092,384	11.9%	0.1903	0.3826	0.4059	0.2116	0.00
2021	2,203,063	12.6%	0.2006	0.3369	0.4711	0.1920	0.00
2022	2,313,695	13.2%	0.1890	0.3478	0.4275	0.2246	0.00
2023	2,556,417	14.6%	0.2262	0.3208	0.4051	0.2741	0.00
2024	2,626,932	15.0%	0.1958	0.3430	0.2945	0.3625	2.32
metro							
DFW	8,407,227	47.9%	0.1761	0.4043	0.3507	0.2450	0.00
Houston	9,145,206	52.1%	0.1987	0.3583	0.3980	0.2437	0.67
age bin							
2-15	6,338,820	36.1%	0.2830	0.3960	0.3361	0.2679	0.57
16-19	5,848,228	33.3%	0.1659	0.3753	0.4053	0.2194	0.25
20-24	5,365,385	30.6%	0.0994	0.3315	0.4512	0.2173	0.18
odo bin							
below_med	8,824,756	50.3%	0.1476	0.3612	0.3748	0.2640	0.69
above_med	8,727,677	49.7%	0.2286	0.3902	0.3790	0.2308	0.00
k							
k = 0	3,002,936	17.1%	0.5607	0.4360	0.2321	0.3318	2.03
k = 1	2,354,706	13.4%	0.3257	0.4064	0.3417	0.2519	0.00
k = 2	1,944,902	11.1%	0.1592	0.4936	0.3270	0.1793	0.00
k = 3	1,737,913	9.9%	0.0791	0.4872	0.3585	0.1543	0.00
k = 4	1,577,046	9.0%	0.0503	0.3457	0.5171	0.1372	0.00
k = 5	1,399,429	8.0%	0.0491	0.0000	0.9030	0.0970	0.00
k = 6	1,246,694	7.1%	0.0476	0.0000	0.8993	0.1007	0.00
k > 6 (flat S)	4,288,807	**24.4%**	0.0449	0.0000	0.8870	0.1130	0.00

Note: Lapsed (emissions-noncompliant, non-appearing) vehicle-years. N is gap mass (n_quiet_rows); S is survival to entry of k. The three shares are the per-period trinomial, conditional on not yet scrapped at entry to k. Ratio of weighted means throughout, never mean of ratios: S weighted by n_quiet_rows; the three shares weighted by n_quiet_rows * w, where $w = S(1 - c)$ is the Item-6 demand/coverage weight, which is why they sum to 1 in every row. $g > 1$ mass% is the share of the row's n_quiet_rows sitting in cells with $g = f/S > 1$. Collapsed over origin_type, which is a row dimension of the trinomial grain but not a reporting block (spec §1.5). $k^* = 6$; rows past it are pooled into the flagged tail row. **Flat-survival tail.** Past $k^* = 6$, S is held CONSTANT in both models and all four duration forms – i.e. no vehicle is assumed to scrap after lapse duration 6. The $k > 6$ row carries 24.4% of all lapsed vehicle-years, so this convention governs roughly a third of the table's mass, not a residual. Source: data/processed/direct_fraud_signal_extension/onroad_trinomial_v2.parquet. Generated 2026-09-02 18:57 by src/analysis/vehicle_panel/27_marginal_exhibit.R (arm current).

Figure G.1 shows this composition over lapse duration, and Figure G.2 shows it over calendar time. The trend in the year figure is the result that matters for the paper's policy question: the share of the lapsed pool driving unregistered rises from 21.1 % in 2016 to 36.2 % in 2024, while survival is essentially flat and the scrappage share falls. The rise is concentrated in the years following the crackdown, and it is the quantitative counterpart to the substitution documented in Section 4.3: as fraudulent passes became harder to buy, a growing share of the drivers who would have bought one instead kept driving without registering.

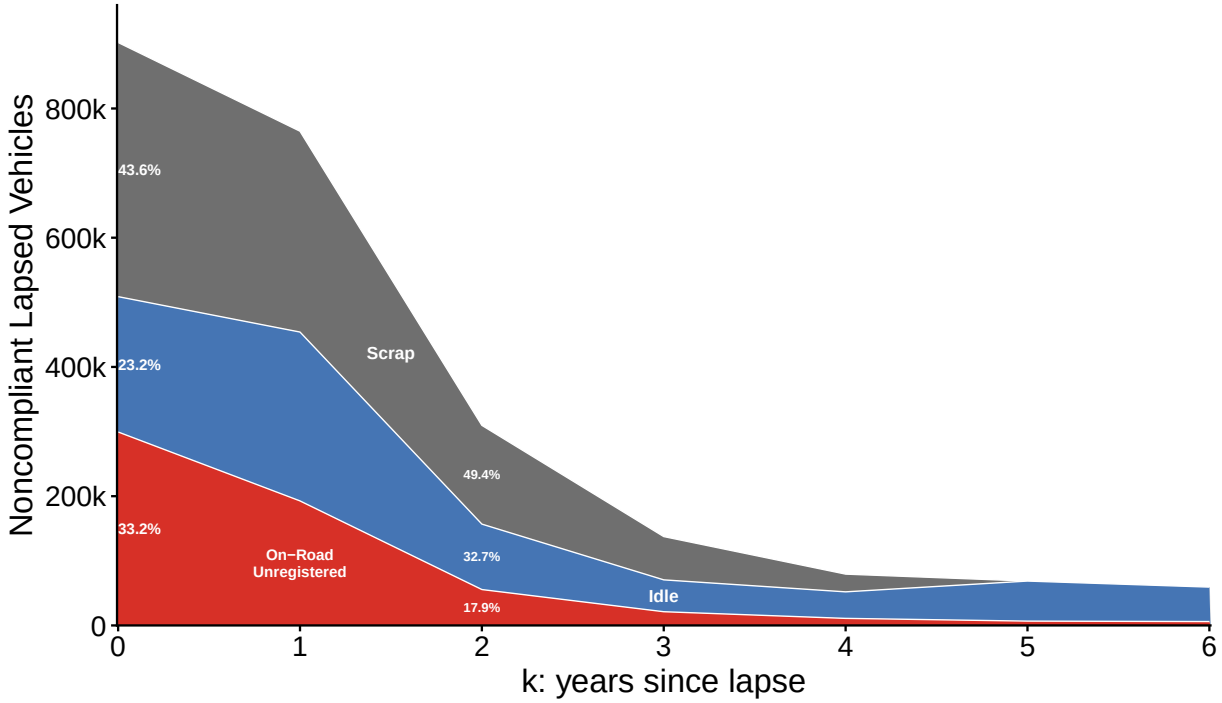


Figure G.1: Composition of the Lapsed Pool by Lapse Duration

Note: Survival- and coverage-weighted stock of lapsed vehicles by lapse duration k, split by state (scrap, idle, unregistered driving). The display ends at $k^* = 6$: scrappage is held at zero from that point on, as we do not have enough observations to estimate the survival curve with precision beyond it.

The early-year estimates deserve one caution: the sensing network in 2016 and 2017 was thin, so those years' estimates carry little information and should not be read as a sharp starting level for the trend.

Robustness to the duration form. The duration profile enters both components through a shared basis, and the choice of that basis is the only substantive functional-form decision in the chain. Re-estimating the entire model with a cubic profile, and with a fully saturated profile that places a free coefficient at every duration, moves the aggregate survival and unregistered-driving shares by at most 0.7 percentage points, and the idle share by at most 1.2 percentage points. The decomposition is therefore not an artifact of the duration parameterization.

Figure G.3 makes the same point at the level of the survival profile itself. The three arms are estimated on identical data with the same year effects; the duration basis is the

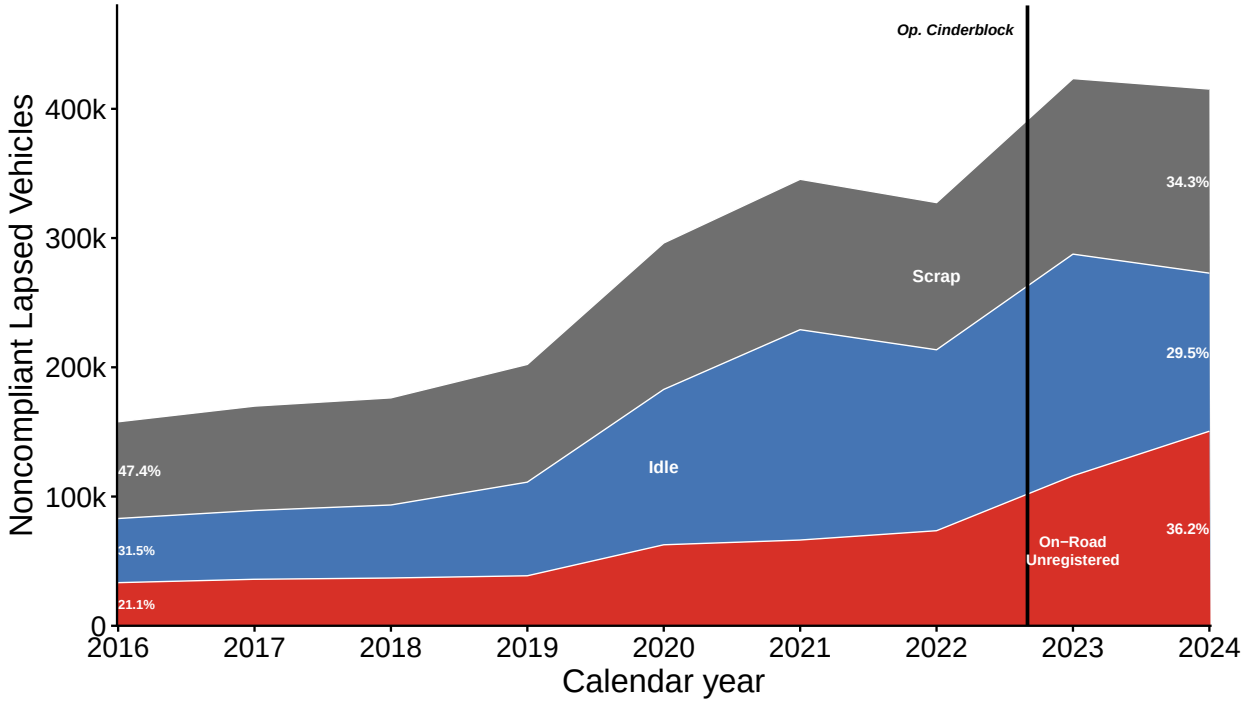


Figure G.2: Composition of the Lapsed Pool by Calendar Year

Note: Stacked area, by calendar year, of the lapsed pool split into scrap, idle, and unregistered driving. The height of the stack is the decision stock defined in the text. It is not a raw count of lapsed vehicles, which is 7.0 times larger. Percentage labels give the share of the decision stock in each state in the first and last sample years. The vertical rule marks the onset of the crackdown in August 2022.

only thing that changes between them. The three profiles trace the same steep decay out of the first lapse year and settle onto the same level by the end of the estimation support. Where they separate at all it is in the middle durations, where the saturated profile spends its extra freedom and holds a little more mass in the pool for a year or two before rejoining the other two; the two parametric arms are nearly indistinguishable throughout. The basis choice moves the shape of the decay, not its beginning or its end.

G.5 Estimation and Implementation Details

Sample and joint estimation. The on-road share model is fitted on 17,552,506 lapsed vehicle-years, grouped into 36,298 vehicle groups defined by metro-year, vehicle age bin, odometer bin, lapse duration and sensing overlap, stacked with the reference group. Detection and the on-road share are estimated in a *single* complementary log-log likelihood with 36 parameters rather than in two stages: a two-stage procedure would pin the detection level in as if known, understating the uncertainty in every on-road share coefficient. Within the joint fit, sampling uncertainty in the detection parameters accounts for a median factor of 1.0815 on the on-road share standard errors—about 8.1 % of a typical standard error— relative to a counterfactual in which those parameters were treated as known. Standard errors are clustered by vehicle identification number, since a vehicle contributes one observation per year of its lapse spell; clustering leaves the median standard error essentially unchanged (a

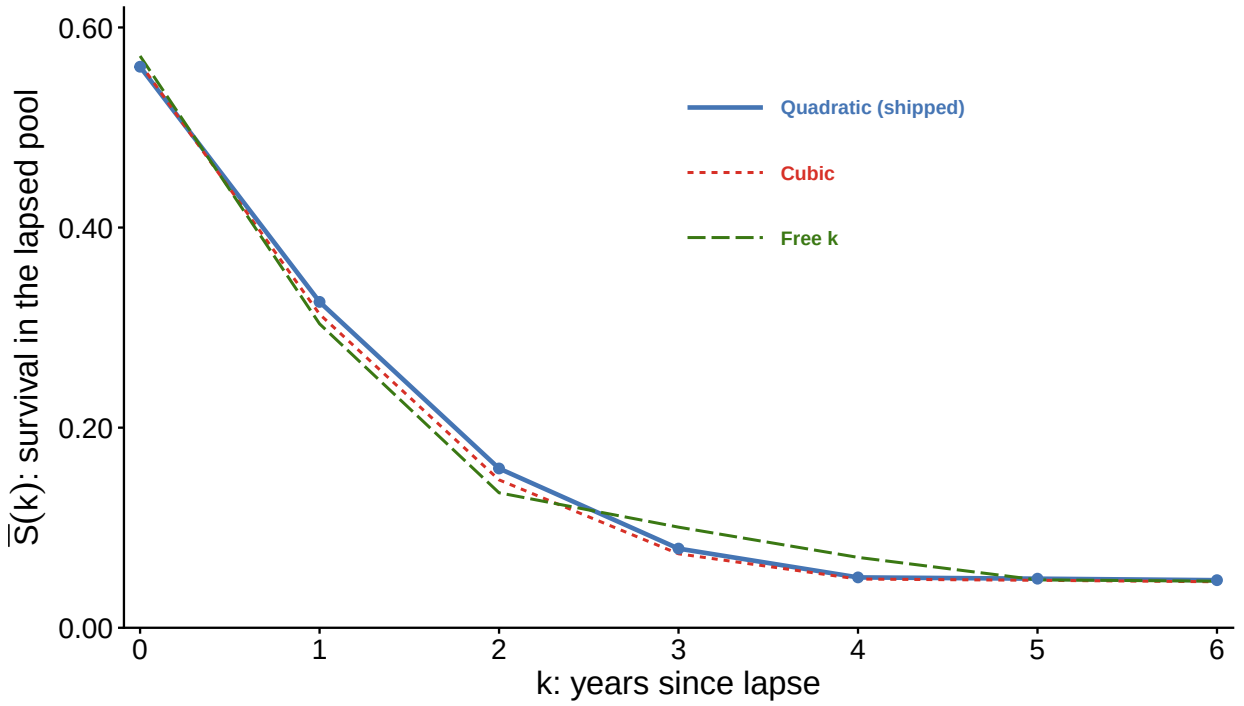


Figure G.3: Survival by Lapse Duration Under Three Duration Forms

Note: Mass-weighted average survival $\bar{S}(k)$ of the lapsed pool by years since lapse, compared across the baseline quadratic duration profile and the cubic and fully saturated (free k) alternatives. All three are re-estimations of the same model on the same data with the same year effects, the same objective and the same survival anchor; only the duration basis differs. Vehicle groups are averaged with a single fixed weight, the raw count of lapsed vehicle-years in the group, taken once from the baseline decomposition and applied to all three arms: it is a property of the data rather than of any fit, so the three lines are averages over the same population and the only thing that differs between them is the duration form. Under this weight the baseline line reproduces the survival column of the lapse-duration block of Table G.3 exactly. The panel is cut at $k^* = 6$, the edge of the estimation support, beyond which the zero-hazard convention holds every arm flat and any comparison between them is mechanical.

ratio of 1.0058), which is what one expects when the within-vehicle correlation is already absorbed by the duration and vehicle-group structure.

The lapse window and the home-metro restriction. A vehicle enters the panel at its first failed or fraudulent inspection and contributes one vehicle-year row per calendar year thereafter, whether or not it ever returns; nothing conditions on eventual return, since doing so would build the survival result into the on-road share estimate. For vehicles whose lapse originates in a fraudulent pass, the first lapse year begins when the purchased term expires, and only sightings after that date count, with the sensing overlap ω recomputed over the same post-expiry window. Sightings that occur while a vehicle is still covered by a term it paid for are not evidence of unregistered driving, and including them would have attributed a large volume of legally registered driving to the lapsed pool. Qualifying sightings are restricted to monitor sites in the vehicle's own metro, which is also the metro whose sensing calendar defines ω; a small share of vehicle-years sighted only at the other metro's sites is dropped, both because no matching on-road share denominator exists for them and because

"sighted only in the other metro" is the signature of a misassigned home market.

Duration basis and the flat region. The same duration columns are supplied to the on-road share index and to the cumulative baseline hazard of (33), top-coded at $k^* = 6$ for both. Top-coding is what implements the flat region: every observation beyond k^* receives the value at k^*. The alternative—letting each functional form extrapolate freely past the last duration the data support—is not neutral. The candidate forms disagreed severalfold on the tail level of survival when their tails were free, and agree to within 0.0022 once the tail level is pinned by the terminal anchor. The flat region is thus an explicit and shared convention that replaces a form-dependent extrapolation, rather than an additional restriction imposed on informative data.

The terminal anchor. The moments pin the shape of the duration profile and the covariate contrasts, but not the level of survival at the longest durations. Vehicle groups clear the minimum-sighted bar only out to $k = 4$ years of lapse, short of k^*, so the level of the frozen tail is set by a terminal anchor rather than by whichever functional form was chosen. Let j index the covariate groups the surface is fitted on, $\eta(j)$ their fitted indices, and q_j group j's share of lapsed vehicle-year mass over the fitting set; the anchor solves

$$\sum_j q_j \exp\big(-\kappa\, e^{\eta(j)}\big) \;=\; \bar{S}_{\text{target}}(k^*)$$

for the terminal level $\kappa = \Lambda_0(k^*)$, where $\bar{S}_{\text{target}}(k^*)$ is the return-survival ratio at k^* pooled over the estimation cohorts (278 sighted lapsed vehicles, of which 28 returned).[65] Identification therefore divides cleanly: the group moments pin the shape of $\Lambda_0(k)$ over the durations the sighted sample reaches and every covariate contrast in η, the terminal anchor pins the one level the moments cannot reach, and Assumption G.3 is what makes any of it a statement about survival rather than about return behavior. The surface is fitted on 216 vehicle groups spanning $k = 0$ to 4 and covering 83 of the 108 covariate groups, and predicted on all 108 ; the remaining 25 groups—the oldest age bins in the early sensing years, where sightings are scarcest—carry no moment of their own and receive index predictions.

Inference for the survival parameters. The survival surface is fitted by weighted minimum distance rather than by maximum likelihood, so there is no likelihood-based variance matrix. Standard errors come from the criterion sandwich $(G'WG)^{-1}$, where G is the Jacobian of the fitted moments in the parameters and W the minimum-distance weight matrix, evaluated at the fitted parameters; this treats the fitted vehicle groups as independent and takes the nonparametric moment standard errors as known. The criterion prices all but one parameter direction. Because the terminal anchor fixes the level of survival, a common shift

[65] The anchor is a moment condition on the fitted surface, not on the baseline curve, and it has to be: the metro indicators span the constant direction, so a common shift in them multiplies exactly the same product that a shift in the baseline level multiplies and a pin on $\Lambda_0(k^*)$ alone would be undone by it. Because the anchor carries the level, $\eta(x)$ carries no intercept, and the level direction is not priced by the fitting criterion—the rank note reported with Table G.2.

in the covariate index $\eta(x)$ is offset exactly by a shift in the level of the cumulative baseline, leaving that direction unidentified by the criterion; it loads on the metro level terms, which are absorbed in the fixed-effect block and not reported. Every coefficient reported in Table G.2 is orthogonal to the unpriced direction.

Fit of the survival surface. How well the surface reproduces the moments it was matched to has to be read in standard-error units rather than in levels. The parametric surface is meant to depart from any individual group's nonparametric ratio—pooling across duration, metro, year and vehicle characteristics is the entire purpose of fitting one—so an absolute deviation says nothing without knowing how precisely the group it is measured against was estimated, and the thinnest groups are precisely where the largest absolute gaps sit. Scaled by each group's own standard error and weighted by the same $1/\text{se}^2$ the criterion uses, the root mean squared deviation of fitted from nonparametric survival is 0.050 standard errors, and 7 of the 216 fitted groups deviate by more than two standard errors—about what independent sampling noise would produce. The surface is therefore not straining against its moments anywhere in bulk; where it does deviate materially, it deviates in groups the criterion had almost no information about, which is the pattern the following paragraph traces.

The bound on g and its residual. The share of survivors on the road, $g(x, k)$ of (34), is a probability and must lie in the unit interval in every vehicle group; that it does is a genuine over-identifying check, since the two components in the ratio are estimated on different margins and nothing in the estimation imposes it. The check passes at every lapse duration in the aggregate and in all but 6 of the estimated vehicle groups. The exceptions lie in three vehicle groups—one in the first lapse year, two at three years of lapse—all in the below-median odometer bin and among the thinnest groups the fit reaches, and together they carry 0.3478 % of lapsed mass. They reflect group-level lack of fit rather than a contradiction in the data: in all but one of the affected groups the gap from the raw counterpart is within sampling error, with the fitted on-road share sitting above and the fitted survival below their raw values, the two misfits compounding in the ratio. The cause is that both components are multiplicative in age, odometer and metro-year with no interactions, so the odometer effect is forced constant across markets and years. We report the main specification with the affected probabilities set to the boundary of the unit interval, and a robustness specification that drops the affected groups; neither is dominant—boundary values fabricate a moment, dropping groups understates lapsed mass—and the two give the same answer at every margin we report.

Registration coverage as a weight. A vehicle in its first lapse year may spend part of that year on a registration term it had already purchased. Driving on a term one has paid for is compliant, so coverage is not a fourth state; it enters as an observation weight that scales each lapsed vehicle-year by the probability that the vehicle is surviving and off any purchased term. The weight binds only in the first lapse year, where the uncovered share averages 0.54 , and is one everywhere from the second lapse year onward, since a one-year term cannot reach two years out. Both figures in this appendix plot this weighted decision

stock, which is why their heights are well below raw lapsed counts.

The demand likelihood weights the later years of a lapse in the same spirit (Section 6.1). In each calendar year after the first, a lapsed vehicle that has not returned by the start of the year enters the likelihood with a weight equal to its estimated survival to the start of that year, $S(x, k-1)$, times the share of the year it remains unregistered. A vehicle that returns for a repair four months into a year therefore counts as a third of a lapsed vehicle in that year.

Data construction. The lapsed panel, the sensing-overlap calculation and the three-way decomposition follow the sample definition of Section G.1. The underlying inspection panel is described in Appendix B, and the remote-sensing records and their matching to the inspection panel are described in Appendix D.

Appendix H. Demand: Estimation, Identification and Fit

This appendix gives the details behind the demand estimates of Section 6.1. The nesting parameter λ_F governs where the buyers of a closed station go, and every enforcement counterfactual turns on it, so much of the appendix concerns the variation that pins it down. Every alternative but scrapping carries metro-by-period effects, and fraud carries home-county effects, so λ_F is pinned down by changes in the set of active stations within a metro and period, the largest of them during the 2022 crackdown (Figure 2). Appendix H.1 describes the estimation sample, Appendix H.2 the estimator, and Appendix H.3 the fraud inclusive value. Appendix H.4 reports λ_F under alternative specifications of the time effects, and Appendix H.5 the full estimates and the model's fit. Appendix H.6 discusses how we measure the set of active stations, and Appendix H.7 tests whether drivers delayed their inspections during the crackdown.

H.1 Estimation Sample

Our sample is the population of malfunctioning vehicles in Dallas–Fort Worth and Houston from 2016 to 2024, and each observation is a driver's choice in one inspection cycle (Section 6.1). A vehicle is a new-malfunction vehicle in a cycle if it fails the first inspection of the cycle, or if its driver buys a fraudulent pass without first failing a test. It is a carryover-malfunction vehicle if its previous cycle ended without an honest pass: in a failure that was never cleared, in a fraudulent pass, or in a lapse. We do not distinguish drivers who repaired in the previous cycle from those whose vehicles passed without a repair: both ended the cycle with an honest pass. A vehicle that meets both definitions counts as carryover. A carryover-malfunction vehicle stays in the sample until it passes an honest inspection. Because we never observe scrapping, a lapsed vehicle counts in later years in proportion to its estimated probability of not yet having been scrapped, which comes from the survival estimates of Appendix G.

Classifying a vehicle as carryover in 2016 requires knowing how its earlier cycles ended. We therefore run the fraud detection model of Section 3 on all inspections back to 2007. We

use the years before 2016 only to classify each vehicle's history; no choice before 2016 enters the likelihood.

Three groups of drivers enter the likelihood (Appendix H.2). Testers fail an honest inspection during the cycle, so we observe their outcome and predict their repair bill from their diagnostics. Carryover-malfunction drivers who do not test leave no diagnostics, but their history tells us that the vehicle is malfunctioning. New-malfunction drivers who do not test appear in the data only when they buy a fraudulent pass. To keep the computation manageable, we group drivers who share the same home county, age group, mileage group, metro and year, and who choose in the same month. This gives 9,072 groups. We also group testers' predicted bills into bins. A vehicle's home county is the registration county recorded most often in its inspection records. For vehicles with no recorded registration county, we use the county in which the vehicle was most often inspected. The home county also determines the driver's home metro.

H.2 Estimation

We estimate the 88 parameters of equations (3), (9) and (10) by maximum likelihood in two nested loops. The outer loop searches over three sets of parameters. The first is the distribution of price sensitivity, $\bar{a}$ and σ_α. The second is the coefficients $\boldsymbol{\beta}_k$ on the age-group and mileage terms and the metro-by-period effects in the utilities of repair, fraud, idling and driving unregistered, with home-county effects in the utility of fraud. The third is the scalars λ_F, μ, η_R, η_F, η_L and τ_C. The inner loop, run at every trial value, recovers the station qualities.

The inner loop. We never search over station qualities. Let $\mathcal{J}_{mt}$ be the alternatives inside the fraud nest in market m and year t: each large station, the competitive fringe and the option of buying outside the home metro. Write δ_{jmt} for the quality of alternative j, which is δ_{smt} for a large station, δ^f_{mt} for the fringe and δ^{out}_{mt} for the outside option. Given a trial (μ, λ_F), we invert the qualities: within each market-year, we solve for the vector $\boldsymbol{\delta}$ at which the model's predicted share of each alternative equals its observed share of the market-year's fraud purchases,

$$\hat{s}_j(\boldsymbol{\delta}; \mu, \lambda_F) \;=\; s_j^{\text{obs}}, \qquad j \in \mathcal{J}_{mt}. \tag{35}$$

The predicted share sums each alternative's probability within the nest over the months and groups of drivers in the market-year, weighted by each group-month's observed fraud purchases. This aggregation is exact. Because δ_{smt} absorbs each station's price (Section 6.1), which station a fraud buyer chooses depends on neither the repair bill nor price sensitivity. The choice among stations therefore separates from the rest of the likelihood and needs only (μ, λ_F) from the outer loop. We solve system (35) by a damped Newton iteration on the log ratio of predicted to observed shares, starting from the solution at the previous outer iterate. A station outside the set of active stations in a month has probability zero for drivers who choose in that month, both in the aggregation and in the inclusive value. As in Grieco et al. (2025), this approach combines a likelihood over individual choices with an exact fit to station-level shares, and it removes hundreds of station qualities from the search.

The normalization. Adding a common constant to every quality in a market-year leaves every choice unchanged. Each quality enters the inclusive value only through $\exp(\delta_{jmt}/\lambda_F)$, so a uniform shift c cancels in the choice among stations and moves $\lambda_F \log S_i$ by exactly c, which the fraud fixed effects absorb. The qualities are therefore identified only up to a level, and we fix that level by setting their unweighted mean to zero in every market-year,

$$\frac{1}{|\mathcal{J}_{mt}|} \sum_{j \in \mathcal{J}_{mt}} \delta_{jmt} \;=\; 0 \qquad \text{for every market } m \text{ and year } t. \tag{36}$$

We run the inner solve with the first alternative's quality pinned, and recenter the qualities to satisfy (36) after convergence.

The normalization matters for what λ_F can learn from. It removes the average quality of a market-year's stations from the inclusive value. After (36), $\log S_i$ responds to the number and the composition of the active stations a driver faces, but not to how attractive that market's stations are on average in that year. The fraud fixed effects carry that level instead. As a result, differences across metros in the annual level of fraud inform λ_F only through the number and composition of active stations. The normalization does not make the fraud fixed effects redundant. The two are separately identified from different variation: the fixed effects from the level of fraud purchases in each metro and period, and the qualities from how those purchases are distributed across stations within a market-year.

The normalization also leaves untouched the object the supply side uses. A station's quality θ_{smt} in Section 5.3 adds to δ_{smt} the taste for fraud common to all drivers in the market-year. Shifting every δ_{smt} by c and the fraud fixed effects by $-c$ leaves θ_{smt} unchanged.

The likelihood. The likelihood has three parts, because whether we see a driver depends on whether the driver tests. For testers, we evaluate the probability of the observed choice at the predicted bill, averaging over price sensitivity by Gauss–Hermite quadrature. A lapse enters as a weighted combination of scrapping, idling and driving unregistered, with weights from the lapse decomposition of Appendix G. This is similar to how observations with an unobserved type enter the likelihood in the EM algorithm (Arcidiacono and Miller, 2011), except that our weights are estimated beforehand, from the return rates and the remote-sensing data, rather than updated with the demand parameters. For carryover-malfunction drivers who do not test, we also integrate over the distribution of predicted bills among testers with the same characteristics (Assumption 4; Appendix F.4).

The third part covers new-malfunction drivers who never take an initial inspection. We see such a driver only if the driver buys a fraudulent pass. A driver who has the vehicle repaired before even taking an initial inspection appears in the inspection data as an honest pass, like a vehicle that was never malfunctioning, and a driver who lapses without testing leaves no record of a malfunction at all. Consequently, we need a way to account for these missing repair and lapse choices. As Section 6.1 states, we assume that new-malfunction drivers who do not test choose like new-malfunction testers with the same characteristics. Equivalently, the probability of taking an initial inspection does not depend on the choice the driver goes on to make. The testers' choices then give the choice probabilities of the non-testers. The observed number who buy fraud without testing then implies how many repaired or lapsed without testing.

The parameter τ_C does this calculation inside the likelihood. It is the probability that a new-malfunction driver takes an initial inspection. For a group of new-malfunction drivers, let P_F be the model's probability that a driver in the group buys fraud, averaged over price sensitivity and over the distribution of predicted bills. A driver in the group appears in the data either by testing, with probability τ_C, or by skipping the test and buying a fraudulent pass, with probability $(1-\tau_C)P_F$. We write the group's likelihood conditional on appearing in the data:

$$N_T \log \frac{\tau_C}{\tau_C + (1-\tau_C)P_F} \;+\; n_F \log \frac{(1-\tau_C)P_F}{\tau_C + (1-\tau_C)P_F}, \tag{37}$$

where N_T is the number of the group's drivers who fail an initial inspection and n_F the number who buy a fraudulent pass without testing. Given τ_C and P_F, the implied number of drivers who repaired or lapsed without testing is $n_F(1-P_F)/P_F$. The parameter τ_C enters only this part of the likelihood: the testers' part is conditional on testing, so τ_C cancels there.

The assumption likely understates how many of these non-testers choose fraud. In particular, non-testers are probably more likely to buy fraud than testers,[66] so treating them like testers makes them look more willing to repair than they likely are. As a result, our counterfactuals, if anything, overstate how much fraud falls among these non-testers when stations are shut down.

Computation and standard errors. We integrate over price sensitivity by Gauss–Hermite quadrature on 101 nodes, and over non-testers' bills on 8 nodes. We maximize the likelihood using its analytic gradient, which includes the response of the inverted qualities to (μ, λ_F) through the implicit function theorem, from 13 starting points, and report the estimate with the highest likelihood. Standard errors come from the Hessian of the likelihood with the qualities concentrated out, computed by central differences of the analytic gradient. They are not clustered. They treat the first-stage estimates, namely the fraud classification, the predicted bills, and the lapse and survival shares, as known, and so understate the uncertainty in our estimates.

What the fixed effects absorb. The fixed effects determine which variation is left to pin down λ_F. Metro-by-period effects enter the utilities of repair, fraud, idling and driving unregistered, relative to scrapping, with 2016–2018 pooled into one period and one period for each year after. We pool 2016–2018 because the roadside remote-sensing data that separate the three lapse alternatives cannot distinguish those years, when the network operated at a few sites for a few weeks a year (Appendix G.2). The metro-by-period effects absorb the level of each alternative in each metro and period, including any effect of enforcement on the taste for fraud at that frequency. Home-county effects enter the utility of fraud only, with one omitted county per metro, so they never span a metro level or a period effect. The inertia terms η_R, η_F and η_L are constant over time. What remains for λ_F is the month-to-month movement in the set of active stations within a metro and period. Because we normalize qualities in each calendar year while the fraud level is pooled over 2016–2018, changes in

[66] A driver open to repair loses little by taking the inspection first: the inspection identifies which components fail, and a vehicle repaired at the same station can be retested at no extra charge.

the number and composition of active stations between those three years also contribute. Column (B) of Table H.1 removes them.

H.3 The Fraud Inclusive Value

Equation (10) writes the inclusive value in three parts: the large stations active in the month of the driver's choice, the competitive fringe, and the option of buying outside the home metro. A station is large in a year if it sells at least 1,000 fraudulent passes to vehicles in our sample that year, enough to support a quality of its own. The fringe stands for the many stations that sell in small volume. They share a single quality and enter with a count, $N^f_{m\ell t}$, so entry and exit among small stations still move S_i. The outside option is always available, which keeps the inclusive value from collapsing when a metro has few active stations. All three kinds of quality are annual, one per market-year, and we invert and normalize them together under (36).

We date the set of active stations by month, while qualities are annual. A station is active from the month of its first fraudulent sale to the month of its last, and $N^f_{m\ell t}$ counts the fringe stations active in month ℓ. We date a driver's compliance choice by the month of the passing inspection, or by the month of the first failed inspection if the cycle ends in a lapse. Two drivers in the same metro and year who choose in different months face the same qualities but different sets of active stations, and hence different values of S_i. This is the variation that pins down λ_F, and it is why the monthly dating matters. At the annual level, the set of stations and the fraud level move together, and the metro-by-period effects absorb both.

The match term $\mu\,\bar{w}_{is}$ applies to large stations only; the fringe and the outside option carry none. In equation (10), $\bar{w}_{is}$ indicates that station s is in driver i's home county. In estimation, it is the share of a group's drivers whose home county is station s's county; because we group drivers by home county (Appendix H.1), that share is one or zero. A large μ would mean that buyers are local. Closing stations in a buyer's own county would then lower that buyer's inclusive value by more than closing the same number elsewhere, and the counterfactuals would depend on which stations close, not only on how many.

Travel costs to a station are likely small, since a station can sell a fraudulent pass without the vehicle being present (Section 2.2). The match value can also pick up local advertising, for example if a station advertises on Craigslist only in the county where it operates.

Why $\log S_i$ is not $\log N$. It is tempting to read $\lambda_F \log S_i$ as a term in the log number of stations, and the two do move together. When qualities are identical and matches uniform, differentiating (10) with respect to $\log N$ gives exactly λ_F. The reasons they differ are what make the counterfactuals informative. First, stations are not identical. S_i weights each by $\exp(\delta_{smt}/\lambda_F)$, so losing a large station differs from losing a small one, and a metro that loses half its stations but keeps its best ones loses far less than half its inclusive value. Second, stations are not equidistant. The $\mu\,\bar{w}_{is}$ term makes S_i specific to the driver, so two drivers in the same metro and month facing the same stations can have different inclusive values. Third, and most important for this appendix, the normalization (36) removes the market-year level of quality from S_i but has no counterpart for $\log N$. The count therefore carries the market-year level of fraud supply, and $\log S_i$ does not. A regression of the fraud share

on $\log N$ would be a reduced-form proxy for the model's variation, not a second estimate of λ_F.

H.4 λ_F Across Time-Effect Specifications

Within any set of months over which the time effects on fraud are constant, the log odds of fraud against scrapping, net of the covariates, home county and history, move only through $\lambda_F \log S_i$. Each column of Table H.1 changes which months the time effects hold constant, and so which movements in $\log S_i$ are left to pin down λ_F. Every column uses the same sample and likelihood. Sampling error is small in every column: the standard error of λ_F is 0.0026 in column (A). We therefore compare columns on an economically meaningful object rather than test them against each other. That object is the within-fraud diversion, the share of the fraud buyers displaced by the stations that actually closed who find another fraud seller. We call a column stable if its diversion is within 5 percentage points of column (A).

Column (A) is our main specification, with $\lambda_F = 0.514$ and a diversion of 48.8%. Splitting the 2016–2018 period into its three years, in column (B), leaves both essentially unchanged, at 0.509 and 49.2%. Column (F) adds shifts in the demand for fraud after August 2022 and after March 2023, separately by metro. It addresses the concern that the crackdown deterred fraud directly, through the announcement rather than through the closures. λ_F moves to 0.486 and the diversion to 51.3%. Because these shifts absorb part of the within-year variation across metros that column (A) uses, column (F) over-controls, and we read it as a reference rather than a preferred specification.

The two columns that change the identifying variation lower λ_F. Column (D′) replaces the metro-by-period fraud levels with common period effects, a constant Dallas–Fort Worth gap and two common crackdown windows. λ_F falls to 0.412 and the diversion rises to 57.9%. Column (C) adds common month-by-year effects to every alternative. These absorb every statewide shock within a year, such as the March 2023 announcement, seasonality or a common shift in when drivers inspect. Column (C) pins down λ_F only from differences between the two metros in the month-to-month movement of their active stations, most of which occurs in crackdown months. λ_F falls to 0.221 and the diversion rises to 76.8%. A lower λ_F means that more of a closed station's buyers find another fraud seller. Our main specification is therefore conservative for the conclusion that substitution undercuts closures, and we treat column (C), the most demanding specification, as an upper bound on the diversion.

Column (E) drops the home-county effects, so county differences in the taste for fraud can load onto μ and λ_F. The home-county premium μ rises from 0.118 to 1.016, and independent starting points settle in two separate optima, so column (E) does not give a well-located estimate of λ_F. We keep the home-county effects for this reason.

H.5 Full Estimates and Fit

The model has 88 parameters, of which Table 3 reports the main ones. The log likelihood is −18,796,902 over 9,072 groups of drivers. The home-county premium μ is 0.118 (standard error 0.011).

Table H.1: The Fraud Nest Parameter Across Time-Effect Specifications

	(A)	(B)	(C)	(D′)	(E)	(F)
λ_F	**0.514**	0.509	0.221	0.412	0.523‡	0.486
	(0.0026)	(0.0027)	(0.0046)	(0.0029)	(0.0024)	(0.0031)
μ	**0.118**	0.118	0.041	0.127	1.016	0.090
	(0.0109)	(0.0110)	(0.0096)	(0.0096)	(0.0019)	(0.0109)
Within-fraud diversion (%)	**48.8**	49.2	76.8	57.9	45.2	51.3
difference from (A), pp	–	0.4	28.1	9.1	−3.6	2.5
Parameters	88	104	492	84	76	92
Flag					two basins	

Note: Each column is a separate estimate of the demand model that differs from column (A), in bold, only in the time effects or controls named here. (A) Baseline: metro-by-period levels on each alternative, home-county dummies on fraud. (B) Metro-by-calendar-year levels: the 2016–18 period is split into its three years. (C) Adds common month-by-year effects on each alternative. (D′) Fraud only: common period effects, a constant Dallas–Fort Worth level and two common crackdown windows replace the metro-by-period fraud levels. (E) Drops the home-county dummies on fraud. (F) Adds four dated fraud level shifts (from August 2022 and from March 2023, by metro). Standard errors, in parentheses, are unclustered: they are the inverse of the observed information of the likelihood with the station quality terms re-solved at every parameter value. On the full population of malfunctioning vehicles they are small in every column, so the table does not test columns against each other; it asks whether an economically meaningful object moves. That object is the within-fraud diversion: the share of the fraud buyers displaced by the stations that actually closed in the enforcement wave who find another fraud seller, holding the surviving sellers fixed. The tolerance for calling a column stable, stated before any column was estimated, is 5 percentage points from column (A). ‡ A flag marks a column whose estimate is not a well-located point: a parameter rests at a bound of the admissible region, or independent starting points disagree along a direction the likelihood barely distinguishes (a flat direction) or settle in two separated optima (two basins). § The curvature of the likelihood is not positive definite at the estimate, so no standard error is reported.

Figure H.1 shows the five outcomes as shares of all vehicles, including those that pass their first inspection honestly.

H.6 Measuring the Set of Active Stations

We build each month's set of active stations from sales, as the detection model of Section 3 classifies them. A station counts as active from the month of its first fraudulent sale to the month of its last. A month without sales inside that window therefore leaves the station active, so a slow month does not remove a station from the set. Only the two ends of a station's spell depend on when its sales happen to fall, and so only the ends can be misdated.

We date a station's exit by its last fraudulent sale, not by the date of an enforcement action against it. A station that winds down before it closes therefore leaves the set of active stations before the enforcement date. This dating is a limitation of our measure.

DPS targets stations with high fraud rates. A station's fraud rate is a persistent characteristic, which its annual quality δ_{smt} absorbs. The identifying assumption of Section 6.1 does not require DPS to choose stations at random. It requires only that the months in which stations close are unrelated to month-to-month shifts in the demand for fraud.

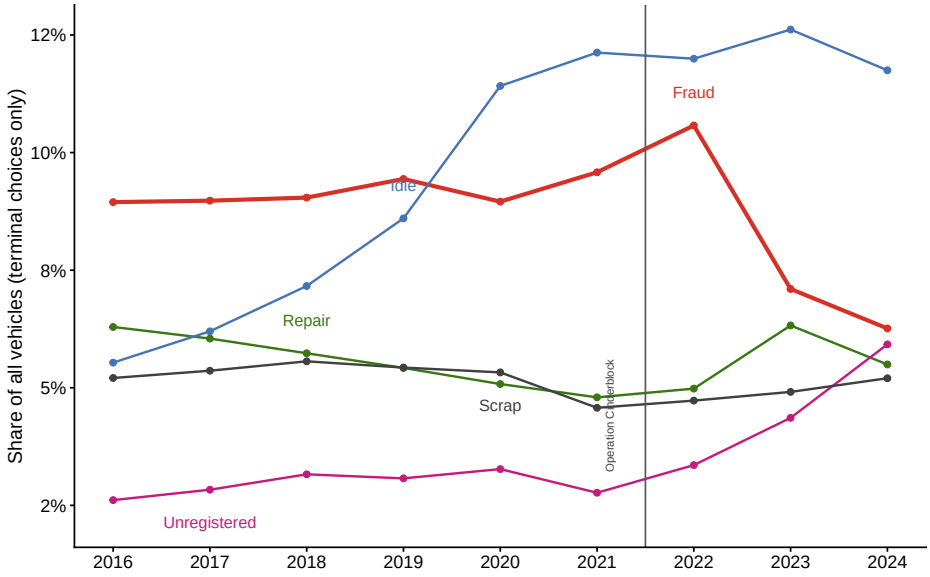


Figure H.1: Outcome Shares of All Vehicles, Dallas–Fort Worth and Houston

Note: Lines show the five outcomes (repair, fraud, driving unregistered, idling and scrapping) as shares of all vehicles in Dallas–Fort Worth and Houston pooled. The denominator includes vehicles that pass their first inspection honestly, but the figure omits their line, which would sit far above the other five and compress them toward the axis. Malfunctioning vehicles enter these shares weighted by their estimated survival (Appendix G), while vehicles that pass honestly enter as a raw count. The horizontal axis is the year of the driver's decision. The vertical line marks the crackdown, which began in August 2022; it sits at 2021.5 because the 2022 decision year is partly after the crackdown began.

H.7 Re-timing, or Substitution?

The main threat to reading the monthly fall in fraud as substitution is delay. If drivers who wanted a fraudulent pass simply waited until they found a seller, the monthly fraud share would fall without any change in what drivers ultimately chose, and λ_F would be overstated.

We test this in two ways. First, we follow cohorts defined by the month of the *initial failure* rather than the month of resolution, so a driver who delays is still counted in the month the driver failed. The fall survives: Houston's April–July cohorts bought fraud at a share of 0.077 in 2023, against 0.142 in 2022 and 0.131 in 2021, a fall of 6.6 percentage points on the year. Over the same year, the lapse share rose from 0.074 to 0.104. Resolution times barely moved: the median number of days to resolution was 3.5 in 2023 against 3.0 in 2022. The re-timing margin, the share of fraud that arrives more than thirty days after the failure, *fell* from 0.032 to 0.024, the opposite of what delay would produce. Table H.2 reports the cohorts.

Second, we look at how long it has been since the buyers of fraudulent passes last passed an inspection. If drivers had deferred fraud and bought it later, the vehicles obtaining fraudulent passes in 2023 would have gone longer since their last pass than those of 2022. We find no evidence of this. The median is 12.36 months in 2023 against 12.28 in 2022, while volume halved, and the upper tail thinned. The 90th percentile fell from 30.4 to 25.8 months, and the share more than eighteen months since the last pass fell from 20.8% to 17.6% (Figure H.2). The fall in fraud reflects substitution, not delay. In any case, the

months a driver spends waiting with an expired registration already count as lapsed in our sample (Section 6.1).

Table H.2: Initial-Failure Cohorts Around the Closures

Cohort (month of initial fail)	N	Fraud share	Fraud ≤30d	Re-timing margin	Repair share	Walk share	Median days	p90 days	Res. ≤30d
Houston, Apr–Jul 2021	72,813	0.131	0.098	0.033	0.788	0.081	3.2	42.4	0.799
Houston, Apr–Jul 2022	70,672	0.142	0.110	0.032	0.783	0.074	3.0	42.0	0.806
Houston, Apr–Jul 2023	87,165	0.077	0.053	0.024	0.820	0.104	3.5	44.6	0.774
DFW, Apr–Jul 2021	73,448	0.113	0.086	0.028	0.811	0.075	3.0	39.7	0.811
DFW, Apr–Jul 2022	70,975	0.130	0.102	0.028	0.795	0.075	3.3	40.7	0.808
DFW, Apr–Jul 2023	78,742	0.103	0.078	0.025	0.814	0.084	3.3	40.0	0.801
Houston, Sep–Dec 2022 (closure window)	66,884	0.124	0.100	0.024	0.799	0.076	3.0	40.7	0.808
DFW, Sep–Dec 2022 (closure window)	65,170	0.118	0.089	0.029	0.807	0.075	4.0	44.8	0.801

Cohorts are defined by the month of the initial failure, so a driver who delays resolution is counted in the month the driver failed. The re-timing margin is the share of fraud resolved more than thirty days after the failure.

Appendix I. The Detection Hazard

We estimate the detection hazard of Section 6.2 by maximum likelihood on DPS's record of enforcement actions. Below we state its specification, describe the sample, report the estimates, and show how well the hazard fits the record and how robust the estimates are.

I.1 Sample and Estimation

The specification follows the two ways in which DPS enforces (Section 2.4). The hazard depends on three pieces of data and on a vector of parameters. The data are the station's fraud rate $\rho_{s\ell t}$, the share of its inspections in the month that the detection model of Section 3 classifies as fraudulent; the rate of covert audits A_{ct} in its county c, the county's covert audits per station in year t; and its type ω_s. The parameters are $\boldsymbol{\vartheta}_r = (\kappa_r, \lambda_r, \psi_A, \psi_S, \gamma_r)$ for the enforcement regime r in force in month ℓ of year t, with the pre-crackdown level κ_{pre} normalized to zero:

$$\begin{aligned} h_{s\ell t} \;&\equiv\; h\big(\rho_{s\ell t},\, A_{ct},\, \omega_s;\, \boldsymbol{\vartheta}_r\big) \;=\; 1 - \exp\big\{-\nu_{s\ell t}\big\}, \\ \nu_{s\ell t} \;&\equiv\; \nu\big(\rho_{s\ell t},\, A_{ct},\, \omega_s;\, \boldsymbol{\vartheta}_r\big) \;=\; e^{\kappa_r + \lambda_r \mathbf{1}\{\omega_s = \text{large}\}} \Big[\, e^{\psi_A}\, A_{ct}\, \rho_{s\ell t} \;+\; e^{\psi_S}\, \rho_{s\ell t}^{\gamma_r} \Big]. \end{aligned} \tag{38}$$

Enforcement arrives at the station at rate $\nu_{s\ell t}$, and the hazard is the probability of at least one arrival in the month. The first term in the brackets is detection through covert audits: an undercover vehicle is one draw from the station's customers, a fraction $\rho_{s\ell t}$ of whom are buying fraud, and the county's audit rate sets how many such draws the station faces. The second is detection through DPS's review of stations' inspection records, which targets stations by their recorded conduct; its intensity is a power of the fraud rate, and the exponent γ_r measures how sharply DPS aims at the worst conduct. Three parameters differ by regime: the level κ_r of both channels, the targeting γ_r, and λ_r, which multiplies the rate for large

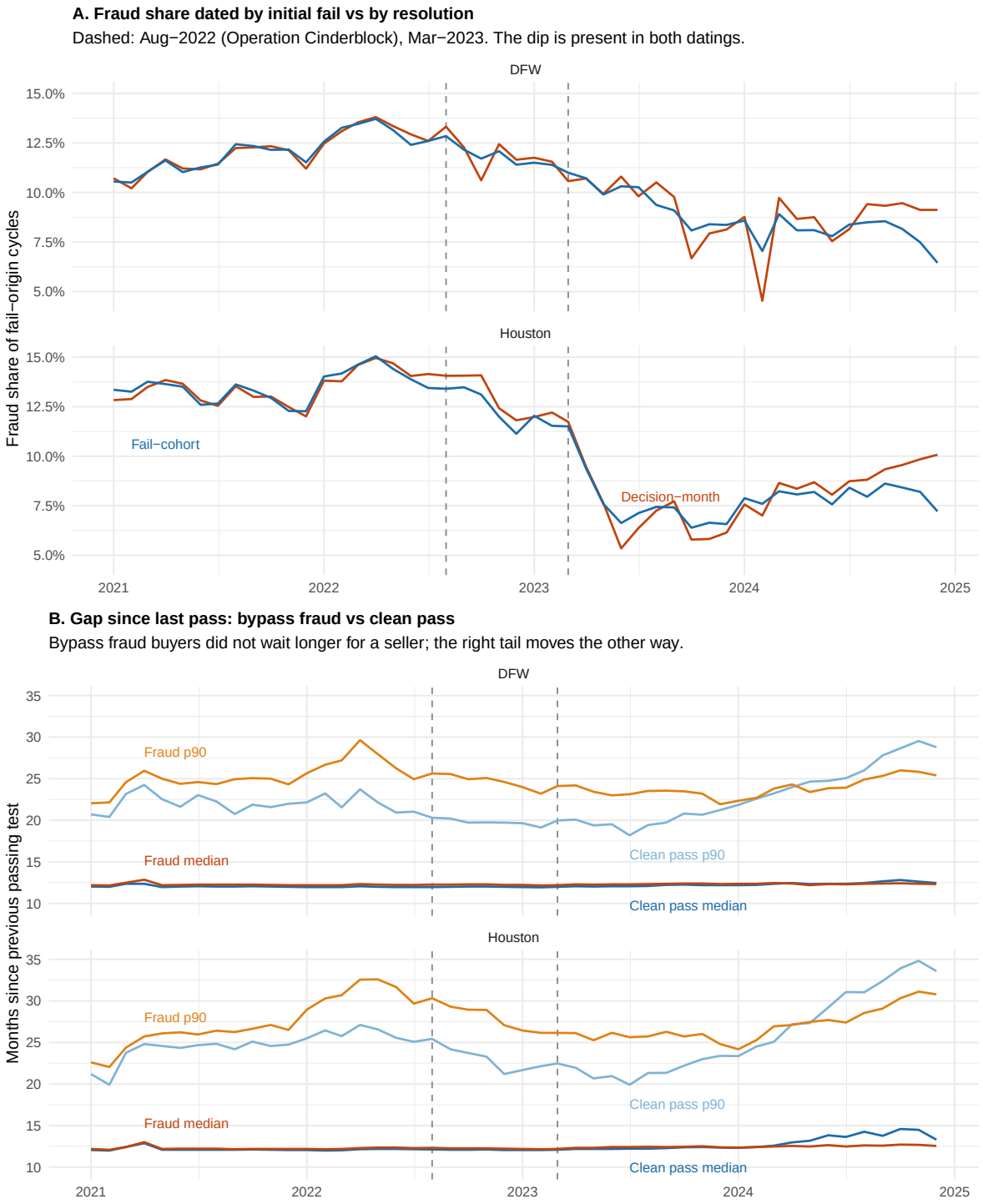


Figure H.2: Re-timing Evidence: Initial-Failure Cohorts and Time Since Last Pass

Note: Left panel: fraud share among drivers who failed an initial inspection, by the month of that failure (cohort basis), against the same share by the month of resolution. Right panel: distribution of months since the vehicle's last passing inspection among buyers of fraudulent passes, by year, against the same distribution for honest passes.

sellers by e^{λ_r}. The channel weights ψ_A and ψ_S are common to both regimes. We estimate $\boldsymbol{\vartheta}_{\text{pre}}$ and $\boldsymbol{\vartheta}_{\text{post}}$ by maximum likelihood on the station-months in which a station has fraud on its record, with standard errors clustered by county and year.

The sample consists of the station-months in which a station could be detected selling fraud. A station enters the risk set in the first month in which the detection model of Section 3 classifies one of its inspections as fraudulent. From then on, it is at risk in each month in which it inspects vehicles, has fraud on its record, and sold fraud that year or the year before.[67] A station has fraud on its record in a month if the detection model classifies at least one of its inspections as fraudulent in that month or in the previous 12 months. Each enforcement action takes the station out of the risk set until it resumes inspections, and for good if it never does. After it resumes, only its conduct since the action counts toward its record. The sample has 150,247 station-months at 4,604 stations in Dallas–Fort Worth and Houston from 2016 to 2024, with 1,356 detections: 273 before the hazard switches regime in October 2022 and 1,083 from then on (Appendix I.4).

We date each detection by the month in which DPS posted the enforcement action. A station may stop inspecting before DPS posts the action against it, so we date an action posted within 2 months of the station's last inspection to that last month. Actions posted later do not count as detections, although they too take the station out of the risk set and reset its record.

DPS often posts an action a month or more after the conduct behind it. In estimation, we therefore replace each term in the brackets of equation (38) by its weighted average over lags of zero to 12 months, counting only the months with fraud on the station's record. The weights follow a discretized gamma distribution whose two parameters we estimate with the rest. They put 47 percent of the weight on the current month and 91 percent on lags of zero to two months. The supply model uses equation (38) itself. Since the weights sum to one, the two forms agree whenever a station's conduct is stable.

The audit rate and the station's type need more detail, and in estimation we add one term to the hazard. The audit rate A_{ct} is the number of covert audits in county c in year t per station licensed on January 1. We normalize it to an average of one across the inspection months of stations after their first inspection classified as fraudulent. We measure a station's type at the start of each year. It counts as a large seller in year t if it sold fraud as one of the large sellers of the demand model in year $t-1$, so that its type does not depend on the year's own sales.[68] Finally, before the crackdown, 38 percent of detections came from a wave of actions against stations in Dallas–Fort Worth in 2016 and 2017. We give the Dallas–Fort Worth station-months of those two years a level term of their own, e^{η}, so that the wave does not shape the estimated pre-crackdown hazard.

[67] A station sells fraud in a year if the detection model attributes at least 12 fraudulent passes to it over the year, and at least 1 percent of its inspections are fraudulent in at least one month. For 2016, we determine whether a station sold fraud in 2015 from the detection model's 2015 classifications. We also require an active license and a county-year covered by DPS's records of covert audits.

[68] For 2016, we use the 2016 large sellers in place of the 2015 ones.

I.2 Estimates

Table I.1 reports the estimates. The crackdown multiplied the level of both channels by $e^{\kappa} = 55$, and it sharpened the records screen's aim at the worst conduct. The exponent γ_r rose from 0.33 to 0.60, so the screen's rate rose by less than e^{κ} at any fraud rate below 100 percent. The estimates also put large sellers at a higher hazard than other stations with the same fraud rate, by a factor of 1.25 before the crackdown and 2.02 after it. Both large-seller shifts are imprecise, and neither is statistically different from zero at the 5 percent level. The gap between the two columns of Panel B is therefore not precisely estimated either.

Panel B summarizes the crackdown as the ratio of the monthly detection probability after it to the probability before it, at a given fraud rate. At a fraud rate of 1 percent, the crackdown raised the hazard 19-fold for other stations and 30-fold for large sellers. The ratio rises with the fraud rate, mainly because the screen became more sharply targeted; the expansion of covert audits, which weigh more at high fraud rates, adds the rest. For a station whose every inspection is fraudulent, the ratio is 70 for other stations and 96 for large sellers.

I.3 Fit

Figure I.1 compares the observed detection rate with the fitted hazard across stations with different fraud rates. We group the station-months by the station's fraud rate in the month and plot, for each group, the share detected against the average fitted hazard. The fitted hazard lies within the 95 percent intervals of most groups in both regimes. The figure omits the groups with no detections (1 in all), for which the hazard predicts 0.4 detections. After the crackdown, the hazard is higher at every fraud rate in the data and rises more steeply in it, as the higher exponent implies.

DPS processes cases in batches, so detections are concentrated in a few months: 62 percent of them fall in the ten busiest months. Figure I.2 therefore compares cumulative counts. Each regime's level is estimated, so the totals agree almost by construction; the figure tests instead the timing of detections within each regime. The predicted count tracks the observed one before the crackdown. After it, the predicted count runs ahead of the observed one until a large batch of actions in the spring of 2023 closes the gap.

The supply model makes one adjustment to this hazard. It scales the post-crackdown level of large sellers' hazard in each market so that the model matches their recorded detections (Section 6.3; Appendix J.1).

I.4 Robustness

The Dallas–Fort Worth wave. The wave of 2016 and 2017 accounts for 38 percent of the detections before the crackdown, so it could drive the pre-crackdown estimates. We therefore drop its station-months and the level term η and estimate the hazard again. The exponents barely move: γ_{pre} is 0.36 and γ_{post} is 0.59. At a 1 percent fraud rate, the crackdown raises the hazard 23-fold for other stations and 30-fold for large sellers. The ratio for large sellers holds only because two changes offset: the ratio for other stations rises, and the pre-crackdown large-seller shift rises to a factor of 1.54.

Table I.1: Estimates of the Monthly Detection Hazard

A. Parameter estimates

	Before the crackdown	After the crackdown
Weight on covert audits, ψ_A	−6.15	
	(0.52)	
Weight on the review of inspection records, ψ_S	−5.97	
	(0.32)	
Level of both channels, κ	0	4.02
		(0.32)
Targeting exponent, γ	0.33	0.60
	(0.08)	(0.06)
Large-seller shift, λ	0.23	0.70
	(0.19)	(0.38)
DFW 2016–17 level, η	1.33	
	(0.28)	
Station-months	150,247	
Stations	4,604	
Detections	1,356	
Log-likelihood	−5,424.1	

B. Ratio of the hazard after the crackdown to the hazard before

	Other stations	Large sellers
Fraud rate of 1%	19	30
	(6)	(13)
Fraud rate of 100%	70	96
	(18)	(28)

Note: Maximum-likelihood estimates of equation (38), with each channel's term averaged over the recording delay (Appendix I.1), on 150,247 station-months with 1,356 detections. The pre-crackdown level is normalized to zero. The level term η applies only to the Dallas–Fort Worth station-months of 2016 and 2017. The table omits the two parameters of the recording-delay distribution. Standard errors, in parentheses, are clustered by county and calendar year (120 clusters) and treat the fraud rates as data. Panel B: the ratio of the fitted monthly detection probability after the crackdown to the probability before it, at the stated fraud rate, for a station outside the 2016–17 wave, with each regime at its own average audit rate. The ratio therefore includes the expansion of covert audits after the crackdown. Probabilities rise less than proportionally with the arrival rate where they are high, so the gap between the columns narrows at high fraud rates. Delta-method standard errors.

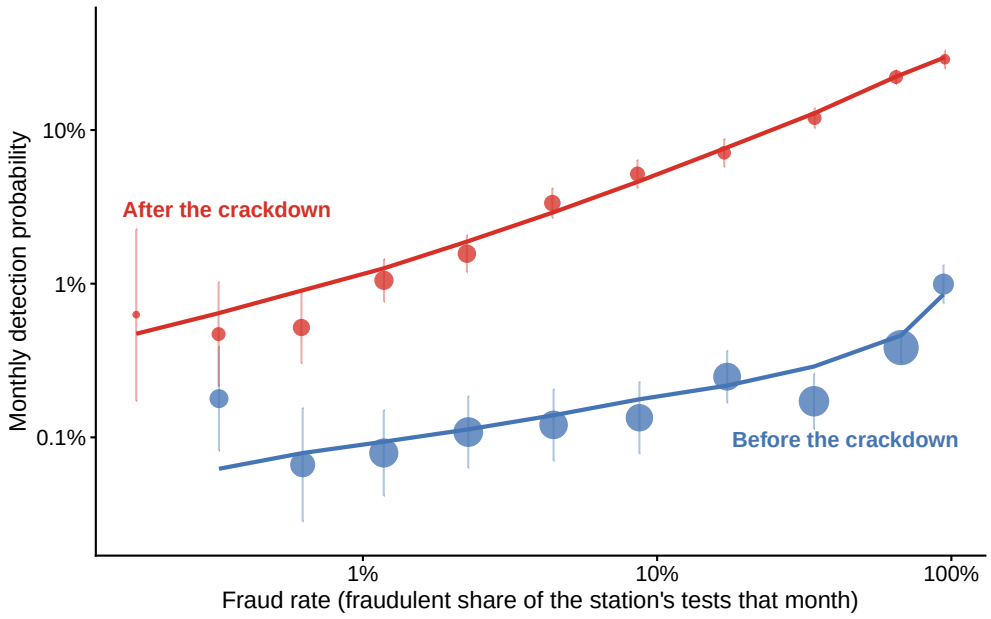


Figure I.1: Observed and Fitted Detection Rates by Fraud Rate

Note: Station-months are grouped by the station's fraud rate in the month; both axes are logarithmic. Markers are each group's observed monthly detection rate, with 95 percent (Wilson) intervals and area proportional to the group's station-months. Lines are the fitted hazard averaged over the same station-months. Before the crackdown: station-months before October 2022; after: from October 2022 on. Groups with no detections are not drawn.

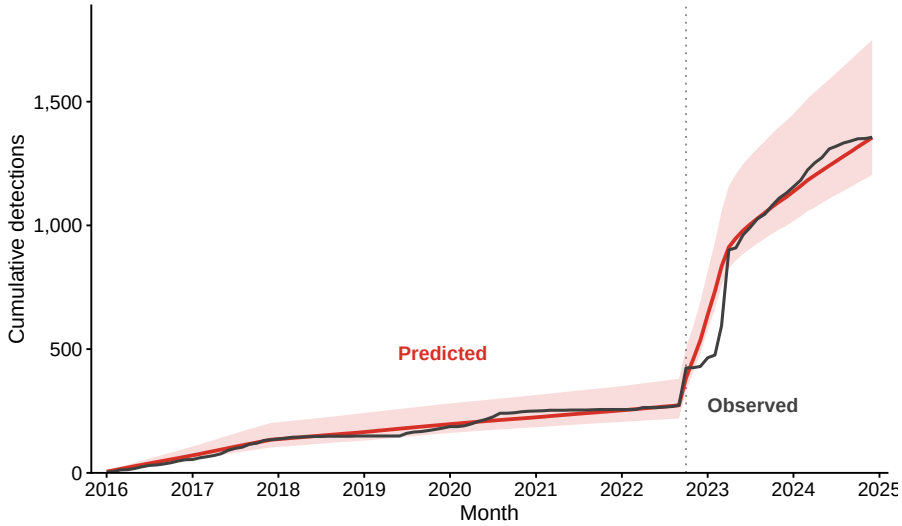


Figure I.2: Cumulative Detections, Observed and Predicted

Note: Cumulative observed detections and the cumulative sum of the fitted monthly hazards over the sample's station-months, 2016–2024. The shaded band is a 95 percent band for the predicted count, from the clustered covariance of the estimates. The dotted line marks October 2022.

The onset month. The crackdown began in August 2022, but DPS posted only 7 enforcement actions against stations in the sample in August and September; the first large batch of actions arrived in October. The hazard therefore takes its post-crackdown form from October 2022. We read this month off the enforcement record, so we also report the fit with the switch at the policy date, August. That fit is worse, by 188 log-likelihood points at the same number of parameters, a comparison that favors October, since we chose that month from the same record. Both exponents are lower, but the crackdown still sharpens the screen's aim, raising γ from 0.24 to 0.49. The large-seller estimates are the least robust. With a switch in August, the post-crackdown large-seller shift falls to a factor of 1.28, and the crackdown raises large sellers' hazard at a 1 percent fraud rate 17-fold rather than 30-fold.

The pre-crackdown exponent. γ_{pre} rests on 273 detections and is the less precise of the two exponents. The rise in the exponent with the crackdown is nonetheless clear: it is 0.27 with a clustered standard error of 0.08. The sharper targeting after the crackdown is therefore not an artifact of the imprecise pre-crackdown estimate.

Appendix J. Details of the Supply Model

This appendix sets out the inputs of the supply model of Sections 6.3 and 7.2, how we solve it, the details of its estimation, its fit year by year, its robustness, and the full welfare ledger of the crackdown.

J.1 Inputs

Demand and quality. The supply model takes demand from Section 6.1: the market-by-period effects, the home-county premium, the nesting parameter, and the qualities of large sellers. In each month, the market is the population of malfunctioning vehicles whose owners make their compliance choice that month. The counterfactuals hold that population fixed while prices and the sellers available change. Demand estimates, once a year, the utility δ_{smt} that each large seller offers drivers, which combines its quality and its price. We separate the two with the quantity condition of equation (40) (Appendix J.2). At the seller's observed sales, the condition gives the price that rationalizes them, and the rest of δ_{smt} is the seller's quality apart from price, θ_{smt}. A large seller that enters draws its quality from the qualities that demand estimates for large sellers in its market, pooled over 2017 to 2024.

Large and fringe sellers. In the data, a station is a large seller in a year if it meets the threshold of Section 6.1, the same rule that makes it a separate alternative for drivers. A station below the threshold is a fringe seller if it sells at least 12 fraudulent passes in the year and is open in at least one month.[69] In the model, types are fixed (Section 5.3). A simulated seller keeps the type of the set of potential sellers it comes from, and we count it

[69] A seller is open in a month if it sells at least one fraudulent pass that month and fraudulent passes are at least 1% of its inspections. "Large sellers open" in Section 7.3 and Table J.3 counts large sellers open in this sense in a typical month, the average over the months of the year.

as open by the same monthly rule. A station is a license, so a business that reopens under a new license enters the model as a new potential seller. A seller is available to drivers in the months it operates.

The sets of potential sellers. The sets of potential sellers of Section 6.3 hold 1,210 potential fringe sellers in DFW and 1,678 in Houston ($\bar{N}_m^f$). Under free entry, the set of potential large sellers matters only where it binds. It does not bind in the estimation windows. In the counterfactual without the crackdown, entry of large sellers in DFW in 2023 comes close to it (Section 7.4).

We form the sets of potential sellers of Section 5.3 by a convention rather than from the licenses we observe, since enforcement shapes which licenses we observe: each set holds two times the largest number of sellers of its type in any year of our sample, and it is the same in every year and every counterfactual.

The detection hazard. The supply model uses the hazard of Section 6.2 at its estimated parameters (Appendix I), evaluated at each seller's fraud rate in each month. A seller's fraud rate is its fraudulent passes as a share of its inspections, with its honest inspections fixed at the 2017–2021 median for its type and market. Its audit rate is that of its market's principal county, Dallas or Harris, in the year. A large seller carries the large-seller term e^{λ_r} of equation (38) in full. Appendix I estimates that term on stations that were large sellers in the previous year; the model applies it to every large seller, new or continuing. The hazard takes its post-crackdown form in October 2022.

We make one adjustment to the hazard, which we call the large-seller adjustment. After the crackdown, the estimated hazard predicts more detections of large sellers than DPS recorded. We scale the post-crackdown level of large sellers' hazard in each market by the ratio of recorded to predicted detections, 0.97 in DFW and 0.61 in Houston. The adjustment adds no free parameter and leaves the fringe's hazard and the pre-crackdown hazard unchanged. It applies wherever the model uses large sellers' post-crackdown hazard, in estimation and in the counterfactuals. Appendix J.5 reports the model without it.

J.2 Solving the Model

The value of a spell. Section 6.3 derives a station's best plan when its monthly profit pq and its hazard $\bar{h}$ are constant: it plans $T^* = \min\{12,\, 1/(\chi\bar{h})\}$ months (equation (12)). At an interior plan, $\bar{h}T^* = 1/\chi$, and the spell is worth pq times

$$A \;=\; \frac{1}{\bar{h}}\left[(1-\chi)\left(1-e^{-1/\chi}\right) + e^{-1/\chi}\right] \tag{39}$$

months of profit. The bracket depends on χ alone, so A is inversely proportional to the hazard: a seller that faces twice the hazard keeps half as many months of profit. Divided by twelve, A is the share of a year's profit without detection risk that the spell keeps, the retained share of Table 4.

Plans and detection. In the model, a large seller's expected profit and hazard change from month to month with demand's monthly number of drivers and with its beliefs about which rivals operate. At the start of the year, once it learns its quality, a large seller chooses its plan, the months it will operate and the quantity it will sell in each. It chooses both to maximize equation (5) over the profits and hazards it expects in each month. A seller that plans the full year starts in January; one that plans a shorter spell starts in a month drawn uniformly over the year. A spell ends at the end of the year at the latest. Whether the station operates the next January is part of the next year's entry decision and plan (Assumption 2), so a spell's value counts only the months of the current year. In each month it operates, DPS detects it with the probability the hazard gives at that month's fraud rate. The same monthly hazard governs its rivals' beliefs about when it stops.

Quantities. A large seller chooses each month's quantity against its beliefs about which rivals operate that month, how much they plan to sell, and the prices they post. Its planned quantity q enters both its expected monthly revenue, $\pi(q) = p\,q$, where p is the price at which it expects that quantity to sell, and its hazard. The value of its spell is $\pi(q)\,A(T, \bar{h}(q))$. Since the seller chooses T optimally, the envelope theorem removes the effect of q through T, and the first-order condition for q is

$$p\left(1 - e_A\, e_h\right) \;=\; -\,q\,\frac{\partial p}{\partial q}, \qquad e_h \;=\; \frac{\partial \ln \bar{h}}{\partial \ln q}, \qquad e_A \;=\; -\,\frac{\partial \ln A}{\partial \ln \bar{h}}, \tag{40}$$

where the right side is the seller's Cournot markup, with its rivals' plans held fixed. At an interior plan, equation (39) gives $e_A = 1$, since the station's months of revenue fall in proportion to its hazard, and the condition becomes

$$p \;+\; q\,\frac{\partial p}{\partial q} \;=\; e_h\,p, \qquad e_h \;=\; \frac{\partial \ln \bar{h}}{\partial \ln q}, \tag{41}$$

where e_h is the elasticity of the station's hazard in its quantity, which the exponent γ_r of equation (38) governs. The seller then acts as a Cournot competitor whose marginal cost is the share e_h of its price. The condition has no parameters of its own: the slope of demand comes from Section 6.1 and the hazard's elasticity from Section 6.2. When the plan runs the full year, e_A is below one. Before the crackdown, the marginal cost of detection would lower large sellers' annual volume by at most 3.9%, and the model sets it to zero in those years. In 2024, it lowers their volume by 2.1% in DFW and 12.0% in Houston.

We use the quantity condition in two ways. In the model, it sets each seller's planned quantities against the demand and rivals the seller expects. In estimation, it recovers the prices we do not observe (Appendix J.1).

Prices, rationing, and the fringe. Each large seller posts one price for the year, the price at which it expects its planned quantities to sell (Section 5.3). Within the year, it sells what drivers demand at that price, up to its planned quantity. A driver whom a seller at capacity turns away chooses again with that seller unavailable. The driver chooses among the other large sellers at their posted prices, the fringe, the stations outside the metropolitan area, repair, and lapse, in the proportions that demand's substitution patterns imply. The

model repeats this until no seller sells more than it planned. The fringe and the stations outside the metropolitan area sell at the going rate (Section 5.3), and the fringe's sales are split equally among the fringe sellers operating that month. Fringe sellers follow the same stopping rule, but at their small scale their planned spells run the full year after the crackdown as well as before it.

Capacity sold. Within the year, the model predicts how much of the capacity stations planned actually sells, because they post prices and plan capacity before they know which rivals will operate (Assumption 3). Before the crackdown, large sellers sold 99.3% of the passes they planned in DFW and 99.7% in Houston, because most of them operated all year and met the rivals they expected.[70] In 2024 they sold 92.9% and 95.0%. A seller whose rivals the regulator detects early turns drivers away, while one that meets more rivals than it expected is left with unsold capacity. We do not observe stations' capacity, so this is a prediction of the model rather than a fit to the data. Table J.1 reports it by market and year.

Computing the equilibrium. We solve each market-year by iterating on stations' beliefs. Given the number of large sellers that enter and the fringe entry probability $\phi_{mt}(\text{fringe})$, each seller's beliefs about which rivals operate in each month follow from the entry probabilities, the distribution of quality, rivals' plans, and the detection hazard. We solve each type's best plan against those beliefs and compute the value of entering for each type. We then move the number of large sellers toward free entry and the fringe entry probability toward equation (15), and repeat until beliefs and choices agree, as in equation (8) (Seim, 2006). Given the equilibrium plans and posted prices, we simulate which sellers operate each month and each seller's sales. Model statistics average 48 simulated paths.

J.3 Estimation

The estimation windows. Each market contributes two windows, 2017–2021 pooled and 2024, because the model has one entry cost per market and so one steady state per regime. Within the pre-crackdown window, DFW's market consolidated. The number of large sellers that sold during the year fell from 107 in 2017 to 90 in 2021, while the value of a spell at the observed number of sellers rose by a factor of 2.15. A static model with one entry cost matches the window's average but not this path.

The entry cost of large sellers. We set $c_{\text{large}}(m)$ equal to the expected value of a spell at the observed number of large sellers, averaged over the years of the pre-crackdown window. We compute this value on the model's own simulated market, drawing rivals' qualities, starts, and detections as the model does. It lies within 6% of the value computed on the observed market.

[70] Capacity sold is realized sales over planned sales, summed over large sellers' months in the year and averaged over simulated paths.

Table J.1: Capacity Planned and Sold by Large Sellers

		Large sellers' volume (000)			Share (%)		
Market	Year	Planned	Demand	Sales	Of planned sold	Rationed	Unsold
Panel A. Model with the crackdown							
DFW	2017	319.8	321.4	316.8	99.1	50.9	49.1
DFW	2018	346.8	348.4	343.8	99.1	59.7	40.3
DFW	2019	411.5	413.4	408.2	99.2	39.9	60.1
DFW	2020	397.0	399.4	394.9	99.5	55.9	44.1
DFW	2021	491.7	494.6	489.9	99.6	58.1	41.9
DFW	2023	277.4	286.9	256.8	92.6	55.6	44.4
DFW	2024	288.2	295.3	267.9	92.9	55.1	44.9
Houston	2017	344.4	345.2	343.4	99.7	67.1	32.9
Houston	2018	335.1	336.0	333.8	99.6	42.0	58.0
Houston	2019	362.2	363.0	361.1	99.7	57.6	42.4
Houston	2020	316.9	317.7	315.8	99.6	57.8	42.2
Houston	2021	352.5	353.3	351.1	99.6	60.6	39.4
Houston	2023	0.7	0.7	0.6	98.4	50.5	49.5
Houston	2024	193.5	194.3	183.8	95.0	51.7	48.3
Panel B. No-crackdown counterfactual							
DFW	2023	824.5	825.6	803.2	97.4	51.3	48.7
DFW	2024	771.6	773.0	761.8	98.7	44.5	55.5
Houston	2023	362.7	363.3	357.2	98.5	47.3	52.7
Houston	2024	545.8	546.5	539.8	98.9	44.0	56.0

Note: By market and year, large sellers' planned sales, the demand for their passes at their posted prices, and their sales, each summed over the year and averaged over simulated paths. The table also reports the share of planned sales that sold and the shares of large sellers' months in which the seller sold its planned quantity and turned drivers away (rationed) or sold less (unsold). Panel A is the estimated model; Panel B is the counterfactual without the crackdown (Section 7.4), whose pre-crackdown years coincide with Panel A and are not repeated. Sales fall short of plans when more rivals operate than a seller expected, and demand exceeds them when fewer do.

The entry likelihood. Let $\bar{N}_m^\omega$ be the number of potential sellers of type ω in market m, Y_τ the number of years in window τ (five before the crackdown, one in 2024), and $n_{m\tau}^\omega$ the number of sellers of type ω that entered market m over those years. With the entry probability $\phi_{m\tau}(\omega)$ of equation (15), the entry costs and scales maximize

$$\mathcal{L}(c,\sigma) \;=\; \prod_{m} \prod_{\tau\in\{\text{pre},\,2024\}} \prod_{\omega} \binom{Y_\tau \bar{N}_m^\omega}{n_{m\tau}^\omega} \phi_{m\tau}(\omega)^{\,n_{m\tau}^\omega} \left[1-\phi_{m\tau}(\omega)\right]^{\,Y_\tau \bar{N}_m^\omega - n_{m\tau}^\omega}, \tag{42}$$

in which the expected payoff of entering, $\bar{\Pi}_\omega(m,\tau)$, is the window's average at the equilibrium beliefs. In the main specification, the scale of large sellers' payoff shocks is zero. Their factor then reduces to the zero-profit condition of Assumption 6, and the likelihood is maximized over the fringe parameters alone.

The full model. With a scale parameter for each type, the entry stage has six parameters, the four entry costs and the two scale parameters, which we call the full model (Table J.2). It places the scale of large sellers' payoff shocks near zero, at 3.0% of their entry cost, and a scale twice as large lowers the log-likelihood by 29.0. Our main specification therefore sets it to zero. Setting the scale to zero changes the entry costs by 7.4% in DFW and 4.1% in Houston (Tables 4 and J.2).

Table J.2: Estimates of the Full Entry Model

	DFW	Houston
Entry cost of a large seller, $c_{\text{large}}(m)$ (\$ thousands)	375.7	205.0
	(1.1)	(0.9)
Scale of large sellers' payoff shocks, σ_{large} (\$ thousands)	11.4	
	(1.0)	
Entry cost of a fringe seller, $c_{\text{fringe}}(m)$ (\$ thousands)	20.0	22.0
	(0.8)	(0.6)
Scale of fringe sellers' payoff shocks, σ_f (\$ thousands)	15.4	
	(0.9)	
Log-likelihood, large-seller and fringe blocks	−13,625.5	

Note: Maximum likelihood estimates of the entry stage with a free scale for the payoff shocks of each type (the full model): the likelihood of equation (42) maximized over the four entry costs and the two scale parameters, given $\hat{\chi}$. Each scale parameter is common to the two markets. Standard errors are in parentheses. The main specification sets the scale of large sellers' payoff shocks to zero (Table 4).

The penalty rate. A spell is a run of consecutive months in which a station is open as a large seller. It ends in detection if DPS detects the station during it, with detections dated as in Appendix I.1, and voluntarily otherwise. We treat a spell still open within 3 months of the end of our data as censored, and we estimate the caught share by the cumulative incidence of detection, with voluntary stops as a competing risk. The model's side of the two moments takes the large sellers that entered each market in 2023. Each draws a quality from its market's distribution and chooses its plan at χ in 2023's market, with the sellers that operated in it, and we classify its simulated spell by the same rule. Under a constant hazard, the share of planned spells that end in detection is exactly $1 - e^{-1/\chi}$ (equation (13)). In the model, spells start throughout the year, so the simulated share differs from that closed form, and we match the simulated share. The weighting matrix of equation (14) is diagonal in the inverses of the two moments' variances rather than optimal, so we read the test of the overidentifying restriction as a specification diagnostic rather than as an exact chi-squared test.

Before the crackdown, large sellers operated for the full year, and a spell that runs to the end of the year carries no information on χ.

Inference. Standard errors condition on the demand and hazard estimates. The standard errors of the fringe's entry cost and scale are clustered by station. The standard error of χ is the GMM standard error under the diagonal weighting matrix, with spells as independent

observations. That of $c_{\text{large}}(m)$ follows from it by the delta method. The crackdown's effects in Table 5 carry simulation standard errors from paths paired across the two scenarios.

Prices. We do not observe prices. For each large seller and year, the price we recover from the observed market is the one at which the quantity condition rationalizes its observed sales (Section 6.3). The fringe sells at the going rate of Section 5.3, and its quality follows in the same way. Table J.3 compares the model's price at large sellers with this recovered price by market and year. In the estimation windows, the two differ by at most 25%, in Houston in 2024. In Houston, the model reproduces only part of the rise in the recovered price from the years before the crackdown to 2024.

Houston is the weaker fit. The model's large sellers lose more of the Houston market than the data's, and its 2024 price at large sellers sits 25% below the price we recover from the observed market, against 9% below in DFW.

J.4 The Model and the Data by Year

Our estimates use the steady states before the crackdown and in 2024 and leave 2022 and 2023 out of the fit. The model has no path between the two regimes. Its 2022 is the pre-crackdown equilibrium at 2022's demand. Stations make their entry decisions and plans in January under the old hazard, and the hazard switches in October 2022. Its 2023 is the post-crackdown equilibrium at 2023's demand. Table J.3 compares both years with the data, together with the years of the estimation windows.

The model misses both years, in opposite directions. In 2022, free entry at the year's demand brings in far more large sellers than operated: 223.8 open in a typical month in DFW, against 73.8 in the data. In 2023 the market was still adjusting. In DFW, 28 large sellers were open in January 2023 and 5 in December. 17 spells began in January 2023, against 7 in January 2024.[71] The model's 2023 equilibrium has 11.2 large sellers open in DFW and 0.2 in Houston, against 16.9 and 20.1 in the data.

The model's prediction for 2024 also overstates the number of large sellers, which plays no part in the entry cost (Section 6.3). It has 14.2 open in a typical month in DFW, against 5 in December 2023 and 4.8 in a typical month of 2024 (Section 7.3). Two features of the 2024 record bear on this. First, recorded enforcement fell in the second half of the year. In DFW, detections at all stations fell from 155 in the first half to 53 in the second. Yet the number of large sellers open hardly changed, from 5.3 in a typical month of the first half to 4.3 in the second.[72] This is what decisions made in January imply. Stations planned 2024 on the hazard they expected in January, so the model holds the 2024 hazard at its post-crackdown level rather than following the fall in recorded enforcement. Second, one long-lived seller in each market sold a large share of 2024's fraud (Section 7.3), over a spell far longer than the model's planned spell. The model cannot produce such a seller. It matches large sellers' combined share of the 2024 market with several times as many sellers, each selling less (Table J.3).

[71] In Houston, 35 large sellers were open in January 2023 and 5 in December. 22 large sellers' spells began in January 2023, and 4 in January 2024.

[72] In Houston, detections fell from 172 to 48, while large sellers open went from 8.3 to 9.2.

Table J.3: The Model and the Data by Year

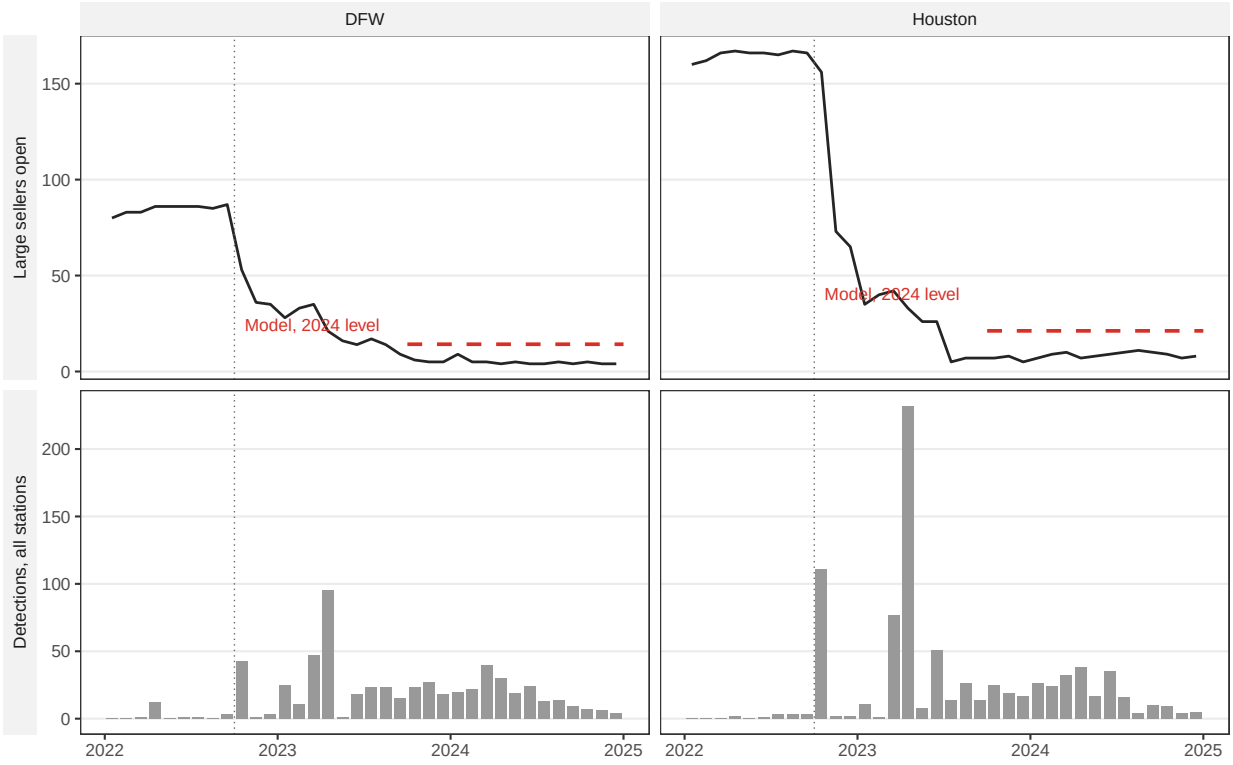
	Large sellers open		Large sellers' share (%)		Price at large sellers ($)		Fraud volume inside the metro (000)		Purchases outside the metro (000)	
Year	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
Panel A. DFW										
2017	89.0	95.2	37.6	43.2	89	93	351	374	9	8
2018	99.4	107.5	42.2	44.4	92	93	390	403	11	10
2019	101.9	111.7	46.1	48.2	94	94	459	477	11	10
2020	77.5	88.8	46.4	44.2	98	96	478	483	10	10
2021	74.7	87.2	51.3	49.0	103	99	559	564	14	14
2022†	73.8	223.8	52.2	66.9	101	100	627	753	20	10
2023†	16.9	11.2	42.3	24.3	120	136	500	326	9	26
2024	4.8	14.2	29.1	28.1	150	136	300	292	100	109
Panel B. Houston										
2017	146.1	150.4	36.8	37.6	88	88	486	491	8	8
2018	141.8	145.0	36.1	36.1	89	89	493	494	9	9
2019	148.3	153.4	36.3	37.3	89	90	513	517	12	12
2020	141.0	147.7	34.1	34.1	90	91	441	454	15	14
2021	149.4	153.0	35.6	34.5	91	92	476	494	41	38
2022†	148.2	228.8	35.1	39.9	91	96	516	560	60	50
2023†	20.1	0.2	11.7	0.1	98	119	216	81	103	167
2024	8.8	21.2	25.7	21.3	145	110	255	226	57	66

Note: By market and year: the number of large sellers open in a typical month, large sellers' share of the market (Section 7.3), the price at large sellers, fraud volume inside the metro (at large sellers and the fringe), and the fraudulent passes the metro's drivers buy at stations outside it, each in the data and in the model. Fraud volume in Section 7.4 is the sum of the two. We do not observe prices. The data's price is the one at which the quantity condition of equation (40) rationalizes large sellers' observed sales. Years marked † lie outside the estimation windows. Before the crackdown, the model's number of large sellers in each year is the observed number; free entry sets it from 2022 on. The model's 2022 is the pre-crackdown equilibrium at 2022's demand, with the hazard switching in October 2022; its 2023 is the post-crackdown equilibrium at 2023's demand.

The record of 2024 supports treating it as a steady state (Assumption 5). The number of large sellers open and the concentration of the market did not change as recorded enforcement fell, and the number open had stopped falling by the autumn of 2023. The two steady states share the penalty rate, the entry costs, and the scale parameters (Section 5.4), so the crackdown changes supply only through the detection hazard, and the counterfactual of Section 7.4 removes that change. If the crackdown also raised stations' costs, for example the cost of finding buyers, the model attributes that rise to detection risk.

How Figure 3 is built. Panel (a) of Figure 3 shows the demand model's fitted shares for every year of its 2016–2024 sample. The supply panels show the supply model's equilibrium in each year from 2017 at that year's estimated demand, solved year by year rather than simulated forward from a starting point; the model is estimated on the five years before the crackdown and on 2024. The two sides differ in what the figure tests. Demand's metro-by-period effects on every alternative pin down the shares in panel (a) closely by construction,

Figure J.1: Large Sellers Open and Detections, 2022–2024



Note: By market, the number of large sellers open in each month from 2022 to 2024 and the model's post-crackdown level. The second series counts detections at all stations in each month, dated as in Appendix I.1.

so that panel documents the reallocation the demand model carries into the counterfactuals rather than testing it; the tests of demand are the alternative specifications in Appendix H.5. The supply estimator uses the lengths of selling spells and the numbers of sellers of each type, not shares or volumes, so the supply panels are predictions rather than targets. Prices of fraudulent passes are not observed, so the figure does not draw them; Appendix J.3 compares the model's price with the price the observed shares imply, and Table J.3 reports the supply model's outcomes by year, including 2022 and 2023, which the model does not target. The model also reproduces large sellers' share of the market, which the body's figure omits: in DFW, large sellers' share fell from 45.2% before the crackdown to 29.1% in 2024, and the model predicts a fall from 46.0% to 28.1%; in Houston, the share fell from 35.8% to 25.7%, against 35.9% to 21.3% in the model.

J.5 Robustness

Table J.4 reports the model under each change below. It gives the 2024 share of the market held by large sellers in Houston, the 2024 price at large sellers in DFW, and the crackdown's effect on total fraud volume in 2024 in each market.

Table J.4: Robustness of the Supply Estimates and the Counterfactual

	Houston large-seller share, 2024 (%)	DFW large-seller price, 2024 ($)	Crackdown effect on total fraud (%)	
			DFW	Houston
Data	25.7	150	–	–
Main: large-seller adjustment to the hazard, $\chi = 1.188$	21.3	136	-48	-48
Hazard without the large-seller adjustment	4.9	136	-49	-60

Note: Each row re-estimates or re-solves the model under one change. The first row is the data and the second the main specification.

The hazard without the large-seller adjustment. The adjustment treats the gap between predicted and recorded detections of large sellers as an error in measuring their risk. If stations instead planned on the risk that the estimated hazard implies, the main specification overstates the value of a spell in Houston, where the gap is largest. It then also overstates the number of large sellers the market supports. We therefore re-solve the model with the estimated hazard in both markets, at the same penalty rate and entry parameters. Large sellers' share of Houston's market in 2024 then falls from 21.3% to 4.9%, against 25.7% in the data. The crackdown's effect on Houston's fraud volume grows from a fall of 48% to a fall of 60%.

The penalty rate from the caught share alone. The two moments of equation (14) pull in different directions. The caught share alone implies $\chi = 0.94$, the months open alone $\chi = 2.05$, and the test of the overidentifying restriction rejects a single rate ($J = 20.7$; Appendix J.3). The months-open moment is the more exposed of the two: a station that stops because it learns that DPS is investigating it appears as a voluntary stop, which shortens the months of undetected stations and pushes that moment toward a higher penalty rate. The stopping rule compresses two facts, how often DPS catches a station and how long the others stay open, into one rate, and the estimate is the compromise between them.

The crackdown's effect by equilibrium margin. Table J.5 reports the decomposition of Section 7.4, which adds the equilibrium margins one at a time from the no-crackdown market to the model with the crackdown: step 0 removes the large sellers the crackdown drove out and sends every displaced buyer to repair or lapse, step 1 lets the displaced buyers choose among the remaining options at unchanged plans and prices, step 2 lets the surviving sellers re-plan under the crackdown's hazard with entry frozen, and step 3 adds free entry and is the estimated model with the crackdown.

The enforcement levers. Table J.6 reports the 2024 market when we halve or double each lever of enforcement after the crackdown, the detection hazard and the penalty rate (Section 7.4).

J.6 The Welfare Ledger

Table J.7 reports the full welfare ledger behind Panel C of Table 5, which pools the two metros. By metro, the crackdown's net is a loss in DFW and a gain in Houston, and the break-even damage is lower in Houston. We also price the avoided damages under two other

Table J.5: The Crackdown's Effect by Equilibrium Margin

	DFW					Houston				
	Fraud (000)	Change (%)	Large share (%)	Large open	Price ($)	Fraud (000)	Change (%)	Large share (%)	Large open	Price ($)
Fraud volume, 2024										
No-crackdown market	771	–	99	228	110	557	–	97	250	94
0. Removal, no substitution	212	−73 (0.6)	95	79	109	281	−49 (0.4)	94	136	93
1. Removal with demand	401	−48 (0.4)	51	79	109	369	−34 (0.3)	73	136	93
2. + Re-planning (entry held)	541	−30 (0.3)	89	54	123	405	−27 (0.3)	88	99	114
3. + Entry (the model)	401	−48 (0.4)	67	14	136	292	−48 (0.4)	63	21	110

Note: Each row is a market in 2024, by metro: total fraud volume (every fraudulent pass the metro's drivers buy, including those bought outside the metro), its change from the no-crackdown market with simulation standard errors in parentheses, large sellers' share of that volume and the number open in the average month, and the volume-weighted price at large sellers. The no-crackdown market keeps the pre-crackdown detection hazard; the paragraph above defines steps 0 to 3, and step 3 is the estimated model with the crackdown. All statistics average 48 simulated paths, paired across rows by path.

Table J.6: The Crackdown's Effect Under Halved and Doubled Enforcement Levers, 2024

	DFW				Houston			
	Fraud (000)	Change vs no crackdown (%)	Large open	Price ($)	Fraud (000)	Change vs no crackdown (%)	Large open	Price ($)
Panel A. Intensity: every seller's detection hazard after the crackdown, scaled								
$\times 0.5$	522	−32 (0.5)	38	124	401	−28 (0.4)	79	112
$\times 1$ (the model)	401	−48 (0.4)	14	136	292	−48 (0.4)	21	110
$\times 2$	265	−66 (0.3)	0	154	190	−66 (0.3)	1	118
Panel B. Severity: the penalty rate χ after the crackdown, scaled								
$\times 0.5$	484	−37 (0.5)	30	127	367	−34 (0.4)	56	110
$\times 1$ (the model)	401	−48 (0.4)	14	136	292	−48 (0.4)	21	110
$\times 2$	286	−63 (0.3)	2	154	208	−63 (0.3)	4	116

Note: Each panel scales one lever of the estimated model with the crackdown by one half, one, or two, after the crackdown: every seller's detection hazard (Panel A) or the penalty rate (Panel B). The change is against the market without the crackdown, on fraud volume, every fraudulent pass a metro's drivers buy at large sellers, at the fringe, or at stations outside the metro, with simulation standard errors in parentheses. Large sellers open is the number in the average month, and the price is the volume-weighted price at large sellers.

conventions: unregistered vehicles at half the mileage, and no adjustment for the worse defects of fraudulently passed vehicles (Appendix D.3). Each raises the avoided damages and the net. The memo rows price the avoided damages at the fraud the supply model sells rather than at the fraud drivers would buy at its prices, and on that basis the net is a loss in both metros.

Table J.7: The Crackdown's Welfare Ledger, 2024

	DFW	Houston	Both
Driver surplus	-271.5	-135.9	-407.4
Repair-sector profit	81.9	48.9	130.8
Fraud-station rents	1.1	1.8	2.9
Avoided damages, unregistered at 100% VMT	182.0	96.0	278.0
Avoided damages, unregistered at 50% VMT	190.3	101.9	292.2
Avoided damages, no severity adjustment (100% VMT)	216.9	114.2	331.1
Net (unregistered at 100% VMT)	-6.4	10.8	4.4
Net (unregistered at 50% VMT)	1.8	16.7	18.5
Net, no severity adjustment (100% VMT)	28.4	29.0	57.4
Net excluding fraud rents (100% VMT)	-7.6	9.0	1.4
Net excluding fraud rents (50% VMT)	0.7	14.9	15.6
Break-even damage per fraud-passed vehicle-year ($)	264	199	239
Memo: avoided damages and net at the supply model's sales (100% VMT)			
Avoided damages	82.3	47.8	130.1
Net	-106.2	-37.4	-143.6

Note: Each row is the crackdown's effect in 2024 in millions of dollars, the market with the crackdown less the market without it, except the break-even damage, at which the net at 100% VMT would be zero. Every row holds the population of drivers fixed at its size in the market with the crackdown, and the avoided damages follow the choices the demand model predicts for those drivers. VMT is vehicle miles traveled: at 100% an unregistered vehicle is assumed to be driven as much as a registered one, and at 50% half as much. The severity adjustment lowers the damages for the worse defects of fraudulently passed vehicles (Appendix D.3). The memo rows price the avoided damages at the fraud the supply model sells rather than at the fraud drivers would buy at its prices.