Title
Essays on Energy and Environmental Economics

Author
Dunkle Werner, Karl W

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays on Energy and Environmental Economics

by

Karl W. Dunkle Werner

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor James Sallee, Chair
Professor Severin Borenstein
Associate Professor Meredith Fowlie

Spring 2021

Essays on Energy and Environmental Economics

Abstract

Essays on Energy and Environmental Economics

by

Karl W. Dunkle Werner

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Associate Professor James Sallee, Chair

Over the past decades, two things have become increasingly apparent: first, climate change and associated environmental impacts are pressing issues, and second, despite this growing threat, existing policies still fall far short. The goal of my research, and what I hope for the field more broadly, is to achieve effective, efficient, and equitable policy. My dissertation research covers a wide range of topics, focusing on three different areas of energy and environmental economics: methane emissions from oil and gas production; flooding on agricultural land; and energy utility regulatory rates of return. The common thread is using applied economic tools and answering policy-relevant questions with data and analysis. Often, the data that are available are far from the ideal dataset, or the policies that are on the table are far from the first best. Here, my coauthors and I adopt the "economist as plumber" mindset, using the tools that are available to address the challenges at hand (Duflo 2017).

In my first chapter, my coauthor Wenfeng Qiu and I study emissions of methane, a powerful greenhouse gas, from oil and gas wells in the US. These emissions contribute significantly to climate change—they are approximately as large as the emissions of all fuel burned in the western US electricity grid. 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 that 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. Our approach highlights the value of information in designing policy, centering the regulatory innovation that is possible when additional information becomes available.

My second chapter is coauthored with Oliver Browne, Alyssa Neidhart, and Dave Sunding. We study high-frequency flood risk on agricultural land. Floods destroy crops and lower the value of agricultural land. Economic theory implies that the hedonic discount on the value of a parcel of flood-prone land should scale with the expected probability flooding. Most empirical studies of the impact of flood risk on property values in the United States focus on the relatively small risk posed by the 100-year or 500-year floodplains, as reported in maps produced by the Federal Emergency Management Agency (FEMA). These studies consequently find a relatively small corresponding discount in property values. However, a significant amount of agricultural bottom-land lies in floodplains that flood more frequently. We estimate the hedonic discounts on with agricultural land that floods at these higher frequencies along the Missouri River. As flood risk increases, the value of flood-prone land decreases, with a hedonic discount ranging from close to zero in the 500-year floodplain to approximately 17% in the 10-year floodplain. To illustrate the importance of characterizing these higher frequency flood risks, we consider a climate change scenario, where properties that already face some flood risk are expected to flood more frequently.

My third chapter, coauthored with Stephen Jarvis, examines the regulated rate of return on equity utility companies are allowed to collect from their customers. Utilities recover their capital costs through regulator-approved rates of return on debt and equity. The US costs of risky and risk-free capital have fallen dramatically in the past 40 years, but these utility rates of return have not. We estimate the gap between what utilities are paid now, and what they would have been paid if their rate of return had followed capital markets, using a comprehensive database of utility rate cases dating back to the 1980s. We estimate that the current average return on equity is 0.5–4 percentage points higher than historical relationships would suggest, and consumers pay an average of $2–8 billion per year more than they would otherwise. We then revisit the effect posited by Averch and Johnson (1962), estimating the consequences of this incentive to own more capital: a 1 percentage point increase in the return on equity increases new capital investment by about 5% in our preferred estimate.

REFERENCES

Averch, Harvey, and Leland L Johnson. 1962. "Behavior of the firm under regulatory constraint." *The American Economic Review* 52 (5): 1052–1069. (Cited on page 2).

Duflo, Esther. 2017. "The Economist as Plumber." *American Economic Review* 107, no. 5 (May): 1–26. https://doi.org/10.1257/aer.p20171153. (Cited on page 1).

Chapters

Acknowledgments

This dissertation, and my entire PhD process, would not have been possible with an immense amount of support from a long list of people. You all have supported me intellectually, socially, emotionally, and financially, from before I started my PhD to today. You've helped me grow as a person and a researcher, helped me survive the PhD program, and helped me find a job. I haven't thanked any of you enough, and I won't be able to start now. You all know what you've done for me, and hopefully you know how much it means to me.

- Alyssa
- Andy C
- Andy H
- Arianna
- Ben
- Carmen
- Catie
- Claire
- Cristina
- Dave
- Derek
- Diana
- Earl
- Eleanor
- Elisa
- Eva
- Evan
- Gabe
- Hal
- Hannah
- Jenya
- Jim
- Josh
- Karen
- Kate
- Louis
- Mariko
- Matt
- Max
- Megan
- Meredith
- Oliver
- Peiley
- Ruth
- Ryan
- Scott
- Severin
- Shaun
- Shuyi
- Sofia
- Song
- Stephen
- Susanna
- Wenfeng

I'm also grateful for the generous support of the Alfred P. Sloan Foundation Pre-doctoral Fellowship on Energy Economics, awarded through the NBER.

Chapter One

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

Coauthor: Wenfeng Qiu

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 (US EPA 2020a; 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 magnitude as the emissions from all fuel used in the western US electricity grid (US EPA 2020). 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 emissions 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 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, 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

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.

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 determine which policies are feasible. In recent decades, remote sensing, administrative data, and other indirect information have improved dramatically, raising 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 a promising case where they could be applied, while acknowledging the measurements' limits for enforcing policy. We integrate a theoretical model, tailored to our data setting, with a newly constructed dataset and Bayesian structural estimation to evaluate the gains a constrained regulator could achieve with additional information.

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. Well operators can reduce the duration of leaks by checking wells more frequently and the occurrence of new leaks by investing in routine maintenance and better equipment. We assume well operators abate expected emissions by reducing the *probability* a well is leaking at any given moment, rather than reducing the size of leaks. 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 data on leaks at the well pad level.[2] 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. 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 the distribution of sizes when leaks occur, as well as the well operator' 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 well operators 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 measure-

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

ments may be necessary, even when leaks can be measured remotely, because of noise in the remote sensing or regulatory constraints.

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.

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 could achieve the first-best under additional assumptions. Each of these policies implies some expected value of the fee a well operator would 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 use 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).

These are relatively simple policies. Although they do not offer the theoretical gains of dynamic enforcement or sophisticated mechanism design, they do allow us to focus on the gains that are possible under feasible policies, and how outcomes change as technological innovation makes more information available (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.

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 of carbon dioxide equivalent (CO_2e), which leads to improving deadweight loss (DWL) by 3.39 percent of the difference between the no-fee outcome and the Pigouvian first best.[3] 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

3. 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. 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. Alternatively, 100-year GWP is approximately consistent with a 3% discount rate on climate damages (Sarofim and Giordano 2018; Mallapragada and Mignone 2019).

the same as the uniform policy), but is now heterogeneous across wells, with an interdecile range of \$0–0.527. Now the improvement in DWL is 5.43 percent, and the average fall in emissions is 55.5 tons.

We can also consider targeting on observed leaks. We focus on the realistic case where the remote measurement has a high detection threshold, so only the largest leaks are detected. 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 budget and allowed penalty, there's a much higher expected fee: \$8.34 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 62.3 percent, and average emissions declines of 683 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 fee 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 arise in many environmental questions, such as regulating nonpoint pollution, as well as other economic topics such as tax evasion, teacher value-added and principal–agent problems (Segerson 1988; Allingham and Sandmo 1972; Chetty, Friedman, and Rockoff 2014). In all of these areas, if regulators 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.[4] 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. Thus, we contribute to a line of innovation literature that includes Nagaraj (2020), which considers a private-sector case where satellite imagery enabled entry by small firms and changed the structure of the market.

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 that estimates 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

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

elasticity of methane emissions– reported in the EPA's greenhouse gas reporting program (GHGRP) – with respect to the commodity price of natural gas of methane emissions. That paper, like ours, estimates abatement costs in the absence of any direct price on emitting methane. To get around the lack of existing policy, that paper notes that the private incentive to sell natural gas into the commodity market can proxy for a fee on methane emissions. We rely on the same argument. Because natural gas prices move in a limited range and are always much lower than the social cost of methane emissions, 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 per ton CO_2e as the low-end fee a regulator might 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 \$8.34 per ton of CO_2e leads to emissions reductions of 683 tons CO_2e per well per year, from a baseline of 1500 tons (a 45% reduction).

This similarity is notable as Marks (2018) uses a different source of identifying variation and a different measure of methane emissions. That paper uses variation in the price of natural gas to identify the change in reported quantity emitted, while we use the variation in leak sizes and occurrence (see details in section 5). 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). Finally, Marks (2018) estimates abatement of aggregate operator-basin emissions, whereas we estimate abatement in the probability of a leak at each well pad.

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 the 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 million tons of CO_2e (Alvarez et al. 2018 provides emissions estimates for 2015).[5] 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

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

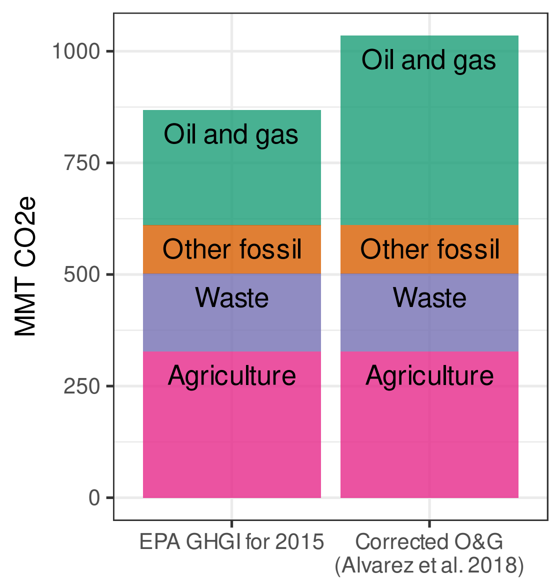
leaks or emissions from burning the fuel.[6] For comparison, the contribution to gross domestic product (GDP) for the entire oil and gas sector, less wages and depreciation, averages $34 billion per year.

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.[7] We focus on audit policies because they're a traditional tool of environmental, health, and safety regulation. 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 (US EPA 2020b). Other estimates are lower, but easily over $100 per well pad for on-the-ground audits.[8] Second, the fines that the regulator charges are limited.

Remote sensing may provide valuable but imperfect information. We consider the role satellite measurements 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

6. With a conversion factor of 34, $58.82 per ton CO_2e is $2 per kg CH_4, or about $35 per 1000 cubic feet (mcf) of 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.

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

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

is the expected result of equipment operating properly. Large intentional vents are rare, as the operator would typically set up a flare to burn the gas.[9] 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), has 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 include 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.

The report ICF INTERNATIONAL (2014), as well as the academic literature Lyon et al. (2016), Rutherford et al. (2020), Ravikumar et al. (2020), and Robertson et al. (2020) provide more detail on the sources of emissions and what well operators may do to reduce them. These academic papers typically highlight the distinction between separator tanks and all other sources.[10] Tanks may have large leaks from flashing (where dissolved gas escapes as oil decompresses), and operators may choose to vent the gas or collect it. Large tank leaks can also come from abnormal conditions, such as a stuck separator valve (which could leak the well's entire gas production), thief hatches left open, or rusted-through holes. In addition to emissions from tanks, large leaks may come from diverse sources, ranging from unlit flares to liquids unloadings. These sources vary in their causes and appropriate abatement method, ranging from LDAR effort to equipment choices. Emissions from abnormal conditions are expected to occur some fraction of the time. For this project, we don't differentiate between normal and abnormal operations – beyond noting that the large leaks we consider are rare – since in both cases the well operator can reduce their expected quantity of emissions at some cost.

9. Flares burn natural gas and produce CO_2, dramatically lowering the GHG output. They're over 90% effective (98% when operating properly), though intentional venting of unburnt gas may increase when flaring is penalized (Calel and Mahdavi 2020).

10. In contrast, the EPA's greenhouse gas inventory (GHGI) does not recognize tank emissions other than flashing, which results in Rutherford et al. (2020) estimating emissions from tanks more than 20-times larger than GHGI does.

2.2 TRADITIONAL ECONOMICS SOLUTIONS

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. As discussed above, the feasibility of imposing fines is limited in this context. The mechanism of Segerson (1988), 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 do 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, the 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

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 of audits they can do and the opportunities for targeting. 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 apply 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:

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

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 in 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 $C_i'(q_i) = p_i \cdot e_i$. 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. t_i is the expected value of the fee the operator will pay when well i has a leak. This expected value will typically be lower than the fee the well pays when audited and leaking, as the well is not guaranteed to be audited. t_i and p_i are both costs of emitting gas – from the regulator and from the opportunity cost of selling the gas – and under expected profit maximization, operators will treat them equivalently. Therefore, the first order condition is $C_i'(q_i) = (p_i + t_i) \cdot e_i$. 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:

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

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

$$\begin{aligned}&\text{DWL}_i(t_i) = \\ &\left((p_i + \delta)\, e_i - \frac{(p_i + t_i)\, e_i}{\alpha_i + 1}\right)\left(\frac{(p_i + t_i)\, e_i}{A_i}\right)^{\frac{1}{\alpha_i}}\\ &\quad - \frac{\alpha_i}{1 + \alpha_i}(p_i + \delta)\, e_i\left(\frac{(p_i + \delta)\, e_i}{A_i}\right)^{\frac{1}{\alpha_i}}\end{aligned}$$

Mathematical details are in appendix section 1.1.

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 which wells 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 they're covered by and what is the probability of an audit.

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 level of 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 won't fix 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:

$$\min_{r_i} \sum_i \text{DWL}_i(r_i) \quad s.t. \ \sum_i r_i \le M, \ r_i \in [0,1]$$

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:

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

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

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

3.2.3 Policy 2: targeting on covariates

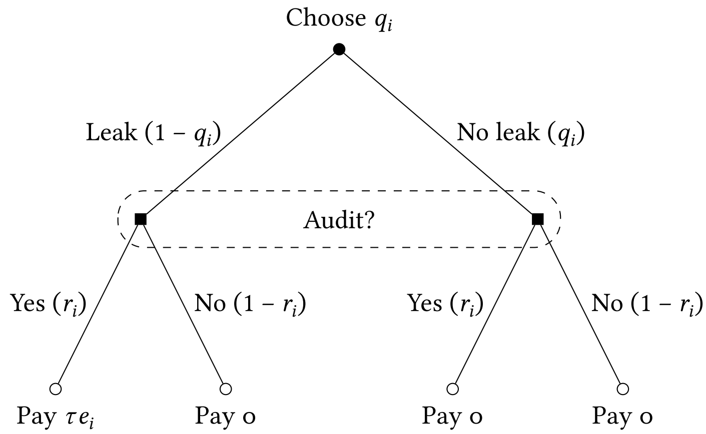
Moving to the case where audits are targeted on well covariates (policy 2), it's helpful to have an extensive form game tree. See figure 2 for the game each well faces under this policy. Recall that the well operator knows e_i, the size of their leak if a leak happens, their cost function $C(q_i; A_i, \alpha_i)$ 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 evaluating the policy. The regulator's problem is now:

$$\min_{\{r_i\}_{i=1}^N} \sum_{i=1}^N \text{DWL}_i$$

$$s.t. \ \sum_{i=1}^N r_i \le M \text{ and } \forall i : r_i \in [0,1]$$

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 a 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.

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

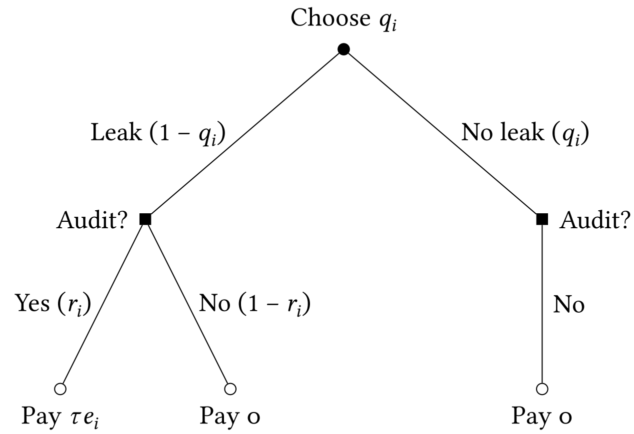
As before, λ is the shadow price of having an additional audit. $\lambda < 0$ because increasing audits lowers DWL. a_i and b_i correspond to the constraints $r_i \geq 0$ and $r_i \leq 1$. The feasible set under the constraint is a compact subset of $\mathbb{R}^N$ so a solution exists. Because DWL_i 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 is able to 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.

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 indepen-

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 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.

dent 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, the probability of being audited when leaking and when not leaking were the same, since audit probabilities depended only on the X, not on the realized leaks. Now, with every leak observable, the probability of being audited when not leaking falls to zero. The budget constraint becomes:

$$\sum_i q_i \cdot 0 + (1-q_i) r_i \le M$$

$$\sum_i \left(\frac{(p_i + \tau r_i)\, e_i}{A_i} \right)^{\frac{1}{\alpha_i}} r_i \le M$$

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:

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

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

As before, λ is the shadow price of having an additional audit, and 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. 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 well operators' 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 the detection threshold into account. As the game tree suggests, if a well is *not* detected 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; because they know they will always be audited with probability s_i, 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 depend only 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 operator, so the large-e wells 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 optimizes 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:

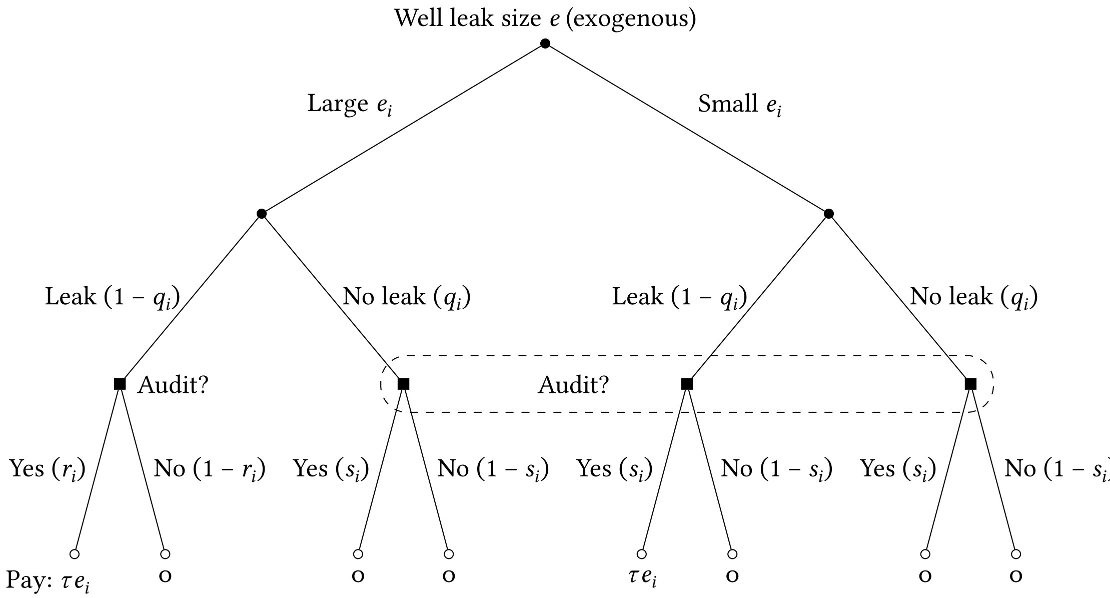
$$\min_{\{r_i\}_{i=1}^N, \{s_i\}_{i=1}^N} \sum_i z_i \text{DWL}_i(r_i) + (1 - z_i)\text{DWL}_i(s_i)$$

$$\textit{s.t.} \sum_i \Big\{ z_i[(1 - q_i(r_i))r_i + q_i(r_i)s_i] + (1 - z_i)s_i \Big\} \leq M$$

$$\forall i : r_i \in [0, 1], s_i \in [0, 1]$$

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 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.

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 operator knows e_i; the regulator can only form expectations. The well operator 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.

Compared to the audit policies, the remote fee policy is simple. In this scenario, the regulator is able to measure leaks remotely, and to 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 because 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 1.3.

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 leak size when a well is 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.

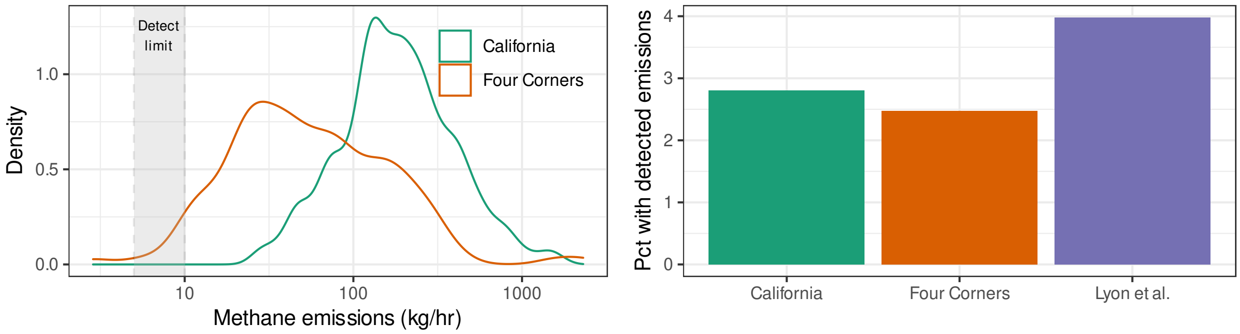
There are a number of studies that measured well leaks on the ground; however, because they decline to publish well identifiers or covariates beyond contemporaneous gas production, we are not able to use their data for our analysis. A comparison is plotted in appendix figure 9. These ground studies often note that large leaks are from valves left open, or, most frequently, separator tanks with open hatches.

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 don't want these multiple revisits to

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.

affect 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 its 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, South" for California wells and "El Paso, San Juan Basin" and "Transwestern, San Juan Basin" for the Four Corners wells.[11] These are midstream prices,

11. 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 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)	176.5	1240.1	1.6	261.8
Oil (bbld)	36.0	230.7	0.2	43.4

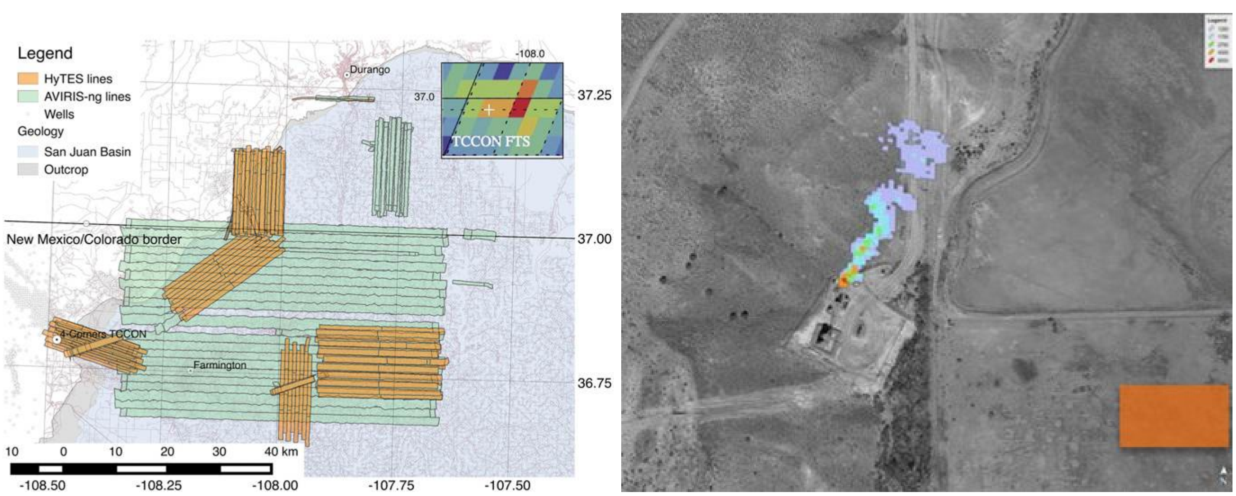
NOTES: Only wells that report positive gas production are included. In all three panels, wells are grouped into pads; see text for details. mcfd means thousands of standard cubic feet per day. bbld means barrels of oil per day. Well data from Enverus (2019). Flight data from Duren et al. (2019) and Frankenberg 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.

Table 2: Balance comparison: well pads 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]

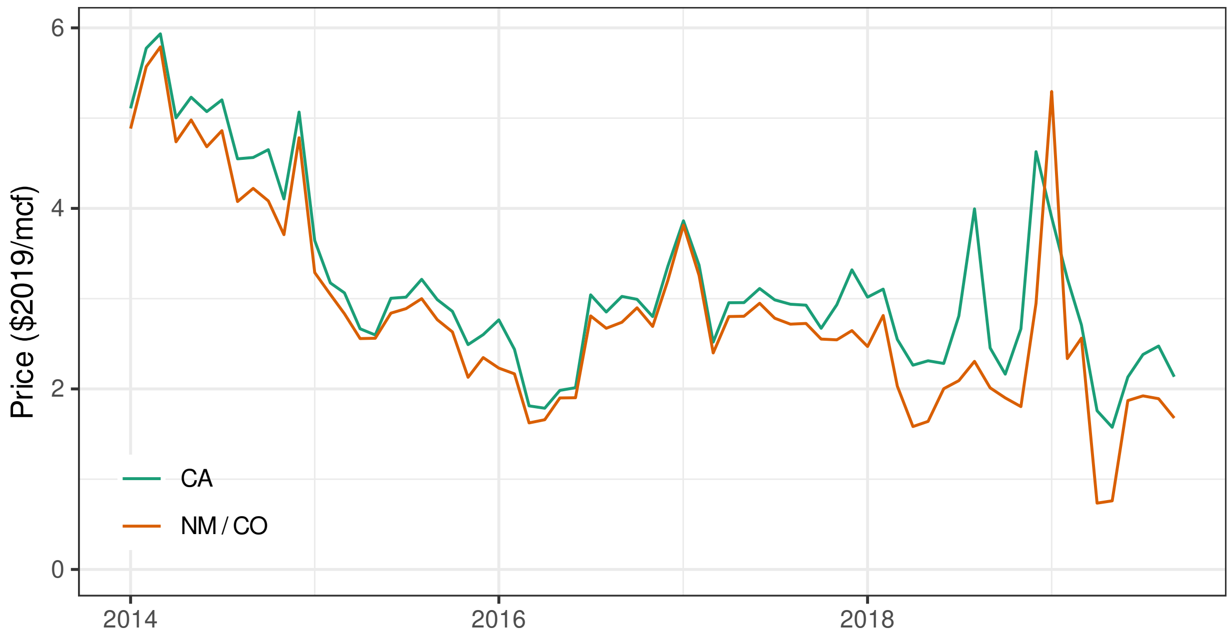
Note: Values are means, with the 10th to 90th percentile value in brackets.
California and Four Corners data are from the AVIRIS-NG sample (panel A of table 1). Lyon et al. (2016) data cover basins throughout the US. All values are well-pad aggregates.

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.

and they will tend to be slightly higher than the prices well operators actually receive. This difference 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 possible 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 misspecification.

5.1 MODELS

Our modeling approach follows closely from our theory. We begin by estimating the distribution of leak sizes, then we use the well operators' 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.[12] 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,

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

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 fee on emissions, 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 $\underline{e}\, p_i H$.[13]

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

13. A more intuitive approach would be to multiply by $\hat{e}_i p_i H$. We estimate e_i and A_i simultaneously; including the predicted value $\hat{e}_i$ in the product leads to poor joint estimation.

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, with 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 were 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 well operators' 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.[14] This correlation is not included in our model and would generate biased estimates. We do include basin indicators, but those are at a large geographic scale.

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, which would be another form of measurement error. We believe these issues of selection and measurement 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.

14. 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.

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 it faces a fee t_i.

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

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'.

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

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 depending on 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 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

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	0.00512
	[0.038,0.038]	[0,0]	[0,0]	[0,0.018]
Target leaks, low threshold	2.16	3.17	0	3.84
	[2,2.4]	[2.3,3.8]	[0,0]	[3.8,3.8]
Target leaks, high threshold	2.1	3.84	0	3.84
	[2.1,2.2]	[3.8,3.8]	[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	0.934	0.737	0	1.92
	[0.85,1]	[0,1.9]	[0,0]	[1.9,1.9]
Target leaks, high threshold	0.958	1.09	0	1.92
	[0.87,1.1]	[0,1.9]	[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	0.0676	0	0	0.163
	[0.062,0.074]	[0,0]	[0,0]	[0.16,0.16]
Target leaks, high threshold	0.0686	0	0	0.163
	[0.062,0.075]	[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	1.61
	[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	28.4	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	0.896
	[0.25,0.25]	[0,0]	[0,0]	[0.84,0.99]
Target leaks, low threshold	25	25	25	25
	[25,25]	[25,25]	[25,25]	[25,25]
Target leaks, high threshold	14.2	25	0	25
	[13,15]	[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	1.05	1.02	0	2.12
	[0.96,1.2]	[0.0000006,2.1]	[0,0]	[2.1,2.1]
Target leaks, high threshold	1.08	1.31	0	2.12
	[0.99,1.2]	[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.)

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 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 the regulator is 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 24.8 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 71.5 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 820 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) does well relative to the other audit policies, which was expected from the fact that the regulator is able to use the limited audit budget more effectively. The high-threshold policies – both the infeasible remote fee and the feasi-

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	0.526	0.259	0.0128
	[0.517,0.57]	[0.254,0.281]	[0.0123,0.015]
Target covariates	1.06	0.566	0.0455
	[0.866,1.46]	[0.449,0.81]	[0.033,0.0728]
Target leaks, low threshold	25.9	14	1.27
	[23.3,31.1]	[12.4,16.5]	[1.12,1.47]
Target leaks, high threshold	24.8	13.9	1.29
	[23.6,26.5]	[12.8,15.7]	[1.16,1.48]
Remote, low threshold	36.4	21.7	2.24
	[35.9,38.7]	[21.4,23.2]	[2.21,2.43]
Remote, high threshold	25.6	15.2	1.56
	[24.8,26.7]	[14.7,15.9]	[1.51,1.64]
B: $\mathbb{E}$[emiss] improvement (%)			
Uniform	0.442	0.221	0.0189
	[0.432,0.487]	[0.216,0.244]	[0.0184,0.0208]
Target covariates	0.885	0.477	0.0458
	[0.719,1.23]	[0.377,0.684]	[0.0351,0.0681]
Target leaks, low threshold	21.7	11.6	1.05
	[19.4,26.6]	[10.3,14]	[0.926,1.24]
Target leaks, high threshold	20.8	11.5	1.07
	[19.7,22.6]	[10.5,13.3]	[0.958,1.26]
Remote, low threshold	30.5	18	1.85
	[29.9,33]	[17.7,19.7]	[1.81,2.04]
Remote, high threshold	21.4	12.6	1.29
	[20.8,22.5]	[12.2,13.4]	[1.25,1.38]

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 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	6.6	3.39	0.288
	[6.49,7.12]	[3.34,3.67]	[0.283,0.312]
Target covariates	9.69	5.43	0.621
	[8.89,11.2]	[4.87,6.61]	[0.496,0.885]
Target leaks, low threshold	96.4	85.7	15.3
	[96.2,97]	[85.2,87.4]	[13.6,18.2]
Target leaks, high threshold	71.5	62.3	15.2
	[69.2,72.7]	[60.8,63.7]	[14,17.1]
Remote, low threshold	96.4	85.7	23.6
	[96.3,97]	[85.2,87.4]	[23.2,25.2]
Remote, high threshold	69	61	16.5
	[66.6,70.6]	[59.2,62.8]	[16,17.3]
B: $\mathbb{E}$[emiss] improvement (%)			
Uniform	5.46	2.81	0.245
	[5.34,6]	[2.75,3.09]	[0.24,0.271]
Target covariates	8.07	4.52	0.523
	[7.34,9.47]	[4.02,5.56]	[0.415,0.747]
Target leaks, low threshold	90	76.3	12.8
	[89.5,91.5]	[75.5,79.1]	[11.3,15.4]
Target leaks, high threshold	66.6	55.5	12.7
	[65.1,67.8]	[54.2,57]	[11.6,14.5]
Remote, low threshold	90	76.3	19.6
	[89.5,91.5]	[75.5,79.1]	[19.2,21.4]
Remote, high threshold	64.5	54.4	13.7
	[62.6,66.2]	[52.8,56.2]	[13.3,14.5]

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 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.

ble 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.

7 CONCLUSION

The five policies we consider, ranging from no audits to fully remotely assessed fees, offer a sampling of feasible policies. All of these policies fall short of the first best, some dramatically so. However, we can also see the importance of the fee level – a \$5 per ton CO_2e fine is not very effective, particularly when audit probabilities are low. Importantly, we found that 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 the 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 leak detection and repair (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 well operators' costs. We then consider how different policies translate into incentives the wells face, and how 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.

Our findings highlight the importance of thinking about measurement and 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.

Appendices

A PROOFS

1.1 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.

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

□

1.2 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 $G_i(r_i)$:

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

1.2.1 *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

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

$$= \int_{r_k^*}^{r_j^*} \frac{\partial \text{DWL}_k}{\partial r} - \frac{\partial \text{DWL}_j}{\partial r} \, dr$$

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, $\frac{\partial^2 \mathscr{L}}{\partial r_i \partial e_i} = \frac{\partial^2 \text{DWL}_i}{\partial r_i \partial e_i} < 0$. The inequality follows directly from the specific choice function we chose and the expected fee t_i being an increasing function of r_i,

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

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

1.3 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.

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

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 $\sum r_i \leq M$.

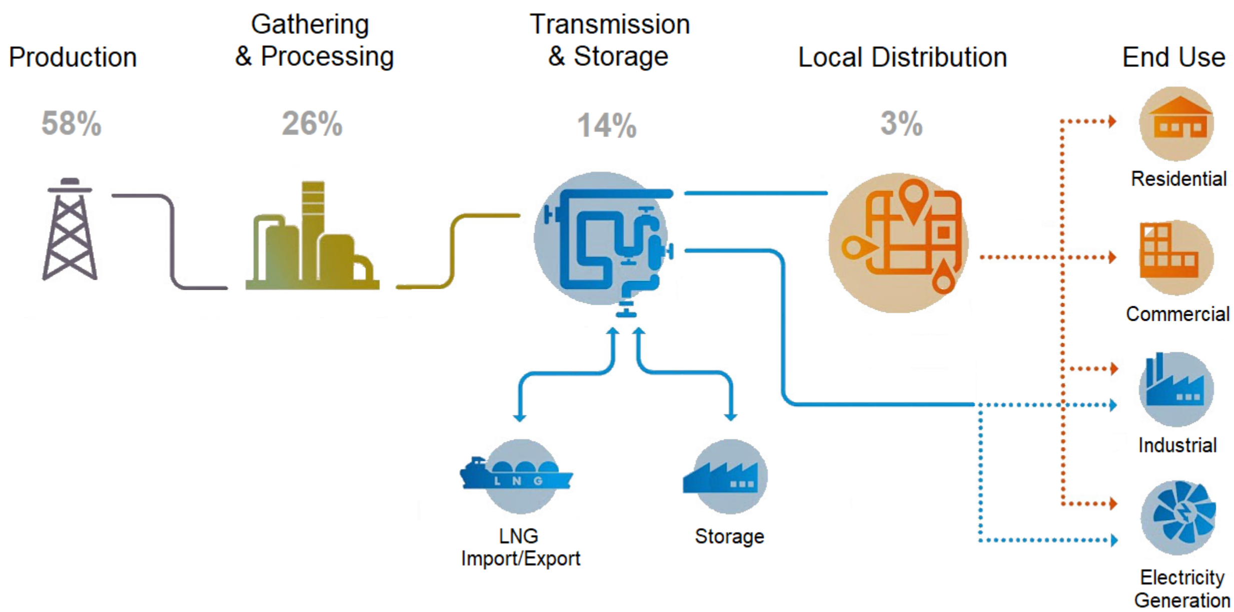
The budget constraint for the measure-then-audit policy (3) with no detection threshold becomes:

$$\underline{e} = 0 : \quad \sum_i \left(\frac{(p_i H + \tau T r_i)\, e_i}{A_i}\right)^{\frac{1}{\alpha_i}} r_i \leq M$$

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

B METHANE MEASUREMENT

Figure 8: Production segment is responsible for 58% of leaks from natural gas supply chain



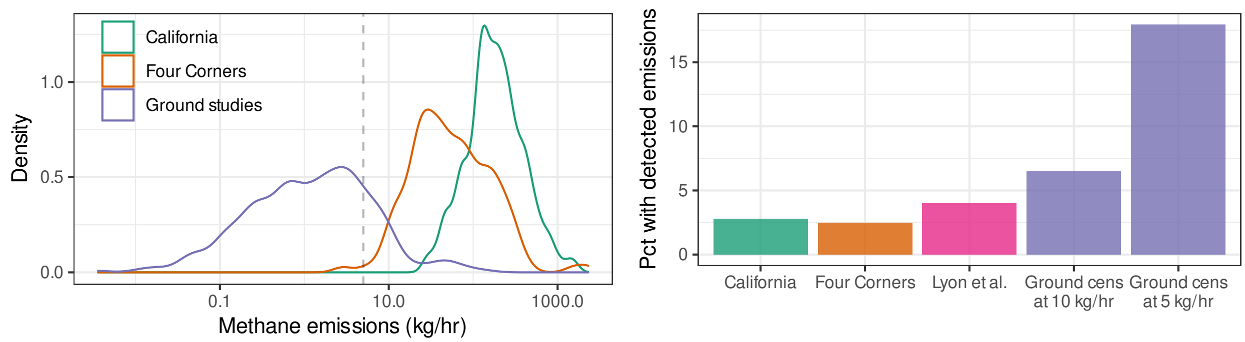
Source: Marks (2018) figure 1, from estimates in Alvarez et al. (2018). Excludes end-use leaks.

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 9: 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.

Table 8: Parameters of methane leak size models

	LogNormal	LogNormal meas. err.	Cost param. (common)	Cost param. (heterog.)
Intercept	4.4	4.03	6.15	4.7
	[2.8,6.1]	[3,5]	[4.4,7.7]	[3.1,6.3]
IHS of gas prod (mcfd)	0.112	0.152	−0.19	0.0469
	[−0.065,0.28]	[0.039,0.26]	[−0.37,0.024]	[−0.13,0.21]
IHS of oil prod (bbld)	−0.0339	−0.0817	0.181	0.0148
	[−0.22,0.16]	[−0.21,0.047]	[0.0042,0.37]	[−0.17,0.23]
Basin: San Joaquin	−0.196	−0.0975	−0.0821	−0.085
	[−0.94,0.51]	[−0.56,0.37]	[−0.77,0.49]	[−0.81,0.58]
Basin: San Juan	−1.15	−1.27	−0.358	−0.994
	[−2.1,−0.24]	[−1.9,−0.64]	[−1.2,0.39]	[−1.9,−0.16]
Oil prod share	0.736	0.999	−1.35	0.263
	[−0.49,1.9]	[0.19,1.8]	[−2.6,0.098]	[−1,1.4]
IHS of age (yr)	0.0623	0.0678	0.00505	0.0749
	[−0.12,0.25]	[−0.051,0.19]	[−0.15,0.16]	[−0.095,0.25]
Drill: Horizontal	0.263	0.229	0.132	0.165
	[−0.51,1.1]	[−0.35,0.82]	[−0.55,0.98]	[−0.6,1]
Drill: Unknown	−0.622	−0.714	−1.25	−0.742
	[−1.2,−0.0013]	[−1.1,−0.29]	[−1.8,−0.76]	[−1.4,−0.12]
Drill: Vertical	−0.0847	−0.167	−0.0318	−0.149
	[−0.48,0.31]	[−0.45,0.12]	[−0.37,0.28]	[−0.53,0.23]
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	0.14	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	−5.84	−0.785
	[−7.4,−4.3]	[−6.9,−4.8]	[−5.1,3.6]
IHS of gas prod (mcfd)	0.327	0.329	−0.0538
	[0.17,0.49]	[0.21,0.45]	[−1.1,1]
IHS of oil prod (bbld)	−0.177	−0.172	0.428
	[−0.37,0.011]	[−0.31,−0.035]	[−0.53,1.2]
Basin: San Joaquin	−0.161	−0.154	0.384
	[−0.77,0.58]	[−0.6,0.33]	[−0.57,1.4]
Basin: San Juan	−0.327	−0.308	−0.575
	[−1.1,0.48]	[−0.89,0.31]	[−1.6,0.41]
Oil prod share	2.16	2.19	0.475
	[1.1,3.2]	[1.4,3]	[−0.51,1.5]
IHS of age (yr)	0.124	0.117	−0.207
	[−0.072,0.35]	[−0.019,0.25]	[−1.2,0.64]
Drill: Horizontal	0.126	0.161	−0.0374
	[−0.88,0.92]	[−0.49,0.75]	[−0.97,0.91]
Drill: Unknown	1.52	1.53	0.375
	[0.91,2.1]	[1.1,2]	[−0.62,1.3]
Drill: Vertical	−0.0692	−0.0632	−0.49
	[−0.44,0.33]	[−0.34,0.23]	[−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.

Bibliography

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 4).

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 3).

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 4).

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 1, 5, 7, 16, 35).

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 8).

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 3).

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 1).

Calel, Raphael, and Paasha Mahdavi. 2020. "Opinion: The unintended consequences of antiflaring policies—and measures for mitigation." *Proceedings of the National Academy of Sciences* 117, no. 23 (June 9, 2020): 12503–12507. https://doi.org/10.1073/pnas.2006774117. (Cited on page 7).

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 7).

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 4).

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 3).

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 pages 3, 36).

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 17–19, 22, 37).

Enverus. 2019. September 1, 2019. https://drillinginfo.com. (Cited on page 19).

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 17–20, 37).

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 23).

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 1).

Huggins, Jonathan H., and Jeffrey W. Miller. 2019. "Robust Inference and Model Criticism Using Bagged Posteriors," arXiv: 1912.07104 [stat.ME]. (Cited on page 22).

ICF International. 2014. *Economic Analysis of Methane Emission Reduction Opportunities in the U.S. Onshore Oil and Natural Gas Industries.* Technical report. May. https://www.edf.org/sites/default/files/methane_cost_curve_report.pdf. (Cited on page 7).

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 2, 7, 17–20, 37).

Mallapragada, Dharik S., and Bryan K. Mignone. 2019. "A theoretical basis for the equivalence between physical and economic climate metrics and implications for the choice of Global Warming Potential time horizon." *Climatic Change* 158, no. 2 (September 9, 2019): 107–124. https://doi.org/10.1007/s10584-019-02486-7. (Cited on page 3).

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 page 22).

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 4, 5, 10, 35).

Marks, Levi. 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 5).

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 22).

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 3).

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 4).

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 3).

Omara, Mark, Melissa R. Sullivan, Xiang Li, R. Subramanian, Allen L. Robinson, and Albert A. Presto. 2016. "Methane Emissions from Conventional and Unconventional Natural Gas Production Sites in the Marcellus Shale Basin." *Environmental Science & Technology* 50, no. 4 (November): 2099–2107. https://doi.org/10.1021/acs.est.5b05503. (Cited on page 37).

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 pages 18, 37).

Pigou, Arthur Cecil. 1932. *The Economics of Welfare.* 4th ed. Chap. 9, bk. 2. Macmillan / Co. (Cited on page 8).

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 8).

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 1).

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 pages 4, 7).

Rella, Chris W., Tracy R. Tsai, Connor G. Botkin, Eric R. Crosson, and David Steele. 2015. "Measuring Emissions from Oil and Natural Gas Well Pads Using the Mobile Flux Plane Technique." *Environmental Science & Technology* 49, no. 7 (March): 4742–4748. https://doi.org/10.1021/acs.est.5b00099. (Cited on page 37).

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 pages 5, 7).

Robertson, Anna M., Rachel Edie, Dustin Snare, Jeffrey Soltis, Robert A. Field, Matthew D. Burkhart, Clay S. Bell, Daniel Zimmerle, and Shane M. Murphy. 2017. "Variation in Methane Emission Rates from Well Pads in Four Oil and Gas Basins with Contrasting Production Volumes and Compositions." *Environmental Science & Technology* 51, no. 15 (July): 8832–8840. https://doi.org/10.1021/acs.est.7b00571. (Cited on page 37).

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 7).

Sarofim, Marcus C., and Michael R. Giordano. 2018. "A quantitative approach to evaluating the GWP timescale through implicit discount rates." *Earth System Dynamics* 9, no. 3 (August 17, 2018): 1013–1024. https://doi.org/10.5194/esd-9-1013-2018. (Cited on page 3).

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 4, 8).

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 6).

US EPA. 2020. "Emissions & Generation Resource Integrated Database (eGRID) 2018." https://www.epa.gov/sites/production/files/2020-01/documents/egrid2018_summary_tables.pdf. (Cited on pages 1, 5).

US EPA. 2020a. *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 1).

———. 2020b. "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 6).

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 2).

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.

Wickham, Hadley. 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.

Transition

Now I turn from the challenge of creating methane policy to flood risks on agricultural land. These topics share the broad umbrella of applied environmental economics. Though the setting and policy response are very different, they share the core economic principal of economic agents making choices to maximize their outcomes. These papers also share data challenges: the size and quality of the datasets is far from ideal. Consequently, estimates are sometimes noisy, and my coauthors and I need to frame our research questions carefully to ensure we ask questions the data can answer.

Chapter Two

Hedonic Valuation of High-Frequency Flood Risk on Agricultural Land

Coauthors: Oliver Browne, Alyssa Neidhart, and David Sunding

1 INTRODUCTION

Natural hazards such as flood, drought, fire and earthquake pose risks to both property and agricultural production. These risks increase the possibility that the property owner will have to make costly repairs, or that a season's crops will be lost. These hazards lower the expected benefits of owning the land, as they impose both monetary and personal costs. In economic theory, with efficient markets for land, these risks should be capitalized into property values (Rosen 1974). Our work builds on this theory, studying flooding on agricultural land, including areas that have high risks of floods.

Using data from parcels sold between 1988 and 2019 in counties that border the Missouri River, we find that parcels in the 10-year floodplain are 3% less valuable (95% CI: [−18.7, 15.8]) per acre than comparable properties that face minimal risk of flooding. Our estimates for areas that flood less frequently are similar in magnitude: all in the range between −10% and zero, and not statistically insignificant at the 95% level.

This analysis is timely, as agricultural land in and near our study area has seen substantial flooding several times over the past decade. It's also novel, incorporating detailed data on flood risks and finding that the additional granularity substantially affects our understanding of hedonic valuation. Finally, this analysis is important looking forward. Flood risk is expected to increase in these agricultural lands and others, as climate change increases the variance of rainfall patterns. Valuing these changes provides a richer understanding of the costs of climate change for agriculture, and the benefits from additional flood protections.

Our analysis contributes to the large literature on the value of risks from natural hazards. Often, papers in this literature find that these risks are not properly capitalized into property values, or that these risks only become properly capitalized after the realization of a disaster, when they become more salient. Most of the papers in this literature have two limitations: they focus on hazards that are realized relatively infrequently, and they focus almost exclusively on *residential* property sales.

There are a couple of reasons to think agricultural land might respond more sharply to flood risk. Unlike living at a residential property, operating a farm is a business. The farm operator might try to maximize profits, which provides a strong incentive to properly capitalize risks into property values. Rational inattention may explain some of the lack of response in residential properties, while farm owners have a stronger incentive to consider flooding risk. Additionally, we're considering flood probabilities much higher than the residential flooding literature typically includes. Farm owners (and other landowners) might have more experience dealing with higher frequency risks and therefore might be more likely to value the discount in their decisions to buy or sell land.

Feasible adaptation and appropriate policy differ for residential and agricultural land. Residential flooding risk can be mitigated through both building design and relocating to a less flood-prone area. Agricultural land cannot be put on stilts. The feasible forms of abatement on this land are changes in flood barriers (e.g. levees), river flow management, changing crops, or ceasing to use the land for agriculture. This smaller set of private abatement options makes government policy relatively more important in the agricultural setting.

Our hedonic analysis compares parcels in areas that face a risk of flooding with similar, control parcels that face minimal flood risk. We use data on property sales, as compiled by CoreLogic, using flood map contours calculated from US Army Corps of Engineers (USACE) flood stage modeling. These flood contours are more detailed than the Federal Emergency Management Agency (FEMA) flood maps typically used in this literature. We control for differences in the assessed value of improvements (i.e. buildings) on the land, as well as a rich set of covariates relevant to the land's agricultural value including soil quality and drainage characteristics. We also include proximity to urban areas, controlling for the land's potential development value, as well as county fixed effects. These controls allow us to estimate causal effect of flood risk, disentangling it from correlated variables like soil quality.

We consider how these estimated hedonic flooding discounts intersect with predicted changes in river flows and flood risks as the climate changes. These climate predictions are drawn from a report U.S. Bureau of Reclamation (2012).[1]

Using detailed flooding data, we estimate the hedonic discount in the per-acre value of agricultural properties along the Missouri River. We estimate regressions for parcels that are expected to flood at least every 500, 100, 50, 20, and 10 years, comparing each group with similar parcels that face minimal risk of flooding. Results are in table 2. We then aggregate these results, estimating an average slope of the hedonic discount across different flood groups. We focus on elasticity of property value with respect to flood frequency, finding a slope of −0.5%, with a 95% confidence interval of [−1.5%, 0.48%]. Taking the point estimate literally, a doubling in flood period (e.g. the 50-year group vs the 100-year) is expected to be 0.5% *less* valuable on a per-acre basis. We provide a more formal theoretical treatment in section 2.

1.1 LITERATURE REVIEW

Our research is almost unique in this literature in looking at the hedonic value of flooding on agricultural land. Struyk (1971) and Wang (2020) are the only papers we are aware of in the literature that relies on agricultural property values. Much of the existing residential literature does not have sufficient gradation in flood risk to consider the slope of the hedonic valuation curve. Bin, Kruse, and Landry (2008) is an exception; like our work, that paper finds that the magnitude of the flood risk matters, though they find a steeper slope than we do.

The literature on hedonic valuation of flooding on residential properties is more densely populated. That literature generally finds zero

1. That report uses bias corrected and spatially downscaled coupled model intercomparison project phase 3 (CMIP3) climate projections, using a range of future greenhouse gas emissions (Intergovernmental Panel on Climate Change (IPCC) low, medium, and high values: B1, A1b, and A2 from Nakićenović et al. 2000).

or small effects, but typically looks at much smaller flood risks than our study. For instance, Bialaszewski and Newsome (1990), Bin and Kruse (2006), Bin and Landry (2013), Atreya and Czajkowski (2019), and Hino and Burke (2021) find little to no change in capitalization due to flood risk. Some studies find a price effect, but typically smaller than the future cost of flood insurance premiums mandated by government-backed mortgages. Speyrer and Ragas (1991), Schwartz (2001), and Daniel, Florax, and Rietveld (2009) are examples here. On the other hand, MacDonald, Murdoch, and White (1987) finds that flood risks reflect future discounted insurance premiums, for some reasonable value of the homeowner's discount rate. Knowing the correct discount rate – and incorporating uninsurable losses – are significant challenges in this type of analysis.

Other papers consider the dynamic response to realized flooding, finding that the hedonic discount is larger or insurance takeup is higher following a flood. Similarly, seeing a flood nearby, without directly experiencing it, tends to increase takeup. These effects often fade over time. Bin and Polasky (2004), Atreya, Ferreira, and Kriesel (2013), Atreya and Ferreira (2015), Hallstrom and Smith (2005), Carbone, Hallstrom, and Smith (2006), Morgan (2007), Daniel, Florax, and Rietveld (2009), Kousky (2010), Nyce et al. (2015), and Beltrán, Maddison, and Elliott (2018) provide nice examples of this result. Our work differs from this context because we look at parcels in much higher flood risk areas. (As well as the residential vs agricultural difference mentioned above).

Some studies are inconclusive, such as Chao, Floyd, and Holliday (1998), while a small number of studies find substantial capitalization of flood risk: Fridgen and Shultz (1999), Troy and Romm (2004), Pope (2008), and Posey and Rogers (2010). Flood risk may also increase search frictions, making transactions harder (Turnbull, Zahirovic-Herbert, and Mothorpe 2013).

Another point the literature emphasizes is that spatial amenities matter, may not be fully observed, and may be highly correlated with flood risk. This correlation can make econometric identification challenging (Bin et al. 2008). While the parcels in our study may not have an ocean view, they do have amenities in the form of soil quality and other characteristics. Soil quality is strongly affected by historical flooding (floods brought nutrients and improved the soil), and controlling for these characteristics is an important piece of our identification argument.

The literature makes a couple of other important points about flood maps and their inexact relationship with flood risk. Gibson, Mullins, and Hill (2019) finds re-drawn maps produced larger belief changes than being directly hit by a hurricane or by insurance reform, and Shr and Zipp (2019) finds prices decrease when a property is reclassified into a flood zone, but not when it's reclassified out. Wang (2020) considers the discontinuity between areas with different FEMA flood zones – the map changes discontinuously, while the risk of flooding may or may not change as sharply – and finds substantial price effects.

In addition to the flooding literature, we contribute to the literature on agriculture and climate change. A large majority of those works focus on changes in temperature, rather than flooding. A few prominent examples of hedonic valuation in this climate–temperature space are Mendelsohn, Nordhaus, and Shaw (1994), Fisher et al. (2012), and Severen, Costello, and Deschênes (2018). As with the flooding literature, casual identification remains a challenge.

In addition to the hedonic valuation literature, we rely on studies of climate change and associated shifts in flood risk. For instance, van der Wiel et al. (2018) integrates a global climate model with models of surface water flow to estimate the probabilities of anomalously high river levels in the Mississippi basin. River levels are informative of flooding, but many other factors determine whether a specific parcel away from the river will flood. We discuss these challenges of estimating flood frequency and projecting future climate change in greater depth in sections 3 and 5.

2 HEDONIC FRAMEWORK

We consider a hedonic valuation model in the Rosen–Roback framework (Rosen 1974; Roback 1982). In our setting, the hedonic variable of interest is the probability of flooding. We consider how the valuation of agricultural land, on a per-acre basis, varies with flood risk.

The value of a parcel to a farm owner is the discounted stream of future profits, $\{\pi_t\}_{t=0}^{\infty}$. For simplicity, we omit other, non-monetary benefits or costs of owning a farm. If these can be translated to dollars, they can be included in π_t without changing the expressions. Let f_t be the flood frequency in each year. For instance, f_t might be 500, with a flood probability of 1/500 in year t. Assume that farm owners are maximizing the *expected value* of the future profit stream. If a flood occurs, profits for that year are $\theta\pi_t$, otherwise they're π_t. Here, π_t is an expectation over other factors that may affect profits, such as crop prices or farm policy changes. (For exposition, we assume $\theta < 1$. It can be negative.) We assume these other factors are uncorrelated with flood probability. Therefore, expected profits in year t are $(1 + (\theta - 1)/f_t)\pi_t$. f_t and π_t do not need to be constant over time, though in our hedonic analysis we will assume they are. Let ρ be the owner's discount rate.[2]

$$\Pi_0 = \sum_{t=0}^{\infty} \rho^t \pi_t \quad \text{with minimal flood risk}$$

2. This expression is somewhat simplified. We could, for instance, include a more complicated Bellman equation or consider the resale value of the farm. Doing so would add mathematical complexity, but would not increase economic insight nor guide our empirical specification.

$$\Pi = \sum_{t=0}^{\infty} \rho^t \left(1 + \frac{\theta - 1}{f_t}\right)\pi_t \quad \text{with } f_t \geq 1 \text{ freq.}$$

Where the properties of Π and Π_0 are assumed to be identical except for their flood risk. Define the hedonic discount for having more than minimal risk of flooding as the difference in logs between Π and Π_0.

$$\begin{aligned}
\log(\Pi_0) - \log(\Pi) &= \log \frac{\sum_{t=0}^{\infty} \rho^t \pi_t}{\sum_{t=0}^{\infty} \rho^t \left(1 + \frac{\theta-1}{f_t}\right)\pi_t} \\
&= \log \frac{\sum_{t=0}^{\infty} \rho^t \pi_t}{\left(1 + \frac{\theta-1}{f_t}\right)\sum_{t=0}^{\infty} \rho^t \pi_t} \quad \text{if } f_t \text{ is constant} \\
&= \log \frac{f_t}{f_t + \theta - 1} = \log f_t - \log(f_t + \theta - 1)
\end{aligned}$$

In the empirical section, we will first estimate the hedonic discount for each flood frequency, then fit an average slope. That is, we estimate $\log(\Pi_0) - \log(\Pi) \sim \hat{\beta} \log f_t$. (Since flood probability is the reciprocal of frequency, the coefficient here will be the negative of the value we would estimate if we were to use the log of flood probability.) As in a standard Rosen–Roback framework, this estimation leads to a bid curve for flood frequency.

The Rosen–Roback model also includes a supply side, which, combined with the demand side, gives a hedonic price schedule. A hedonic price schedule is the intersection of the buyers' bid functions and the sellers' supply functions. Each parcel has a continuous supply curve for the flood probability, even if it's quite steeply sloped, since reductions in flood risk (e.g. building additional levees around the property) are available and cost different amounts at different properties.

To take this theoretical model to the data, we make additional assumptions. These are strong, though fairly standard in the hedonic valuation literature. Specifically, we assume that the properties in our dataset are representative of all properties – there's no bias in coming up for sale.

Instead of using repeat sales to address omitted variables bias, we rely on county fixed effects and controls for variables that are important for the properties' current agricultural output and future potential development value. We discuss these controls further in section 3.

3 DATA

The setting for our analysis are counties adjacent to the Missouri River between Gavin's Point Dam and Kansas City. This region covers 29 counties in five states: Kansas, Missouri, Iowa, Nebraska and South Dakota. Our analysis relies primarily on three sources of data. First, the outcome variable, the sale prices for agricultural land, is constructed from county recorder and assessor data. Second, our independent variable of interest, the flood risk associated with the land in each property sale, is based on combination of a US Bureau of Reclamation (USBR) study of Missouri River flood patterns that predicts quantifies of Flood Stages and modeling of the predicted inundation associated with each flood stage based on topological modeling. Third, as control variables, we use a variety of hydrologic and soil characteristics, constructed from gridded soil survey geographic (GSSURGO) database maintained by the US Geological Survey (USGS), as well as county and year fixed effects.

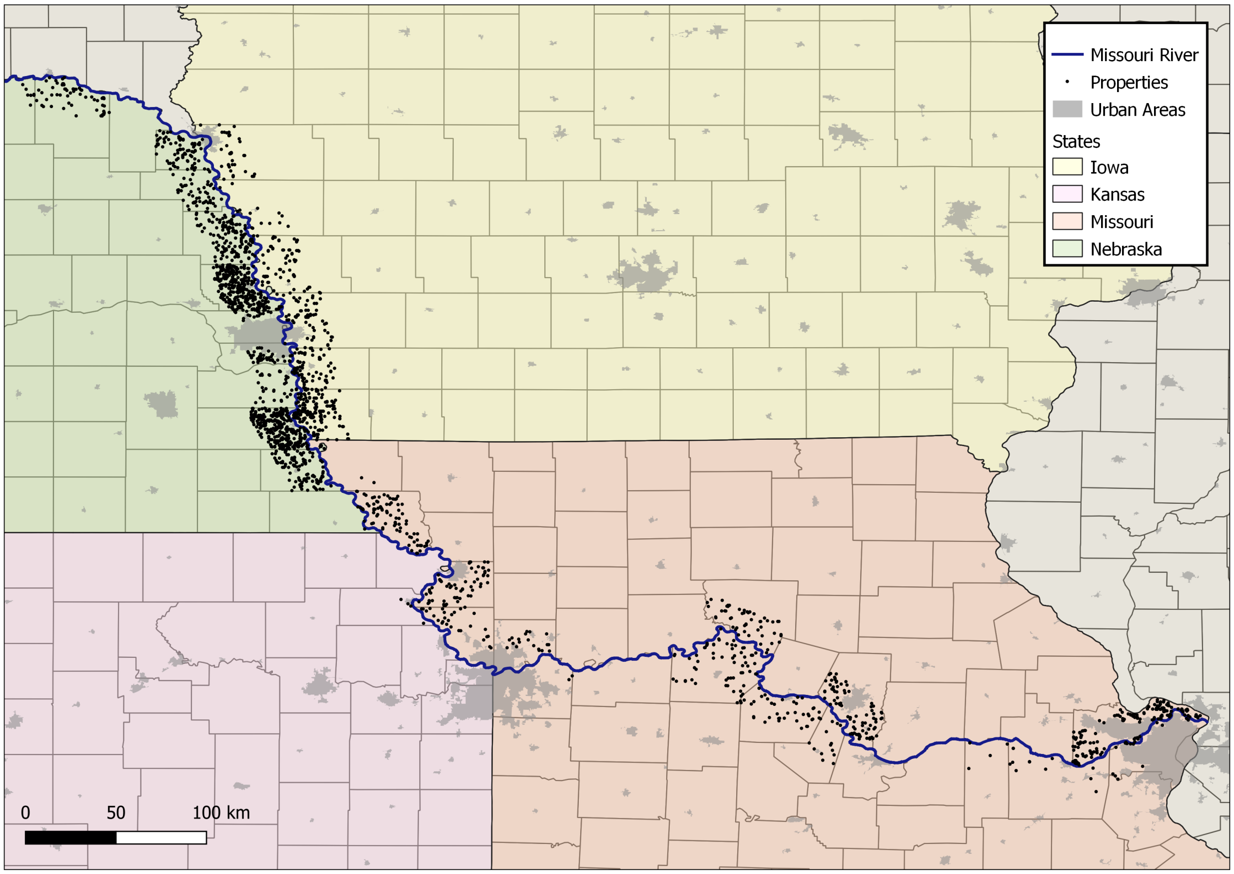
The outcome variable in our analysis is the per-acre sale price of agricultural land. We construct this variable from a dataset that compiles information from the assessors' and recorders' offices in each county adjacent to the Missouri River. Starting with the universe of sales from these counties' recorders, we restrict our data to arms-length sales of parcels that are classified by either the state or county as being in agricultural land-use, and parcels are least 40 acres. We restrict our data to sales after 1988, because records of property sales are not consistently digitized prior to this date. We restrict our data to sales for which we can identify the location and acreage of the transacted plots of land and for which the land is outside urban-areas. Sale price per acre is based on the growing acreage reported in county records. In order to ensure that the data in our analysis reflects the sale of land that is primarily agricultural, we excluded all transactions with a sale price less than \$200 or greater \$15,000 per acre (2018 dollars). The logic here is that properties with very high prices per acre are likely deriving their value from improvements.[3] Very low prices per acre may not be arms-length transactions (despite being recorded as arms-length in the CoreLogic data).

We note that a number of the counties in adjacent to the Missouri River are non-disclosure counties, which do not require the disclosure of sales price in real estate sales, which we exclude from our analysis. However, we do also include information supplied by assessors in these counties on sales in these counties. Because our measure of flood risk is based on a study of the Missouri River, we do not have flood risks on key tributaries of the Missouri, so we restrict our data to properties within 20 kilometers of the main-stem of the Missouri River.

We measure flood risk using the 2004 Upper Mississippi River System Flow Frequency Study, produced by the US Army Corps of Engineers (*Upper Mississippi River System Flow Frequency Study* 2003). This study estimates flood stages at every mile of the Mississippi and Missouri Rivers above Themes, IL, using gauge data from 1989 and 1998 to calibrate a regional stream-flow model. This study estimates the flood stage, the elevation of the river surface above sea level, for floods of frequencies ranging between two-year floods and five-hundred-year floods. In our analysis, we

3. We include an improvement-adjusted version of the regressions in table 4, but do not make this our primary specification because of concerns that assessed value may differ systematically from true value.

Figure 1: Map of study area, flood zones and the locations of property sales



calculate approximate flood zones by comparing parcel elevations to the elevation of river flood stages. This comparison is a first-order approximation that neglects local topology and the location of flood-control levees. Other studies that consider the relationship between flood risk and property value typically rely on maps produced by the Federal Emergency Management Agency. FEMA's flood maps are produced in a more ad-hoc manner by FEMA field agents who visit each county and produce maps that identify 100-year and 500-year flood zones and areas protected by federal flood levees based on historical claims and a county-specific study. Although they account for federally constructed levees, FEMA is prohibited by law from accounting for the impact of levees constructed by private landowners or local flood-control districts. FEMA's flood maps may also be prone to political influence due to their primary role in determining flood insurance rates paid by local landowners.

Figure 1 contains a map of the study area, including inundation zones, shown in blue shades and the locations of property sales included in our statistical analysis, shown as black dots.

Our analysis also controls for a variety of characteristics that may confound estimation the direct estimation of the impact of flood risk on sale price. To account for the impact that proximity to urban areas may have on property values, we calculate the distance of each property to the nearest urbanized area of more than 10,000 people.[4] We calculate the elevation of each property sold from topological maps.

4. In the context of our study area, the relevant urban areas are Sioux City IA, Omaha, NE/Council Bluffs IA,

We construct controls for soil characteristics from GSSURGO database. Differences in soil quality can potentially confound our estimates since it is both a key determinant of agricultural productivity and is highly correlated with flood risks. When land floods, fertile alluvial soils high in nutrients and organic matter are deposited, resulting in bottomland that can be more productive than lands that flood less frequently. To account for this difference in soil type and productivity, we control for the land capability classification. Land capability classes are a commonly used eight-point measure of soil productivity used by the USGS to group soils primarily based on their capability to produce commonly cultivated crops and pasture plants without deteriorating over a long period time. We also include soil quality controls that describe the ability of soils to cope with frequent flooding. We control for soil water storage, a measure of the volume of water soil can hold; we also control for permeability, a measure of how quickly water moves through the soil, and we control for erodability, a measure of the susceptibility of soil to erosion by runoff and flood impacts. The USGS estimates these characteristics as a function of the composition of soils and the distribution of size and the density of soil particles. For each of the soil characteristics that we control for, we divide the soils into terciles of quality (the top, middle, and bottom thirds of the distribution), and include in our regressions a dummy variable for each tercile of each soil quality variable.

We also include a dummy variable that describes whether the sale took place during the growing season, defined as May to October, and whether the property has improvements, such as barns, homes, grain-silos or other buildings.

Like much of the climate literature, but unlike the more recent hedonic valuation literature, we're unable to use repeat sales at the same property to identify causal effects. The fundamental reason is that flood risk has changed very little over our sample period, so there's no within-property valuation to identify the marginal price of flood risk. An additional logistical difficulty is that farm properties sell much less frequently than houses, so we have very few repeat sales in our panel, even if we had suitable variation in flood risk.

Table 1 shows the summary statistics by flood zone. A couple features of the table are worth highlighting. First, some of the flood zones have very few properties. Second, there's substantial variation in price per acre, even within this sample of agricultural properties that have had the price limited to the range of $200–15,000 per acre.

4 ESTIMATED RELATIONSHIP BETWEEN FLOOD RISK AND PROPERTY VALUE

Using the datasets we've built, we now turn to our hedonic regressions. In these analyses, we consider how similar properties differ in their per-acre sales prices when they face different flood frequencies. There are a number of ways one could define both the price and the flood frequency variables. We have a small number of properties that match our criteria, since agricultural properties tend to turn over less frequently. For each flood zone f, we estimate models of the form:

$$\begin{aligned}\log \text{price per acre}_i &= \beta_f D_i^f \\ &\quad + \gamma^f \text{distance to city}_i + \phi_i^f + \psi_i^f + \varepsilon_i\end{aligned}$$

Where the left-hand side is the sales price per acre, in $2018. The estimate of interest is β^f for flood category f. (Here, subscript i indexes observations within a regression, and superscript f indexes the separate regressions for different flood zones.) D_i^f is an indicator for

Lincoln NE, St Joseph MO, and Kansas City, MO/Kansas City, KS

Table 1: Summary statistics by flood frequency – USACE flood zones

Flood Frequency	Minimal	500-year	100-year	50-year	20-year	10-year	5-year	2-year
No. Obs.	2,181	25	34	5	26	57	172	198
Property Characteristics:								
Price per Acre	5.175	6.69	4.596	5.124	5.5	6.221	5.222	5.511
	(3.66)	(4.69)	(3.79)	(3.05)	(3.55)	(3.62)	(3.35)	(3.47)
Sale Amount	423.3	468.1	487	421.6	440.4	507.1	573.1	604.1
	(403)	(327)	(543)	(290)	(394)	(314)	(560)	(691)
Acreage	89.81	92.54	84.86	85.01	82.96	110.2	111.1	114.3
	(56.6)	(58.8)	(57.6)	(32.2)	(39.3)	(123)	(79.9)	(106)
Elevation (m)	332.3	287.9	291.1	274.5	248.1	249.9	219.7	209
	(60.9)	(46.1)	(41.5)	(78.6)	(61.2)	(66)	(71.7)	(67.7)
Distance to Urban	15.43	16.08	10.05	15.86	16.29	13.66	16.43	16.7
Center (km)	(9.37)	(9.96)	(5.94)	(8.08)	(7.25)	(7.75)	(9.45)	(9.06)
Soil Characteristics:								
Soil Water Storage	0.1884	0.1456	0.1442	0.1246	0.1657	0.1644	0.1609	0.154
	(0.0293)	(0.0417)	(0.0361)	(0.03)	(0.0397)	(0.0376)	(0.0367)	(0.0322)
Irrigation	3.292	2.937	2.354	2.8	2.533	2.195	2.408	2.837
Capacity Class	(1.41)	(1.26)	(0.929)	(0.447)	(1.09)	(1.17)	(1.14)	(1.15)
Permeability	6.332	3.159	7.848	1.784	9.722	12.3	17.2	10.29
	(7.17)	(5.05)	(16.3)	(2.93)	(18.7)	(26.2)	(24.8)	(17.7)
Erodability	0.424	0.3365	0.3136	0.3195	0.3834	0.3837	0.3904	0.3591
	(0.0717)	(0.0952)	(0.085)	(0.123)	(0.0962)	(0.0975)	(0.102)	(0.0898)
Growing Season	0.453	0.52	0.5	0.6	0.5385	0.5614	0.5	0.5556
(0/1)	(0.498)	(0.51)	(0.508)	(0.548)	(0.508)	(0.501)	(0.501)	(0.498)
Improvements	0.2581	0.12	0.05882	0	0.1538	0.05263	0.03488	0.0202
(0/1)	(0.438)	(0.332)	(0.239)	(0)	(0.368)	(0.225)	(0.184)	(0.141)

Source: CoreLogic, FEMA, USACE, and GSSURGO.
Table values are sample means and standard deviations. Price figures are in thousands of 2018 dollars. Statistics are for exclusive flood groups: 100-year is not included in 500-year.

whether property i is in category f. The ϕ_i and ψ_i represent fixed effects for soil characteristics as well as county and year fixed effects. We include distance to the nearest city as a control in all of our specifications. We estimate each flood category separately, and the coefficients β^f, γ^f, ϕ^f, and ψ^f are allowed to vary across regressions.

Our preferred specification uses inclusive flood categories. This definition means a parcel expected to flood once in a 100 years is *included* in the indicator for properties that are expected to flood at least every 500 years. We estimate results in table 2. In the 500-year column, we have an indicator variable that is set to one if the property is in the 500-year zone or above, and zero if the property faces a lower risk of flooding. In the 100-year column, we exclude the properties that are only in the 500-year bin, then have an indicator variable that is set to one if the property is in the 100-year zone or above, and zero if the property faces a lower risk of flooding. The pattern repeats for the other flood frequencies, up to the 10-year group, which excludes properties that are expected to flood once every 20–500 years, and includes observations from properties that are expected to flood every 10 years or more frequently, or less than once every 500 years. We also estimate regressions for the 5-year and 2-year flood frequencies. They are qualitatively similar to the 10-year. For the sake of space, we include those regressions in figure 2, but exclude them from the table.

This inclusive definition of flood zones is somewhat atypical for treatment effects regressions, but closely mirrors the standard way of doing this analysis in the flooding literature. For comparison, we also include the exclusive flood categories in appendix table 5. However, some of the groups are quite small, making the estimates noisy.

In table 2, we consider three sets of fixed effects. Specification A is our preferred specification, and includes fixed effects for each county, each year, and terciles of the soil variables, as well as a linear control for distance to the nearest city. (See our discussion in section 3 for more on the soil measures, their construction, and why they're important controls.) Specification B removes the soil fixed effects, but retains the county and year. Specification C removes all of the fixed effects. Comparing specification A to B tells us a bit about the importance of the soil controls, an important aspect of the heterogeneity in value in this context. The estimated coefficients are not statistically different from one another. While some of the differences in point estimates are economically significant, it's hard to make strong claims from such noisy estimates. Comparing specifications B and C highlights the importance of our county and year fixed effects.

Following standard practices, we estimate our standard errors using two-way clustering at the year and county level. We have 26 years and 29 counties in our sample. Clustered standard errors only consistent as the number of clusters approaches infinity. Because we have a limited number of clusters, we expect the errors we estimate to be somewhat too small.

We then take these estimated coefficients and estimate the slope relative to log flood frequency. The individual coefficients, and the line fit through them, are shown in figure 2. (The fit weights by the inverse of the variance of each estimate.) The slope of the line is −0.5%, with a 95% confidence interval of [−1.5%, 0.48%]. The interpretation of this elasticity is that if the flood frequency years were to double, e.g. from 10-years to 20-years, our predicted valuation would *decrease* by −0.5%. The sign of the coefficient is not consistent with economic intuition, though the confidence interval includes zero. The interval is narrow enough that we're able to rule out large values of the slope. (For instance, moving from the 5-year bin to above the 500-year bin requires seven doublings. Taking our confidence interval literally, seven doublings in flood frequency years leads to a predicted price change of [−11%, 3.4%]).

Additionally, we consider robustness checks in flood zone definition and outcome variable. These results are presented in the appendix. Across specifications, our results are fairly similar. We repeat our regressions using a narrower range of acceptable prices per acre: $645–12,000 instead of $200–15,000. These results are recorded in table 3.

We consider regressions that adjust for the assessed value of improvements on the property. (To calculate the dependent variable, we multiply the observed sales price by the ratio of assessed improvement value to total assessed value.) These results are in table 4.

Tables 6 and 7 have the version of this regression you would get if you only had FEMA 500- and 100-year flood information. (The two tables are the wide and narrow price cap versions, respectively. Both use log price per acre as the dependent variable, like the other regressions.) The FEMA results are qualitatively different (different sign), but also noisy and not statistically distinct from the results that use other flooding data.

Finally, we repeat our regressions using exclusive flood zone definitions. Unlike the previous tables, where e.g. the 100-year flood indicator includes properties in the 10-year bin, these regressions consider mutually exclusive bins. Table 5 estimate exclusive flood bin regressions. Some of these bins have very few observations – see counts in table 1) – so we don't take their wildly varying coefficient estimates too seriously. (More formally, we are far from a point where the asymptotic assumptions of clustered standard errors are valid.)

5 PREDICTED IMPACT OF CLIMATE CHANGE ON PROPERTY VALUE

Climate change will affect rainfall patterns, and therefore how frequently these properties flood. A 2012 report estimated that mean monthly flows on the Missouri river will increase perhaps 60–80% in June, already the peak month (U.S. Bureau of Reclamation 2012). With higher peak monthly flows, we expect flood frequency to increase at some properties. For example, a parcel that currently floods once every 100 years may begin to flood once every 10 years. For historical river flows, the difference between a 1% event and a 10% event is about a 49% increase in flows, leading to about 6 feet of river elevation.[5]

Changes in flooding are an active area of research in the climate literature – modeling high-variance, spatially uneven events like rain storms is much more challenging than temperature changes. These challenges arise on top of challenges of knowing what future greenhouse gas (GHG) emissions will be and precisely how sensitive the climate is GHG.

Taking the change in future flooding as given, our hedonic analysis is informative for valuing future adaptation. Specifically, our estimates tell us the value of climate damages if no further adaptation occurs. The other side of the same coin is that our estimates tell us the gross benefits from adaptation. We don't know the costs of adaptation – though we note that opportunities for adaptation are much more limited than in residential flooding.

In future work, we plan to quantitatively estimate the value of these gross benefits. For the current analysis, we note that properties that flood with moderate to high frequency bear the brunt of the hedonic discount. As flood probabilities increase, they're also the properties that will experience the highest additional damages, absent further adaptation.

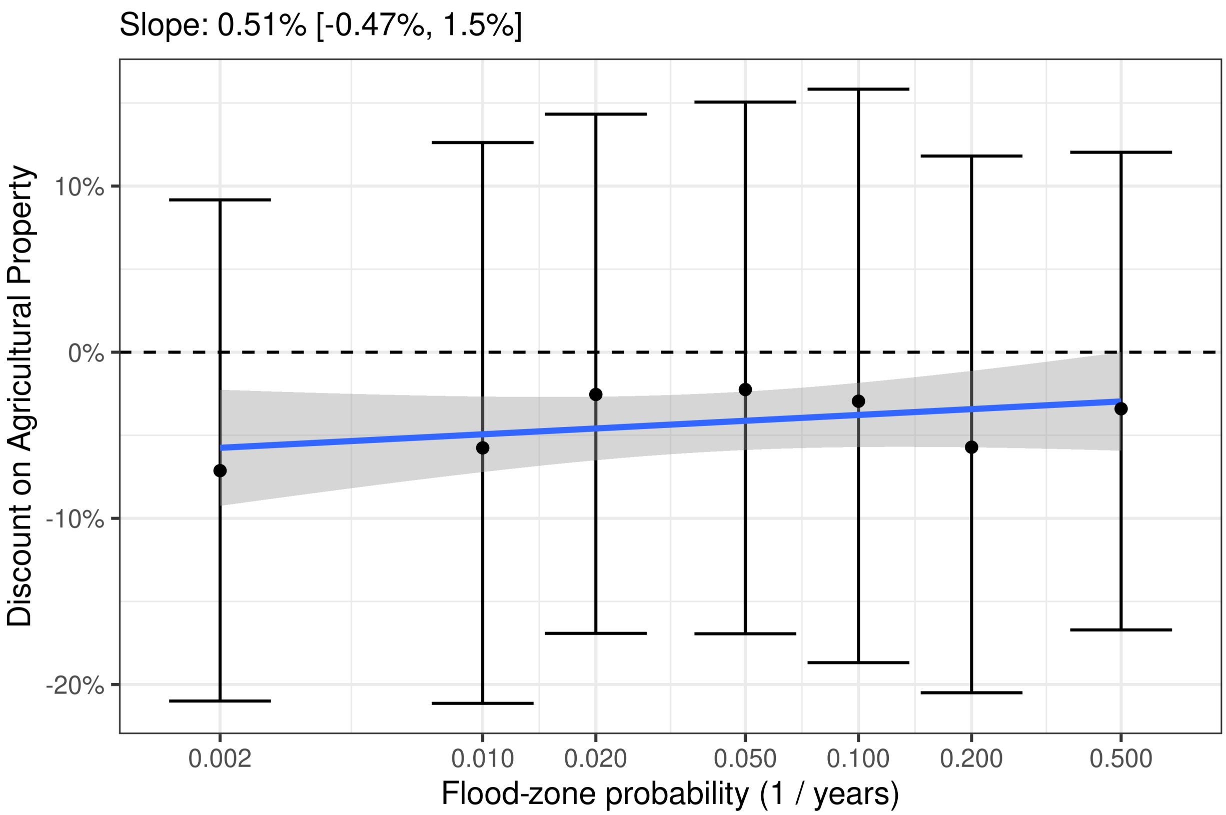
5. These numbers are for the flood gauge at St. Joseph, Missouri. These probabilities relationships vary somewhat along the length of the river see *Upper Mississippi River System Flow Frequency Study* 2003.

Table 2: Hedonic discount by flood zone – dependent variable is log price per acre

Flood bin:	500 yr	100 yr	50 yr	20 yr	10 yr
Spec A					
County + year FE: ✓	−7.13%	−5.76%	−2.54%	−2.25%	−2.95%
Soil FE: ✓	[−21.0%, 9.2%]	[−21.1%, 12.6%]	[−16.9%, 14.3%]	[−16.9%, 15.1%]	[−18.7%, 15.8%]
R^2	0.154	0.153	0.160	0.160	0.164
Within R^2	0.016	0.017	0.017	0.017	0.018
Spec B					
County + year FE: ✓	−7.67%	−7.12%	−3.60%	−3.36%	−3.79%
Soil FE: ✗	[−21.1%, 8.1%]	[−21.0%, 9.2%]	[−17.1%, 12.0%]	[−17.2%, 12.7%]	[−18.5%, 13.6%]
R^2	0.143	0.142	0.149	0.149	0.152
Within R^2	0.019	0.019	0.019	0.019	0.020
Spec C					
County + year FE: ✗	3.78%	4.28%	6.37%	6.37%	5.95%
Soil FE: ✗	[−8.0%, 17.1%]	[−6.8%, 16.6%]	[−5.8%, 20.1%]	[−6.2%, 20.6%]	[−7.0%, 20.7%]
R^2	0.038	0.037	0.038	0.038	0.039
N treat	552	492	458	453	427
N total	2733	2673	2639	2634	2608

Note: Results for indicator regressions on inclusive flood bins using USACE flood data. Each estimate is from a separate regression. Sample includes agricultural properties with price per acre between $200 and $15,000. Bolded estimates are statistically significantly different from zero. Square brackets indicate 95% CI, two-way clustered by county and year. Each column has the same number of treated and total observations.

Figure 2: Hedonic discounts across flood bins



Results correspond to table 2, specification A.

6 CONCLUSION

Agricultural land faces significant flooding risk, but almost all economic studies of flooding focus on residential property. In this study, we bring novel, though noisy, evidence of the value of hedonic discounts on agricultural land near the Missouri River. We estimate regressions of the land value of agricultural parcels using flexible bins of flood zones. Our analysis considers a much wider range of flood probabilities than previous work. We then aggregate those results to an overall hedonic slope. Our point estimate is fairly small, at a −0.5% *increase* for a doubling in flood risk. Our estimate is not significantly different from zero, and our confidence intervals exclude large changes in value.

These estimates, and the topic more broadly, are important to study empirically. Flood risk to agricultural land is expected to increase with climate change, and developing a deeper understanding of the relationship between flooding and property valuation is an essential piece of analysis to inform both flood risk mitigation and the value of climate change mitigation. We leave a more detailed climate projection for future work.

Appendices

A ESTIMATION ROBUSTNESS

In the following tables, we consider alternative data samples and estimation strategies. Table 3 and is similar to 2 except they use a narrower range of prices per acre: [645, 12000]. Table 4 adjusts the price per acre for the assessed value of improvements. Table 5 estimates *exclusive* flood bin regressions. Some of these bins have very few observations see counts in tables 1, so we don't take their wildly varying coefficient estimates too seriously. (More formally, we are far from a point where the asymptotic assumptions of clustered standard errors are valid.) Tables 6 and 7 have the version of this regression you would get if you only had FEMA 500- and 100-year flood information. (The two tables are the wide and narrow pricecap versions, respectively. Both use log price per acre as the dependent variable, like the other regressions.)

Table 3: Specification comparison (narrow pricecap, USACE flood zones). Dep var is log price per acre.

Flood bin:	500 yr	100 yr	50 yr	20 yr	10 yr
Spec A					
County + year FE: ✓	−1.97%	0.13%	3.36%	3.75%	2.02%
Soil FE: ✓	[−13.2%, 10.7%]	[−11.9%, 13.8%]	[−9.0%, 17.4%]	[−9.0%, 18.3%]	[−11.9%, 18.1%]
R^2	0.120	0.119	0.124	0.124	0.125
Within R^2	0.003	0.004	0.004	0.004	0.004
Spec B					
County + year FE: ✓	−3.32%	−1.66%	1.93%	2.26%	0.62%
Soil FE: ✗	[−15.0%, 10.0%]	[−13.6%, 11.9%]	[−9.1%, 14.3%]	[−9.2%, 15.1%]	[−11.5%, 14.4%]
R^2	0.112	0.110	0.114	0.114	0.115
Within R^2	0.004	0.004	0.004	0.004	0.005
Spec C					
County + year FE: ✗	3.60%	5.30%	7.61%	7.62%	7.01%
Soil FE: ✗	[−7.6%, 16.1%]	[−4.7%, 16.4%]	[−2.3%, 18.5%]	[−2.6%, 18.9%]	[−3.6%, 18.8%]
R^2	0.012	0.012	0.013	0.013	0.013
N treat	491	444	412	407	385
N total	2379	2332	2300	2295	2273

Note: Results for indicator regressions on inclusive flood bins using USACE flood data. Each estimate is from a separate regression. Sample includes agricultural properties with price per acre between $645 and $12,000. Bolded estimates are statistically significantly different from zero. Square brackets indicate 95% CI, two-way clustered by county and year. Each column has the same number of treated and total observations.

Table 4: Specification comparison (using assessed improvements, wide pricecap, USACE flood zones). Dep var is log price per acre.

Flood bin:	500 yr	100 yr	50 yr	20 yr	10 yr
Spec A					
County + year FE: ✓	−8.37%	−6.62%	−1.26%	−0.77%	0.54%
Soil FE: ✓	[−25.0%, 12.0%]	[−26.0%, 17.9%]	[−19.4%, 21.0%]	[−19.7%, 22.5%]	[−21.4%, 28.6%]
R^2	0.211	0.211	0.219	0.219	0.222
Within R^2	0.069	0.071	0.072	0.073	0.073
Spec B					
County + year FE: ✓	−7.78%	−7.05%	−1.91%	−1.58%	−0.09%
Soil FE: ✗	[−25.1%, 13.5%]	[−25.8%, 16.4%]	[−19.2%, 19.1%]	[−19.5%, 20.3%]	[−21.0%, 26.4%]
R^2	0.200	0.201	0.207	0.207	0.210
Within R^2	0.072	0.075	0.076	0.076	0.076
Spec C					
County + year FE: ✗	1.02%	0.97%	4.13%	4.09%	4.92%
Soil FE: ✗	[−13.4%, 17.9%]	[−13.3%, 17.6%]	[−11.2%, 22.1%]	[−11.7%, 22.7%]	[−12.7%, 26.1%]
R^2	0.108	0.109	0.111	0.111	0.111
N treat	423	365	334	329	306
N total	2558	2500	2469	2464	2441

NOTE: Results for indicator regressions on inclusive flood bins using USACE flood data. The dependent variable is log of sales price, adjusted by the assessed value of improvements. Each estimate is from a separate regression. Sample includes agricultural properties with price per acre between $200 and $15,000. Bolded estimates are statistically significantly different from zero. Square brackets indicate 95% CI, two-way clustered by county and year. Each column has the same number of treated and total observations.

Table 5: Specification comparison (exclusive USACE flood bins, wide-pricecap data). Dep var is log price per acre.

	Spec A	Spec B	Spec C
500 yr	2.45%	5.50%	15.90%
	[−28.29%, 46.36%]	[−27.74%, 54.03%]	[−22.51%, 73.34%]
200 yr	−27.63%	−25.49%	−10.17%
	[−57.18%, 22.30%]	[−57.81%, 31.58%]	[−48.87%, 57.80%]
100 yr	−29.93%	−29.61%	−20.30%
	[−71.00%, 69.28%]	[−70.61%, 68.57%]	[−63.04%, 71.82%]
50 yr	−7.38%	−6.30%	6.61%
	[−40.95%, 45.26%]	[−41.34%, 49.69%]	[−31.99%, 67.11%]
20-25 yr	13.25%	10.43%	11.18%
	[−21.54%, 63.47%]	[−21.91%, 56.15%]	[−21.03%, 56.53%]
10 yr	10.99%	8.70%	**29.77%**
	[−8.44%, 34.55%]	[−8.45%, 29.06%]	**[11.61%, 50.89%]**
5 yr	−11.57%	−13.18%	−0.81%
	[−30.49%, 12.49%]	[−32.15%, 11.10%]	[−17.85%, 19.76%]
2 yr	−1.64%	−0.89%	6.07%
	[−14.78%, 13.53%]	[−11.51%, 11.01%]	[−8.55%, 23.04%]
R^2	0.158	0.147	0.041
Within R^2	0.021	0.023	
N	2733	2733	2733
County + year FE:	✓	✓	✗
Soil FE:	✓	✗	✗

NOTE: Each column is one regression. The omitted category is flood risk below 500-yr. Sample includes agricultural properties with price per acre between $200 and $15,000.

Table 6: Specification comparison (inclusive FEMA flood bins, wide-pricecap data). Dep var is log price per acre.

Flood bin:	500 yr	100 yr
Spec A		
County + year FE: ✓	9.17%	11.91%
Soil FE: ✓	[−18.8%, 46.7%]	[−15.2%, 47.7%]
R^2	0.164	0.152
Within R^2	0.016	0.014
Spec B		
County + year FE: ✓	5.10%	10.27%
Soil FE: ✗	[−17.3%, 33.5%]	[−13.9%, 41.3%]
R^2	0.153	0.140
Within R^2	0.018	0.016
Spec C		
County + year FE: ✗	14.08%	13.79%
Soil FE: ✗	[−12.3%, 48.4%]	[−13.8%, 50.2%]
R^2	0.037	0.036
N treat	71	533
N total	1880	2342

NOTE: Results for indicator regressions on inclusive flood bins using FEMA flood data. Each estimate is from a separate regression. Sample includes agricultural properties with price per acre between $200 and $15,000. Bolded estimates are statistically significantly different from zero. Square brackets indicate 95% CI, two-way clustered by county and year. Each column has the same number of treated and total observations.

Table 7: Specification comparison (inclusive FEMA flood bins, narrow-pricecap data). Dep var is log price per acre.

Flood bin:	500 yr	100 yr
Spec A		
County + year FE: ✓	−1.72%	−2.06%
Soil FE: ✓	[−23.2%, 25.8%]	[−22.1%, 23.2%]
R^2	0.148	0.126
Within R^2	0.002	0.003
Spec B		
County + year FE: ✓	−6.37%	−3.36%
Soil FE: ✗	[−26.2%, 18.7%]	[−23.7%, 22.5%]
R^2	0.139	0.115
Within R^2	0.003	0.003
Spec C		
County + year FE: ✗	2.17%	2.25%
Soil FE: ✗	[−21.7%, 33.3%]	[−21.5%, 33.2%]
R^2	0.012	0.012
N treat	66	464
N total	1643	2041

NOTE: Results for indicator regressions on inclusive flood bins using FEMA flood data. Each estimate is from a separate regression. Sample includes agricultural properties with price per acre between $645 and $12,000. Bolded estimates are statistically significantly different from zero. Square brackets indicate 95% CI, two-way clustered by county and year. Each column has the same number of treated and total observations.

Bibliography

Atreya, Ajita, and Jeffrey Czajkowski. 2019. "Graduated Flood Risks and Property Prices in Galveston County." *Real Estate Economics* 47 (3): 807–844. https://doi.org/10.1111/1540-6229.12163. (Cited on page 49).

Atreya, Ajita, and Susana Ferreira. 2015. "Seeing is believing? Evidence from property prices in inundated areas." *Risk Analysis* 35 (5): 828–848. https://doi.org/10.1111/risa.12307. (Cited on page 49).

Atreya, Ajita, Susana Ferreira, and Warren Kriesel. 2013. "Forgetting the flood? An analysis of the flood risk discount over time." *Land Economics* 89 (4): 577–596. https://doi.org/10.3368/le.89.4.577. (Cited on page 49).

Beltrán, Allan, David Maddison, and Robert J.R. Elliott. 2018. "Is Flood Risk Capitalised Into Property Values?" *Ecological Economics* 146 (January): 668–685. https://doi.org/10.1016/j.ecolecon.2017.12.015. (Cited on page 49).

Bialaszewski, Dennis, and Bobby A. Newsome. 1990. "Adjusting Comparable Sales For Floodplain Location: The Case of Homewood, Alabama." *The Appraisal Journal* 58 (1): 114. (Cited on page 49).

Bin, Okmyung, Thomas W. Crawford, Jamie B. Kruse, and Craig E. Landry. 2008. "Viewscapes and flood hazard: Coastal housing market response to amenities and risk." *Land Economics* 84 (3): 434–448. https://doi.org/10.3368/le.84.3.434. (Cited on page 49).

Bin, Okmyung, and Jamie Brown Kruse. 2006. "Real estate market response to coastal flood hazards." *Natural Hazards Review,* https://doi.org/10.1061/(ASCE)1527-6988(2006)7:4(137). (Cited on page 49).

Bin, Okmyung, Jamie Brown Kruse, and Craig E. Landry. 2008. "Flood hazards, insurance rates, and amenities: Evidence from the coastal housing market." *Journal of Risk and Insurance* 75 (1): 63–82. https://doi.org/10.1111/j.1539-6975.2007.00248.x. (Cited on page 48).

Bin, Okmyung, and Craig E Landry. 2013. "Changes in implicit flood risk premiums: Empirical evidence from the housing market." *Journal of Environmental Economics and management* 65, no. 3 (May): 361–376. https://doi.org/10.1016/j.jeem.2012.12.002. (Cited on page 49).

Bin, Okmyung, and Stephen Polasky. 2004. "Effects of flood hazards on property values: Evidence before and after hurricane Floyd." *Land Economics* 80 (4): 490–500. https://doi.org/10.2307/3655805. (Cited on page 49).

Carbone, Jared C., Daniel G. Hallstrom, and V. Kerry Smith. 2006. "Can natural experiments measure behavioral responses to environmental risks?" *Environmental and Resource Economics* 33, no. 3 (March): 273–297. https://doi.org/10.1007/s10640-005-3610-4. (Cited on page 49).

Chao, P. T, J. L Floyd, and W Holliday. 1998. "Empirical Studies of the Effect of Flood Risk on Housing Prices." *Institute For Water Resources,* 72. https://www.iwr.usace.army.mil/portals/70/docs/iwrreports/98ps2.pdf%0Ahttp://www.iwr.usace.army.mil/Portals/70/docs/iwrreports/98ps2.pdf. (Cited on page 49).

Daniel, Vanessa E., Raymond J.G.M. Florax, and Piet Rietveld. 2009. "Flooding risk and housing values: An economic assessment of environmental hazard." *Ecological Economics* 69 (2): 355–365. https://doi.org/10.1016/j.ecolecon.2009.08.018. http://dx.doi.org/10.1016/j.ecolecon.2009.08.018. (Cited on page 49).

Fisher, Anthony C, W Michael Hanemann, Michael J Roberts, and Wolfram Schlenker. 2012. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment." *American Economic Review* 102, no. 7 (December): 3749–3760. https://doi.org/10.1257/aer.102.7.3749. (Cited on page 49).

Fridgen, Patrick M, and Steven D Shultz. 1999. "The Influence of the Threat of Flooding on Housing Values in Fargo , North Dakota and Moorhead , Minnesota." *Agricultural Economics,* no. 417, (cited on page 49).

Gibson, Matthew, Jamie T Mullins, and Alison Hill. 2019. "Climate Risk and Beliefs: Evidence from New York Floodplains." (Cited on page 49).

Hallstrom, Daniel G., and V. Kerry Smith. 2005. "Market Responses to Hurricanes." *Journal of Environmental Economics and Management* 50 (3): 541–561. https://doi.org/10.1016/j.jeem.2005.05.002. (Cited on page 49).

Hino, Miyuki, and Marshall Burke. 2021. "The effect of information about climate risk on property values." *Proceedings of the National Academy of Sciences* 118, no. 17 (April 27, 2021): e2003374118. https://doi.org/10.1073/pnas.2003374118. (Cited on page 49).

Kousky, Carolyn. 2010. "Learning from extreme events: Risk perceptions after the flood." *Land Economics* 86 (3): 395–422. https://doi.org/10.3368/le.86.3.395. (Cited on page 49).

MacDonald, Don N., James C. Murdoch, and Harry L. White. 1987. "Uncertain Hazards, Insurance, and Consumer Choice: Evidence from Housing Markets." *Land Economics,* https://doi.org/10.2307/3146293. (Cited on page 49).

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *The American Economic Review* 84 (4): 753–771. http://www.jstor.org/stable/2118029. (Cited on page 49).

Morgan, Ash. 2007. "The Impact of Hurricane Ivan on Expected Flood Losses, Perceived Flood Risk, and Property Values." *Journal of Housing Research* 16 (1): 47–60. https://doi.org/10.1080/10835547.2007.12091977. (Cited on page 49).

Nakićenović, Nebojša, Ogunlade Davidson, Gerald Davis, Arnulf Grübler, Tom Kram, Emilio Lebre La Rovere, Bert Metz, et al. 2000. *Special Report on Emissions Scenarios.* Technical report. Intergovernmental Panel on Climate Change. https://escholarship.org/uc/item/9sz5p22f. (Cited on page 48).

Nyce, Charles, Randy E. Dumm, G. Stacy Sirmans, and Greg Smersh. 2015. "The Capitalization of Insurance Premiums in House Prices." *Journal of Risk and Insurance* 82 (4): 891–919. https://doi.org/10.1111/jori.12041. (Cited on page 49).

Pope, Jaren C. 2008. "Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model." *Land Economics* 84, no. 4 (November 1, 2008): 551–572. https://doi.org/10.3368/le.84.4.551. (Cited on page 49).

Posey, John, and William H. Rogers. 2010. "The impact of special flood hazard area designation on residential property values." *Public Works Management and Policy* 15 (2): 81–90. https://doi.org/10.1177/1087724X10380275. (Cited on page 49).

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90, no. 6 (December): 1257–1278. https://doi.org/10.1086/261120. (Cited on page 50).

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *Journal of political economy* 82 (1): 34–55. (Cited on pages 47, 50).

Schwartz, Arthur L, Jr. 2001. "Environmental Determinants of Housing Prices: The Impact of Flood Zone Status." *Journal of Real Estate Research* 21 (1/2): 3–20. (Cited on page 49).

Severen, Christopher, Christopher Costello, and Olivier Deschênes. 2018. "A Forward-Looking Ricardian Approach: Do land markets capitalize climate change forecasts?" *Journal of Environmental Economics and Management* 89 (December): 235–254. https://doi.org/10.1016/j.jeem.2018.03.009. (Cited on page 49).

Shr, Yau Huo, and Katherine Y. Zipp. 2019. "The aftermath of flood zone remapping: The asymmetric impact of flood maps on housing prices." *Land Economics* 95 (2): 174–192. https://doi.org/10.3368/LE.95.2.174. (Cited on page 49).

Speyrer, Janet Furman, and Wade R. Ragas. 1991. "Housing prices and flood risk: An examination using spline regression." *The Journal of Real Estate Finance and Economics* 4 (4): 395–407. https://doi.org/10.1007/BF00219506. (Cited on page 49).

Struyk, Raymond J. 1971. "Flood Risk and Agricultural Land Values: A Test." *Water Resources Research* 7 (4): 789–797. https://doi.org/10.1029/WR007i004p00789. (Cited on page 48).

Troy, Austin, and Jeff Romm. 2004. "Assessing the price effects of flood hazard disclosure under the California natural hazard disclosure law (AB 1195)." *Journal of Environmental Planning and Management* 47, no. 1 (January): 137–162. https://doi.org/10.1080/0964056042000189844. (Cited on page 49).

Turnbull, Geoffrey K., Velma Zahirovic-Herbert, and Chris Mothorpe. 2013. "Flooding and liquidity on the bayou: The capitalization of flood risk into house Value and Ease-of-Sale." *Real Estate Economics* 41 (1): 103–129. https://doi.org/10.1111/j.1540-6229.2012.00338.x. (Cited on page 49).

U.S. Bureau of Reclamation. 2012. "Climate Change Analysis for the Missouri River Basin." *Technical Memorandum No. 86-68210-2012-03,* no. 86, (cited on pages 48, 56).

Upper Mississippi River System Flow Frequency Study: Appendix E. 2003. Technical report. US Army Corps of Engineers, Rock Island District, November. https://www.mvr.usace.army.mil/Portals/48/docs/FRM/UpperMissFlowFreq/App.%20E%20Kansas%20City%20Dist.%20Hydrology_Hydraulics.pdf. (Cited on pages 51, 56).

van der Wiel, Karin, Sarah B. Kapnick, Gabriel A. Vecchi, James A. Smith, P. C.D. Milly, and Liwei Jia. 2018. "100-year lower mississippi floods in a global climate model: Characteristics and future changes." *Journal of Hydrometeorology* 19 (10): 1547–1563. https://doi.org/10.1175/JHM-D-18-0018.1. (Cited on page 49).

Wang, Haoying. 2020. "Estimating Flood Risk Impact on Farmland Values Using Boundary Discontinuity: Evidence from Lancaster County, Pennsylvania." *Risk Analysis* (November). https://doi.org/10.1111/risa.13623. (Cited on pages 48, 49).

Transition

The next chapter focuses on state policies governing electricity and natural gas utility companies. These state-level decisions determine how utilities are paid for their investments, and how much utility customers have to pay for their service. Capital investments, from pipelines to solar farms, play an enormous role in shaping future US greenhouse gas emissions. While chapter 2 considered future changes in flood risk due to climate shifts, chapter 3 considers these very important capital investments. We focus on how much utilities are paid for their capital, the incentives utilities have to own more, and the effect of these incentives on capital ownership.

Chapter Three

Rate of Return Regulation Revisited

Coauthor: Stephen Jarvis

1 INTRODUCTION

In the two decades from 1997 to 2017, real annual capital spending on electricity distribution infrastructure by major utilities in the United States has doubled (EIA 2018a). Over the same time period annual capital spending on electricity transmission infrastructure increased by a factor of seven (EIA 2018b). The combined total is now more than $50 billion per year. This trend is expected to continue. Bloomberg New Energy Finance predicts that between 2020 and 2050, North and Central American investments in electricity transmission and distribution will likely amount to $1.6 trillion, with a further $1.7 trillion for electricity generation and storage (Henbest et al. 2020).[1]

These large capital investments could be due to the prudent actions of utility companies modernizing an aging grid. However, it is noteworthy that over this time period, utilities have earned sizeable regulated rates of return on their capital assets, particularly when set against the unprecedented low interest rate environment post-2008. As the economy-wide cost of capital has fallen, utilities' regulated rates of return have not fallen nearly as much. The exact drivers for this divergence are unclear, though we rule out large changes in riskiness in section 3. Whatever the underlying cause, the prospect of utilities earning excess regulated returns raises an age-old concern in the sector: the Averch–Johnson effect. When utilities are allowed to earn excess returns on capital, they will be incentivized to over-invest in capital assets. The resulting costs from "gold plating" are then passed on to consumers in the form of higher bills. Capital markets and the utility industry have undergone significant changes over the past 50 years since the early studies of utility capital ownership (Joskow 1972, 1974). In this paper we use new data to revisit these issues. We do so by exploring two main research questions. First, what can we say about the return on equity utilities are allowed by their regulators? Second, how has this return on equity affected utilities' capital investment decisions?

To answer our research questions, we use data on the utility rate cases of all major electricity and natural gas utilities in the United States spanning the past four decades (Regulatory Research Associates 2021). We combine this with a range of financial information on credit ratings, corporate borrowing and market returns. To examine possible sources of over-investment in more detail we also incorporate data from annual regulatory filings on

1. North and Central American generation/storage are reported directly. Grid investments are only reported globally, so we assume the ratio of North and Central America to global is the same for generation/storage as for grid investments.

individual utility capital spending.

We start our analysis by estimating the size of the gap between the allowed rate of return that utilities earn and the correct return on equity. A central challenge here, both for the regulator and for the econometrician, is estimating the correct cost of equity. We proceed by considering a range of approaches to simulating the correct cost of equity based on the observed rates of return and available measures of capital market returns. For the most part, our simulations ask "if approved RoE rates hadn't changed relative to some benchmark index since some baseline year, what would they be today?" We examine a number of benchmark indexes. None of these are perfect comparisons; the world changes over time, and different benchmarks may be more or less appropriate. Taken together, our various estimation approaches result in a consistent trend of excess rates of return. We find that the weighted median of the approved return on equity is 0.5–4 percentage points too high.[2] Applying these additional returns to the existing capital base we estimate excess costs to US customers of \$2–8 billion per year. The majority of these excess costs are from the electricity sector, though natural gas contributes as well.[3]

However, excess regulated returns on equity will also distort the incentives to invest in capital. To consider the change in the capital base, we turn to a regression analysis. Here we aim to identify how a larger RoE gap translates into over investment in capital. Identification is challenging in this setting, so we again employ several different approaches, with different identifying assumptions. In addition to a fixed effects approach, we examine an instrumental variables strategy. We draw on the intuition that when a rate case is decided a utility's RoE is *fixed* at a particular nominal percentage for several years. The cost of capital in the rest of the economy, and therefore the true RoE, will shift over time. We use these shifts in the timing and duration of rate cases as an instrument for changes in the RoE gap. We argue that the instrument is valid, after controlling for an appropriate set of fixed effects. Across the range of specifications used, we find a broadly consistent picture. In our preferred specification we find that an additional percentage point increase in the RoE gap leads to the allowed increase in capital rate base to be about 5 percent higher.

2 BACKGROUND

Electricity and natural gas utility companies are regulated by government utility commissions, which allow the companies a geographic monopoly and, in exchange, regulate the rates the companies charge. These utility commissions are state-level regulators in the US. They set consumer rates and other policies to allow investor owned utilitys (IOUs) a designated rate of return on their capital investments, as well as recovery of non-capital costs. This rate of return on capital is almost always set as a nominal percentage of the installed capital base. For instance, with an installed capital base worth \$10 billion and a rate of return of 8%, the utility is allowed to collect \$800 million per year from customers for debt service and to provide a return on equity to shareholders. State utility commissions typically update these nominal rates every 3–6 years.

Utilities own physical capital (power plants, gas pipelines, repair trucks, office buildings, etc.). The capital depreciates over time, and the

2. Here we weight by the utilities' ratebase, so our results are not over-represented by very small utilities.

3. For comparison, total 2019 electricity sales by investor owned utilities were \$204 billion, on 1.89 PWh of electricity (US Energy Information Administration 2020a). Natural gas sales to consumers are \$146 billion on 28.3 trillion cubic feet of gas (These gas figures include sales to residential, commercial, industrial, and electric power, but not vehicle fuel. They include including all sales, not just those by investor owned utilities. US Energy Information Administration 2020b.)

set of all capital the utility owns is called the ratebase (the base of capital that rates are calculated on). Properly accounting for depreciation is far from straightforward, but we will not focus on that challenge in this paper. This capital ratebase has an opportunity cost of ownership: instead of buying capital, that money could have been invested elsewhere. IOUs fund their operations through issuing debt and equity, typically about 50%/50%. (For this paper, we focus on common stocks. Utilities issue preferred stocks as well, but those form a very small fraction of utility financing.) The weighted average cost of capital is the weighted average of the cost of debt and the cost of equity.

Utilities are allowed to set rates to recover all of their costs, including this cost of capital. For some expenses, like fuel purchases, it's easy to calculate the companies' costs. For others, like capital, the state public utilities commissions are left trying to approximate the capital allocation at a cost competitive capital markets would provide, if the utility was a competitive company, rather than a regulated monopoly. The types of capital utilities own, and their opportunities to add capital to their books, vary across states and time. Utilities in vertically integrated states might own a large majority of their own generation, the transmission lines, and the distribution infrastructure. Other utilities are "wires only," buying power from independent power producers and transporting it over their lines. Natural gas utilities are typically pipeline only – the utility doesn't own the gas well or processing plant.

In the 1960s and 70s, state public utilities commissions (PUCs) began adopting automatic fuel price adjustment clauses. Rather than opening a new rate case, utilities used an established formula to change their customer rates when fuel prices changed. The same automatic adjustment has not happened for capital costs, despite large swings in the nominal cost of capital over the past 50 years. We're aware of one state (Vermont) that has an automatic update rule; we'll discuss that rule in more detail in section 4.1, where we consider various approaches of estimating the RoE gap.[4]

The cost of debt financing is by no means simple, particularly for a forward-looking decision-maker who isn't allowed to index to benchmark values, but is easier to estimate than the cost of equity financing. The cost of debt is the cost of servicing historical debt, and expected costs of new debt that will be issued before the next rate case. The historical cost is known, and can serve a direct basis for future expectations. In our data, we see both the utilities' requested and approved return on debt. It's notable that the requested and approved amounts are very close for debt, and much farther apart for equity.

The cost of equity financing is more challenging. Theoretically, it's the return shareholders require on their investment in order to invest in the first place. The Pennsylvania Public Utility Commission's ratemaking guide notes this difficulty (Cawley and Kennard 2018):

> Regulators have always struggled with the best and most accurate method to use in applying the [*Federal Power Commission v. Hope Natural Gas Company* (1944)] criteria. There are two main conceptual approaches to determine a proper rate of return on common equity: "cost" and "the return necessary to attract capital." It must be stressed, however,

4. At least one other state, California, had an automatic adjustment mechanism that has since been abandoned. Regulators at the California PUC feel that the rule, called the cost of capital mechanism (CCM), performed poorly. "The backward looking characteristic of CCM might have contributed to failure of ROEs in California to adjust to changes in financial environment after the financial crisis. The stickiness of ROE in California during this period, in the face of declining trend in nationwide average, calls for reassessment of CCM." (Ghadessi and Zafar 2017)

> that no single one can be considered the only correct method and that a proper return on equity can only be determined by the exercise of regulatory judgment that takes all evidence into consideration.

Unlike debt, where a large fraction of the cost is observable and tied to past issuance, the cost of equity is the ongoing, forward-looking cost of holding shareholders' money. Put differently, the RoE is applied to the entire ratebase – unlike debt, there's typically no notion of paying a specific RoE for specific stock issues.

Regulators employ a mixture of models and subjective judgment. Typically, these formal models, as well as the more subjective evaluations, benchmark against other US utilities (and often utilities in the same geographic region). There are advantages to narrow benchmarking, but when market conditions change and everyone is looking at their neighbors, rates will update very slowly.

In figure 1 we plot the approved return on equity over 40 years, with various risky and risk-free rates for comparison. The two panels show nominal and real rates. Consistent with a story where regulators adjust slowly, approved RoE has fallen slightly (in both real and nominal terms), but much less than other costs of capital. This price stickiness by regulators also manifests in peculiarities of the rates regulators approve. Rode and Fischbeck (2019) notes the fact that regulators seem reluctant to set RoE below a nominal 10%.

That paper, Rode and Fischbeck (2019), is the closest to ours in the existing literature. The authors use the same rate case dataset we do, and note a similar widening of the spread between the approved return on equity and 10-year Treasury rates. That paper, unlike ours, dives into the financial modeling, using the standard capital asset pricing model (CAPM) to examine potential causes of the increase the RoE spread. In contrast, we consider a wider range of financial benchmarks (beyond 10-year Treasuries) and ask more pointed questions about "what should rates be today if past relationships held?" and "how much has this RoE gap incentivized utilities to own more capital?"

Using CAPM, Rode and Fischbeck (2019) rule out a number of financial reasons we might see increasing RoE spreads. Possible reasons include utilities' debt/equity ratio, the asset-specific risk (CAPM's β), or the market's overall risk premium. None of these are supported by the data. A pattern of steadily increasing debt/equity could explain an increasing gap, but debt/equity has fallen over time. Increasing asset-specific risk could explain an increasing gap, but asset risk has (largely) fallen over time. (They use the Dow Jones Utility Average as a measure of utility asset risk.) An increasing market risk premium has could explain an increased spread between RoE and riskless Treasuries, but the market risk premium has fallen over time. Appendix figure 8, reproduced from Rode and Fischbeck (2019), shows the evolution of asset risk and the market risk premium over time.

Prior research has highlighted the importance of macroeconomic changes, and that these often aren't fully accounted for in utility commission ratemaking (Salvino 1967; Strunk 2014). Because rates of return are typically set in fixed nominal percentages, rapid changes in inflation can dramatically shift a utility's real return. This pattern is visible in figure 1 in the early 1980s. Inflation has lower and much more stable in recent years,

Many authors have written a great deal about modifying the current system of investor-owned utilities. Those range from questions of who pays for fixed grid costs to the role of government ownership or securitization (Borenstein, Fowlie, and Sallee 2021; Farrell 2019). For this project, we assume the current structure of investor-owned utilities, leaving aside other questions of how to set rates across different groups of customers or who owns the capital.

Finally, we note that a utility's approved rate of return or return on equity might differ from the realized return. In this paper, we focus on approved values. Other recent work, e.g. Hausman (2019), highlights important differences between approved costs and realized prices that customers face.

3 DATA

To answer our research questions, we use a database of resolved utility rate cases from 1980 to 2021 for every electricity and natural gas utility that either requested a nominal-dollar ratebase change of $5 million or had a ratebase change of $3 million authorized (Regulatory Research Associates 2021). Summary statistics on these rate cases can be seen in table 1.

We transform this panel of rate case events into an unbalanced utility-by-month panel, filling in the rate base and rate of return variables in between each rate case. There are some mergers and splits in our sample, but our SNL Financial (SNL) data provider lists each company by its present-day (2021) company name, or the company's last operating name before ceased to exist. With this limitation in mind, we construct our panel by (1) not filling data for a company before its first rate case in a state, and (2) dropping companies five years after their last rate case. In contexts where a historical comparison is necessary, but the utility didn't exist in the benchmark year, we use average of utilities that did exist in that state, weighted by ratebase size.

We match with data on S&P credit ratings, drawn from SNL's *Companies (Classic) Screener* (2021) and Wharton Research Data Services (WRDS)' *Compustat S&P legacy credit ratings* (2019). Most investor-owned utilities are subsidiaries of publicly traded firms. We use the former data to match as specifically as possible, first same-firm, then parent-firm, then same-ticker. We match the latter data by ticker only. Then, for a relatively small number of firms, we fill forward.[5] Between these two sources, we have ratings data are available from December 1985 onward. Approximately 80% of our utility–month observations are matched to a rating. Match quality improves over time: approximately 89% of observations after 2000 are matched.

These credit ratings have changed little over 35 years. In figure 2 we plot the median (in black) and various percentile bands (in shades of blue) of the credit rating for utilities active in each month. We note that the median credit rating has not changed much over time. The distribution of ratings is somewhat more compressed in 2021 than in the 1990s. While credit ratings are imperfect, we would expect rating agencies to be aware of large changes in riskiness.[6] Instead, the median credit rating for electricity utilities is A−, as it was for all of the 1990s. The median credit rating for natural gas utilities is also A−, down from a historical value of A.

Beyond credit ratings, we also use various market rates pulled from Federal Reserve Economic Data (FRED). These include 1-, 10-, and 30-year treasury yields, the core CPI, bond yield indexes for corporate bonds rated by Moody's as Aaa or Baa, as well as those rated by S&P as AAA, AA, A, BBB, BB, B, and CCC or lower.[7]

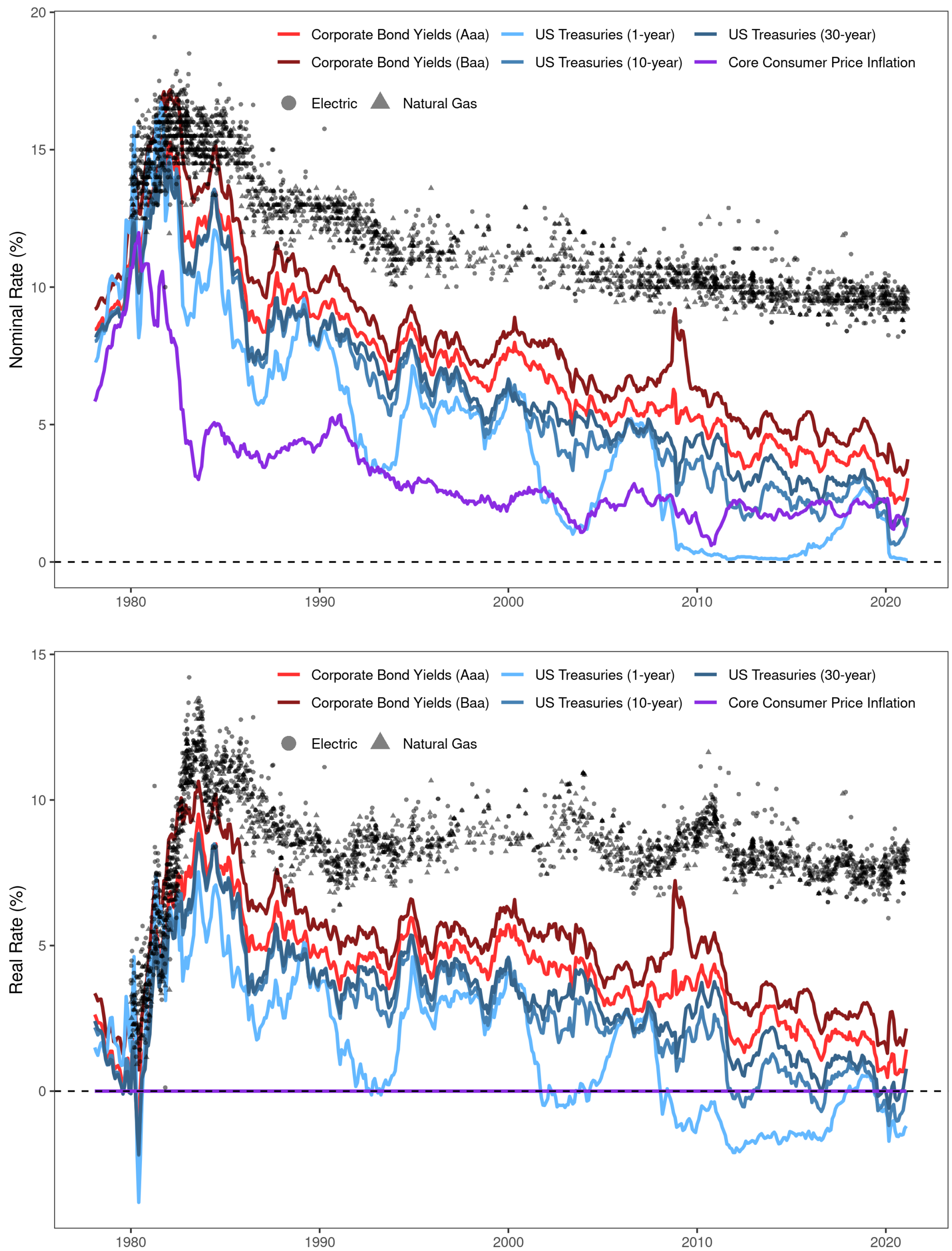
Matching these two datasets – rate cases and macroeconomic indicators – we construct the

5. When multiple different ratings are available, e.g. different ratings for subsidiaries trading under the same ticker, we take the median rating. We round down in the case of an even number of ratings, both here and in figure 2.

6. For utility risk to drive up the firms' cost of equity but not affect credit ratings, one would need to tell a very unusual story about information transmission or the credit rating process.

7. Board of Governors of the Federal Reserve System (2021a, 2021b, 2021c), US Bureau of Labor Statistics (2021), Moody's (2021a, 2021b), and Ice Data Indices, LLC (2021b, 2021a, 2021f, 2021d, 2021c, 2021g, 2021e).

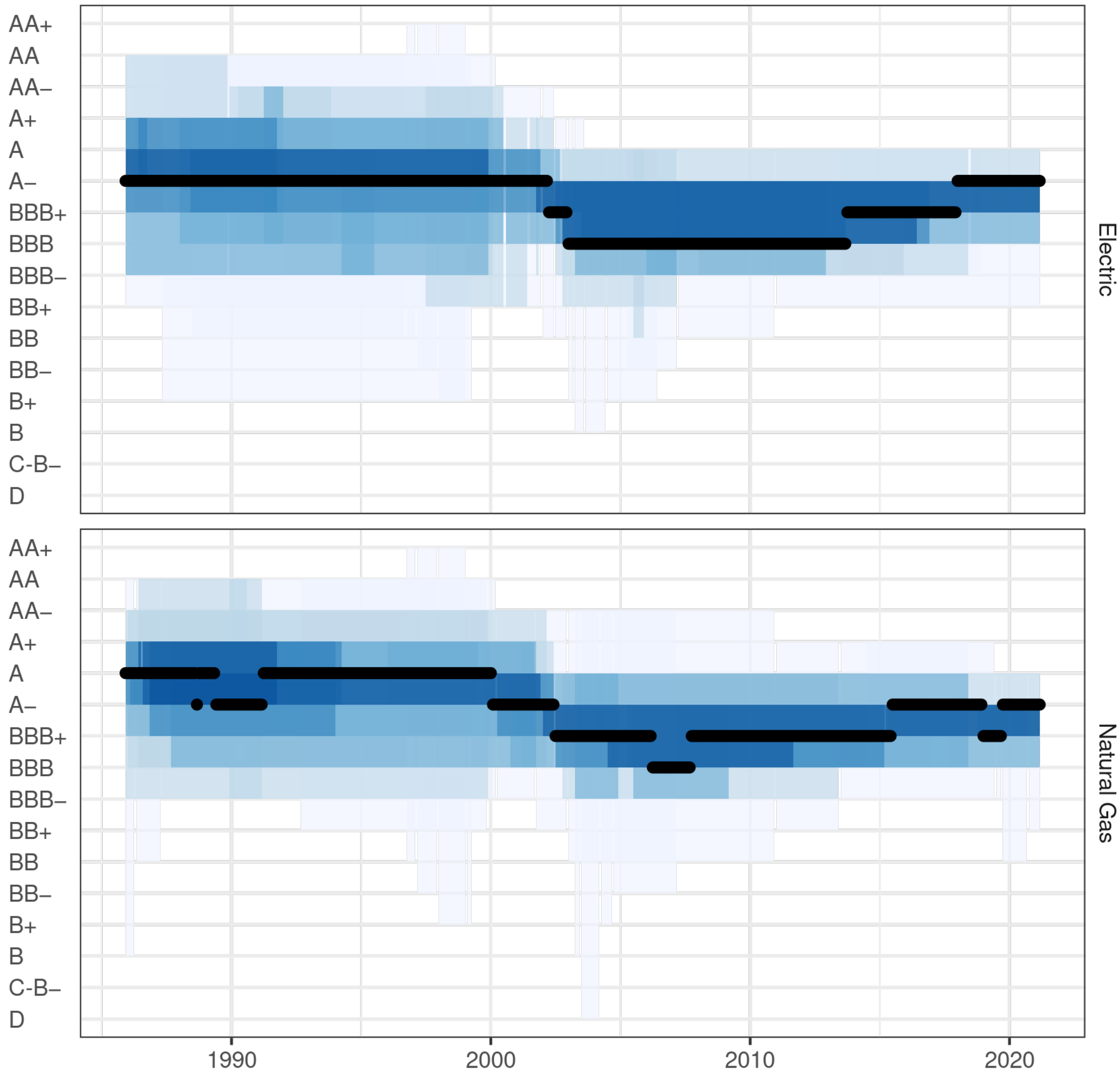
Figure 1: Return on Equity and Financial Indicators: Nominal and Real



Notes: These figures show the approved return on equity for investor-owned US electric and natural gas utilities. Each dot represents the resolution of one rate case. Real rates are calculated by subtracting consumer price index (CPI). Between March 2002 and March 2006 30-year Treasury rates are interpolated from 1- and 10-year rates.

Sources: Regulatory Research Associates (2021), Moody's (2021a, 2021b), Board of Governors of the Federal Reserve System (2021a, 2021b, 2021c), and US Bureau of Labor Statistics (2021).

Figure 2: Credit ratings have changed little in 35 years



Note: Black lines represent the median rating of the utilities active in a given month. We also show bands, in different shades of blue, that cover the 40–60 percentile, 30–70 percentile, 20–80 percentile, 10–90 percentile, and 2.5–97.5 percentile ranges. (Unlike later plots, these *are not* weighted by ratebase.) Ratings from C to B– are collapsed to save space.
Source: *Companies (Classic) Screener* (2021) and *Compustat S&P legacy credit ratings* (2019).

Table 1: Summary Statistics

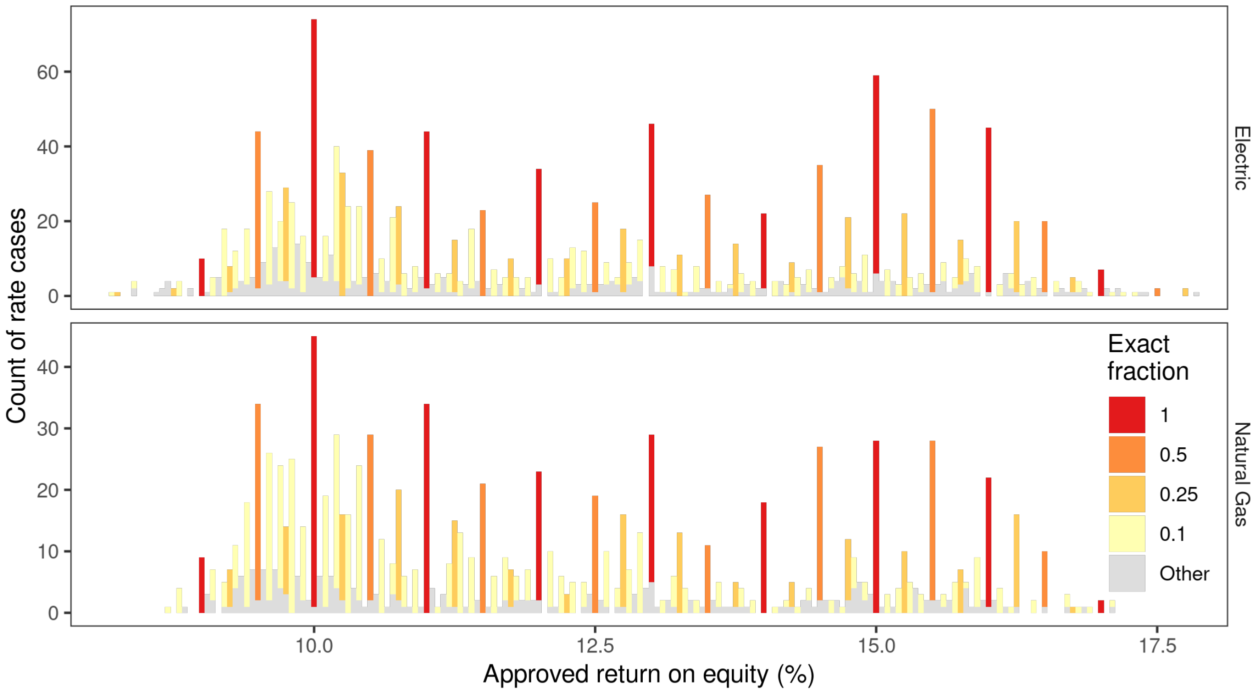
Characteristic	N	Electric	Natural Gas
Rate of Return Proposed (%)	3,324	9.95 (1.98)	10.07 (2.07)
Rate of Return Approved (%)	2,813	9.59 (1.91)	9.53 (1.95)
Return on Equity Proposed (%)	3,350	13.22 (2.69)	13.06 (2.50)
Return on Equity Approved (%)	2,852	12.38 (2.40)	12.05 (2.24)
Return on Equity Proposed Spread (%)	3,350	6.72 (2.18)	6.95 (1.99)
Return on Equity Approved Spread (%)	2,852	5.62 (2.27)	5.68 (2.10)
Return on Debt Proposed (%)	3,247	7.48 (2.11)	7.47 (2.16)
Return on Debt Approved (%)	2,633	7.54 (2.06)	7.44 (2.16)
Equity Funding Proposed (%)	3,338	45 (7)	48 (7)
Equity Funding Approved (%)	2,726	44 (7)	47 (7)
Rate Case Duration (mo)	3,713	9.1 (5.1)	8.1 (4.3)
Rate Base Increase Proposed ($ mn)	3,686	84 (132)	24 (41)
Rate Base Increase Approved ($ mn)	3,672	40 (84)	12 (25)
Rate Base Proposed ($ mn)	2,366	2,239 (3,152)	602 (888)
Rate Base Approved ($ mn)	1,992	2,122 (2,991)	583 (843)

NOTES: This table shows the rate case variables in our rate case dataset. Values in the Electric and Natural Gas columns are means, with standard deviations in parenthesis.
Approved values are approved in the final determination, and are the values we use in our analysis. Some variables are missing, particularly the approved rate base. The RoE spread in this table is calculated relative to the 10-year Treasury rate.
SOURCE: Regulatory Research Associates (2021) and author calculations.

timeseries shown in figure 1. A couple of features jump out, as we mentioned in the introduction. The gap between the approved return on equity and other measures of the cost of capital have increased substantially over time. At the same time, the return on equity has decreased over time, but much more slowly than other indicators. We quantify these observations in section 5.

We note that there are other distortions or ad-hoc evaluations in the PUC process. Rode and Fischbeck (2019) note a hesitancy for PUCs to set RoE below a nominal 10% level. We replicate this finding. In addition, we also note a bias toward round numbers, where regulators tend to approve RoE values at integers, halves, quarters, and tenths of percentage points. This finding is demonstrated in figure 3. We believe the true, unknown, cost of equity is smoothly distributed. If for instance, a PUC rounds in a way that changes the allowed RoE by 10 basis points (0.1%), the allowed revenue on the existing ratebase for the average electric utility in 2019 would change by $114 million. (The median is lower, at $52 million.) Small deviations have large implications for utility revenues and customer payments, though we don't know if rounding has a systematic bias toward higher or lower RoE. Of course, RoE values that aren't set at round numbers might not be any closer to the correct RoE. We leave this round number bias, as well as the above-10% stickiness, for future research.

Figure 3: Return on equity is often approved at round numbers



Colors highlight values of the nominal approved RoE that fall exactly on round numbers. More precisely, values in red are integers. Values in dark orange are integers plus 50 basis points (bp). Lighter orange are integers plus 25 or 75 bp. Yellow are integers plus one of {10, 20, 30, 40, 60, 70, 70, 80, 90} bp. All other values are gray.
Histogram bin widths are 5 bp. Non-round values remain gray if they fall in the same histogram bin as a round value. In that case, the bars are stacked.

Source: Regulatory Research Associates (2021).

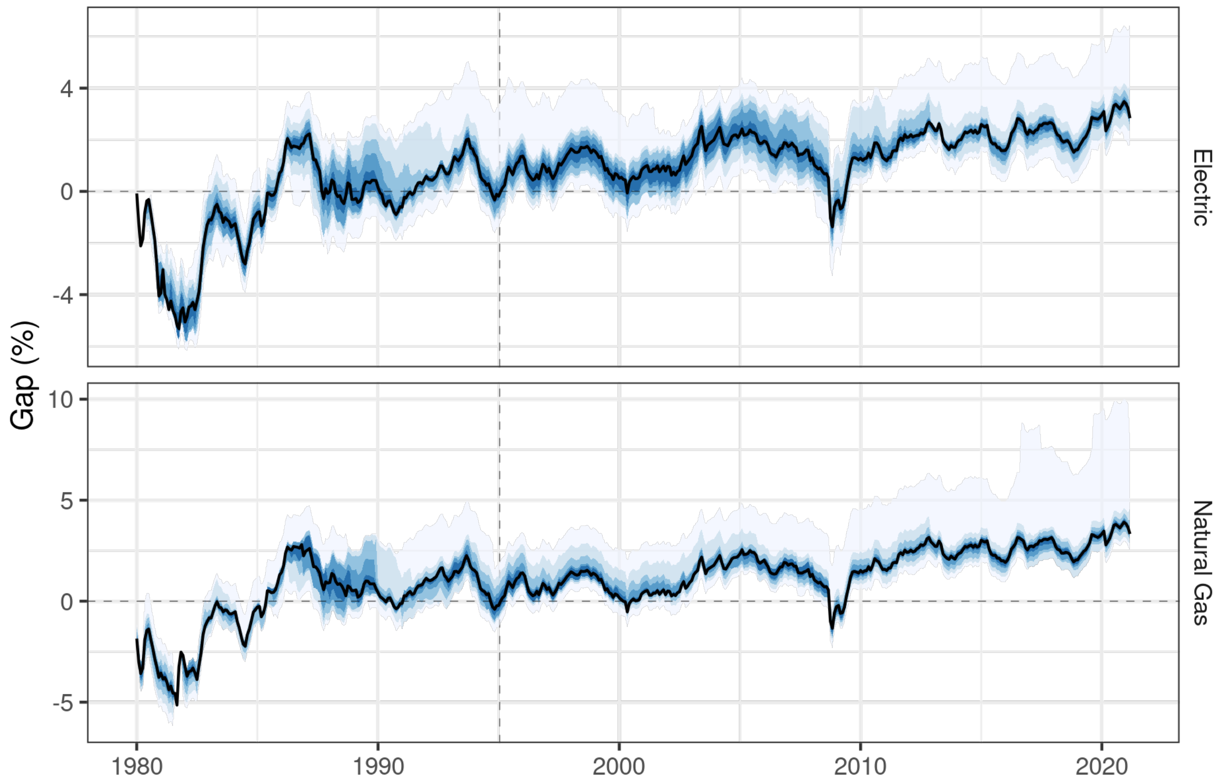
4 EMPIRICAL STRATEGY

4.1 RETURN ON EQUITY GAP

Knowing the return on equity (RoE) gap size is a challenge, and we take a couple of different approaches. None are perfect, but collectively, they shed light on the question. For each of the strategies we outline below (in sections 4.1.1, 4.1.2, 4.1.3, and 4.1.4) we plot the timeseries of the RoE gap. These are plotted in figures 4, 5, 6, and 7. Many of these strategies pick a specific time period as a benchmark. For all of these, we use January 1995. For the most part, our RoE gap results are flat over time (in the case of CPI) or steadily upward sloping (in the case of corporate bonds). The choice of baseline date determines where zero is, so changing the baseline date will shift the overall magnitude of the gap. As long as the baseline date isn't in the middle of a recession, our qualitative results don't depend strongly on the choice.

In each plot, we present the median of our RoE gap estimates, weighting by the utility's ratebase (in 2019 dollars). Our goal is to show the median of ratebase dollar value, rather than the median of utility companies, as the former is more relevant for understanding the impact of the RoE gap. We also show bands, in different shades of blue, that cover the 40–60 percentile, 30–70 percentile, 20–80 percentile, 10–90 percentile, and 2.5–97.5 percentile (all weighted by ratebase).

Figure 4: Return on equity gap, benchmarking to Baa-rates corporate bonds



Base year is 1995. Line represents median; shading represents ranges that cover the central 20, 40, 60, 80, and 95% of total IOU ratebase. See calculation details in section 4.1.1.

4.1.1 Indexed to Corporate Bonds

We first consider a benchmark index of corporate bond yields, rated Baa by Moody's.[8] The idea here is to ask if the *average* spread against the Baa rating hadn't changed since the baseline, what would the RoE be today? The results are plotted in figure 4. Moody's Baa is approximately equivalent to S&P's BBB, which is at or slightly below our most of the utilities in our data. We use January 1995 as our baseline. Our findings are qualitatively the same for other dates, though the magnitude differs.

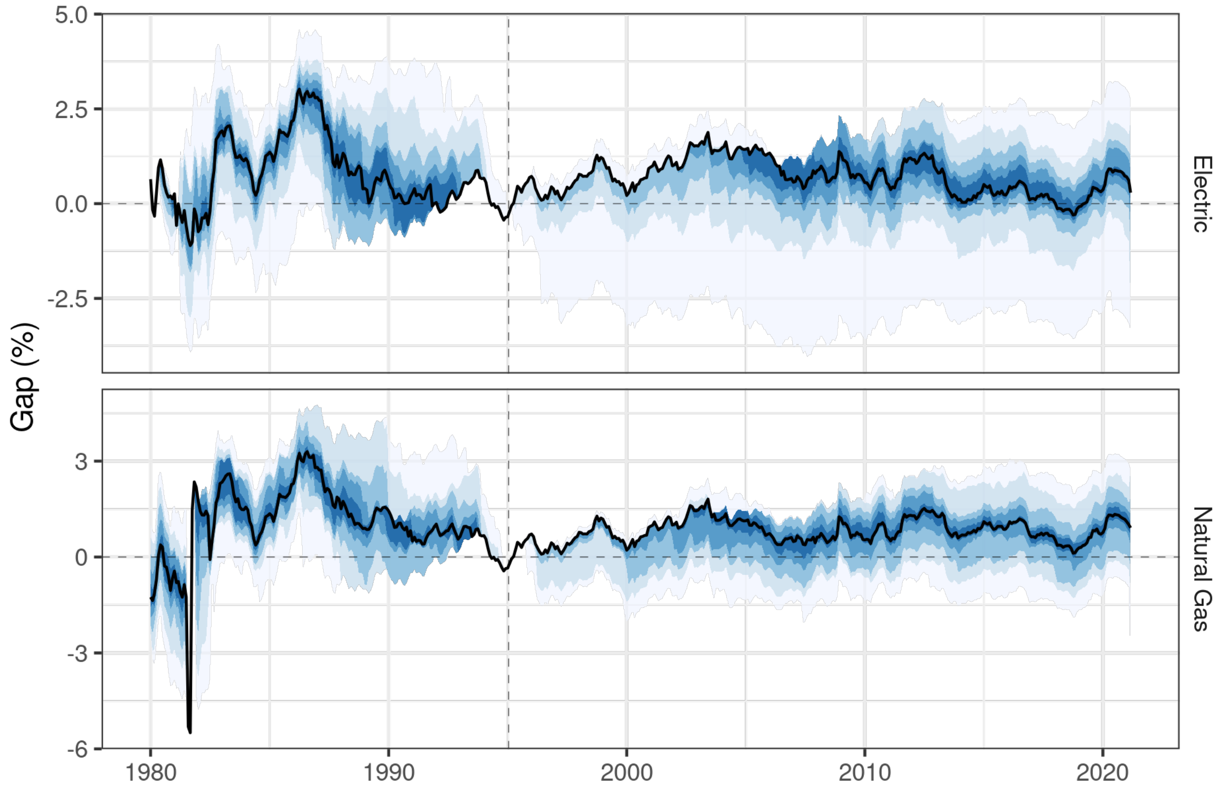
Making comparisons to debt instruments in this way, rather than benchmarking to some economy-wide cost of equity, means the measure of the RoE gap likely understates the gap. Rode and Fischbeck (2019) points out that (1) the market-wide equity risk premium has declined over the period and (2) the same is true for the utility sector.[9] Therefore, we would expect the mean spread against Baa bond yields to have declined, but instead, the spread has increased.

To calculate these results we first find the spread between the approved return on equity and the Moody's Baa rate for each utility in each state in each month. We then take the average at our baseline and simulate what that spread would be if the overall average

8. This index is one of two rating-specific corporate bonds indexes that's available for our entire study period. The other is Moody's Aaa.

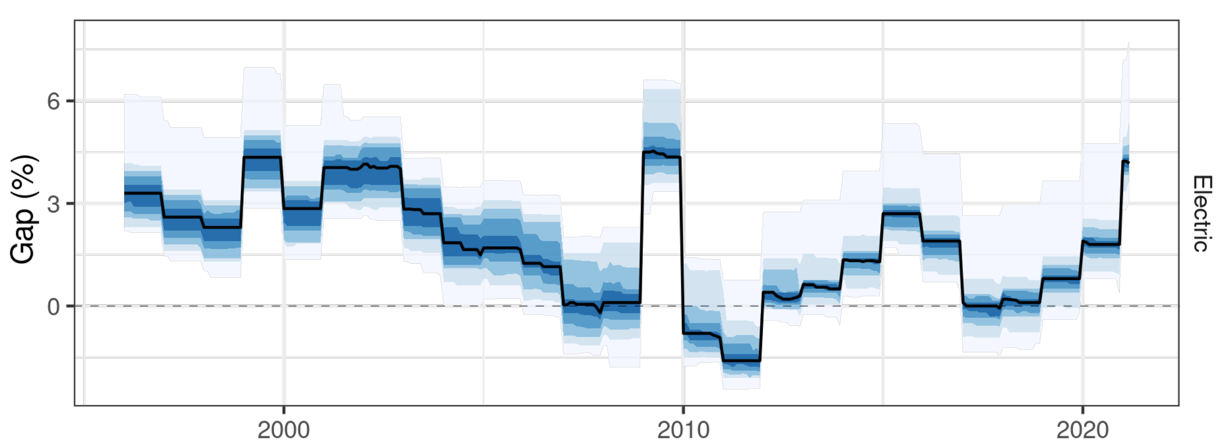
9. To the extent that observed utility stock returns are endogenous to the approved RoE, point #2 might be biased (Werth 1980).

Figure 5: Return on equity gap, using Vermont's update rule



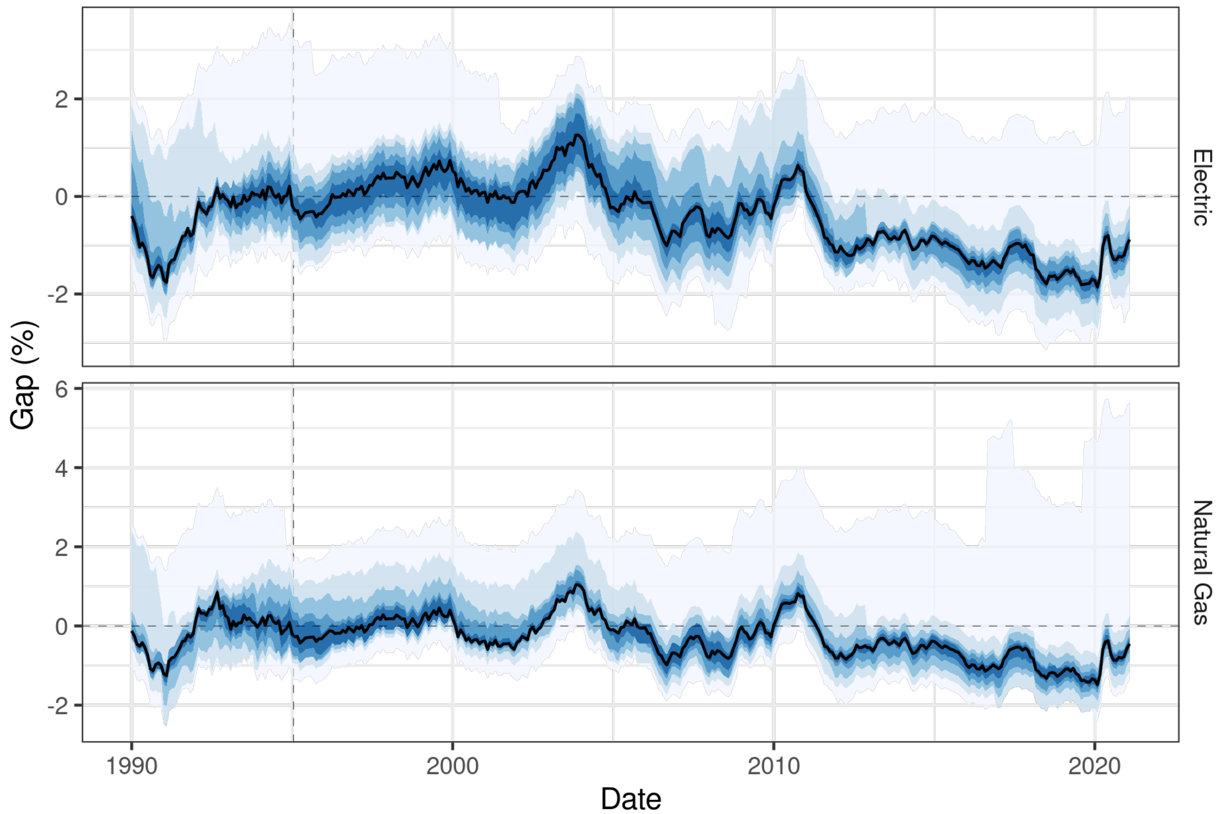
Line represents median; shading represents ranges that cover the central 20, 40, 60, 80, and 95% of total IOU ratebase. See calculation details in section 4.1.2.

Figure 6: Return on equity gap, compared to UK utilities



Base year is 1995. Line represents median; shading represents ranges that cover the central 20, 40, 60, 80, and 95% of total IOU ratebase. See calculation details in section 4.1.3.

Figure 7: Return on equity gap, benchmarking to CPI



Base year is 1995. Line represents median; shading represents ranges that cover the central 20, 40, 60, 80, and 95% of total IOU ratebase. See calculation details in section 4.1.4. Dates before 1990 are omitted for better axis scaling.

spread hadn't changed. One advantage of this approach is that we can still allow utilities to move around in their relative rankings and RoE. For example if a particular utility gets riskier and has correspondingly high RoE, our measure allows for that change in individual riskiness.

4.1.2 *Indexed to Treasuries*

Our next measure uses the RoE update rule recently implemented by the Vermont PUC. This rule is the only one we're aware of, from any PUC, that currently does automatic updating. Define R' as the baseline RoE, B' as the baseline 10-year Treasury bond yield, and B_t as the 10-year Treasury bond yield in year t. The update rule says the RoE in year t is then:

$$R_t = R' + \frac{B_t - B'}{2}$$

In the graph, we set the baseline to January 1995. In reality the commission set the baseline period as December 2018, for their plan published in June 2019. (*Green Mountain Power: Multi-Year Regulation Plan 2020–2022* 2020). We simulate the gap between approved RoE and what RoE would have been if every state's utilities commission followed this rule from 1995 onward. (Pre-1995 values are not particularly meaningful, but we can calculate them with the same formula.) We plot results in figure 5.

4.1.3 International Benchmark

We also consider an international benchmark. Here we ask, "what if US utilities faced a return on equity that was the same as return on equity in the UK?" Unlike the previous cases, we're not considering some benchmark year. Instead, we're considering the contemporaneous gap between the US and UK. Of course many things are different between these countries, and it's not fair to say all US utilities should adopt UK rate making, but we've think this benchmark provides an interesting comparison. Our results are in figure 6.

4.1.4 Indexed to Inflation

We also consider a calculation where we benchmark against core CPI. The mechanics of this calculation are identical to the Baa comparison above, where we calculate the gap between approved RoE and what the RoE would be if the mean spread against core CPI were unchanged. In this analysis, we find a small negative gap: real approved values RoE have declined, but by less than other costs of capital.

4.2 RATE BASE IMPACTS

Next, we turn to the ratebase the utilities own. A utility with a positive RoE gap will have a too-strong incentive to have capital on their books. In this section, we investigate the change in ratebase utilities request and receive. For our purposes, change in ratebase is more relevant than the total ratebase, as the change is a flow variable that changes from rate case to rate case, while the total ratebase is the partially-depreciated stock of all previous ratebase changes. We consider both the requested change and the approved change, though the approved value is our preferred specification. We estimate $\hat{\beta}$ from the following:

$$\log(RBI_{i,t}) = \beta RoE_{i,t}^{gap} + \gamma X_{i,t}\theta_i + \lambda_t + \epsilon_{i,t} \quad (3.1)$$

where an observation is a utility rate case for utility i in year-of-sample t. The dependent variable, $RBI_{i,t}$, is the increase in the rate base, and we take logs. (Cases where the ratebase shrinks are rare, but do happen. We drop these cases.) The independent variable of interest, $RoE_{i,t}^{gap}$, is the gap between the allowed return on equity and the true return on equity over the length of the rate case, where each rate case has a duration of D years.

$$RoE_{i,t}^{gap} = RoE_{i,t}^{allowed} - \frac{1}{D}\sum_{t}^{t+D} RoE_{i,t}^{correct} \quad (3.2)$$

Unlike section 4.1, for this analysis we care about differences in the gap between utilities or over time, but do not care about the overall magnitude of the gap. For ease of implementation, we begin by considering the gap as the spread between the approved rate of return and the 10-year Treasury bond yield. We do not expect the correct return on equity to be equal to the 10-year Treasury yield, but our fixed effects account for any constant differences. Future research will consider a richer range of gap calculations.

4.2.1 Fixed Effects Specifications

Our goal is to make causal claims about $\hat{\beta}$, so we are concerned about omitted variables that are correlated with both the estimated RoE gap and the change in ratebase. We begin with a fixed-effects version of the analysis. Our preferred version includes time fixed effects, λ_t, at the year-of-sample level and the unit fixed effects, θ_i, are at the utility company and state level.[10] Here, the identifying assumption is that after controlling for state and year effects, there are no omitted variables that would be correlated with both our estimate of the RoE gap and the utility's change in ratebase. The identifying variation is the differences in the RoE gap within the range of rate case decisions

10. Many utilities operate within only on state, but some span multiple. These company and state fixed effects are only partially nested.

Table 2: RoE gap, by different benchmarks

A: Electric		Baa yield	VT rule	UK	CPI
Gap (%)	2000	0.796	0.21	3.17	0.531
	2020	3.26	0.485	2.03	−1.06
Excess payment ($bn)	2000	0.581	0.23	4.54	0.142
	2020	6.54	1.43	3.92	−2.61
B: Natural Gas					
Gap (%)	2000	0.969	0.142		0.704
	2020	3.9	1.15	1.89	−0.421
Excess payment ($bn)	2000	0.0896	0.0183		0.0212
	2020	2.14	0.658	0.975	−0.361

NOTE: Gap percentage figures are an unweighted average across utilities. Excess payments are totals for all IOUS in the US, in billions of 2019 dollars per year, *for the observed ratebase.*
For cases where it's relevant (Baa yield, VT rule, and CPI), the benchmark date is January 1995. See text for details of each benchmark calculation.

for a given utility, relative to the annual average across all utilities. These fixed effects handle some of the most critical threats to identification, such as macroeconomic trends, technology-driven shifts in electrical consumption, or static differences in state PUC behavior. In columns 1–3 of our results tables (3 and 4), we consider different specifications for our fixed effects.

In this case the identification hinges on looking at variation in the RoE gap within the range of rate case decisions for a given utility, relative to the annual average across all utilities. The identifying assumption is that after controlling for state, year, and company effects, there are no omitted variables that would be correlated with both our estimate of the RoE gap and the utility's change in ratebase. These fixed effects handle many of the stories one could tell, such as macroeconomic trends, technological shifts in electrical consumption, or static differences in state PUC behavior. However, there are certainly other avenues for omitted variables bias to creep in, so next we turn to an instrumental variables strategy.

4.2.2 Instrumenting with Rate Case Timing and Duration

To try and further deal with concerns regarding identification, we examine an instrumental variables approach based on the timing and duration of rate cases.

Our IV analysis takes the idea that rates move around in ways that aren't always easy for the regulator to anticipate. So for instance if the allowed return on equity is set in year 0 and financial conditions change in year 2 such that the real allowed return on equity increases, then we would expect the utility to increase their capital investments in ways that

Table 3: Relationship Between Proposed Rate of Return and Proposed Rate Base

	Fixed effects specs.			IV
Model:	(1)	(2)	(3)	(4)
Variables				
RoE gap (%)	0.0670***	0.0436*	0.0672***	0.0353
	(0.0134)	(0.0217)	(0.0151)	(0.0215)
Fixed-effects				
State	Yes	Yes	Yes	Yes
Year		Yes	Yes	Yes
Company			Yes	Yes
Fit statistics				
Observations	3,210	3,210	3,210	3,210
R^2	0.37	0.39	0.73	0.73
Within R^2	0.24	0.23	0.29	0.29
Wald (1st stage)				50.9
Dep. var. mean	63.69	63.69	63.69	63.69

Two-way (Year & Company) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

NOTES: The dependent variable in the first panel is log of the utility's proposed rate base increase. Columns 1–3 show varying levels of fixed effects. Column 4 is the IV discussed in section 4.2. Our preferred specification is column 4 of table 4. First-stage *F*-statistic is Kleibergen–Paap robust Wald test. All regressions control for an indicator of electricity or natural gas.

are unrelated to other aspects of the capital investment decision. For this instrument to work, it needs to be the case that these movements in bond markets or the like are conditionally independent of decisions that the utility is making, except via this return on equity channel. We control for common year fixed effects, and then the variation that drives our estimate is that different utilities will come up for their rate case at different points in time.

5 RESULTS

Beginning with the RoE gap analysis from section 4.1, table 2 summarizes the graphs, using 2000 and 2020 as example points in time. The table highlights the RoE gap and the excess payment on the existing ratebase. Our results on the RoE gap can largely be guessed from a close inspection of figure 1. Approved RoE has not changed much in real terms (i.e. relative to core CPI), but the gap has increased between RoE and various financial benchmarks. Of our various imperfect estimates of the gap, we believe the Baa benchmark is the most credible.

Table 4: Relationship Between Approved Rate of Return and Approved Rate Base

	Fixed effects specs.			IV
Model:	(1)	(2)	(3)	(4)
Variables				
RoE gap (%)	0.0551***	0.0752***	0.0867***	0.0523**
	(0.0200)	(0.0240)	(0.0225)	(0.0252)
Fixed-effects				
State	Yes	Yes	Yes	Yes
Year		Yes	Yes	Yes
Company			Yes	Yes
Fit statistics				
Observations	2,491	2,491	2,491	2,491
R^2	0.33	0.36	0.69	0.69
Within R^2	0.21	0.20	0.22	0.22
Wald (1st stage)				69.1
Dep. var. mean	38.63	38.63	38.63	38.63

Two-way (Year & Company) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

NOTES: The dependent variable in the first panel is log of the utility's approved rate base increase. Columns 1–3 show varying levels of fixed effects. Column 4 is the IV discussed in section 4.2. Our preferred specification is column 4. First-stage F-statistic is Kleibergen–Paap robust Wald test. All regressions control for an indicator of electricity or natural gas.

Totalling up the 2020 excess payments gives us \$8.7 billion in the Baa benchmark, or \$2.1 billion in the Vermont benchmark. The UK benchmark falls between these, at \$4.9 billion.

We also consider how the RoE gap affects capital ownership. Tables 3 and 4 show our regression results for proposed and approved values, respectively. Our preferred specification is column 4, the IV specification, in table 4. These results find that a 1 percentage point increase in the approved RoE gap leads to a 5.2% increase in the increase in approved rate base. These results have a strong first stage (Kleibergen–Paap F-stat of 69).

As a caveat, we note that an IOU can increase their capital holdings in two distinct ways. One option is to reshuffle capital ownership, either between subsidiaries or across firms, so that the IOU ends up with more capital on its books, but the total amount of capital is unchanged. The second option is to actually buy and own more capital, increasing the total amount of capital that exists in the state's utility sector. We do not differentiate between these two cases. Because we don't differentiate, we consider excess payments by utility customers, but we remain agnostic about the socially optimal level of capital investment.

6 CONCLUSION

Utilities invest a great deal in capital, and need to be compensated for the opportunity cost of their investments. Getting this rate of return, particularly the return on equity, correct is challenging, but is a first-order important task for state PUCs.

Our analysis shows that the RoE that utilities are allowed to earn has changed dramatically relative to various financial benchmarks in the economy. Across relevant benchmarks, we found that current rates are perhaps 0.5–4 percentage points too high, resulting in $2–8 billion in excess rate collected per year, given the existing ratebase.

We then turned to the Averch–Johnson effect, and estimated the additional capital this RoE gap generates. In our preferred specification, we estimate that an additional percentage point in the RoE gap leads to 5% higher rate base increases.

We hope that policymakers and regulators consider these changes and these benchmarks in future rate making and the role that a wider variety of metrics benchmarks and adjustments can play in utility rate cases. We close by echoing Rode and Fischbeck (2019) and the Vermont PUC. Just as PUCs adopted fuel adjustment clauses in the 1960s and 1970s, RoE adjustment clauses are a tool that would allow rates to automatically adjust to changing market conditions. It would, of course, be possible to change the formula from time to time, but by default, the PUC wouldn't need to, even as the cost of raising capital changes. If such a scheme was implemented, it would be necessary to think hard about the baseline rate. As we demonstrated, the approved RoE has grown over time, so the choice of baseline period is crucial.

Figure 8: Figures 8 and 9 from Rode and Fischbeck (2019), showing CAPM β and market risk premium

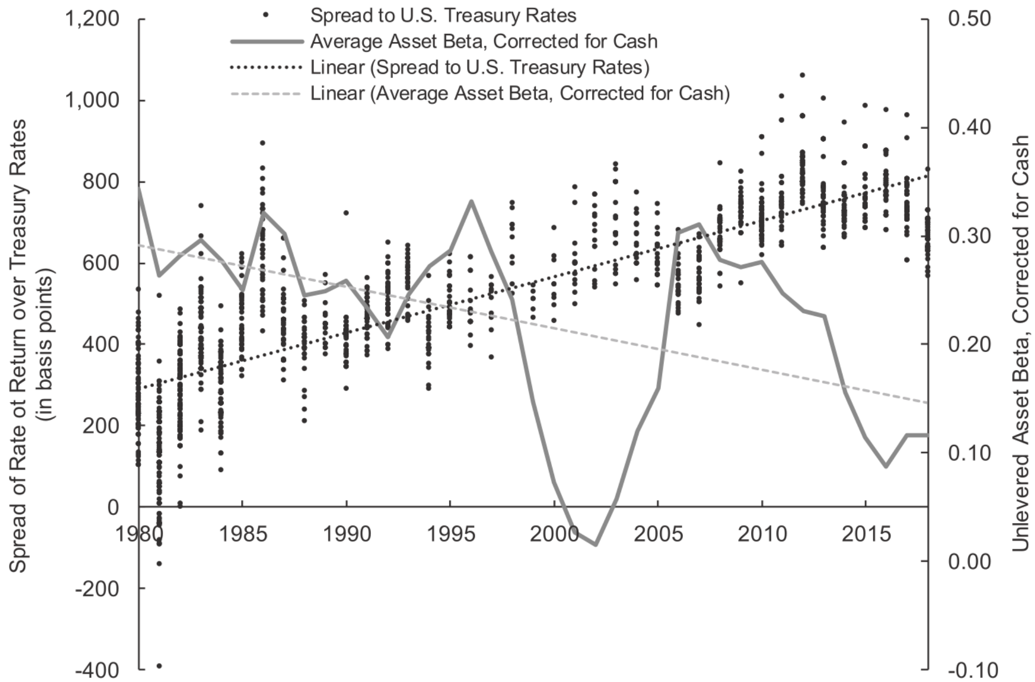


Fig. 8. Authorized return on equity premium vs. industry average asset beta.

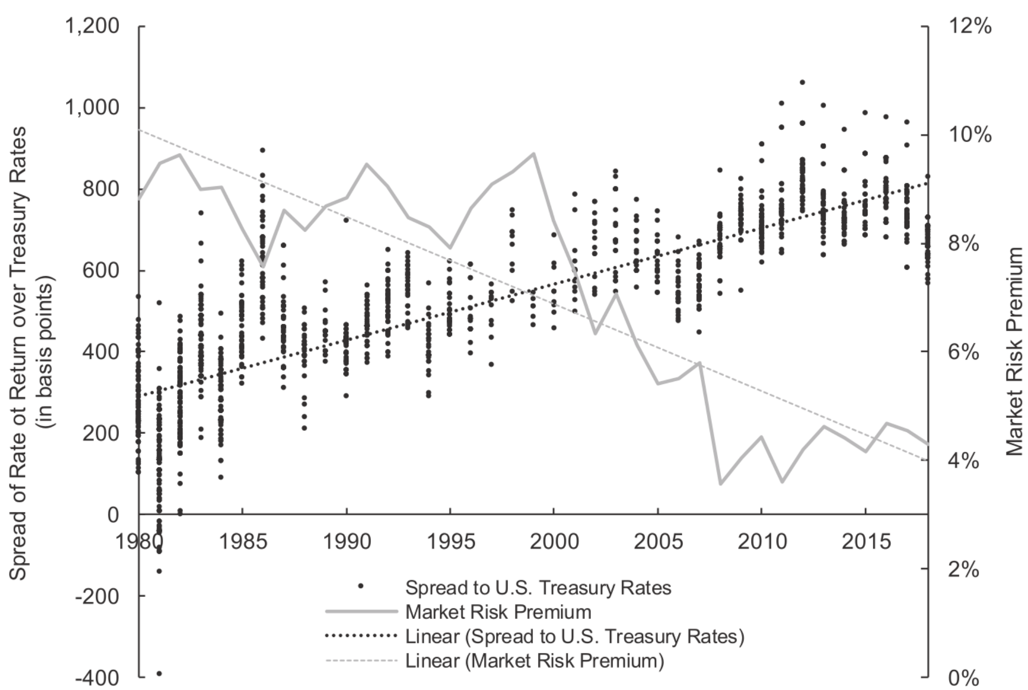


Fig. 9. Authorized rate-of-return premium vs. *ex ante* estimated market risk premium.

Bibliography

Borenstein, Severin, Meredith Fowlie, and James Sallee. 2021. *Designing Electricity Rates for An Equitable Energy Transition.* Technical report 314. February. https://haas.berkeley.edu/wp-content/uploads/WP314.pdf. (Cited on page 73).

Cawley, James H, and Norman J Kennard. 2018. *A Guide to Utility Ratemaking.* Technical report. Pennsylvania Public Utility Commission. https://www.puc.pa.gov/General/publications_reports/pdf/Ratemaking_Guide2018.pdf. (Cited on page 72).

EIA. 2018a. *Major utilities continue to increase spending on U.S. electric distribution systems.* Report, Today In Energy. Energy Information Administration. https://www.eia.gov/todayinenergy/detail.php?id=36675. (Cited on page 70).

———. 2018b. *Utilities continue to increase spending on transmission infrastructure.* Report, Today In Energy. Energy Information Administration. https://www.eia.gov/todayinenergy/detail.php?id=34892. (Cited on page 70).

Farrell, John. 2019. "Power Plant Securitization: Coming to a State Capitol Near You." *Institute for Local Self-Reliance* (May 13, 2019). https://ilsr.org/power-plant-securitization-coming-to-a-state-capitol-near-you/. (Cited on page 73).

Ghadessi, Maryam, and Marzia Zafar. 2017. *An Introduction to Utility Cost of Capital.* Technical report. April 18, 2017. https://www.cpuc.ca.gov/uploadedFiles/CPUC_Public_Website/Content/About_Us/Organization/Divisions/Policy_and_Planning/PPD_Work/PPD_Work_Products_(2014_forward)/PPD-An-Introduction-to-Utility-Cost-of-Capital.pdf. (Cited on page 72).

Green Mountain Power: Multi-Year Regulation Plan 2020–2022. 2020. Technical report. September 3, 2020. Accessed March 17, 2021. https://puc.vermont.gov/sites/psbnew/files/doc_library/green-mountain-power-multi-year-regulation-plan.pdf. (Cited on page 81).

Hausman, Catherine. 2019. "Shock Value: Bill Smoothing and Energy Price Pass-Through." *The Journal of Industrial Economics* 67, no. 2 (December 4, 2019): 242–278. https://doi.org/10.1111/joie.12200. (Cited on page 74).

Henbest, Seb, Matthias Kimmel, Jef Callens, Tifenn Brandily, Meredith Annex, Julia Attwood, Melina Bartels, et al. 2020. *New Energy Outlook 2020 Executive Summary.* Technical report. October. Accessed April 1, 2021. https://assets.bbhub.io/professional/sites/24/928908_NEO2020-Executive-Summary.pdf. (Cited on page 70).

Joskow, Paul L. 1972. "The Determination of the Allowed Rate of Return in a Formal Regulatory Hearing." *The Bell Journal of Economics and Management Science* 3 (2): 632–644. https://doi.org/10.2307/3003042. (Cited on page 70).

———. 1974. "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation." *The Journal of Law and Economics* 17, no. 2 (October): 291–327. https://doi.org/10.1086/466794. (Cited on page 70).

Rode, David C., and Paul S. Fischbeck. 2019. "Regulated equity returns: A puzzle." *Energy Policy* 133 (October): 110891. https://doi.org/10.1016/j.enpol.2019.110891. (Cited on pages 73, 77, 79, 86, 87).

Salvino, S. M. 1967. "Rate of Return Dilemma of Public Utilities under Rising Cost of Money Conditions." *Financial Analysts Journal* 23 (6): 45–49. http://www.jstor.org/stable/4470243. (Cited on page 73).

Strunk, Kurt G. 2014. *The Decoupling of Treasury Yields and the Cost of Equity for Public Utilities.* Technical report. June 13, 2014. https://www.nera.com/content/dam/nera/publications/archive2/PUB_Equity_Risk_Premium_Utilities_0614(1).pdf. (Cited on page 73).

Werth, Alix Elaine. 1980. "The effects of regulatory policy on the cost of equity capital and the value of equity in the electric utility industry." PhD diss., Massachusetts Institute of Technology. (Cited on page 79).

DATA CITATIONS

Board of Governors of the Federal Reserve System. 2021a. *1-Year Treasury Constant Maturity Rate.* FRED, Federal Reserve Bank of St. Louis, April 1, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/GS1. (Cited on pages 74, 75).

———. 2021b. *10-Year Treasury Constant Maturity Rate.* FRED, Federal Reserve Bank of St. Louis, April 1, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/GS10. (Cited on pages 74, 75).

———. 2021c. *30-Year Treasury Constant Maturity Rate.* FRED, Federal Reserve Bank of St. Louis, April 1, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/GS30. (Cited on pages 74, 75).

Companies (Classic) Screener. 2021. S&P Global Market Intelligence, January. Accessed March 17, 2021. https://platform.marketintelligence.spglobal.com/web/client?auth=inherit#office/screener. (Cited on pages 74, 76).

Compustat S&P legacy credit ratings. 2019. Wharton Research Data Service, October 18, 2019. Accessed March 12, 2021. (Cited on pages 74, 76).

Ice Data Indices, LLC. 2021a. *ICE BofA AA US Corporate Index Effective Yield.* FRED, Federal Reserve Bank of St. Louis, April 6, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAMLC0A2CAAEY. (Cited on page 74).

———. 2021b. *ICE BofA AAA US Corporate Index Effective Yield.* FRED, Federal Reserve Bank of St. Louis, April 6, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAMLC0A1CAAAEY. (Cited on page 74).

———. 2021c. *ICE BofA BB US High Yield Index Effective Yield.* FRED, Federal Reserve Bank of St. Louis, April 6, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAMLH0A1HYBBEY. (Cited on page 74).

———. 2021d. *ICE BofA BBB US Corporate Index Effective Yield.* FRED, Federal Reserve Bank of St. Louis, April 6, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAMLC0A4CBBBEY. (Cited on page 74).

———. 2021e. *ICE BofA CCC & Lower US High Yield Index Effective Yield.* FRED, Federal Reserve Bank of St. Louis, April 6, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAMLH0A3HYCEY. (Cited on page 74).

———. 2021f. *ICE BofA Single-A US Corporate Index Effective Yield.* FRED, Federal Reserve Bank of St. Louis, April 6, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAMLC0A3CAEY. (Cited on page 74).

Ice Data Indices, LLC. 2021g. *ICE BofA Single-B US High Yield Index Effective Yield.* FRED, Federal Reserve Bank of St. Louis, April 6, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAMLH0A2HYBEY. (Cited on page 74).

Moody's. 2021a. *Moody's Seasoned Aaa Corporate Bond Yield.* FRED, Federal Reserve Bank of St. Louis, April 1, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/AAA. (Cited on pages 74, 75).

———. 2021b. *Moody's Seasoned Baa Corporate Bond Yield.* FRED, Federal Reserve Bank of St. Louis, April 1, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/BAA. (Cited on pages 74, 75).

Regulatory Research Associates. 2021. *Rate Case History.* S&P Global Market Intelligence, March. Accessed March 18, 2021. https://platform.marketintelligence.spglobal.com/web/client?auth=inherit#industry/pastRateCases?Type=1. (Cited on pages 70, 74, 75, 77, 78).

US Bureau of Labor Statistics. 2021. *Consumer Price Index for All Urban Consumers: All Items in U.S. City Average.* FRED, Federal Reserve Bank of St. Louis, March 10, 2021. Accessed April 7, 2021. https://fred.stlouisfed.org/series/CPIAUCSL. (Cited on pages 74, 75).

US Energy Information Administration. 2020a. *2019 Utility Bundled Retail Sales – Total.* October 2, 2020. Accessed April 28, 2021. https://www.eia.gov/electricity/sales_revenue_price/xls/table10.xlsx. (Cited on page 71).

———. 2020b. *Summary Statistics for Natural Gas in the United States, 2015–2019.* September 30, 2020. Accessed April 28, 2020. https://www.eia.gov/naturalgas/annual/csv/t2019_01.csv. (Cited on page 71).

SOFTWARE CITATIONS

Bache, Stefan Milton, and Hadley Wickham. 2020. *magrittr: A Forward-Pipe Operator for R.* 2.0.1. https://cran.r-project.org/package=magrittr.

Dowle, Matt, and Arun Srinivasan. 2021. *data.table: Extension of 'data.frame'.* 1.14.0. https://cran.r-project.org/package=data.table.

Fabri, Antoine. 2020. *safejoin: Join safely and Deal with Conflicting Columns.* 0.1.0. August 19, 2020.

François, Romain, Jeroen Ooms, Neal Richardson, and Apache Arrow. 2021. *arrow: Integration to 'Apache' 'Arrow'.* 3.0.0. https://cran.r-project.org/package=arrow.

Gaure, Simen. 2013. "lfe: Linear group fixed effects." User documentation of the 'lfe' package, *The R Journal* 5, no. 2 (December): 104–117. https://journal.r-project.org/archive/2013/RJ-2013-031/RJ-2013-031.pdf.

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

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

———. 2020b. *rlang: Functions for Base Types and Core R and 'Tidyverse' Features.* 0.4.10. https://cran.r-project.org/package=rlang.

———. 2020c. *tidyselect: Select from a Set of Strings.* 1.1.0. https://cran.r-project.org/package=tidyselect.

Hester, Jim. 2020. *glue: Interpreted String Literals.* 1.4.2. https://cran.r-project.org/package=glue.

Iannone, Richard, Joe Cheng, and Barret Schloerke. 2020. *gt: Easily Create Presentation-Ready Display Tables.* 0.2.2. https://cran.r-project.org/package=gt.

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.

Kuhn, Max. 2020. *caret: Classification and Regression Training.* 6.0-86. https://cran.r-project.org/package=caret.

Microsoft Corporation and Steve Weston. 2020. *doParallel: Foreach Parallel Adaptor for the 'parallel' Package.* 1.0.16. https://cran.r-project.org/package=doParallel.

Müller, Kirill. 2020. *here: A Simpler Way to Find Your Files.* 1.0.1. https://cran.r-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2021. *tibble: Simple Data Frames.* 3.1.0. https://cran.r-project.org/package=tibble.

Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes.* 1.1-2. https://cran.r-project.org/package=RColorBrewer.

Ooms, Jeroen. 2019. *curl: A Modern and Flexible Web Client for R.* 4.3. https://cran.r-project.org/package=curl.

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.* 4.0.3. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2021. *broom: Convert Statistical Objects into Tidy Tibbles.* 0.7.5. https://cran.r-project.org/package=broom.

Teetor, Nathan. 2018. *zeallot: Multiple, Unpacking, and Destructuring Assignment.* 0.1.0. https://cran.r-project.org/package=zeallot.

Ushey, Kevin. 2021. *renv: Project Environments.* 0.13.0. https://cran.r-project.org/package=renv.

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.

———. 2020. *tidyr: Tidy Messy Data.* 1.1.2. https://cran.r-project.org/package=tidyr.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

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. 2021. *dplyr: A Grammar of Data Manipulation.* 1.0.4. https://cran.r-project.org/package=dplyr.

Wickham, Hadley, and Jim Hester. 2020. *readr: Read Rectangular Text Data.* 1.4.0. https://cran.r-project.org/package=readr.

Wickham, Hadley, Jim Hester, Winston Chang, Kirill Müller, and Daniel Cook. 2021. *memoise: Memoisation of Functions.* 2.0.0. https://cran.r-project.org/package=memoise.

Zeileis, Achim, and Gabor Grothendieck. 2005. "zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. https://doi.org/10.18637/jss.v014.i06.

Conclusion

These three papers cover a variety of topics in applied environmental economics. The first chapter addresses methane emissions from oil and gas wells, and considers the potential gains from policies that target these emissions. These gains could be large, but depend a great deal on the information the regulator has available and the details of the policy they enact. The second chapter considers the loss in value caused by flooding on agricultural land, examining losses over a wide range of flood frequencies. We contextualize these results in a world with changing climate, as properties that now flood occasionally are expected to flood more frequently in the future. The third chapter focuses on the rates of return utility companies are allowed to earn. These rates determine the profitability of investing in capital, the rates customers pay, and the amount of capital the utilities end up owning. All three of these chapters investigate policy-relevant economic topics, and all three use applied econometric tools to bring data to the question.