Hard to Measure Well: Can Feasible Policies Reduce Methane Emissions?

Karl Dunkle Werner and Wenfeng Qiu

(Karl's job market paper)

November 21, 2020

Click here for most recent version

Abstract

Oil and gas wells emit large quantities of methane, a greenhouse gas 34 times more potent than carbon dioxide. Methane emissions are rarely priced and lightly regulated—in part because they are hard to measure—leading to a large climate externality. However, measurement technology is improving, with remote sensing and other techniques opening the door for policy innovation. We present a theoretical model of emissions abatement at the well level and a range of feasible policy options, then use data constructed from cross-sectional scientific studies to estimate abatement costs. We simulate audit policies under realistic constraints, varying the information the regulator uses in choosing wells to audit. These policies become more effective when they can target on well covariates, detect leaks remotely, and charge higher fees for leaks. We estimate a policy that audits 1% of wells with uniform probability achieves less than 1% of the gains of the infeasible first best. Using the same number of audits targeted on remotely sensed emissions data achieves gains of 30–60% of the first best. These results demonstrate that because leaks are rare events, targeting is essential for achieving welfare gains and emissions reductions. Auditing a small fraction of wells can have a large impact when properly targeted.

JEL: C54, H23, K32, Q55, Q58

Corresponding author: Karl Dunkle Werner <karldw@berkeley.edu>

Author affiliations: UC Berkeley Department of Agricultural and Resource Economics (ARE)

We are deeply grateful for the insightful advice and feedback we've received from faculty and students in the Energy Institute at Berkeley, Daniel Cusworth, Kelsey Foster, Christian Frankenberg, Catie Hausman, Larry Karp, Ryan Kellogg, Susan Powell, Evan Sherwin, Andrew Thorpe, Hannah Yung, and our wonderful PhD cohort. We also appreciate great research assistance from Shuyi Deng. All errors are, of course, our own. Karl is grateful for the generous support of the Alfred P. Sloan Foundation Pre-doctoral Fellowship on Energy Economics, awarded through the NBER. Revision: ba740f9.

1 Introduction

Oil and gas wells emit large quantities of methane, a powerful greenhouse gas with the second largest impact after carbon dioxide. Methane accounts for roughly one tenth of total greenhouse gas (GHG) emissions, though its contribution is measured much less precisely than carbon dioxide's. Fossil fuels, particularly oil and natural gas, are the largest human-driven sources of methane (EPA 2020b; Alvarez et al. 2018). As fracking has dramatically increased US oil and natural gas production, methane emissions have followed, and these emissions are now roughly the same size as the emissions from all fuel used in the western US electricity grid (EPA 2020a). Natural gas has been heralded as a cleaner substitute for coal and a bridge fuel in the transition to a low carbon economy. However, if these methane leaks are large enough, natural gas may emit more GHG than coal.[1] Beyond debates over coal and natural gas, these leaks increase both the lifecycle GHG emissions of gasoline and the relative value of renewable energy.

Measuring methane is costly—it's infeasible to put a continuous emissions monitor on every well—so pricing emissions is challenging. The standard economic prescription in this case would be to audit infrequently and charge a high fine, so that the expected penalty is still equal to the social cost (plus enforcement costs). This approach has theoretical appeal, but is infeasible because of legal and logistical constraints. The constraints on fees range from the backstop of bankruptcy, to legal doctrine limiting punitive damages and to political pushback (Boomhower 2019; *Exxon Shipping Co. v. Baker* 2008). Currently, no US jurisdiction charges a price for methane emissions (Rabe, Kaliban, and Englehart 2020).

This paper combines an economic model with empirical estimates in order to quantify the potential gains from *feasible* audit policies and to demonstrate the value of remotely sensed data that could improve audit targeting. We account for real-world constraints on policies that can be enacted and the information available to the regulator. These constraints take the form of limits on the fees that can be charged, the regulator's capacity to conduct audits, and the fidelity and detection threshold of the remotely sensed measurements. Policies under these constraints offer some improvement over the benchmark of no policy, but the gains vary dramatically, depending on the fee the regulator can charge and the remote sensing information available.

We note that imperfect measurement is not isolated to methane, or even environmental economics. Enforcing any policy requires measurement. The quality and cost of these measurements determines which policies are feasible. In recent decades, remote sensing, administrative data, and other indirect information have improved dramatically, raising

1. The lifecycle GHG emissions of natural gas may be lower than coal as long as the total leakage rate is below 5–10% (Hausfather 2015). We focus on upstream leakage from wells, where 1–4% of gas leaks out. Emissions from pipelines and end users also contribute significantly, and further quantifying all of these remains an active field of research.

the possibility that policy can be based on or informed by these measures. At the same time, and despite a great deal of excitement about remote sensing, policies that make direct use of these tools are rare. Our work highlights one case where they could be applied, while acknowledging the measurements' limitations.

To start our analysis, we develop a theoretical model of abatement and welfare. Using the model, we consider how well operators would change their behavior in response to a feasible but imperfect audit policy—one where the expected fee for emitting differs from the social cost, and may be zero for some wells because of measurement or auditing limitations.

In our model, and consistent with the scientific literature, large leaks are the result of stochastic process failures (Lyon et al. 2016; Zavala-Araiza et al. 2017). These leaks are rare and hard to predict, but large sources of GHG. Therefore we assume well operators abate by reducing the probability a well is leaking at any given moment, rather than reducing leak size. Our stylized model yields closed-form solutions for welfare and abatement as functions of the leak size distribution and the well operators' cost parameters. We parameterize the model flexibly, using scientifically measured data on well-level leaks. To construct the distribution of emissions, we combine several datasets from different scientific teams. We match these leakage measures to specific well pads, and we estimate the fraction that have detectable emissions.[2] Our main dataset uses emissions estimates collected by airplanes flying over approximately 15,000 well pads in California, New Mexico, and Colorado. We use the variation in leak size and presence to infer both the distribution of sizes when leaks occur, as well as the wells' costs of preventing those leaks.

In addition to being a greenhouse gas, methane is also the primary component of natural gas. To leak methane into the atmosphere is to lose the commodity value of the gas, which provides a private incentive to abate. However, because the commodity price is less than one tenth the value of social damages from leaking, well operators don't face a strong enough incentive to abate to the socially optimal level. We use this private incentive to learn as much as possible in the absence of policy. We build our model assuming wells are avoiding leaks optimally, given this weak private incentive, then consider how behavior would change if the well operator faced some expected fee for emissions.

When we discuss audits, we consider on-the-ground measurements. When a well is audited, we assume the regulator has to drive to the well and take downwind measurements of the well's emissions using standard methods approved by the Environmental Protection Agency (EPA). These on-the-ground measurements may be necessary, even when leaks can be measured remotely, because of noise in the remote sensing or regulatory constraints.

2. A well pad is a group of one or more closely spaced wells, typically within a few yards of one another.

To think about a range of audit policies, we compare five cases:

(0) no audits, the status quo,

(1) audit every well with equal probability,

(2) target audits based on well covariates,

(3) measure leaks remotely and target audits, and

(4) measure leaks remotely and assess fines based on measurements.

Assessing fines on the basis of remote measurements (policy 4) is infeasible in current legal structures, but provides an interesting point of comparison.

These are relatively simple policies, and they set aside the potential gains from dynamic enforcement or sophisticated mechanism design (cf. Blundell, Gowrisankaran, and Langer 2020; Cicala, Hémous, and Olsen 2019; Oestreich 2017). We use these policies as tools to think about the space of potential pricing options, and how that space is changed with the availability of remotely sensed measurements.

Comparing each of the audit policies (1–3) with the status quo (policy 0) allows us to think about the gains available from an audit-type policy. Comparing policy 3, which uses remote sensing, with policies 1 and 2, which don't, provides information about the scope for policy innovation with these new tools. Charging fees based on remote sensing alone (policy 4) provides an infeasible benchmark, and can achieve the first-best under additional assumptions. Each of these policies implies some expected fee the well operator will pay. In our model, we will consider the deadweight loss that arises from the regulator not being able to set the expected fee to the social cost of emissions.

In the policies that used remotely sensed data, and in any policy that depends on measurement, the effectiveness of the policy depends on the quality of the measurement. For our context, the detection threshold is an important concern—with a high threshold, only the largest leaks will be detected remotely. In our analysis, we assume these measurements are available from methane-observing satellites, using realistic values of their detection capacity (Cusworth et al. 2019).

We estimate that audit policies yield gains over the status quo of no fees. These gains vary dramatically with the fee amount and the specifics of the policy. The different policies we consider translate into different expected fees for emitting. For instance, the uniform audit policy has the same expected fee for every well, and that fee increases as the audit probability or allowed penalty increases. For a mid-level penalty and 1% audit budget, the average of the expected fee for emitting is \$0.147 per ton CO_2e, which leads to improving deadweight loss (DWL) by 3.37 percent of the way from the no-fee outcome to the Pigouvian first best. In this scenario, average emissions would fall by 34.5 tons CO_2e per well per year. With the same audit budget and penalty, we can also consider a policy that targets on well covariates, rather than auditing with uniform probability. In this case, the average fee is \$0.147 (mechanically the same as the uniform policy), but now heterogeneous across wells, with an interdecile range of \$0–0.507. Now the improvement in DWL is 5.8 percent, and the average fall in emissions is 59.8 tons.

We can also consider targeting on observed leaks. As we mentioned, in this case the regulator can prioritize wells that were observed leaking, saving some audit effort of auditing wells when they're not leaking. For the same audit probability and allowed penalty, there's a much higher expected fee: \$8.54 per ton CO_2e, with an interdecile range of \$0–14.7 across wells.

This stronger incentive leads the well operators to abate more, leading to a DWL improvement of 62.8 percent, and average emissions declines of 692 tons. We also consider other policy options, including different penalties, different fixed audit probabilities, audit probabilities that depend on the cost of auditing, and other policy benchmarks.

These results highlight the importance of both measurement and regulatory constraints. If there were no limits on the size of fees for leaks, a sufficiently high fine could be employed to induce efficient abatement without targeted audits. Given realistic constraints on fees amounts and the rarity of leaks, untargeted audits produce very small welfare gains compared to audits that are targeted based on remotely sensed information. If the allowed fees are severely constrained, even targeted audits yield small gains.

Though the estimates in our paper focus on methane emissions, we view our results as a contribution to several broader literatures. First, and most directly, we contribute to the discussion of designing and evaluating policy with imperfect measurement. Second, we contribute to the innovation literature, considering how technological progress in measurement allows for policy innovation. Third, we compare our results with the small literature on methane abatement.

The challenges of imperfect measurement and imperfect targeting are first-order important to many areas of economics. These challenges arise in environmental questions, like regulating non-point pollution, as well as other fields such as tax evasion, teacher value-added and most principal–agent problems (Segerson 1988; Allingham and Sandmo 1972; Chetty, Friedman, and Rockoff 2014). All of these areas, if the regulator could accurately observe individual actions, they would be able to achieve their goals much more directly. However, lack of accurate, individual measurement leads to more complicated policies that draw inferences from indirect evidence. Our research highlights the value of one type of indirect measurement: remote sensing measures that guide on-the-ground audits.[3] We compare the policies that are achievable with and without these additional data. These additional measurements are a form of innovation, enabling policies that were not previously feasible. Nagaraj (2020) considers a private-sector case, where satellite imagery enabled entry by small firms and changed the structure of the market.

3. Alix-Garcia and Millimet (2020) provides a relevant guide on real-world challenges of satellite data, particularly for measuring binary outcomes.

Finally, we're contributing to a relatively small literature on policies to address methane emissions. Ravikumar et al. (2020) is the only study we're aware of to estimate the observed change in methane emissions from a change in policy. The authors performed repeated surveys of a small number of facilities before and after a leak detection and repair (LDAR) program began, and estimate a 44% reduction in emissions.

Two recent working papers by Levi Marks provide a point of comparison for pricing methane emissions. Marks (2018) estimates the elasticity with respect to the commodity price of natural gas of methane emissions reported in the EPA's greenhouse gas reporting program (GHGRP). To get around the lack of existing policy, that paper makes the same argument we do, noting that the private incentive to sell natural gas can proxy for a tax on methane emissions. That paper, unlike ours, uses variation in the price of natural gas to identify the change in reported quantity emitted. (We use the variation in leak sizes and occurrence; see details in section 5.)

Because natural gas prices move in a limited range, Marks (2018) considers a \$5 per ton CO_2e tax on emissions to avoid extrapolating too far from the data. The paper estimates that this tax would result in an emissions reduction of 56%. To estimate a comparable policy, we will consider \$5 as the low-end fee a regulator may charge. We find sharply different results for our \$5 fee, largely because low audit probabilities lead to an expected fee much lower than \$5. When we consider *expected* fees of approximately \$5, we find results in a similar range, as long as the regulator is able to detect leaks remotely. For instance, an average fee of \$8.54 per ton CO_2e leads to emissions reductions of 692 tons CO_2e per well per year, from a baseline of 1520 tons (a 45% reduction).

We note that we're looking at a very different type of emissions than Marks (2018) did. That paper uses reported emissions at the operator-basin level from the EPA inventory, which undercount the large, rare leaks that make up our dataset (Robertson et al. 2020). That paper also uses a different source of identifying variation and different mechanism of abatement (a more traditional abatement in quantity, rather than our abatement in probability).

Marks (2019) uses the same abatement figures as Marks (2018) to consider the welfare gains from a sampling-based tax: some fraction of a firm's facilities are randomly sampled with a ground-level measurement, and the firm is assessed a tax based on the sample. That paper takes a similar approach to our audit designs, particularly our consideration of targeting on covariates (policy 2). In contrast to our work, that paper focuses on firm-level emissions, auditing a subset of the firm's facilities and charging a fee based on the firm's estimated total. We consider each well pad individually and focus on the challenge of using measurement to target audits. In future research we hope to consider a variety of more sophisticated audit policies, including ones that integrate well ownership.

2 Background

2.1 Institutional Setting

We first provide background on the institutional details of our setting, a discussion of the more traditional economic approaches to regulation, and a sampling of the relevant literature. These details motivate the approach we take in our theoretical modeling, as well as the constraints we consider for the regulator.

The upstream production of US onshore oil and gas sector emits approximately 6–10 million tons of methane per year (as of 2015), which is approximately 25% of total US methane emissions or 200–325 tons of CO_2e (Alvarez et al. 2018 provides emissions estimates for 2015).[4] Using a low-end $58.82/ton social cost of carbon ($2/kg methane), these upstream emissions work out to $12–19 billion per year in climate damages, before downstream leaks or emissions from burning the fuel.[5] For comparison, the contribution to gross domestic product (GDP) for the entire oil and gas sector, less wages and depreciation, averages $34 billion per year.

There is little policy addressing methane emissions, either in the US or globally. The most active current regulations are in Colorado, which requires well operators to visit wells and look for leaks. Other states and the US federal government have considered or begun to implement similar regulations. In these policies, the well operator is required to visit the well at some frequency. In Colorado, this ranges from once in the well's lifetime to every month, depending on the well's size and location. Well operators need to record, report, and repair their leaks. There's no penalty for reporting leaks. In fact, the Colorado regulator views a high number of found-and-fixed leaks as a success. These LDAR policies, like the audit policies we consider in this paper, reflect the policymaker's limited resources and measurement challenges.

This paper considers audit policies as a compelling alternative.[6] We focus on audit policies because they're the traditional tool of environment, health, and safety regulation.

4. CO_2 emissions from the western US electricity grid were about 245 million tons in 2018 (EPA 2020a).

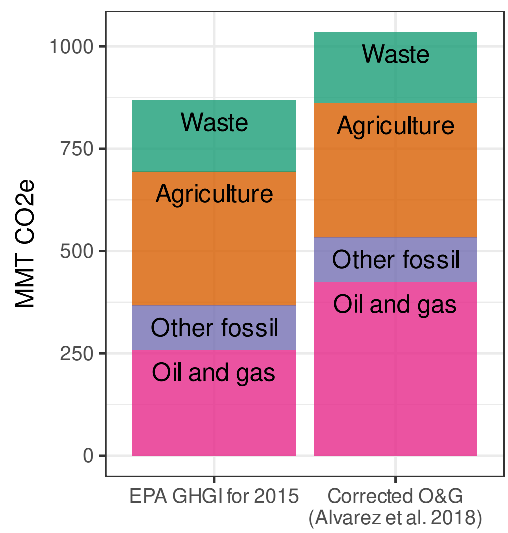
5. Throughout this paper we use a CO_2e conversion factor of 34, the standard 100-year global warming potential (GWP) from the Intergovernmental Panel on Climate Change (IPCC) (Myhre et al. 2013). There is a lively debate, e.g. Allen et al. (2018), about the correct way of comparing emissions of different greenhouse gases. The most internally consistent approach would be to use a social cost of carbon (SCC) and social cost of methane. With a conversion factor of 34, $58.82 per ton CO_2e is $2 per kg CH_4, or about $35 per mcf natural gas. We describe $58.82 as "low end" because the widely used EPA $42/ton number from 2013 is $55 in 2019 dollars. However, numerous studies have found that the damage numbers are too low in the underlying models used by US EPA (2016), so we feel it would be undesirable to rely on them too heavily. Using the existing SCC estimates, the implied conversion between one ton of CH_4 and one ton of CO_2 ranges from 25.8 to 45.

6. We do not consider the horse-race between audits and a stringent LDAR program; we don't have the data to make that comparison.

Audits also set an expected price on emissions, which can be helpful when the regulator doesn't know the optimal abatement technology or behavior for each well operator. However, these audit policies face challenges. First, visiting wells is expensive and time-consuming. The EPA estimates that it costs $450–600 per visit (EPA 2020). Other estimates are lower, but easily over $100 per well pad for on-the-ground audits.[7] Second, the fines that the regulator charges are limited.

Remote sensing may provide valuable but imperfect information. We consider the role satellite measures may play in an audit or pricing policy. Our depiction of remote sensing is somewhat stylized. We assume that the remote measurement is perfectly accurate, except for a detection threshold. Because we assume well operators respond purely to the expected value of a fee, any measurement error in assessing the fee doesn't matter, as long as the measurement is unbiased. Large enough errors could, for instance, inefficiently force the well to declare bankruptcy, but we put these concerns aside. Appendix table 7 provides more detail on satellite measurement error.

Figure 1: Scientific literature finds oil and gas emissions 65% higher than EPA inventory



Emissions at oil and gas wells come from a variety of sources. There are a large number of small, intentional vents. This venting is the expected result of equipment operating

7. Personal correspondence with Arvind Ravikumar (Assistant Professor of Energy Engineering, Harrisburg University of Science and Technology), May 22, 2020.

properly. Large intentional vents are rare, as the operator would typically set up a flare. There are a large number of small, unintentional leaks from various pieces of equipment.

Finally, there are a relatively small number of large leaks. These large leaks are responsible for the majority of emissions. At any point in time, a small fraction of wells are leaking—in our dataset, it's on the order of 1–3%. These large leaks are often from separator tanks left open or other valves that weren't sealed (Lyon et al. 2016). They also occur during the drilling and fracking process (completions), and when wells blow out. The small vents and leaks are easier to measure and predict. As a result, they're better represented in the emissions inventories. Rutherford et al. (2020) finds that underestimated emissions from large vents and malfunctions explains the difference between the official inventory and the estimates in the scientific literature. The difference highlighted in figure 1 represents emissions that the US is not measuring in official inventories, much less charging for emissions. The US is not alone; another recent study found a similar underestimate in the Canadian GHG inventory (Chan et al. 2020).

Other research, such as Alvarez et al. (2018), have estimated methane leakage at the basin, state, or nation level. These estimates are essential to know the overall leakage rate. However, to think about leakage abatement by individual well operators, we need to focus on individual well pads.

Mitigating these leaks depends on finding them, as well as taking care in not creating them. This care can be additional attention to closing tank hatches, or more frequent visits to reduce the duration of a leak. When we consider policies that increase the expected cost of having a leak, we assume the well operator will try to have fewer leaks, or to have the leaks last shorter amounts of time. These efforts could be anything from additional employee training to smarter tank hatches to additional LDAR visits.

2.2 Traditional Economics Solutions

Because measurement and fines are limited, the traditional economics solutions to reach the first best fall short in our constrained context. The Pigouvian prescription would be to charge well operators for the damages of their emissions (Pigou 1932). Without accurate measures of those emissions, a Pigouvian tax can't be implemented. The Becker (1968) or Polinsky and Shavell (1979) approach would be to audit a small fraction of wells and charge them large fines if they are in violation. Unfortunately, fines are limited through court precedent, the backstop of bankruptcy, and political feasibility. The Segerson (1988) approach is maybe the most promising here. Segerson's idea, originally developed for non-point pollution, is a tax and dividend approach. Each source pays the full social cost for *all* emissions in their area beyond the socially optimal level, giving everyone the incentive to fully internalize their emissions, even when individual emissions can't be measured. Unfortunately, the payments are implausibly large, and well heterogeneity makes partitioning responsibility a challenge. Beyond these efficiency concerns, policies in the style of Segerson (1988) are politically unpopular, even relative to direct emissions pricing. We'll instead consider policies that make due with limited information and enforcement capacity.

In other information-constrained contexts, the regulator often uses indirect information as a guide, but can't act on it directly. For instance, Occupational Safety and Health Administration (OSHA) may decide to audit a workplace when there are high rates of worker injury, but the OSHA inspectors still need to conduct the audit before they're able to assess a penalty. In pollution contexts, from particulate matter to NO_x, satellite measures regularly detect that regions are out of compliance with the US Clean Air Act. However, satellite measurements are noisier than ground-based measures, and only the official, ground-based measurement network is used for compliance status.

3 Theory

Motivated by background, we begin by developing a theory of well abatement and the regulator's response. In section 3.1 we start with a model of the well operator's problem. Solving this model gives us an expression for well operator behavior—including DWL and change in emissions—as a function of the expected fee the operator faces. Using these results, we turn to the planner's problem in section 3.2. The planner or regulator wishes to maximize welfare, subject to constraints on the number and targeting of audits they're able to do. We consider the five policies discussed above, from status quo to auditing plus remote sensing.

3.1 Well Operator's Problem: Choosing Abatement

Well operators abate by reducing the probability that a well is leaking, rather than reducing leak size. We say each well i has a fixed potential leak size e_i. The probability of not leaking, q_i, is chosen by the well operator at a cost $C_i(q_i)$. We present results for a general C_i, assuming that marginal costs are positive and convex ($C_i'(q) > 0$, $C_i''(q) > 0$). We assume C_i'' is continuous and C_i' is invertible.

We consider counterfactual policies that would weakly increase q_i. Without marginal costs that increase sharply as $q_i \longrightarrow 1$, we run the risk of assuming that wells will choose $q_i = 1$. We view this corner solution of perfect abatement as unrealistic, so we further assume $\lim_{q \to 1} C_i'(q) = \infty$.

The idea of continuous marginal cost may seem strange, when we often think of abatement as requiring costly one-time capital investments. Recall that $1 - q_i$ is the probability of leaking at any particular point in time (*not* the probability of developing a leak). We're thinking of abatement as largely about additional effort and monitoring by the well operator: checking tanks and valves, training employees to close hatches, and so on.

To take the cost functions to the data, we need a specific functional form. We assume a cost function C, parameterized by well-specific values A_i and α_i. Broadly, A_i scales the cost and α_i determines the elasticity with respect to $1 - q$. We specifically assume:

$$\text{Total cost: } C_i(q_i) = C(q_i; A_i, \alpha_i) = -\frac{A_i}{\alpha_i + 1}(1 - q_i)^{\alpha_i + 1}$$
$$\text{Marginal cost: } C'(q_i; A_i, \alpha_i) = A_i(1 - q_i)^{\alpha_i}$$

This cost function $C(q)$ has a constant elasticity of $\alpha_i + 1$ with respect to the probability of having a leak, $1 - q$. (In turn, the marginal cost function has an elasticity of α_i.) The parameters $A_i > 0$ and $\alpha_i < -1$ allow for well-level heterogeneity, which we discuss further in the estimation section (5). We choose this form because it is relatively simple, allows for rich heterogeneity, and satisfies our cost function assumptions.

The assumptions of fixed e_i and the specific cost function are by far the strongest in this section. While the fixed leak size assumption may seem restrictive, the important feature for our model is that the leak size is not a choice variable. We present a static model here, but the intuition can be extended to the case where the leak size is periodically drawn from a distribution conditional on the well's covariates, rather than fixed for the well's lifetime.

Even without any leak regulation, the operator has a private benefit of capturing leaks, since the methane can be sold into the natural gas commodity market at price p_i. Without any policy, a firm chooses its abatement effort level to maximize its expected profit, $\mathbb{E}[\pi_i] = q_i \cdot p_i \cdot e_i - C_i(q_i)$. The first-order condition (FOC) for an interior solution is given by $C_i'(q_i) = p \cdot e_i$. As long as $p_i > 0$, we can invert the cost function to solve for q_i.

Now consider some expected fee, t_i, which may vary across wells. A firm maximizing expected profit will treat this t_i the same as the commodity price, giving the FOC $C_i'(q_i) = (p + t_i) \cdot e_i$. Let δ be the external social cost of methane, so $p_i + \delta$ is the total social loss from emitting one additional unit. For all the reasons detailed above, we assume the regulator is constrained, leading to the second-best case with a fee lower than social marginal cost ($t_i < \delta$).

We assume a utilitarian objective, where the regulator tries to minimize the ex-ante DWL before leaks are realized. We can write the DWL from setting $t_i < \delta$ for well i as the following expression. The intuition here is the same as the Harberger triangle in public finance—because of increasing marginal costs, the first unit of abatement provides a lot of social value, and the last unit of abatement provides almost none. The general expression for the DWL from well i is:

$$\mathrm{DWL}_i(t_i) = \int_{q_i = C_i'^{-1}((p_i + t_i) e_i)}^{C_i'^{-1}((p_i + \delta)\, e_i)} (p_i + \delta)\, e_i - C_i'(q) dq$$

Substituting in the specific marginal cost and evaluating the integral gives us:

$$\mathrm{DWL}_i(t_i) = \left((p_i + \delta)\, e_i - \frac{(p_i + t_i)\, e_i}{\alpha_i + 1} \right) \left(\frac{(p_i + t_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} - \frac{\alpha_i}{1 + \alpha_i} (p_i + \delta)\, e_i \left(\frac{(p_i + \delta)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}}$$

Mathematical details are in appendix section A.1.

In considering this DWL, we make a number of additional assumptions. We assume there are no other market failures beyond these methane leaks. We use methane and natural gas interchangeably in the theory section, but in the estimation we acknowledge that natural gas is not entirely methane. Finally, we assume the population of wells is fixed, and wouldn't change under different policies.

3.2 Regulator's Problem: Choosing Audits

In the well operator's problem, we considered how t, the expected fee, affects abatement effort. Here we consider how different policies lead to different t.

Recall that we are considering five different policies. (0) no auditing, the status quo; (1) audit every well with equal probability; (2) target audits based on well covariates; (3a and 3b) measure leaks remotely to target audits; and (4a and 4b) measure leaks remotely and assess fines based on measurements. The a and b variants consider the cases where all of our large leaks can be detected, or only leaks above some high detection threshold, as would be the case for satellite measures. The game theory of the regulator's problem will be somewhat different when they can only observe the largest leaks. We consider a range of allowed fees, where the regulator may not be able to charge for the full social cost, as is the case for almost every GHG pricing policy.

In the audit cases, we consider a regulator choosing how to allocate a fixed budget of M audits, and then consider how the shadow price on the audit budget compares to engineering estimates of the cost of conducting audits. We focus on this audit-budget approach, rather than directly choosing the number of audits based on the cost of conducting audits. We've never heard of a government agency where the audit budget was set to equalize the marginal cost of auditing with the marginal benefit. We consider different levels of the audit budget (changing M) within each audit policy.

It's important to be clear about the timing of events. First, the regulator commits to an audit policy. Second, firms learn their e_i and choose their q_i. Third, emissions are realized:

each well i has a leak of size e_i with probability $1 - q_i$. Fourth, the regulator measures emissions (if applicable). Fifth and finally, the regulator conducts audits and assesses fees. When choosing who to audit, we assume the regulator can observe well covariates and can form expectations of leak size, but can't observe abatement effort or actual leak size. We also assume the regulator can't keep a secret—the well operator knows exactly what policy regime they're in and what the probability of an audit is.

Define τ as the fee per kilogram of methane, when a leak is detected, and r_i as the probability well i is leaking when it is audited. Therefore the expected fee when leaking is $t_i = \tau r_i$, with units of dollars per kilogram of methane. We don't know what τ would be feasible, so we consider a few different values to cover a range of possibilities. We test $\tau = \{\$5 \text{ per ton } CO_2e, \delta, 2\delta\}$. \$5 per ton CO_2e is the low-end value we consider for comparison with Marks (2018). δ is the social cost. If every well could be targeted, $\tau = \delta$ would be the Pigouvian prescription. $\tau = 2\delta$ is loosely motivated by the result of the *Exxon Valdez* US Supreme Court ruling. In that case, the Court limited punitive fees to a 1:1 ratio with economic damages (for a total of twice the economic damages; *Exxon Shipping Co. v. Baker* 2008.

We assume the regulator wants to choose the probability each well is audited, r_i, to minimize the ex-ante DWL. The regulator isn't fixing leaks when they're found. If the regulator's audit provides valuable ex-post information to the well operator, then our estimates will be a lower bound on the gains of the audit policy. Note that the well operator's choice variable, q_i, has been integrated out so we can express the DWL from the previous section directly in terms of the regulator's parameters. Substituting in τr_i for t_i, the regulator's general problem is:

$$\min_{r_i} \sum_i \mathrm{DWL}_i(r_i) \qquad s.t. \sum_i r_i \leq M, \; r_i \in [0, 1]$$

3.2.1 *Policies 0 and 4: no auditing and remotely assessed fees*

In the business-as-usual case (0) of no auditing, $r_i = 0$. In the case with fees assessed directly from remote measurements (4), there are still no audits, but wells face some fee t_i. Policy 4 provides an infeasible benchmark; feasible policies require on-the-ground audits. The DWL goes to zero—and the first-best can be achieved—if the remotely sensed fee can be set to the social cost of emissions (δ) for every well. If there's a high detection threshold where wells with small e_i can't be detected, $t_i = 0$ for wells with e_i below the threshold. We assume they know they're below the threshold, and choose their level of abatement accordingly.

3.2.2 Policy 1: uniform audit probabilities

The audit policies are somewhat more complicated. The ideal audit probability, absent any constraint or cost of auditing, would be $r_i = \delta/\tau$. For instance, if the well operator must pay two times the social cost when found leaking, then it's optimal to audit them with a 50% probability. In general, a constraint of M audits will be binding when the fraction of audits, M/N, is less than δ/τ.

For the uniform audit policy (1), the constrained maximization problem with shadow price λ is:

$$\min_{r\in[0,1]} \mathscr{L} = \sum_i \mathrm{DWL}_i(r) + \lambda\left(M - \sum_i r\right)$$

$$r = \begin{cases} \frac{M}{N} & \text{if the audit budget constraint binds} \\ \frac{\delta}{\tau} & \text{if not} \end{cases}$$

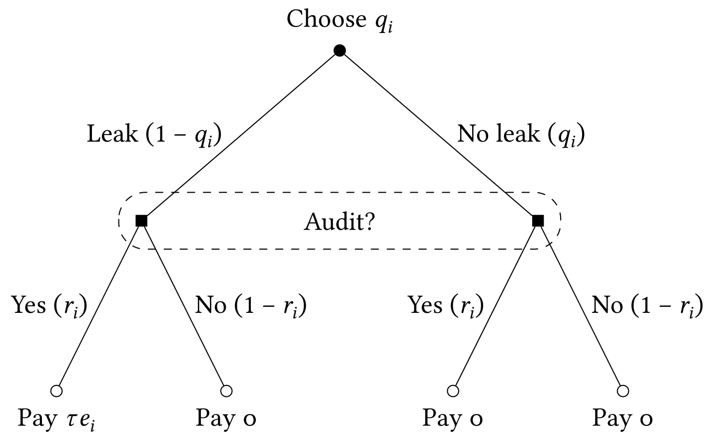
$$\lambda = \begin{cases} \frac{1}{N}\sum_i \frac{\partial \mathrm{DWL}_i(r)}{\partial r} & \text{if the constraint binds} \\ 0 & \text{if not} \end{cases}$$

3.2.3 Policy 2: targeting on covariates

Moving to the case where audits are targeted on well covariates (policy 2), it's helpful to have an extensive form game tree. See figure 2 for the game each well faces under this policy. Recall that the well operator knows e_i, the size of their leak if a leak happens, their cost function $C(q_i; A_i, \alpha_i)$ and the regulator's audit rule $r_i = r(X)$. Based on these inputs, they choose their probability of not leaking, q_i. After that, leaks are realized, and the well has a leak with probability $1 - q_i$. If the well is leaking and is audited, the well operator pays τe_i. If they're not audited, or audited but not leaking, they pay no fee. We assume that the X variables the regulator uses to target audits cannot be changed. In our empirical analysis, the variables we include would be difficult to change without making the well significantly less profitable.

In choosing the audit rule $r(X)$, the regulator tries to set the optimal ex-ante incentives for the choice of q. The natural question is what functional form $r(X)$ should take. In the empirical implementation, we choose r based on the predicted DWL_i, which is itself a function of our X variables. Therefore, we simply allow the regulator to choose a vector of r_i, one for every well. This vector is implicitly a function of X. We discuss the implications further in the estimation section 5, particularly the distinction between targeting r_i and

Figure 2: Game tree: targeting on covariates (policy 3)



In this figure, the well chooses its probability of not having a leak, q_i, with full knowledge of the probability they will be audited, r_i. Then nature determines whether a leak occurs or not. The regulator does not know whether leak has occurred—their information set is indicated in a dashed oval. If the well is leaking and is audited, the well operator pays τe_i. In all other cases, they pay zero.

evaluating the policy. The regulator's problem is now:

$$\min_{\{r_i\}_{i=1}^N} \sum_{i=1}^N \mathrm{DWL}_i \qquad s.t. \sum_{i=1}^N r_i \le M \text{ and } r_i \in [0,1] \text{ for all } i$$

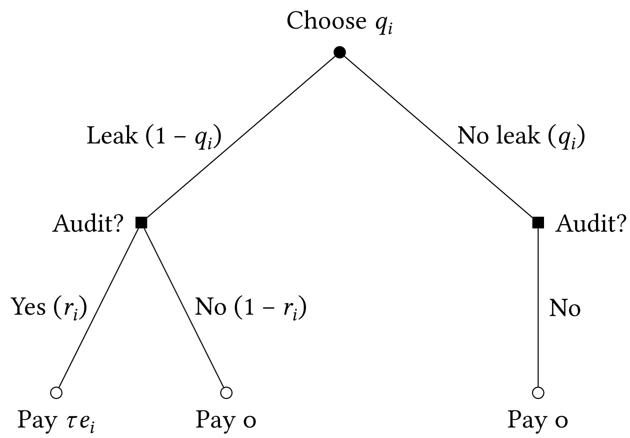
$$\min_{\{r_i\}_{i=1}^N} \mathscr{L} = \sum_i \mathrm{DWL}_i(r_i) + \lambda\left(M - \sum_i r_i\right) + \sum_i (r_i - 0)\, a_i + (r_i - 1)\, b_i$$

As before, λ is the shadow price on having an additional audit. $\lambda < 0$ because increasing audits lowers DWL. a_i and b_i correspond to the constraints $r_i \ge 0$ and $r_i \le 1$. The feasible set under the constraint is a compact subset of $\mathbb{R}^N$ so a solution exists. Because DWL_i only depends on r_i and DWL_i is strictly increasing, convex in r_i and the constraints are linear inequalities, we conclude a unique solution exists and can be characterized by the standard Karush–Kuhn–Tucker (KKT) conditions. We solve for $\{r_i\}_{i=1}^N$ and λ numerically using Ipopt. For notation simplicity, we don't detail the case where the audit budget isn't binding ($M/N > \delta/\tau$). The solution for that case is the same as in the non-binding uniform case, $r_i = \delta/\tau$.

3.2.4 Policy 3a: targeting on leak observations

We next consider a policy that targets audits based on ex-post observed leaks (policy 3a). Here, the regulator gets to observe whether wells are leaking—and the leak size—before choosing whether to audit. Even though the leak is measured remotely, on-the-ground measurements may be required for legal reasons or because the remote measurement is noisy. Since leaks are rare, the regulator can now use the audit budget much more efficiently, as they waste less effort auditing wells that aren't leaking. Figure 3 details the simple case with no detection threshold in measuring leaks. In this case, the only reason not to audit a well that was found leaking would be if the audit budget was extremely small. The regulator would never audit a well that was measured not leaking.

Figure 3: Game tree: target leaks (policy 4) with no additional censoring



In this figure, the well chooses its probability of not having a leak, q_i, with full knowledge of the probability they will be audited, r_i. Then nature determines whether a leak occurs or not. The regulator knows when a leak has occurred, and they will never audit a well that isn't leaking. If the well is leaking and is audited, the well operator pays τe_i. In all other cases, they pay zero.

Until now, in the uniform and target-on-covariates policies (1 and 2), the probability a given well was audited was the same whether or not a leak actually occurred. That is, each well's audit probability is statistically independent of its leak probability. When we consider targeting on realized emissions (3a), that's no longer true. The expected number of audits is now the sum of: audits when the well is leaking, times the probability it is leaking *plus* audits when the well is not leaking, times the probability it is not leaking. In the target-on-covariates case, we forced the probability of being audited when leaking and when not leaking to be the same, since audit probabilities depended only on the X, not the realized leaks. Now, with every leak observable, the probability of being audited when not leaking falls to zero. The budget constraint becomes:

$$\sum_i q_i \cdot 0 + (1 - q_i) r_i \leq M$$
$$\sum_i \left(\frac{(p_i + \tau r_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} r_i \leq M$$

Note that the q_i here is $q_i(r_i)$ after responding to the audit policy, not the status-quo $\tilde{q}_i$. Using the well operator's FOC, we substitute in $q_i(r_i)$ in the second line above.

The problem in choosing r_i is:

$$\min_{\{r_i\}_{i=1}^N} \mathscr{L} = \sum_i \mathrm{DWL}_i + \lambda \left(M - \sum_i \left(\frac{(p_i + \tau r_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} r_i \right) + \sum_i (r_i - 0)\, a_i + (r_i - 1)\, b_i$$

$$\frac{\partial \mathscr{L}}{\partial r_i} = \underbrace{\frac{A^{-\frac{1}{\alpha}} e^{\frac{\alpha+1}{\alpha}} \tau(\delta - \tau r)(p + \tau r)^{\frac{1-\alpha}{\alpha}}}{\alpha}}_{\frac{\partial \mathrm{DWL}_i}{\partial r_i}} - \lambda \cdot \underbrace{A^{-\frac{1}{\alpha}} e^{\frac{1}{\alpha}} (p + r\tau)^{\frac{1}{\alpha} - 1} (\alpha(p + r\tau) + r\tau) \frac{1}{\alpha}}_{\frac{\partial (1-q_i) r_i}{\partial r_i}} + a_i - b_i$$

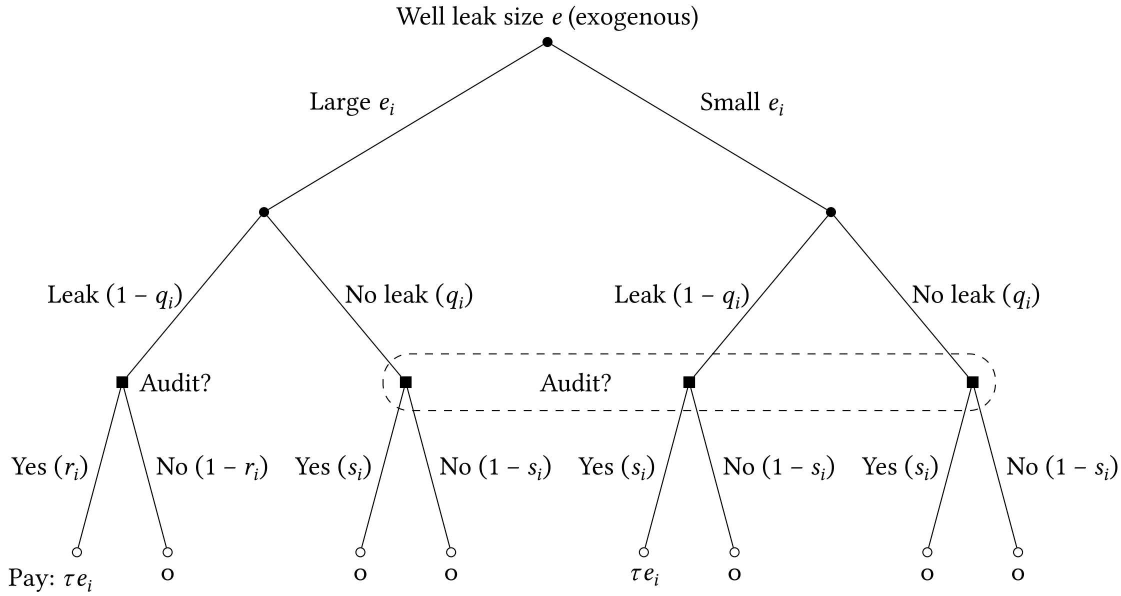
As before, λ is the shadow price on having an additional audit, and a_i and b_i correspond to the constraints $r_i \geq 0$ and $r_i \leq 1$. The feasible set under the constraint is a compact subset of $\mathbb{R}^N$ so a solution exists. Unfortunately, the problem is no longer monotonic or convex. We are able to solve numerically, but without guarantees of a unique global maximum. In contrast to the previous cases, whether the audit budget binds now depends endogenously on the wells' abatement.

3.2.5 Policy 3b: targeting on leak observations, only observing the largest leaks

The last audit scheme we consider is the case where we're targeting on observed leaks (policy 3b), but small values of e_i aren't detected. In this case, failing to observe a leak might mean that the well isn't leaking, or it might mean that the well has a leak below the detection threshold. Figure 4 has a game tree for the regulator's problem. Each well operator knows whether they're on the left branch (large e_i) or right branch (small e_i), since e_i is not a choice variable. The dashed oval indicates the regulator's information set—they cannot tell whether a well has a large e_i and isn't leaking, a small e_i and isn't leaking, or a small e_i and is leaking.

The regulator sets audit probabilities based on whether the leakage is detected, taking into account of the detection threshold. As the game tree suggests, if a well is *not* detected

Figure 4: Game tree: target leaks (policy 4) with censoring



In this figure, nature determines the well's potential leak size, e_i. It is not a choice variable. The well knows e_i; the regulator can only form expectations. The well chooses their probability of not having a leak, q_i. If a leak happens at a large well, it is detected. If a leak happens at a small well, it is not. If a leaking well is audited, it pays τe_i. The dashed oval indicates the regulator's information set.

with leakage, there is no way to distinguish between whether it is actually not leaking or the leakage is small. As a result, the regulator can only specify an audit probability $r_i(X_i)$ for a well i with detected leakage and an audit probability $s_i(X_i)$ for a well with *no* detected leakage (the covariates X included in the brackets means the r, s can depend on these covariates).

A well operator's response to this policy will depend on their e_i. For small-e wells, their response is straightforward: as they know they will always be audited with probability s_i so the DWL will be $\text{DWL}_i(s_i)$. But for large wells, the incentives are more complicated. A large well i will have q_i probability of *not* leaking. But even when it is not leaking, it will be audited with probability $s_i(X_i)$. Since it is not leaking, the audit will not lead to any penalty (the auditing effort is wasted here). Large-e wells will not care about s_i. q_i and DWL_i will be functions that only depends on r_i. The *ex ante* DWL for a large-e well will be $\text{DWL}_i(r_i)$.

The budget is similar to the case with no detection threshold. As with the DWL, it's the probability of being audited *when leaking* that matters to the well, so the large-e choose $q_i(r_i)$ (s_i does not enter). Unlike the previous cases, large-e wells now have some s_i probability of being audited when not leaking. Therefore, the audit costs are $(1 - q_i(r_i))r_i + q_i(r_i)s_i$ for each large-e well, and s_i for each small-e well.

Given these DWL and budget components, the regulator needs to pick r_i and s_i. The regulator still does not know which wells are large-e or small-e, so they optimize a weighted average, where the weights are the probability the well's leak is above the threshold. For a detection threshold $\underline{e}$, define $z_i \equiv \Pr(e > \underline{e} \mid X_i)$. The regulator then optimizes the problem:

$$
\begin{aligned}
\min_{\{r_i\}_{i=1}^N, \{s_i\}_{i=1}^N} \quad & \sum_i z_i \mathrm{DWL}_i(r_i) + (1 - z_i)\mathrm{DWL}_i(s_i) \\
s.t. \quad & \sum_i z_i[(1 - q_i(r_i))r_i + q_i(r_i)s_i] + (1 - z_i)s_i \leq M \\
& \forall i : \ r_i \in [0, 1], s_i \in [0, 1]
\end{aligned}
$$

We can compare this minimization problem with the previous one, where $\underline{e} = 0$, to confirm that the previous problem was a special case of this one. In the previous problem, z_i converges to 1 so we do not need to worry about s_i in the objective function. Moreover, in the budget constraint, it is then obvious to see why s_i should be set to zero to save audit effort. Lowering the detection threshold leads to a lower DWL. Intuitively, all the actions with a high detection threshold remain feasible under the lower threshold, but the lower threshold provides additional information to the regulator. Of course, other audit strategies are possible, and may be preferred if the regulator can easily distinguish large-e wells without leaks from small-e wells.

Compared to the audit policies, the remote fee policy is simple. In this scenario, the regulator is able to measure leaks remotely, and charge a fee for emissions *without* doing an on-the-ground audit. The regulator can measure every well, with zero marginal cost. We assume, as before, that this measurement is accurate, at least in expectation. We do not think this is a feasible policy, but it provides a useful benchmark to think about the possible gains of the audit policies. When the regulator is able to measure all leaks with no detection limit (4a), and is able to charge the full social cost for detected leaks, this policy recreates Pigouvian taxation. When the fee is lower than the social cost, the policy implements second-best Pigouvian taxation, much like a standard carbon tax. As in the measure-then-audit (3) case, we consider the possibility that only the largest leaks can be detected (4b). In that case, wells without a detected leak will not be charged any fee.

3.3 Adding Time to the Model

The model we present is static: a one-shot game where the regulator sets incentives and the wells respond. This static model captures the essence of the problem we're interested in, and adding strategic dynamics would complicate things without adding insight. However, the real world is dynamic. To present welfare results as dollars per year, we present a simple extension of our static model into a world with time.

There are H (8760) hours in a year. We assume, with minimal loss of generality, that the well operator pays $C(q)$ once per year to have an average no-leak probability of q across all hours of the year. The probability across hours need not be independent and identically distributed (I.I.D.). q is the probability of not having a leak, averaged across all hours. It is *not* the probability of a leak starting or stopping. This distinction is important, both for the way we include time in the model, and because our data provides very little information on leaks beginning and ending.

We can think of the abatement in $C(q)$ as any adjustment that reduces the probability the well is leaking. These adjustments can be capital investments that reduce the probability of a leak beginning. Or there may be increased operator monitoring that leaves the probability of a leak beginning unchanged, but reduces the length of leaks when they occur. Setting H to be a year is a normalization that makes it easier to discuss annualized figures, but has little impact beyond that.

The expected quantity of leaks from well i over the H-hour period is $H(1 - q_i)e_i$. If the regulator knew q_i and observed a leak of size e_i, the first-best fee would be $\delta H(1 - q_i)e_i$. The length of an individual leak is irrelevant to the expected value—it could be one leak that lasts $H(1 - q)$ hours or $H(1 - q)$ separate leaks that each last one hour.

However, the regulator is constrained. They don't know q_i. Instead, they can charge a fee on the expected emissions, even though they took measurements at a snapshot in time. Define T hours as the regulator's expectation of a well's emissions—when detecting a leak of size e_i kilograms per hour, they charge a fee for Te_i kilograms of emissions. This expectation *does not* need to be correct.

The fee the regulator charges is $\tau T e_i$. It's the product τT that determines the fine magnitude. (Indeed, our implementation uses a single variable τT, with units of dollar-hours per kilogram). We don't want to use an implausibly large T to back our way into a very high penalty for leaking. Instead, we consider a few different values $T = \{1 \text{ day}, 1 \text{ week}, 1 \text{ month}\}$. We focus on $T = 1$ week as our main case, since $T/H \approx \sum(1-q_i)/N$. Recall that we already consider $\tau = \{\$5, \delta, 2\delta\}$, so for instance, $T = 2$ weeks, $\tau = \delta$ is already covered by $T = 1$ week, $\tau = 2\delta$. Considering different values of T would provide further variation in τT, but doesn't add any other robustness to the analysis.

With the addition of time the mathematical expressions we provided earlier are minimally changed. Specifically, in the DWL expressions, p_i is replaced with Hp_i, δ is replaced by $H\delta$, and τr_i is replaced with $T\tau r_i$. The full expressions are provided in appendix section A.3.

This theory section provides a set of expressions that characterize the audit probabilities and DWL for a set of second-best audit policies. If the regulator did not face constraints on the number of audits they could conduct and the fees they could charge, the regulator could achieve the first best with high fees or ubiquitous audits. The remote sensing element relaxes the audit budget, allowing the regulator to target audits more effectively, but faces its own limitations in terms of which leaks can be detected.

4 *Well and Leak Data*

To estimate the theory models we discussed above, we need data on leaks. In particular, we need to estimate the wells' leak size when leaking (e_i), and the cost parameters of abatement A_i and α_i. These estimates are based on the distribution of observed leak sizes and the observations of whether wells leak or not. We build this dataset using scientific studies of large leaks, matched with a database of all US wells and commodity natural gas prices. We discuss each of these sources in turn.

4.1 *Methane measurements*

There's a robust, ongoing effort in the scientific community to measure leakage from all parts of the oil and gas supply chain. Alvarez et al. (2018) provide a thorough discussion of leaks from different sources.

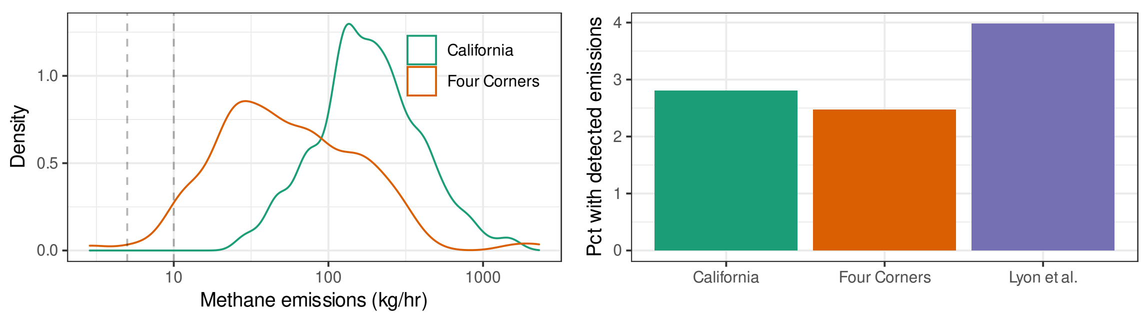
Our work relies on these scientific measurements. We primarily use measurements taken from airplanes using the "next generation airborne visible/infrared imaging spectrometer" (AVIRIS-NG) sensor. These studies surveyed wells in California and the Four Corners region of northern New Mexico and southern Colorado (Duren et al. 2019; Frankenberg et al. 2016). These flights primarily covered the San Joaquin and San Juan basins, with some flights in other California basins. These studies aimed to survey a representative sample of wells in their respective areas, with the goal of characterizing the distribution of leak sizes and estimating total regional emissions. See tables 1 and 2 for summary statistics and covariate comparisons. The airplanes used in these studies are able to detect leaks of approximately 5–10 kg/hr, depending on wind conditions. By having flights over tens of thousands of wells, these studies are able to capture the right tail of the leak distribution in a way smaller studies cannot.

The methane measurements report measured methane plumes, including their time and location, as well as a guess of the associated infrastructure. The scientists also report the plane's flight path, which will be important for defining the sample of wells without detected leaks. Both the California and Four Corners studies include some leak measurements from non-well sources, such as landfills, coal mines, pipelines, and gas processing facilities. We include all plumes that are either unidentified or identified as related to an oil or gas well.

We also use evidence from Lyon et al. (2016). That study surveyed a large number of wells for leak presence, but did not quantify the leak size. The detection threshold in that study was roughly equivalent to the AVIRIS-NG studies. These data corroborate the AVIRIS-NG studies, finding that leaks of this size are rare. When leaks are found, they're often from separator tanks. In these data, as well as the AVIRIS-NG studies, leaks are hard to predict. Some variables, like well size, are statistically significant, but overall prediction quality is poor when considering cross-validated mean squared error (MSE) or logistic loss.

There are a number of studies that measured well leaks on the ground, however, because they decline to publish well identifiers or covariates beyond contemporaneous gas production, we are not able to use their data for our analysis. A comparison is plotted in appendix figure 8. These ground studies often note that large leaks are from valves left open, or most frequently, separator tanks with open hatches.

Figure 5: Distribution of detected methane leaks



SOURCES: California and Four Corners distributions come from aircraft studies (Duren et al. 2019; Frankenberg et al. 2016). Lyon et al. (2016) provides information about leak prevalence (with a detection threshold roughly similar to the California and Four Corners studies), but not leak size.

As mentioned above, the focus of this paper is larger leaks, those detectable by the AVIRIS-NG airplane measurements. This definition is convenient, in the sense that it means we can sidestep the censoring in measurements and model abatement as a reduction in leak probability rather than a change in the distributions of sizes. This focus on larger leaks also maintains the focus on the more important abatement opportunities. In the ground studies data, more than three quarters of the total leakage is from leaks larger than 5 kg/hr.

We interpret the studies' repeated notes about leaking tanks to mean that these leaks are often accidents or process failures, rather than venting that occurs in the course of normal operations. Of course, additional monitoring is not free; it requires additional personnel and training. Unlike some of the pollution control literature, we're not thinking of abatement as a one-time capital expense, though additional capital investments may play some role.

A number of the wells in California are flown over multiple times. These revisits do not occur often enough or for enough wells that we can use panel methods and consider the evolution of leaks over time. At the same time, we want to avoid these multiple revisits affecting our cross-sectional data analysis, so we currently consider only the first time each well was flown over. In future research, we will use these repeat observations to provide more complete information on leak occurrence and persistence.

4.2 Well Data

Well data are from Enverus, formerly known as DrillingInfo. These data cover all wells in California, New Mexico, and Colorado, the primary states in our analysis. As mentioned in the previous section, the methane measurement flights have recorded flight paths (see an example in figure 6). We use these paths to determine which wells the plane flew over and could have measured. We exclude wells that did not have any gas production during the month of the flight. (While wells are designated either oil or gas wells, a large majority produce both.) (In these states, 98.9% of well pads report nonzero gas production.) See table 1 for summary statistics on wells in our analysis.

We match observed plumes to well pads based on geospatial location. Before matching with leaks, we aggregate individual wells to well pads. We define well pads as groups of wells that share an operator and geologic basin, and are nearby one another. Following Omara et al. (2018), we consider a 50 m radius around each well's surface location, and take the union of any circles that intersect. In the more densely packed San Joaquin basin, we use a radius of 20 m. We match the detected methane plumes to wells, matching each plume to the nearest well within 500 m—we assume plumes farther away are from non-well sources and we leave them unmatched. 31% of methane measures are dropped in this matching.

4.3 Price Data

We use the private incentive generated by the commodity price of natural gas to estimate our cost coefficients. The ideal price data would tell us what each well operator was paid for their gas production. That information isn't available, so we instead use gas prices at trading hubs near the wells. We use the average of the SNL series "SoCal Gas" and "PG&E,

Table 1: Well summary data

	Mean	Std. dev.	p10	p90
Panel A: Well pads included in flyover studies (N = 14,399)				
Age (yr)	18.1	12.7	3.5	38.7
Gas (mcfd)	115.1	1039.5	0.7	243.5
Oil (bbld)	17.9	196.7	0.0	29.9
Detect leak (%)	2.7	16.1	0.0	0.0
Leak size (kg/hr)	197.1	242.5	26.3	418.3
Gas price ($/mcf)	2.8	0.2	2.6	3.0
Panel B: Well pads checked by Lyon et al. (2016) (N = 8220)				
Age (yr)	9.4	9.3	1.8	22.2
Gas (mcfd)	385.4	1864.9	2.0	678.1
Oil (bbld)	47.9	253.7	0.0	70.5
Detect leak (%)	4.0	19.5	0.0	0.0
Gas price ($/mcf)	3.7	0.7	2.5	4.3
Panel C: All well pads active in June 2018 in CA, NM, and CO (N = 65,644)				
Age (yr)	19.6	13.0	4.8	40.3
Gas (mcfd)	161.6	940.3	1.3	250.6
Oil (bbld)	33.9	220.5	0.2	40.7

NOTES: Only wells that report positive gas production are included. In all three panels, wells are grouped into pads; see text for details. mcfd means thousands of standard cubic feet per day. bbld means barrels of oil per day. Well data from Enverus (2019). Flight data from Duren et al. (2019) and Frankenberg et al. (2016), covering parts of California, New Mexico, and Colorado. In panel A, leak size is for wells with non-zero leaks (N = 384). Prices are local to the month and state of the study, adjusted to 2019 USD.

South" for California wells and "El Paso, San Juan Basin" and "Transwestern, San Juan Basin" for the Four Corners wells.[8] These are midstream prices, and they will tend to be a small amount higher than the prices well operators actually receive. This difference

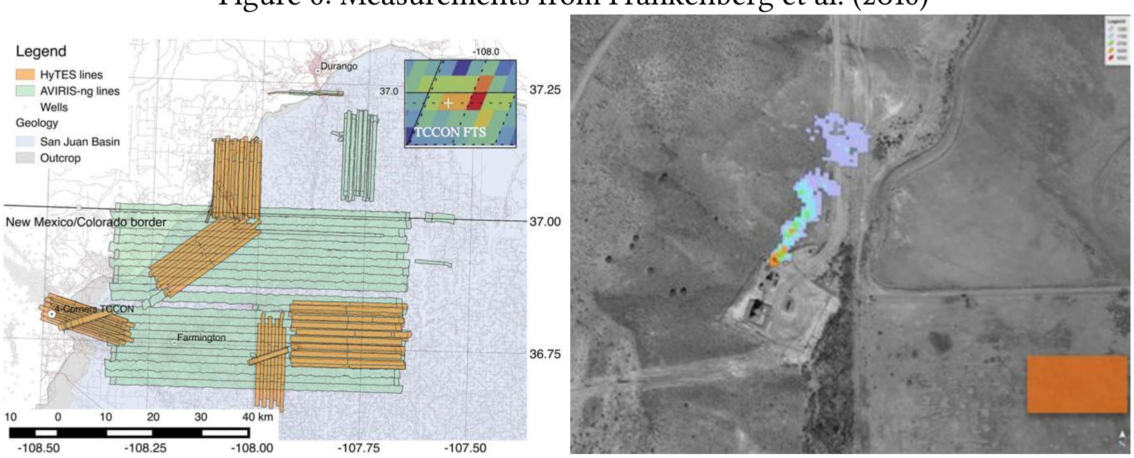
8. Details of index SNL's construction are available in https://www.spglobal.com/platts/plattscontent/_assets/_files/en/our-methodology/methodology-specifications/na_gas_methodology.pdf. Accessed 2020-10-27.

Table 2: Balance comparison: wells with and without detected leaks differ, but with overlapping covariate support

	California		Four Corners		Lyon et al. (2016)	
	Detect	No detect	Detect	No detect	Detect	No detect
Age (yr)	19.3	17	18.4	20.1	4.12	9.66
	[2.6,40]	[2.8,40]	[6.6,28]	[7.3,37]	[0.42,8.8]	[1.9,23]
Gas (mcfd)	248	88.8	323	146	1510	339
	[0.4,140]	[0.35,130]	[48,760]	[22,330]	[11,2800]	[2,610]
Oil (bbld)	83.3	27.2	0.102	0.223	300	37.5
	[0.88,70]	[0.98,44]	[0,0.13]	[0,0.5]	[0,880]	[0,59]
Detect leak (%)	100	0	100	0	100	0
	[100,100]	[0,0]	[100,100]	[0,0]	[100,100]	[0,0]
Gas price ($/mcf)	2.91	2.93	2.56	2.56	3.94	3.72
	[2.8,3]	[2.9,3]	[2.6,2.6]	[2.6,2.6]	[3.9,4.2]	[2.5,4.3]

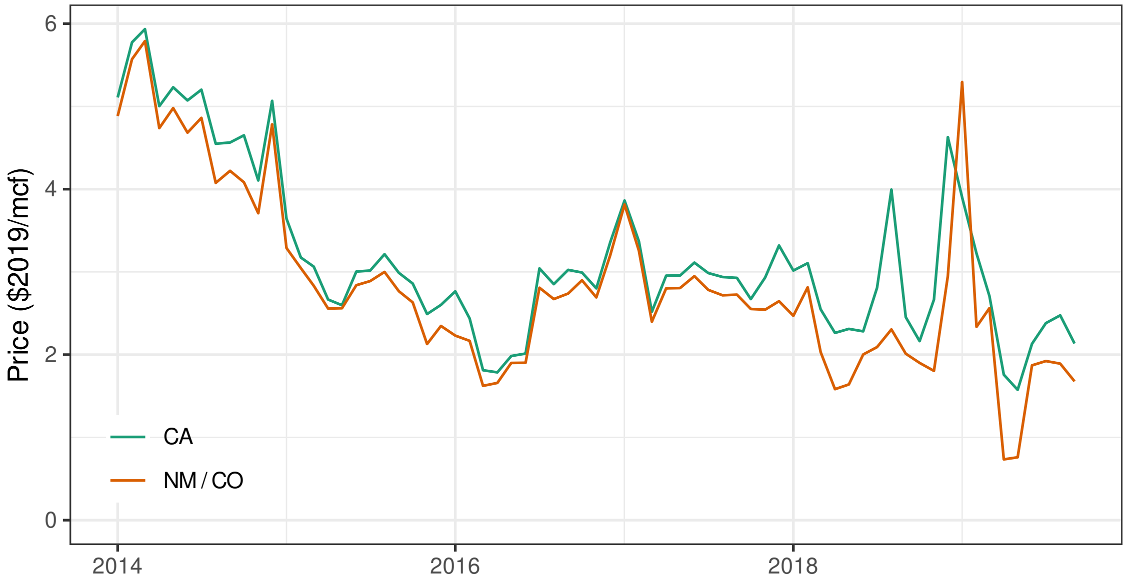
Note: Values are means, with the 10th to 90th percentile value in brackets.
California and Four Corners data are from the aviris-ng sample (panel a of table 1). Lyon et al. (2016) data cover basins throughout the US. All values are aggregated well-pad aggregates.

Figure 6: Measurements from Frankenberg et al. (2016)



Left: Flight lines from Frankenberg et al. (2016). Only aviris-ng flights, in green, were able to quantify emissions.
Right: Gas storage tank, emitting ~146 kg CH_4 per hr, using aviris-ng instrument. Orange bar is 60 m wide.

Figure 7: Natural gas prices



Source: SNL Financial (snl) natural gas price indexes for deliveries near study wells. See text for details.

will lead to a small upward bias in our estimated cost of abatement, and therefore a small downward bias in the estimates of gains from policy.

If we had more data on methane leaks, particularly measures of the same wells or subregions over time, we could use variation in the price of natural gas to identify our coefficients. We do not have this panel structure. While there are a few repeat measurements, the current data is more or less a cross section. There is some variation in prices across states, but we think it would be inappropriate to use this variation to identify cost parameters—too many other things change with geography. In future work, and as more methane measurement data become available, we plan to revisit our analysis using price variation.

5 *Estimation*

The studies mentioned above are primarily focused on total leakage or on the right tail of emissions (termed "super-emitters"). Our goals are closely related, but because we're focused on the effect of an achievable policy, we care more about the distribution across all wells, particularly the probability that a well pad leaks.

We note two important features of the leak distributions. First, the size of estimated leaks has a long right tail, as discussed in original papers that measured these leaks. Like those papers, we find that a lognormal distribution fits the measurements well. Second, we note that in the airplane studies, well over 95% of wells don't have observable leaks. To fit these additional zeros, we estimate a two-part model, estimating a fairly standard logit for whether a leak is detected, and a fairly standard lognormal for size in the sub-sample with detected leaks. We estimate these two pieces, rather than a Heckman-style selection model, because the selection model is a bad fit for our context. Most importantly, we're interested in the predicted values, not the coefficients on particular variables. We want to know the distribution of leaks and leak probabilities, but we're not interested in "undoing the selection" of which wells leak. That is to say, we don't care much about the leak sizes for the leaks that don't occur. Manning, Duan, and Rogers (1987) find that this two-part estimation performs well, even when the true data generating process is the selection model. In the absence of valid instruments, the selection model performs poorly, even when the analyst knows the true specification.

In slight contrast to standard approaches, we estimate this model in a bootstrapped Bayesian framework (Huggins and Miller 2019). We draw 100 bootstrap datasets with replacement, and for each dataset we estimate 400 Markov chain Monte Carlo (MCMC) draws from the Bayesian posterior. We subset those 40,000 draws to 4000 draws by taking every 10th draw. For each draw we calculate outcomes (audit probability, DWL, etc.). To report point estimates and confidence intervals (CIs), we use the means and even-tailed 95% quantile ranges. This approach, called "BayesBag," has several advantages. First, we can easily estimate somewhat unusual models, with measurement error and the predicted leak size entering the leak probability. Second, Bayesian models handle uncertainty much more cleanly than standard frequentist models. (Meager 2019 highlights the benefits of this type of Bayesian modeling in an economics context.) Finally, the bootstrap provides some robustness to model mis-specification.

5.1 Models

Our modeling approach follows closely from our theory. We begin by estimating the distribution of leak sizes, then we use the wells' FOC to estimate the cost parameters.

We model the measured e_i as lognormal. The scientific literature that investigates the distribution of leak sizes tends to land on lognormal. The lognormal distribution expects positive probability everywhere above zero. Our leak measurements are censored at 5–10 kg/hr by the AVIRIS-NG sensitivity, so we actually model $e_i - \underline{e}$ as the dependent variable, where $\underline{e} = 5$ is the approximate censoring threshold. Define $e_i \in (\underline{e}, \infty)$ and X_i as the observed leak size and well covariates.[9] β and σ are parameters to estimate. We can model the leak distribution:

$$e_i \sim \text{LogNormal}(X_i\beta, \sigma) + \underline{e}$$
$$\hat{e}_i = \exp(X_i\hat{\beta}) \cdot \frac{1}{N}\sum_i \exp(\hat{\varepsilon}_i) - \underline{e}$$

The mean of the exponent of the residual ε_i is a semi-parametric smearing estimator (Manning, Duan, and Rogers 1987), designed to be more robust to cases where the outcome distribution is not lognormal.

For robustness, we consider that the AVIRIS-NG methane measurements have some noise. The data from the Duren et al. (2019) study (in California) report the estimated standard error for each leak measurement. The other measurements do not report measurement error. We impute the measurement error by taking the mean of the measurement error leak ratio (measurement error divided by monthly production), and apply this mean to the wells in Colorado and New Mexico. Using these reported and imputed measurement errors, we estimate a measurement error model, assuming that leak presence is measured accurately, but the size of the leak is measured with I.I.D. noise. We then estimate the parameters β that explain the distribution of the (unobserved) underlying leak sizes. The estimated parameters from the measurement error model are qualitatively similar to our other models; see table 9 for details.

After modeling the distribution of leak sizes when they occur, we use the expected values to predict A_i and α_i. This is a nonlinear, likelihood-based model that uses the cost function developed in section 3.1. Recall the marginal cost function $C'(q_i; A_i, \alpha_i) = A_i(1 - q_i)^{\alpha_i}$, where q_i is the probability well i will not leak. Define the observation of having a leak $d_i \in \{0, 1\}$. In our data, we have no tax, so wells set their marginal cost equal to $p_i e_i H$. In the equations below, we use the observed p_i and d_i, along with $\hat{e}_i$ from the leak size estimation and the period length H. These let us infer the cost parameters A and α.

We want to allow for heterogeneity in A_i and α_i with our covariates, while enforcing the acceptable ranges of these parameters, with $A_i \in (0, e_i p_i H)$ and $\alpha_i < -1$. There are many ways this could be done; for parsimony we use linear expressions $X\psi$ and $X\phi$ with an inverse logit transformation to scale the values. To be clear, we are not running a logistic regression, just using the inverse logit function. $\text{logit}^{-1}(X_i\psi)$ is in the range $(0, 1)$. We want A_i in the range $(0, e_i p_i H)$ for every i, so we multiply by $\underline{e}\, p_i H$.[10]

$$\text{FOC:} \quad p_i e_i H = A_i(1 - \tilde{q}_i)^{\alpha_i}$$

9. About notation: "~" indicates "is distributed as" and "=" indicates "is equal to." We use logit^{-1} as the inverse logit function, $\text{logit}^{-1}(z) = \exp(z)/(\exp(z) + 1)$.

10. A more intuitive approach would be to multiply by $\hat{e}_i p_i H$. We estimate e_i and A_i simultaneously, and we found that including the predicted value $\hat{e}_i$ in this way leads to poor estimation.

$$
\begin{aligned}
\text{Rearranged:}\quad 1 - \tilde{q}_i &= \left(\frac{p_i \hat{e}_i H}{A_i}\right)^{1/\alpha_i} \\
\text{Def.}\quad A_i &= \operatorname{logit}^{-1}(X_i \psi) \cdot p \cdot \underline{e} \cdot H \\
\text{Def.}\quad \alpha_i &= -1/\operatorname{logit}^{-1}(X_i \phi) \\
d_i &\sim \text{Bernoulli}(1 - \tilde{q}_i)
\end{aligned}
$$

The choice of X variables matters in this analysis, though for a different reason than in many analyses. They have some role in controlling for endogeneity (more below), but their role is at least as important in allowing for well heterogeneity. The variables we include are the inverse hyperbolic sine (IHS) of gas production per month when the measurement occurred, the IHS of oil production that month, geologic basin indicators, drilling direction indicators (to capture fracking), and the fraction of production from oil (in barrel of oil equivalents). We use IHS, rather than logs, for oil production because some wells produce no oil and we do not want to drop them. We use IHS for gas production only for symmetry with oil; we have dropped wells with zero reported gas production. These oil-only wells may still have methane emissions, but our private-benefit abatement model would be a poor fit for abatement behavior at these wells. Summary statistics for these variables are in table 1.

We employ a fully Bayesian model, including priors on the variables. Our goal in choosing priors is that they are very weakly informative on the outcome scale, following current Bayesian standard practices (Gelman et al. 2020). Specifically, we chose priors with mean zero and a standard deviation large enough that the predicted value of the outcomes e_i and q_i could take any reasonable value. For e_i, reasonable values are up to perhaps 100 times larger than the largest leak we see. For q_i we aimed for a roughly uniform prior distribution. The prior standard deviations are much smaller here; because of the logit transformation, making the prior standard deviations larger would put a lot of prior weight on probabilities near zero or one (see Gelman et al. 2020 for much more discussion). We use a Student's t distribution with three degrees of freedom to allow for somewhat more weight in the distributions' tails than the Normal. Specifically, we de-mean all of the X variables and use priors of Student's $t(3, 0, 3)$ for each of the leak size parameters β and σ. We use Normal$(0, 0.5)$ for the A_i coefficients (ψ) and Normal$(0, 0.75)$ for the α_i coefficients (ϕ). The prior covariance between coefficients are all zero.

The identifying variation comes from a couple of different types of heterogeneity in the covariates and outcomes. Let's first imagine we estimated homogeneous costs parameters, A and α the same for all wells. We could do this estimation with no covariates, finding the values of A and α that fit the overall, unconditional leak probability given the unconditional distribution of leak sizes. That estimate would be unbiased if there are no other factors that correlate with both the well's leak probability and leak size. We think it's important to allow for covariates—for example, wells with low levels of production have smaller detected leaks—so we define A_i and α_i. These heterogeneous parameters vary with the covariates X. The identifying variation here is that wells with different levels of the covariates have different leak probabilities, and different $\hat{e}$.

5.2 *Selection Bias*

Our analysis makes causal claims about counterfactual behavior: what would happen to leak probabilities if the cost of leaking increased? To identify these causal effects, we rely on a selection-on-observables assumption. There are two major ways selection bias could arise in our setting. The first is if the set of wells that were flown over are systematically different from other wells in the same basin. The second is if there are omitted variables that affect the wells' cost of avoiding leaks and are correlated with our estimated leak size.

We view the first case, sampling bias, as unlikely. The scientific teams planned their flight routes to sample a large fraction of the wells in the relevant basin. In choosing their flights, their goal was to have representative measurements, not to measure specific wells or find the largest leaks. Of course, there will be differences between geologic basins. We think these measurements are representative of the sampled basins, but more measurements would be necessary to draw conclusions about the national population of wells.

The second case, omitted variables bias, is a larger concern. As we said above, the identifying variation is that wells with different levels of the covariates have different leak probabilities, and different $\hat{e}$. We're assuming that conditional on leak size and the heterogeneity allowed in A_i and α_i, leak probability is independent of other factors affecting cost.

A counterexample, where omitted variable bias could occur, would be if wells near Los Angeles face higher labor costs, so have higher costs of increasing q, but also tend to leak less often because of their geology.[11] This correlation is not included in our model and would generate biased estimates. We do include basin indicators, but those are at a larger geography.

There's also potential for non-classical measurement error. The accuracy of measurement depends partly on the wind speed when the airplane is overhead. If this measurement error is correlated with other factors, such as well operators' costs or the commodity price of natural gas, then our estimates of the operators' abatement costs will be biased. For densely spaced wells, it's unlikely but possible that the leak is matched to the wrong well, another form of measurement error. We believe these issues of selection and measurement

11. To address this specific example we could include Bureau of Labor Statistics (BLS) estimates of county-level labor costs, but omitted variables would remain a concern.

error are small relative to the first-order effects we estimate—particularly because we focus on the relative gains of different policies instead of the specific dollar-value gains.

5.3 Fitted Values

To calculate the policy simulations, we need fitted values for $\hat{e}$, $\hat{A}_i$, $\hat{\alpha}_i$, and $\hat{q}_i(t_i)$. For the most part, we directly plug in our estimated coefficients. For $\hat{e}$, we use the smearing estimator mentioned above, which is a minor change to be more robust to violations of the assumed lognormal distribution. Recall H is the number of hours in the well operator's decision problem, T is the number of hours the regulator can assume a leak lasts, and t_i is the expected fee per hour per kilogram. Combining all of these, in the fourth line we can predict how a well's probability of having no leak will increase when they face a fee t_i.

$$\begin{aligned}
\hat{e}_i &= \exp(X_i\beta) \cdot \frac{1}{N}\sum_i \exp(\hat{\varepsilon}_i) + \underline{e} \\
\hat{A}_i &= \text{logit}^{-1}(X_i\hat{\psi}) \cdot p_i \cdot \underline{e} \cdot H \\
\hat{\alpha}_i &= -1/\text{logit}^{-1}(X_i\hat{\phi}) \\
\hat{q}_i(t_i) &= 1 - \left(\frac{(Hp_i + Tt_i)\hat{e}_i}{\hat{A}_i}\right)^{1/\hat{\alpha}_i}
\end{aligned}$$

In some of our policies, such as targeting on covariates (policy 2), we consider how choosing heterogeneous audits can decrease DWL. However, it would be an unfair comparison to target on $\hat{e}_i$, since this is an expected value, with lower variance than the observed distribution of e_i from wells that leaked. Therefore, we also calculate a random draw from the distribution of leak sizes. Call this draw e_i'. (Note that we do this even for wells with observed leak sizes). The corresponding probability of not having a leak is q_i'.

$$\begin{aligned}
e_i' &\sim \text{LogNormal}(X_i\hat{\beta}, \hat{\sigma}) + \underline{e} \\
q_i' &= 1 - \left(\frac{(Hp_i + Tt_i)e_i'}{\hat{A}_i}\right)^{1/\hat{\alpha}_i}
\end{aligned}$$

When we consider targeting our policies, we will target based on $\hat{e}_i$ and $\hat{q}_i$, but we will score the outcomes using e_i' and q_i'.

6 Policy Simulation

In this section we consider how different policies translate into different expected fees, and how those fees affect the DWL and emissions outcomes. We begin in 6.1 by examining how different policies and values of T and τ translate into the expected fee a well will pay. Then in section 6.2 we translate these fees into DWL and emissions outcomes.

6.1 Audit Probabilities and Expected Fees

Recall the policies we consider are: (0) no audits, the status quo; (1) audit every well with equal probability; (2) target audits based on well covariates; (3a and 3b) measure leaks remotely and target audits (with and without a detection threshold); and (4a and 4b) measure leaks remotely and assess fines based on measurements (with and without a detection threshold).

Tables 3 and 4 provide the expected fee as a fraction of the social cost δ, assuming 1% of wells are audited each year for the uniform, target covariates, and measure-then-audit policies (1, 2, and 3). The tables differ in their assumed value if T, the length of time the regulator is able to charge for emissions. In table 3, T = 1 week. In table 4, T = 3 months. As we mentioned earlier, only the product τT matters, so these two tables can be also be interpreted as keeping T the same and increasing τ by a factor of 13. If all wells could be audited, the first-best τT would be δH, higher than the cases we consider here.

The uniform results are unsurprising. For example, in the first line of the table, if 1% of wells are audited and the fee is $\tau T = 2\delta \cdot$ one week, then the expected fee as a percentage of δH is $100 \cdot 0.02 \cdot$ one week$/H = 0.384\%$. The mean, median, 10th and 90th percentile will all be equal. The confidence interval is a point, since the uniform fee does not depend on anything about the wells. The policy that targets on covariates is more interesting. This policy has the same mean because it has the same budget constraint, but the distribution of audit effort is no longer uniform across wells. Indeed, in these results, the 10th percentile of audit probability is zero—the audit budget constrains the regulator to focus effort on some wells and never audit others. The skew is more pronounced for when the allowed fee is more constrained (τ is lower). In panel C, even the 90th percentile of audit probability is zero. The regulator compensates for the low allowed fee by focusing all of the audit budget on a small fraction of wells.

6.2 Deadweight Loss and Emissions

With the fitted values in hand, we're able to calculate the DWL and emissions under each policy. We solve each policy's optimization problem numerically, choosing audit probabilities to minimize the DWL under the relevant constraints. Recall that the optimization problem is convex for the uniform and target-on-covariates policies, so a local optimum is guaranteed to be a global optimum. The optimization problem is not convex for the target-on-leaks policies, so we don't have the same guarantees. For each policy and each

Table 3: Expected fee, as a percentage of δ (1% annual audit budget and T = 1 week)

	Mean (%)	Median (%)	p10 (%)	p90 (%)
Panel A: High fee ($\tau = 2\delta$)				
Uniform	0.0384	0.0384	0.0384	0.0384
	[0.038,0.038]	[0.038,0.038]	[0.038,0.038]	[0.038,0.038]
Target covariates	0.0384	0	0	0
	[0.038,0.038]	[0,0]	[0,0]	[0,0]
Target leaks, low threshold	2.1	3.06	0	3.84
	[1.9,2.4]	[1.6,3.5]	[0,0]	[3.8,3.8]
Target leaks, high threshold	2.1	3.48	0	3.84
	[1.7,2.2]	[0,3.8]	[0,0]	[3.8,3.8]
Panel B: Medium fee ($\tau = \delta$)				
Uniform	0.0192	0.0192	0.0192	0.0192
	[0.019,0.019]	[0.019,0.019]	[0.019,0.019]	[0.019,0.019]
Target covariates	0.0192	0	0	0
	[0.019,0.019]	[0,0]	[0,0]	[0,0]
Target leaks, low threshold	0.929	0.533	0	1.92
	[0.84,0.98]	[0,1.2]	[0,0]	[1.9,1.9]
Target leaks, high threshold	0.929	0.693	0	1.92
	[0.78,1]	[0,1.6]	[0,0]	[1.9,1.9]
Panel C: Low fee (τ = \$5 per ton CO_2e)				
Uniform	0.00163	0.00163	0.00163	0.00163
	[0.0016,0.0016]	[0.0016,0.0016]	[0.0016,0.0016]	[0.0016,0.0016]
Target covariates	0.00163	0	0	0
	[0.0016,0.0016]	[0,0]	[0,0]	[0,0]
Target leaks, low threshold	0.0688	0	0	0.163
	[0.063,0.072]	[0,0]	[0,0]	[0.16,0.16]
Target leaks, high threshold	0.0666	0	0	0.163
	[0.059,0.074]	[0,0]	[0,0]	[0.16,0.16]

Note: Values are the expected fee per kg emitted, as a percentage of the social cost of emissions ($100\ T\tau r_i/\delta H$). Panels A, B, and C set T = 1 week and consider different values of τ. Each row considers different audit rules to optimally allocate r_i according to the fixed audit budget, which is set to 1% of all well pads. Columns provide distributional statistics across well pads. δ = \$2 per kg methane.
Wells in this table are the sample of wells included in the AVIRIS-NG sample (table 1 panel A). Point estimates and square brackets indicate the mean and 95% CI. (See text for CI details.)

Table 4: Expected fee, as a percentage of δ, with a 1% annual audit budget and $T = 3$ months

	Mean (%)	Median (%)	p10 (%)	p90 (%)
Panel A: High fee ($\tau = 2\delta$)				
Uniform	0.5	0.5	0.5	0.5
	[0.5,0.5]	[0.5,0.5]	[0.5,0.5]	[0.5,0.5]
Target covariates	0.5	0	0	1.64
	[0.5,0.5]	[0,0]	[0,0]	[1.5,1.8]
Target leaks, low threshold	50	50	50	50
	[50,50]	[50,50]	[50,50]	[50,50]
Target leaks, high threshold	29	50	0	50
	[27,30]	[50,50]	[0,0]	[50,50]
Panel B: Medium fee ($\tau = \delta$)				
Uniform	0.25	0.25	0.25	0.25
	[0.25,0.25]	[0.25,0.25]	[0.25,0.25]	[0.25,0.25]
Target covariates	0.25	0	0	0.863
	[0.25,0.25]	[0,0]	[0,0]	[0.68,0.92]
Target leaks, low threshold	24.9	25	25	25
	[25,25]	[25,25]	[25,25]	[25,25]
Target leaks, high threshold	14.5	25	0	25
	[14,15]	[25,25]	[0,0]	[25,25]
Panel C: Low fee (τ = \$5 per ton CO_2e)				
Uniform	0.0213	0.0213	0.0213	0.0213
	[0.021,0.021]	[0.021,0.021]	[0.021,0.021]	[0.021,0.021]
Target covariates	0.0213	0	0	0
	[0.021,0.021]	[0,0]	[0,0]	[0,0]
Target leaks, low threshold	1.04	0.811	0	2.12
	[0.94,1.1]	[0.12,1.5]	[0,0]	[2.1,2.1]
Target leaks, high threshold	1.05	0.999	0	2.12
	[0.87,1.1]	[0,2.1]	[0,0]	[2.1,2.1]

NOTE: Values are the expected fee per kg emitted, as a percentage of the social cost of emissions ($100\ T\tau r_i/\delta H$). Panels A, B, and C set $T = 1$ week and consider different values of τ. Each row considers different audit rules to optimally allocate r_i according to the fixed audit budget, which is set to 1% of all well pads. Columns provide distributional statistics across well pads. δ = \$2 per kg methane.
Wells in this table are the sample of wells included in the AVIRIS-NG sample (table 1 panel A). Point estimates and square brackets indicate the mean and 95% CI. (See text for CI details.)

set of parameter values, we repeat the optimization process 4000 times, once for each draw.

In choosing the audit probabilities, we use the expected value $\hat{e}_i$. However, variance matters here, since we're thinking about heterogeneous targeting, and there's less variance in the expected value $\hat{e}_i$. Therefore to *score* the policy in terms of DWL and emissions, we use a draw from the conditional distribution of e_i conditional on X_i. When relevant, we also use the draw to determine if the leak is above the high detection threshold.

Tables 5 and 6 present results for the same policies, with the same annual audit budget, at different levels of stringency. In table 5, the allowed fee when a leak is detected is $\tau \times 1$ week (with τ different values of τ in the columns). Table 6 presents results of a fee that's 13 times larger, $\tau \times 3$ months, again considering different values of τ. Recall that wells are audited at most once per year. In this setup, the first best could be achieved by auditing every well and charging a fee of $\tau T = \delta H$ when a leak was detected. Both table 5 and 6 are lower than this, but we think they provide realistic values of the range of fees that could be assessed.

In these results, it's clear that the allowed fee matters a great deal. In table 5, with $T = 1$ week, none of the policies do particularly well, even in the infeasible case where they're able to measure all leaks remotely and charge for those emissions—the fee is just too low. In the very best feasible case, policy 3b, these policies move 24.7 percent of the way from the no-policy DWL to the first best. With a higher fee, the same policy in table 6 is able to move 72.1 percent of the way to first best.

Turning to emissions, we see a similar pattern, with an average reduction of 256 tons CO_2e per well per year for policy 3b with $\tau T = 2\delta \cdot 1$ week and 832 tons CO_2e per well per year for policy 3b with $\tau T = 2\delta \cdot 3$ months. In the tables, we present both emissions and DWL outcomes on a scale from zero to 100 for ease of comparison. On this scale, zero is the no-policy result and 100 is the outcome under first-best Pigouvian taxation.

In these results, the uniform policy does worse than we expected, particularly in table 5. Because leaks are difficult to predict, we expected the policies of uniform audits and targeting on covariates to perform similarly, but the targeting does substantially better in relative terms.

Targeting on observed leaks (policy 3a and 3b) do well relative to the other audit policies, which was expected from the fact that they're able to use the limited audit budget more effectively. The high-threshold policies—both the infeasible remote fee and the feasible target-on-observed leaks—capture a large fraction of the gains that could be achieved by the low-threshold policies. Because these are only able to target the largest leaks, we had expected them to do substantially worse, but that seems not to be the case.

Table 5: Policy outcomes: Percent improvement from no-policy baseline (Audit budget = 1% per year, T = 1 week)

	$\tau = 2\delta$	$\tau = \delta$	$\tau = \$5$
A: DWL improvement (%)			
Uniform	0.522	0.257	0.0127
	[0.502,0.537]	[0.246,0.264]	[0.0117,0.014]
Target covariates	1.21	0.656	0.055
	[1,1.54]	[0.531,0.861]	[0.0429,0.0766]
Target leaks, low threshold	25.2	13.6	1.24
	[23.2,26.5]	[12.4,14.2]	[1.11,1.3]
Target leaks, high threshold	24.7	13.5	1.25
	[23.6,26.3]	[12.7,14.5]	[1.16,1.33]
Remote, low threshold	36.2	21.6	2.23
	[35.2,36.9]	[20.9,22.1]	[2.14,2.29]
Remote, high threshold	25.7	15.3	1.57
	[25.2,26.4]	[15,15.7]	[1.54,1.61]
B: $\mathbb{E}$[emiss] improvement (%)			
Uniform	0.438	0.219	0.0187
	[0.417,0.452]	[0.209,0.227]	[0.0178,0.0193]
Target covariates	1.01	0.551	0.0534
	[0.839,1.29]	[0.448,0.721]	[0.0436,0.071]
Target leaks, low threshold	21.1	11.3	1.03
	[19.3,22.3]	[10.2,11.8]	[0.917,1.08]
Target leaks, high threshold	20.7	11.2	1.03
	[19.7,22.1]	[10.5,12.1]	[0.953,1.1]
Remote, low threshold	30.3	17.9	1.83
	[29.2,31.1]	[17.2,18.4]	[1.75,1.89]
Remote, high threshold	21.5	12.7	1.29
	[21.2,22.1]	[12.5,13.1]	[1.28,1.33]

NOTE: Panels A and B show results for DWL and emissions, both on a scale from 0 to 100, where 0 is the no-policy baseline and 100 is the outcome of the infeasible first-best Pigouvian tax (higher is better). Columns show are different policy stringency levels $\tau = \{2\delta, \delta, \$5\}$. Rows are different constrained policy options, listed previously. DWL numbers include the costs of auditing. Wells in this table are the sample of wells included in the AVIRIS-NG sample. Square brackets indicate 95% CI.

Table 6: Policy outcomes: Percent improvement from no-policy baseline (Audit budget = 1% per year, T = 3 months)

	$\tau = 2\delta$	$\tau = \delta$	$\tau = \$5$
A: DWL improvement (%)			
Uniform	6.55	3.37	0.286
	[6.31,6.72]	[3.24,3.46]	[0.274,0.294]
Target covariates	10.2	5.8	0.719
	[9.55,10.9]	[5.28,6.47]	[0.584,0.941]
Target leaks, low threshold	96.4	85.4	15
	[96.1,96.6]	[84.6,86.1]	[13.6,15.6]
Target leaks, high threshold	72.1	62.8	14.8
	[70.2,74.5]	[61.2,64.9]	[14,15.9]
Remote, low threshold	96.4	85.5	23.4
	[96.1,96.6]	[84.8,86.1]	[22.7,23.9]
Remote, high threshold	69.7	61.6	16.6
	[67.5,72.7]	[59.8,64]	[16.3,17]
B: $\mathbb{E}$[emiss] improvement (%)			
Uniform	5.41	2.78	0.243
	[5.17,5.58]	[2.66,2.87]	[0.232,0.251]
Target covariates	8.49	4.83	0.603
	[7.96,9.13]	[4.39,5.43]	[0.492,0.788]
Target leaks, low threshold	89.8	75.9	12.4
	[89.2,90.3]	[74.6,76.9]	[11.3,13]
Target leaks, high threshold	67.1	55.8	12.3
	[65.5,69.1]	[54.5,57.3]	[11.6,13.3]
Remote, low threshold	89.8	76	19.5
	[89.2,90.3]	[74.8,76.9]	[18.7,20]
Remote, high threshold	65.1	54.8	13.8
	[63.2,67.6]	[53.4,56.6]	[13.6,14.2]

NOTE: Panels A and B show results for DWL and emissions, both on a scale from 0 to 100, where 0 is the no-policy baseline and 100 is the outcome of the infeasible first-best Pigouvian tax (higher is better). Columns show are different policy stringency levels $\tau = \{2\delta, \delta, \$5\}$. Rows are different constrained policy options, listed previously. DWL numbers include the costs of auditing. Wells in this table are the sample of wells included in the AVIRIS-NG sample. Square brackets indicate 95% CI.

7 *Conclusion*

The five policies we consider, ranging from no audits to fully remotely assessed fees, offer a sampling of feasible policies a regulator could implement. All of these policies fall short of the first best, some dramatically so. However, we can also see the importance of the fine level—a \$5 per ton CO_2e fine is not very effective, particularly when audit probabilities are low. Importantly, we found that the gains a policy that used remote sensing to guide audits did quite well, almost as well as the infeasible policy where fines are assessed without a corroborating on-the-ground audit. This remote-based policy does somewhat worse when measurements have a high detection threshold, but overall the policy still does quite well, because it still allows the regulator to target their auditing effort much more effectively than targeting on covariates alone.

All of these policies are focused on the large, infrequent leaks that are measured in the AVIRIS-NG datasets. Smaller leaks are more frequent, and while they make up a minority of well emissions, it's worth targeting policy at these leaks as well. Such a policy could use audits, like the ones we consider above, or might use some other tool like a stringent LDAR mandate or a technology standard on components that can leak.

To generate our estimates, we use the observed leaks to estimate the distribution of leak sizes and the well operators costs of abatement, using a Bayesian bagging model to simultaneously estimate the leak size and wells' costs. We then consider how different policies translate into incentives the wells face, and the operators would change their abatement under each policy. We calculate the expected DWL and changes in emissions for policies implemented under a number of different constraints, from limits on the fee that can be charged, to the number of audits that can be performed, to the size of leaks that can be detected remotely. These limits matter a great deal to how effective the policy can be.

Taking a step back, our findings highlight the importance of thinking about measurement an policy together. We found that additional information on leaks can dramatically improve social outcomes. At the same time, the regulatory details matter a great deal—the policies that use more information still perform poorly when the regulator's ability to charge fees is severely constrained. Our work contributes to a broader literature on the role of measurement in determining policy outcomes.

References

Alix-Garcia, Jennifer, and Daniel L Millimet. 2020. "Remotely Incorrect?" August 24, 2020. Accessed October 20, 2020. https://faculty.smu.edu/millimet/pdf/ri.pdf. (Cited on page 5).

Allen, Myles R., Keith P. Shine, Jan S. Fuglestvedt, Richard J. Millar, Michelle Cain, David J. Frame, and Adrian H. Macey. 2018. "A solution to the misrepresentations of CO_2-equivalent emissions of short-lived climate pollutants under ambitious mitigation." *npj Climate and Atmospheric Science* 1, no. 1 (June). https://doi.org/10.1038/s41612-018-0026-8. (Cited on page 7).

Allingham, Michael G., and Agnar Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics,* 323–338. http://www.academia.edu/download/30457905/allingham-sandmo_(jpube72).pdf. (Cited on page 5).

Alvarez, Ramón A., Daniel Zavala-Araiza, David R. Lyon, David T. Allen, Zachary R. Barkley, Adam R. Brandt, Kenneth J. Davis, et al. 2018. "Assessment of methane emissions from the U.S. oil and gas supply chain." *Science* (June): eaar7204. https://doi.org/10.1126/science.aar7204. (Cited on pages 2, 7, 9, 21).

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76, no. 2 (March): 169–217. https://doi.org/10.1086/259394. (Cited on page 9).

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer. 2020. "Escalation of Scrutiny: The Gains from Dynamic Enforcement of Environmental Regulations." *American Economic Review* 110, no. 8 (August): 2558–2585. https://doi.org/10.1257/aer.20181012. (Cited on page 4).

Boomhower, Judson. 2019. "Drilling Like There's No Tomorrow: Bankruptcy, Insurance, and Environmental Risk." *American Economic Review* 109, no. 2 (February): 391–426. https://doi.org/10.1257/aer.20160346. (Cited on page 2).

Chan, Elton, Douglas E. J. Worthy, Douglas Chan, Misa Ishizawa, Michael D. Moran, Andy Delcloo, and Felix Vogel. 2020. "Eight-Year Estimates of Methane Emissions from Oil and Gas Operations in Western Canada Are Nearly Twice Those Reported in Inventories." *Environmental Science & Technology* (November 10, 2020). https://doi.org/10.1021/acs.est.0c04117. (Cited on page 9).

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104, no. 9 (September): 2593–2632. https://doi.org/10.1257/aer.104.9.2593. (Cited on page 5).

Cicala, Steve, David Hémous, and Morten Olsen. 2019. "Advantageous Selection as a Policy Instrument:Unraveling Climate Change." November 6, 2019. Accessed November 7, 2019. http://home.uchicago.edu/~scicala/papers/unraveling/unraveling_climate_change_draft.pdf. (Cited on page 4).

Cusworth, Daniel H., Daniel J. Jacob, Daniel J. Varon, Christopher Chan Miller, Xiong Liu, Kelly Chance, Andrew K. Thorpe, et al. 2019. "Potential of next-generation imaging spectrometers to detect and quantify methane point sources from space." *Atmospheric Measurement Techniques* 12, no. 10 (October): 5655–5668. https://doi.org/10.5194/amt-12-5655-2019. (Cited on page 4).

Duren, Riley M., Andrew K. Thorpe, Kelsey T. Foster, Talha Rafiq, Francesca M. Hopkins, Vineet Yadav, Brian D. Bue, et al. 2019. "California's methane super-emitters." *Nature* 575, no. 7781 (November): 180–184. https://doi.org/10.1038/s41586-019-1720-3. (Cited on pages 21, 22, 24, 28).

Enverus. 2019. September 1, 2019. https://drillinginfo.com. (Cited on page 24).

EPA. 2020a. "Emissions & Generation Resource Integrated Database (eGRID) 2018." https://www.epa.gov/sites/production/files/2020-01/documents/egrid2018_summary_tables.pdf. (Cited on pages 2, 7).

———. 2020b. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2018.* https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2018. (Cited on page 2).

EPA. 2020. "Oil and Natural Gas Sector: Emission Standards for New, Reconstructed, and Modified Sources Reconsideration." *Federal Register* 85, no. 179 (September 15, 2020). Accessed November 1, 2020. https://www.federalregister.gov/d/2020-18115/p-305. (Cited on page 8).

Frankenberg, Christian, Andrew K. Thorpe, David R. Thompson, Glynn Hulley, Eric Adam Kort, Nick Vance, Jakob Borchardt, et al. 2016. "Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region." *Proceedings of the National Academy of Sciences* 113, no. 35 (August): 9734–9739. https://doi.org/10.1073/pnas.1605617113. (Cited on pages 21, 22, 24, 25).

Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. "Bayesian Workflow" (November 3, 2020). arXiv: 2011.01808 [stat.ME]. (Cited on page 29).

Hausfather, Zeke. 2015. "Bounding the climate viability of natural gas as a bridge fuel to displace coal." *Energy Policy* 86 (November): 286–294. https://doi.org/10.1016/j.enpol.2015.07.012. (Cited on page 2).

Huggins, Jonathan H., and Jeffrey W. Miller. 2019. "Robust Inference and Model Criticism Using Bagged Posteriors," arXiv: 1912.07104 [stat.ME]. (Cited on page 27).

Lyon, David R., Ramón A. Alvarez, Daniel Zavala-Araiza, Adam R. Brandt, Robert B. Jackson, and Steven P. Hamburg. 2016. "Aerial Surveys of Elevated Hydrocarbon Emissions from Oil and Gas Production Sites." *Environmental Science & Technology* 50, no. 9 (April): 4877–4886. https://doi.org/10.1021/acs.est.6b00705. (Cited on pages 3, 9, 22, 24, 25).

Manning, Willard G, Naihua Duan, and William H Rogers. 1987. "Monte Carlo evidence on the choice between sample selection and two-part models." *Journal of econometrics* 35 (1): 59–82. (Cited on pages 27, 28).

Marks, Levi. 2018. "The Abatement Cost of Methane Emissions from Natural Gas Production." November 19, 2018. Accessed December 29, 2018. http://www.levimarks.com/marks_ml_abatement.pdf. (Cited on pages 6, 13).

———. 2019. "A Sampling-Based Approach to Emissions Pricing." October 27, 2019. Accessed November 12, 2019. http://www.levimarks.com/marks_sampling_based_pricing.pdf. (Cited on page 6).

Meager, Rachael. 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics* 11, no. 1 (January): 57–91. https://doi.org/10.1257/app.20170299. (Cited on page 27).

Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, et al. 2013. "Anthropogenic and Natural Radiative Forcing." Chap. 8 in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* edited by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley, 659–740. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.018. (Cited on page 7).

Nagaraj, Abhishek. 2020. "The Private Impact of Public Data:Landsat Satellite Maps and Gold Exploration." March 5, 2020. Accessed April 16, 2020. http://abhishekn.com/files/nagaraj_landsat2020.pdf. (Cited on page 5).

Oestreich, Andreas Marcel. 2017. "On optimal audit mechanisms for environmental taxes." *Journal of Environmental Economics and Management* 84 (July): 62–83. https://doi.org/10.1016/j.jeem.2017.02.005. (Cited on page 4).

Omara, Mark, Naomi Zimmerman, Melissa R. Sullivan, Xiang Li, Aja Ellis, Rebecca Cesa, R. Subramanian, Albert A. Presto, and Allen L. Robinson. 2018. "Methane Emissions from Natural Gas Production Sites in the United States: Data Synthesis and National Estimate." *Environmental Science & Technology* 52, no. 21 (September): 12915–12925. https://doi.org/10.1021/acs.est.8b03535. (Cited on page 23).

Pigou, Arthur Cecil. 1932. *The Economics of Welfare.* 4th ed. Chap. 9, bk. 2. Macmillan / Co. (Cited on page 9).

Polinsky, A Mitchell, and Steven Shavell. 1979. "The Optimal Tradeoff between the Probability and Magnitude of Fines." *The American Economic Review* 69, no. 5 (December): 880–891. (Cited on page 9).

Rabe, Barry, Claire Kaliban, and Isabel Englehart. 2020. "Taxing Flaring and the Politics of State Methane Release Policy." *Review of Policy Research* 37, no. 1 (January): 6–38. https://doi.org/10.1111/ropr.12369. (Cited on page 2).

Ravikumar, Arvind P, Daniel Roda-Stuart, Ryan Liu, Alexander Bradley, Joule Bergerson, Yuhao Nie, Siduo Zhang, Xiaotao Bi, and Adam R Brandt. 2020. "Repeated leak detection and repair surveys reduce methane emissions over scale of years." *Environmental Research Letters* 15, no. 3 (February 26, 2020): 034029. https://doi.org/10.1088/1748-9326/ab6ae1. (Cited on page 6).

Robertson, Anna M., Rachel Edie, Robert A. Field, David Lyon, Renee McVay, Mark Omara, Daniel Zavala-Araiza, and Shane M. Murphy. 2020. "New Mexico Permian Basin Measured Well Pad Methane Emissions Are a Factor of 5–9 Times Higher Than U.S. EPA Estimates." *Environmental Science & Technology* (October 15, 2020). https://doi.org/10.1021/acs.est.0c02927. (Cited on page 6).

Rutherford, Jeffrey, Evan Sherwin, Arvind Ravikumar, Garvin Heath, Jacob Englander, Daniel Cooley, David Lyon, Mark Omara, Quinn Langfitt, and Adam Brandt. 2020. "Closing the gap: Explaining persistent underestimation by US oil and natural gas production-segment methane inventories" (November 3, 2020). https://doi.org/10.31223/x5jc7t. (Cited on page 9).

Segerson, Kathleen. 1988. "Uncertainty and incentives for nonpoint pollution control." *Journal of Environmental Economics and Management* 15, no. 1 (March): 87–98. https://doi.org/10.1016/0095-0696(88)90030-7. (Cited on pages 5, 9, 10).

US EPA. 2016. "The Social Cost of Carbon: Estimating the Benefits of Reducing Greenhouse Gas Emissions," August. Accessed January 8, 2018. https://19january2017snapshot.epa.gov/climatechange/social-cost-carbon_.html. (Cited on page 7).

Zavala-Araiza, Daniel, Ramón A Alvarez, David R. Lyon, David T. Allen, Anthony J. Marchese, Daniel J. Zimmerle, and Steven P. Hamburg. 2017. "Super-emitters in natural gas infrastructure are caused by abnormal process conditions." *Nature Communications* 8, no. 1 (January 16, 2017). https://doi.org/10.1038/ncomms14012. (Cited on page 3).

Part I
Appendix

A Proofs

A.1 Proofs for the Well Operator's Problem

Proposition 1 (Properties of DWL)

DWL$_i$ is decreasing and convex in t_i for our assumed cost function and any other that satisfies our basic assumptions (twice continuously differentiable on $q \in (0,1)$, $C'(q) > 0$, $C''(q) > 0$, and $\lim_{q \to 1} C'(q) = \infty$).

Proof: Because C' is strictly increasing and convex, $C'^{-1}(x) =: f(x)$ is strictly increasing and concave.

$$\frac{\partial \text{DWL}_i}{\partial t_i} = -\frac{\partial C'^{-1}(e_i \cdot (p_i + t_i))}{\partial t_i} \cdot e_i(\delta - t_i) = -f' \cdot e_i^2(\delta - t_i) < 0$$

$$\frac{\partial^2 \text{DWL}_i}{\partial t_i^2} = -\frac{\partial f' \cdot e_i^2(\delta - t_i)}{\partial t_i} = -e_i^2(\underbrace{f'' \cdot e_i(\delta - t_i)}_{<0} - f') > 0$$

□

A.2 Audit Probability Proofs

Define $G_i(r_i)$ as the budget consumption function.

$$G_i(r_i) = \begin{cases} r_i & \text{for target-covariates} \\ (1 - q_i(r_i))r_i & \text{for target-leaks} \end{cases}$$

Rewriting the Lagrangians for policies 3 and 4 with $G_i(r_i)$:

$$\min_{\{r_i\}_{i=1}^N} \mathscr{L} = \sum_i \text{DWL}_i(r_i) + \lambda\left(M - \sum_i G_i(r_i)\right) + \sum_i (r_i - 0)\, a_i + (r_i - 1)\, b_i$$

A.2.1 Target Auditing on Well Covariates

Proposition 2 (Monotone Audit Rule)

If $\frac{\partial^2 \mathscr{L}}{\partial r_i \partial e_i} < 0$, then the solution of is monotonically increasing in e_i.

Proof: Let r_i^*s be the solution of the problem. Suppose there exists k, j such that $e_k > e_j$ but $r_k^* < r_j^*$. Consider $\hat{r}_i$s such that $\hat{r}_i = r_i^*$ for all $i \neq k, j$ and $\hat{r}_k = r_j^*, \hat{r}_j = r_k^*$. Clearly, r_i^* also satisfy all the constraints. The difference in the total DWL for $\hat{r}_i$s and r_i^*s is equal to

$$\begin{aligned}
&\mathrm{DWL}_k(\hat{r}_k) + \mathrm{DWL}_j(\hat{r}_j) - (\mathrm{DWL}_k(r_k^*) + \mathrm{DWL}_j(r_j^*)) \\
=&\mathrm{DWL}_k(r_j^*) - \mathrm{DWL}_k(r_k^*) + \mathrm{DWL}_j(r_k^*) - \mathrm{DWL}_j(r_j^*) \\
=&\int_{r_k^*}^{r_j^*} \frac{\partial \mathrm{DWL}_k}{\partial r}\, dr + \int_{r_j^*}^{r_k^*} \frac{\partial \mathrm{DWL}_j}{\partial r}\, dr \\
=&\int_{r_k^*}^{r_j^*} \frac{\partial \mathrm{DWL}_k}{\partial r} - \frac{\partial \mathrm{DWL}_j}{\partial r}\, dr
\end{aligned}$$

Since $\frac{\partial^2 \mathscr{L}}{\partial r_i \partial e_i} < 0$ and $e_k > e_j$, the integrand is negative and hence the whole integral is negative, which implies DWL under $\hat{r}_i$s is small. This a contradiction to r_i^*s being optimal. □

Note that for the target-on-covariates policy, $\frac{\partial^2 \mathscr{L}}{\partial r_i \partial e_i} = \frac{\partial^2 \mathrm{DWL}_i}{\partial r_i \partial e_i} < 0$. The inequality follows directly from the specific choice function we chose and the expected fee t_i being an increasing function of r_i,

$$\frac{\partial^2 DWL_i}{\partial t_i \partial e_i} = (1 + \frac{1}{\alpha}) e_i^{\frac{1}{\alpha}} \frac{(\delta - t_i)}{\alpha(p_i + t_i)} (\frac{(p_i + t_i)}{A})^{\frac{1}{\alpha}} < 0$$

Therefore, for this policy, the optimal values of r_i are monotonic in e_i.

A.2.2 Target Auditing on observed emissions

A.3 Expressions with time

Here we present the DWL and audit problems with variables T and H included. See the main text section 3.3 for details.

$$\begin{aligned}
&C_i'(q_i) = A_i(1 - q_i)_i^{\alpha} \quad \text{(cost function unchanged)} \\
&C'(q_i) = (p_i H + t_i T) e_i \quad \text{(FOC MB changed)} \\
&\mathrm{DWL}_i = \left((p_i + \delta) H e_i - \frac{(H p_i + T \tau r_i)\, e_i}{\alpha_i + 1} \right) \left(\frac{(H p_i + T \tau r_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} - \frac{\alpha_i}{1 + \alpha_i} (p_i + \delta) H e_i \left(\frac{(p_i + \delta) H e_i}{A_i} \right)^{\frac{1}{\alpha_i}}
\end{aligned}$$

We turn to the audit policies with audit budget M. For uniform and targeting on covariates (policies 1 and 2), the budget is binding if $\frac{M}{N} < \frac{\delta H}{\tau T}$. The budget constraint remains $\sum r_i \leq M$.

The budget constraint for the measure-then-audit policy (3) with no detection threshold becomes:

$$\begin{aligned}
\underline{e} = 0 \ : \quad & \sum_i \left(\frac{(p_i H + \tau T r_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} r_i \leq M \\
\underline{e} > 0 \ : \quad & \sum_i z_i \left[\left(\frac{(p_i H + \tau T r_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} r_i + \left(1 - \left(\frac{(p_i H + \tau T r_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} \right) s_i \right] + (1 - z_i) s_i \leq M
\end{aligned}$$

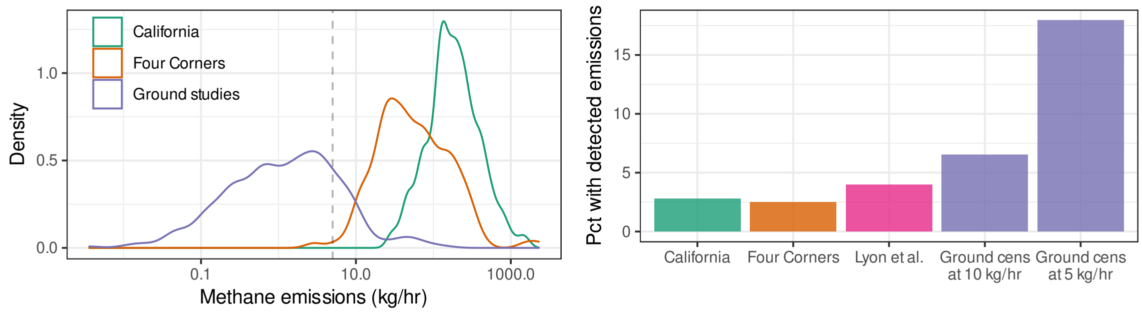
B Methane Measurement

Table 7: Estimated satellite detection varies by leak size and background

Surface type	True emissions (kg/hr)	Estimated emissions (kg/hr)
Grass	100	No detection
Grass	500	279 (101)
Grass	900	542 (38)
Bright	100	93.5 (18.3)
Bright	500	338 (83.1)
Bright	900	577 (115)

NOTE: Table is a subset of Cusworth et al. (2019) table 2 (CC BY 4.0). Results simulate methane retrievals from the EnMAP satellite, expected to launch in 2021. Values in parentheses are standard deviations from five iterations. We exclude the paper's results for images with "dark" or "urban" backgrounds, as these include water and confuse the image processing algorithm. In personal communication, the lead author notes "one could dramatically improve the prediction if there were some sort of decision tree that was based on the underlying surface."

Figure 8: Distribution of detected methane leaks, comparison with ground-based measurement



Left: emissions conditional on detection. Right: fraction of well pads with detected emissions. The "ground cens at" columns are the ground studies' observations with artificial censoring applied, at either 5 or 10 kg/hr, the approximate detection threshold of both the California and Four Corners studies. Without artificial censoring, the ground-based measurements are non-zero approximately 97% of the time. 5 kg/hr is noted with a dashed line in the left plot.

Sources: Ground studies include measurements primarily from Robertson et al. (2017) with additional contributions from Rella et al. (2015), Omara et al. (2016), and Omara et al. (2018).

California and Four Corners distributions come from aircraft studies (Duren et al. 2019; Frankenberg et al. 2016). Lyon et al. (2016) provides information about leak prevalence (with a detection threshold roughly similar to the California and Four Corners studies), but not leak size.

C *Distribution Fitting*

Table 8: Parameters of methane leak size models

	LogNormal	LogNormal meas. err.	Cost param. (common)	Cost param. (heterog.)
Intercept	4.39	4.02	6.15	4.71
	[2.7,6.1]	[3,5]	[4.4,7.7]	[3.1,6.3]
IHS of gas prod (mcfd)	0.111	0.152	−0.19	0.0466
	[−0.066,0.28]	[0.043,0.27]	[−0.37,0.021]	[−0.13,0.21]
IHS of oil prod (bbld)	−0.0329	−0.0828	0.181	0.015
	[−0.22,0.16]	[−0.21,0.051]	[0.0069,0.37]	[−0.16,0.23]
Basin: San Joaquin	−0.193	−0.0967	−0.0806	−0.0853
	[−0.94,0.52]	[−0.57,0.39]	[−0.77,0.49]	[−0.81,0.6]
Basin: San Juan	−1.15	−1.26	−0.357	−0.993
	[−2.1,−0.24]	[−1.9,−0.61]	[−1.1,0.38]	[−1.9,−0.15]
Oil prod share	0.733	1.01	−1.35	0.262
	[−0.5,1.9]	[0.2,1.9]	[−2.6,0.08]	[−1,1.4]
IHS of age (yr)	0.0637	0.0678	0.00477	0.0754
	[−0.12,0.25]	[−0.053,0.19]	[−0.15,0.16]	[−0.094,0.25]
Drill: Horizontal	0.266	0.214	0.128	0.161
	[−0.52,1.1]	[−0.4,0.81]	[−0.55,0.97]	[−0.6,1]
Drill: Unknown	−0.621	−0.712	−1.25	−0.743
	[−1.2,0.00023]	[−1.1,−0.28]	[−1.7,−0.76]	[−1.4,−0.12]
Drill: Vertical	−0.0859	−0.165	−0.0313	−0.151
	[−0.47,0.3]	[−0.43,0.11]	[−0.37,0.28]	[−0.54,0.23]
σ	0.955	0.797	1.05	0.959
	[0.83,1.1]	[0.72,0.88]	[0.9,1.2]	[0.84,1.1]
N	14399	14399	14399	14399
R^2	0.21	0.13	0.13	0.21
Dep. var. mean	198	198	198	198

NOTE: LogNormal coefficients are on the log scale, so numbers are roughly comparable across models. Square brackets are 95% CI. Omitted category for drilling is directional. Omitted category for basin is all of California outside the San Joaquin basin.
SOURCES: See figure 1.

Table 9: Parameters of methane leak occurrence models

	LogNormal	LogNormal meas. err.	Cost param. (heterog)
Intercept	−5.83	−5.87	−0.774
	[−7.4,−4.4]	[−7,−4.8]	[−5.1,3.6]
IHS of gas prod (mcfd)	0.328	0.331	−0.0643
	[0.17,0.49]	[0.21,0.45]	[−1.1,1]
IHS of oil prod (bbld)	−0.177	−0.172	0.433
	[−0.37,0.014]	[−0.31,−0.033]	[−0.51,1.2]
Basin: San Joaquin	−0.16	−0.148	0.389
	[−0.78,0.59]	[−0.61,0.34]	[−0.57,1.4]
Basin: San Juan	−0.328	−0.298	−0.574
	[−1.1,0.48]	[−0.89,0.31]	[−1.6,0.4]
Oil prod share	2.17	2.2	0.474
	[1.1,3.2]	[1.4,3]	[−0.51,1.4]
IHS of age (yr)	0.124	0.119	−0.2
	[−0.069,0.35]	[−0.015,0.26]	[−1.2,0.66]
Drill: Horizontal	0.127	0.169	−0.0426
	[−0.87,0.94]	[−0.49,0.78]	[−0.98,0.9]
Drill: Unknown	1.52	1.53	0.37
	[0.92,2.1]	[1.1,2]	[−0.65,1.3]
Drill: Vertical	−0.0696	−0.0619	−0.49
	[−0.44,0.33]	[−0.35,0.22]	[−1.5,0.59]
N	14399	14399	14399
R^2	0.015	0.014	0.019
Dep. var. mean	0.0267	0.0267	0.0267

Note: Coefficients are on the logit scale. Cost param. model coefficients aren't comparable. Square brackets are 95% CI. Omitted category for drilling is directional. Omitted category for basin is all of California outside the San Joaquin basin.
Sources: See figure 1.

Software Citations

Bache, Stefan Milton, and Hadley Wickham. 2014. *magrittr: A Forward-Pipe Operator for R.* 1.5. https://CRAN.R-project.org/package=magrittr.

Bürkner, Paul-Christian, Jonah Gabry, Matthew Kay, and Aki Vehtari. 2020. *posterior: Tools for Working with Posterior Distributions.* 0.1.2. https://mc-stan.org/posterior.

Dask Development Team. 2016. *Dask: Library for dynamic task scheduling.* https://dask.org.

Dorman, Michael. 2020. *nngeo: k-Nearest Neighbor Join for Spatial Data.* 0.3.8. https://CRAN.R-project.org/package=nngeo.

Dowle, Matt, and Arun Srinivasan. 2020. *data.table: Extension of 'data.frame'.* 1.13.0. https://CRAN.R-project.org/package=data.table.

Dunkle Werner, Karl. 2020. *kdw.junk: Miscellaneous Functions Karl Found Useful.* 0.0.30. https://github.com/karldw/kdw.junk/.

François, Romain, Jeroen Ooms, Neal Richardson, and Apache Arrow. 2020. *arrow: Integration to 'Apache' 'Arrow'.* 1.0.0. https://CRAN.R-project.org/package=arrow.

Gabry, Jonah, and Rok Češnovar. 2020. *cmdstanr: R Interface to 'CmdStan'.* https://mc-stan.org/cmdstanr.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. http://www.jstatsoft.org/v40/i03/.

Henry, Lionel, and Hadley Wickham. 2020. *purrr: Functional Programming Tools.* 0.3.4. https://CRAN.R-project.org/package=purrr.

Hester, Jim. 2020. *glue: Interpreted String Literals.* 1.4.1. https://CRAN.R-project.org/package=glue.

Klik, Mark. 2020. *fst: Lightning Fast Serialization of Data Frames for R.* 0.9.2. https://CRAN.R-project.org/package=fst.

Köster, Johannes, and Sven Rahmann. 2018. "Snakemake—a scalable bioinformatics workflow engine." *Bioinformatics* 34, no. 20 (May): 3600–3600. https://doi.org/10.1093/bioinformatics/bty350.

Langa, Łukasz, et al. 2020. *Black: The uncompromising code formatter.* https://black.readthedocs.io.

McKinney, Wes, et al. 2010. "Data structures for statistical computing in python." In *Proceedings of the 9th Python in Science Conference,* 445:51–56. Austin, TX.

Müller, Kirill. 2017. *here: A Simpler Way to Find Your Files.* 0.1. https://CRAN.R-project.org/package=here.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–446. https://doi.org/10.32614/RJ-2018-009. https://doi.org/10.32614/RJ-2018-009.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2020. *broom: Convert Statistical Objects into Tidy Tibbles.* 0.7.0. https://CRAN.R-project.org/package=broom.

South, Andy. 2017. *rnaturalearth: World Map Data from Natural Earth.* 0.1.0. https://CRAN.R-project.org/package=rnaturalearth.

Stan Development Team. 2020. *CmdStan: The Shell Interface to Stan.* 2.25.0. October 26, 2020. https://mc-stan.org/users/interfaces/cmdstan.

Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace.

Vaughan, Davis, and Matt Dancho. 2018. *furrr: Apply Mapping Functions in Parallel using Futures.* 0.1.0. https://CRAN.R-project.org/package=furrr.

Wächter, Andreas, and Lorenz T. Biegler. 2005. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming." *Mathematical Programming* 106, no. 1 (April 28, 2005): 25–57. https://doi.org/10.1007/s10107-004-0559-y.

Walt, Stéfan van der, S Chris Colbert, and Gaël Varoquaux. 2011. "The NumPy Array: A Structure for Efficient Numerical Computation." *Computing in Science & Engineering* 13, no. 2 (March): 22–30. https://doi.org/10.1109/mcse.2011.37.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2019. *stringr: Simple, Consistent Wrappers for Common String Operations.* 1.4.0. https://CRAN.R-project.org/package=stringr.

Wickham, Hadley, and Jennifer Bryan. 2019. *readxl: Read Excel Files.* 1.3.1. https://CRAN.R-project.org/package=readxl.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *dplyr: A Grammar of Data Manipulation.* 1.0.0. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Lionel Henry. 2020. *tidyr: Tidy Messy Data.* 1.1.0. https://CRAN.R-project.org/package=tidyr.